FRACTAL GEOMETRY AND DYNAMICS

MOOR XU
NOTES FROM A COURSE BY YAKOV PESIN

ABSTRACT. These notes were taken from a mini-course at the 2010 REU at the Pennsylvania State University. The course was taught for two weeks from July 19 to July 30; there were eight two-hour lectures and two problem sessions. The lectures were given by Yakov Pesin and the problem sessions were given by Vaughn Climenhaga.

CONTENTS

1. INTRODUCTION

This course is about fractal geometry and dynamical systems. These areas intersect, and this is what we are interested in. We won't be able to go deep. It is new and rapidly developing.[1]

Fractal geometry itself goes back to Caratheodory, Hausdorff, and Besicovich. They can be called fathers of this area, working around the year 1900. For dynamical systems, the area goes back to Poincare, and many people have contributed to this area. There are so many

[1] 19 June 2010

people that it would take a whole lecture to list. The intersection of the two areas originated first with the work of Mandelbrot. He was the first one who advertised this to non-mathematicians with a book called *Fractal Geometry of Nature*. This tells how the subject can be applied to models in physics. So many people from other areas were interested in it. Until recently, the area has been driven by physicists. They didn't prove anything, and made a bit of a mess; a lot things were not correct. They are responsible for the major ideas, however. They had major ideas, but some things were not correct. First we need to discuss some of each of fractal geometry and dynamics, and show how they interact.

2. Introduction to Fractal Geometry

First, we'll discuss fractal geometry.

Fractal sets are lines ($\mathbb{R}^1$), planes ($\mathbb{R}^2$) etc. Fractal sets are subsets $A \subset \mathbb{R}^n$. Here's a description of what they are. We'll go into the details later.

(1) complicated geometry
(2) self-similarity – the structure of the set repeats itself on different scales. We need to be less rigorous and allow some small distortion – nature isn't perfect. For example, the interval $[0, 1]$ is self-similar. We have to allow simple transformations: translation, rotation, etc.

2.1. Some Examples.

Example 2.1. We give an example of a set that is a fractal by satisfying both of these properties. We start with an interval, divide it into three parts, remove the central part, and replace it with an equilateral triangle so that each segment is of length $\frac{1}{3}$. This is a broken line with length $\frac{4}{3}$. We repeat for each segment to get a curve of length $(\frac{4}{3})^2$. We go on with this procedure. We repeat infinitely many times to get a fractal that looks like a snowflake. This is called a **von Koch curve**.

This has complicated geometry because each iteration has length $\frac{4}{3}^n$, so it has infinite length. But the curve lives in a bounded region, so this is a bounded curve of infinite length. If we pick a small piece and magnify it, we get the same curve. It is obviously self-similar because of the procedure for making it.

It is also not differentiable anywhere: there are a dense set of non-differentiable points. If the curve were differentiable, it would have finite length.

Example 2.2. Weierstrass suggested considering the function

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

If $0 < a < 1$, $b > 0$, $ab > 1+\frac{3}{2}\pi$, the graph of this function is an example of something that is continuous but not differentiable anywhere. Note that the finite sum

$$f_m(x) = \sum_{n=0}^{m} a^n \cos(b^n \pi x),$$

is differentiable, however. This curve has complicated geometry and is self-similar: It is another example of a fractal. This gives an analytic approach while van Koch gave a geometric approach.

Can you have a curve that is differentiable somewhere, but not differentiable somewhere, and have infinite length? Of course, we can cheat and attach a straight segment to a fractal. But this isn't interesting. Here's a more interesting version:

Problem 2.1. Is there a continuous curve which has infinite length and is differentiable everywhere except at a set of points that is nowhere dense? Note that this set has to be fairly complicated: It has to not be finite, but its complement must be dense. Pesin hasn't seen a solution written anywhere.

We consider another example of a fractal that is important in dynamical systems.

Example 2.3. Take the interval $[0, 1]$. Divide it into three parts and remove the middle third. Repeat this by cutting each remaining segment into three parts and removing the middle third of each. We repeat this infinitely. At the end, we obtain the **Cantor set**.

We removed intervals of total length 1, so this is a set of measure zero.

This set is known for a game called the Cantor game. We take an interval, pick a point, measure the distance to the nearest end, and take that distance from the other end. Repeat this process, and you eventually go out of the interval. Take two players, have them each choose a point, and they want to stay in the interval for as long as possible. The Cantor set has infinitely many steps in the interval, so being close to the Cantor set is good.

Here's another game:

Example 2.4. Take a point in an equilateral triangle, connect it to the nearest vertex, and go back double the distance to the vertex. Eventually, we land outside the triangle. What's the winning set?

We can easily see that this set is made by repeatedly removing the central equilateral triangles. After infinitely many steps, we obtain the **Sierpinski gasket**. This game is called the Sir Pinski game. This is another example of a fractal set: complicated geometry and self-similar. There is a three-dimensional version of this.

Fractals often show up in nature.

- branches of a California oak tree
- bronchial tree in lungs
- pictures of coastline taken from a satellite.
- Circle Limits IV by M. C. Escher
- other images from Escher

There exist fractal tilings. Starting from a hexagonal grid, we apply some small change to each side. Doing this repeatedly from the smaller lines, we obtain a tiling. Note that the area of each polygon does not change, but the perimeter becomes bigger. In the end, we get boundaries that are continuous curves of infinite length, but the area never changed.

2.2. **Cantor sets.**

When Cantor discovered his set, he was not aware of fractal geometry. Instead, consider $A \subset [0, 1]$. We want to count the number of points of A. If A is finite, this is easy; we just compare it to a set of known size and find a bijection. What if A is infinite?

We can think of this as a story. We have a party, and everyone is required to wear hats. Everyone left, and there was one hat left. To figure out whose hat it is, he invites everyone back, and asks them to wear hats. And someone doesn't have a hat. How many guests were there? It is infinite countable.

Definition 2.1. A set has countable cardinality if there is a bijection with the integers.

How many infinite sets of different cardinality are there? Countable sets are one example. Another example is the points in the interval $[0, 1]$. Cantor first proved that this set has cardinality bigger than that of countable sets. Cantor's question was if there was a set was cardinality between countable and uncountable. His example was the Cantor set. That was his motivation, but it turns out that he was wrong; the Cantor set is uncountable. Poincare said that this is a disease that mathematics would have to recover from. They ended up with fractal geometry.

Definition 2.2. Let the intervals of the first iteration of the Cantor set be Δ_1 and Δ_2, and let the intervals of the k-th iteration be $\Delta_{i_1,i_2,\dots i_k}$. Note that $\Delta_{i_1,i_2,\dots i_k} \subset \Delta_{i_1,i_2,\dots i_{k-1}}$. Then the Cantor set is

$$\mathcal{C} = \bigcap_{n\geq 0} \bigcup_{i_1,\dots,i_n} \Delta_{i_1,i_2,\dots,i_k}.$$

Each $x \in \mathcal{C}$ is associated to a *coding* determined by which interval it is in for each iteration of the Cantor set; for each iteration, it has a choice of two intervals to land in.

Let $\Sigma_2^+ = \{(i_1, i_2, \dots, i_n, \dots)\}$, $i = 1, 2$. Then consider the map $h\colon \Sigma_2^+ \to \mathcal{C}$ defined by

$$h(i_1, i_2, \dots) = x = \bigcap_{n\geq 0} \Delta_{i_1,i_2,\dots i_k}.$$

This is a coding map and is a one-to-one correspondence between Σ_2^+ and $\mathcal{C}$.

We now repeat the Cantor set construction with some changes to make it more general. Instead of dividing each interval into three equal thirds, we choose two arbitrary disjoint intervals. We go on to repeat this. This has the same structure of the Cantor set; we still have $\Delta_{i_1,i_2,\dots i_k}$ and $\mathcal{C}$ defined in the same way. This is also called a **Cantor set**, and there are lots of these. Note that both location and lengths of the intervals were chosen arbitrarily.

What is the length of such a Cantor set? Is it possible that length $\mathcal{C} > 0$?

To answer this, we make this construction a bit more rigid. Choose numbers $0 < \lambda_1, \lambda_2 < 1$ – these are called **ratio coefficients**. Now, choose intervals so that $|\Delta_{11}| = \lambda_1|\Delta_1|$ and $|\Delta_{12}| = \lambda_2|\Delta_2|$. Then $|\Delta_{i_1,i_2,\dots i_k}| = |\Delta_1|\lambda_{i_1}\lambda_{i_2}\dots\lambda_{i_k}$. To obtain the middle third Cantor set, we use $\lambda_1 = \lambda_2 = \frac{1}{3}$; it is a special case of our generalized Cantor set.

Note that the coding argument from before still holds: we still have a coding map $h\colon \Sigma_2^+ \to \mathcal{C}$. The coding does not distinguish ratio coefficients, and all Cantor sets are in correspondence with Σ_2^+. The coding loses a lot of information about the set by losing the ratio coefficients.

Hausdorff and Caratheodory asked if there are any properties that distinguish Cantor sets. We will discuss this later in the course.

3. Introduction to Dynamical Systems

Now that we've seen fractal geometry, we shall discuss dynamical systems.

3.1. **Some definitions.**

Definition 3.1. A **dynamical system** is a map $f : \mathbb{R}^n \to \mathbb{R}^n$ that maps $x \to f(x) = y$.

We think of this as a motion, or a transformation, and takes one point to another point. The simplest example of this is a **linear transformation** $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$. This can be written as $f(x) = Ax$, where A is a matrix. The simplest case is that of $\mathbb{R}^1$, where we have $f(x) = ax$.

We can iterate the map f to obtain

$$x \to f(x) \to f(f(x)) = f^2(x) \to \cdots \to f^n(x).$$

This is called the **trajectory** of x. We say that $f(x)$ is the image of x.

A **fixed point** is a point such that $x = f(x)$. Similarly, a **periodic point** is a point $f^m(x) = x$ for some m. Then m is the **period** of x. For the linear map, 0 is a fixed point, and there are no periodic points. When $a < 1$, 0 is a stable fixed point, and when $a > 1$, 0 is an unstable fixed point.

If we have a set $A \subset \mathbb{R}^n$, the **image** of A is $f(A) = \{f(x) \colon x \in \mathbb{R}^n\}$, and the **preimage** of A is $f^{-1}(A) = \{x \in \mathbb{R}^n \colon f(x) \in A\}$. By taking preimages, we have both positive and negative trajectories; we allow time to be both positive and negative.

Consider a linear map $f(x) = Ax$, $x \in \mathbb{R}^2$ such that $\det A \neq 0$, so that f is invertible. This matrix has eigenvalues $\lambda_1 = \alpha_1 + i\beta_1$ and $\lambda_2 = \alpha_2 + i\beta_2$. For now, assume that the eigenvalues are real numbers. If $\lambda_1 > 1$ and $\lambda_2 > 1$, the map f moves points away from 0. This is an **expanding fixed point**. Similarly, if $\lambda_1 < 1$ and $\lambda_2 < 1$, f moves points toward 0. This is an **attracting fixed point**. If $\lambda_1 < 1$ and $\lambda_2 > 1$, then f moves points along hyperbolas, and 0 is a **hyperbolic fixed point**.

3.2. **Nonlinear maps.** Consider a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ that is not necessarily linear but is invertible. It has a fixed point $f(0) = 0$. We can expand f as a Taylor series:

$$f(x) = f(0) + f'(0)x + g(x) = Ax + g(x)$$

for x within a small neighborhood of 0, where $g(x)$ is term of higher order, so

$$\frac{|g(x)|}{|x|} \to 0.$$

Since $g(x)$ gives only a small error, we can drop this term. Near the fixed point, we look at only the linear term. We can study the behavior of the nonlinear map near the fixed point by studying linear maps and correcting for the nonlinear term.

If we had another fixed point, we can do the same analysis around that fixed point as well. We want to connect all of our fixed points and get a global picture. Around each fixed point, we can draw the trajectories near each fixed point and get a first impression of what happens everywhere.

So we can analyze any map by looking at fixed points and linear maps. But it's not so simple, because we also have to consider periodic orbits. In order to find periodic points, we need to solve $f^n(x) = x$ for each x. Each is nonlinear and there are lots of equations, so this is actually very complicated. The most interested cases are those with infinitely many periodic orbits.

We can plot trajectories of $x \to f^n(x)$ in the plane $\mathbb{R}^2 = (x_1, x_2)$. Instead, we can also plot each coordinate against the time n. There are three types of behavior: regular periodic behavior, first chaotic and then periodic (intermediate), and forever chaotic. This can be seen by the regular periodic or entirely chaotic behavior of these coordinate plots. What drives this type of behavior?

Cantor sets are invariant sets $f(A) = A$ that are invisible but strongly influence the behavior, producing chaotic behavior. This gives a connection between fractal geometry and dynamical systems.

4. More on Dynamical Systems

4.1. **Definitions.** Last time, we considered Euclidean space $\mathbb{R}^p$ and a ball $\Delta \in \mathbb{R}^p$. [2] Inside this ball, we chose smaller discs $\Delta_1, \dots \Delta_r$. We chose ratio coefficients $\lambda_1, \lambda_2, \dots \lambda_r$ so that $0 < \lambda_1 < 1$ and $\sum \lambda_i = 1$. We continue this process to create balls $\Delta_{i_1,\dots,i_n}$ so that

$$\Delta_{i_1,\dots,i_n} \subset \Delta_{i_1,\dots,i_{n-1}}$$

and

$$\Delta_{i_1,\dots,i_n} \cap \Delta_{j_1,\dots,j_n} = \emptyset.$$

Let the radius of $\Delta_{i_1,\dots,i_n}$ be $r_{i_1,\dots,i_n}$, and we require

$$r_{i_1,\dots,i_n} = r_{i_1} \lambda_{i_2} \dots \lambda_{i_n}.$$

Definition 4.1. We call

$$\mathcal{C} := \bigcap_{n \geq 0} \bigcup_{i_1,\dots,i_n} \Delta_{i_1,\dots,i_n}$$

a **Cantor-like set**.

[2] 20 June 2010

Each of these sets is an example of a fractal set. This is a huge class of fractal sets.

Let $\Sigma_2^+ = \{\omega = (i_1, i_2, \dots i_n) : i_j = 1, 2, \dots, r\}$. This is the one-sided (positive) sequence on two numbers. Similar sets of sequences can be defined.

Now, there is a coding map $h : S_2^+ \to \mathcal{C}$ defined by

$$h(\omega) = h(i_1 \dots i_n \dots) = \bigcap_{n \geq 0} \Delta_{i_1, \dots, i_n} = x.$$

This is a bijection, and it is a **symbolic representation** of the Cantor set.

Remark. We can generalize this construction of the Cantor set. Now, consider $\lambda_i \to \lambda_i^{(n)}$. We would still assume that $\sum \lambda_i^{(n)} < 1$; this gives a construction where the ratio coefficients are functions of time. Generalizing further, we can obtain $\lambda_1 \to \lambda_{i_1, \dots i_n}$; this is called a construction of finite memory. Furthermore, instead of considering balls, we can consider ellipses. We can have two sets of ratio coefficients; one for each axis of the ellipses. We will only consider the simplest case of constant ratio coefficients.

We would like to find distinguishing properties of these sets.

Proposition 4.1. *Here are some properties of Cantor sets:*

(1) $\mathcal{C}$ *is closed*
(2) $\mathcal{C}$ *is nowhere dense in* Δ
(3) $\mathcal{C}$ *has measure zero:* $\mathrm{Vol}(\mathcal{C}) = 0$

4.2. **Connection to Dynamical Systems.** We considered the linear map $x \mapsto ax$. Consider the interval $[0, 1]$ and take two subintervals Δ_1 and Δ_2. Consider two linear functions on the two subintervals, and consider a map $f : \Delta_1 \cup \Delta_2 \to [0, 1]$. We can we can iterate this map repeated and obtain an infinite sequence of points. There is a problem, however: the trajectory might land outside of our two subintervals, in which case we have to stop. We want to find the set of points where this iteration procedure can be continued indefinitely.

Remark. We can look at this as a game. We have two players, and they can each choose a point, and repeat our procedure; the winner is the person with more iterations before stopping.

We can also draw a copy Δ_1 and Δ_2 on the vertical axis and project them down to the horizontal axis by the linear map. This yields four intervals on the horizontal axis: $\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}$; points in each of these four intervals can be iterated at least twice. Repeating this procedure, we will have two intervals inside each previous interval. This is a Cantor-like construction.

In this way, we obtain sets $\Delta_{i_1,\dots i_n}$ such that

$$\Delta_{i_1,\dots,i_n} \subset \Delta_{i_1,\dots,i_{n-1}}$$

and

$$\Delta_{i_1,\dots,i_n} \cap \Delta_{j_1,\dots,j_n} = \emptyset.$$

and $\Delta_{i_1\dots i_n}$ is closed. Then the set

$$\mathcal{C} = \bigcap_{n\geq 0} \Delta_{i_1\dots i_n}$$

is the largest invariant set $f(\mathcal{C}) = \mathcal{C}$. This is the most interesting part of the dynamics. We want to concentrate our attention on $\mathcal{C}$, considering the restriction $f|_{\mathcal{C}}$.

4.2.1. *Periodic orbits.*

Proposition 4.2. *The map $f|_{\mathcal{C}}$ has 2^n periodic orbits of period n. Adding all of them together, we obtain a dense set inside the Cantor set.*

Proof. Consider the coding map $h : \Sigma_2^+ \to \mathcal{C}$, and we have a map $f : \mathcal{C} \to \mathcal{C}$. Drawing a commutative diagram, we see that there is a map $\sigma : \Sigma_2^+ \to \Sigma_2^+$. This map can be expressed as $\sigma = h^{-1} \circ f \circ h$.

$$\begin{array}{ccc} \Sigma_2^+ & \xrightarrow{\sigma} & \Sigma_2^+ \\ \downarrow h & & \downarrow h \\ C & \xrightarrow{f} & C \end{array}$$

Exercise 4.1.

$$\sigma(\omega) = \sigma(i_1 i_2 \dots i_k \dots) = (i_2 i_3 \dots i_k \dots)$$

This is known as the **full shift**.

So if x is a periodic point so that $f^n x = x$ then $w = h(x)$ is a periodic point for σ, so that $\sigma^n \omega = \omega$. This is a 1-1 correspondence between periodic points of f and σ. But it is easy to find periodic points of σ: they just consist of repeating blocks

$$(i_1 i_2 \dots i_n | i_1 \dots i_n | i_1 \dots i_n | \dots).$$

Therefore there are 2^n periodic orbits of period n. □

4.2.2. *Chaotic Behavior.*

Remark. If he discussed chaotic dynamics a hundred years ago, Pesin would have been expelled for saying nonsense. The prevailing philosophy was that everything could be found at any time exactly without numerical error terms. There is no randomness involved. This is intuition coming from classical mechanics. So randomness can only come from outside, as an external force.

In 1963, there was a famous meteorologist Lorenz who was interested in predicting the weather and turbulence. He had a computer, which was a great advantage compared to calculating by hand. He simplified the Navier-Stokes equation to obtain a system of three differential equations in three unknowns, now called Lorenz equations. This is a crude approximation to Navier-Stokes. He used the computer to study solutions to his equations, and he determined that the solution is completely chaotic. He was very surprised, so he did a statistical analysis on the local extrema and concluded that the distribution of the local extrema were completely random. He wrote a paper in a meteorology journal and it was completely lost. A mathematician rediscovered this and saw that chaotic behavior appears in deterministic systems. Yorke called this **deterministic chaos**.

Consider a space and a map f moving points of the space to other points. Cut this space into two pieces. Follow a trajectory and mark which piece it visits (1221...). This is a simple coding of the trajectory, and it can be done for any dynamical system.

Given any arbitrary sequence, can you find a point with a trajectory that reproduces the sequence exactly? It this is the case, we have random behavior. Then we can never predict where any point x will be after n steps. This is an explanation of how chaotic motion appear in a deterministic system.

Consider our linear map again. We split $\mathcal{C}$ into two pieces: $\mathcal{C} \cap \Delta_1$ and $\mathcal{C} \cap D_2$. This gives a sequence as before, and we claim we can obtain any sequence by our symbolic representation. So the behavior is very complicated even though we're actually considering a linear map.

In general, how do we know how to split the region into two pieces? There might not be a good way to do this. In some cases, we might have to split into four pieces. This is one of the greatest problems in dynamical systems: Is it possible to produce a coding so that the system produces random behavior? Sometimes this is easy, sometimes it is hard, sometimes it is not possible.

Example 4.1. Consider the circle S^1 and a map $f : S^1 \to S^1$ defined by $f(x) = x + \alpha \pmod 1$. If α is a rational number then every point is periodic. If α is an irrational number, no points are periodic. This is not a random map. This is an interesting dynamical system that is not chaotic.

What drives this chaotic behavior? Going back to our linear map, take $x, y \in \mathcal{C}$ that are very close: $|x - y| < \varepsilon$. Compare the trajectories of x and y. Then $|f(x) - f(y)|$ is bigger than $|x - y|$ by a factor of the slope of the linear map, which is $\lambda_1^{-1} = |\Delta_1|$ or $\lambda_2^{-1} = |\Delta_2|$. Going on and repeating this process, we see that every point of $\mathcal{C}$ is a repelling point. The trajectories separate, and at some point, they can no longer be compared because they are in different pieces of our space. After enough time, they might come back together because the set is closed. This is chaotic behavior: there is a fight between expansion and coming back. This turns out to be a universal behavior.

How do we characterize this chaotic behavior? We need some mathematical tools. We want to characterize the size of this Cantor set.

4.2.3. *Metrics.* First, we need some general notions. In Euclidean space $\mathbb{R}^p$, for $x = (x_1, \dots, x_p)$ and $y = (y_1, \dots, y_p)$, there are several ways to measure distance:

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

$$d_1(x, y) = \sum_{i=1}^{p} |x_i - y_i|$$

$$d_2(x, y) = \max_{1 \le i \le p} |x_i - y_i|$$

We introduce a metric in Σ_2^+. For $\omega^{(1)} = (i_1^{(1)}, i_2^{(1)}, \dots, i_n^{(1)}, \dots)$ and $\omega^{(2)} = (i_1^{(2)}, i_2^{(2)}, \dots, i_n^{(2)}, \dots)$ we have

$$d(\omega^{(1)}, \omega^{(2)}) = \sum_{n=0}^{\infty} \frac{|i_j^{(1)} - i_j^{(2)}|}{2^j}$$

$$d_1(\omega^{(1)}, \omega^{(2)}) = \frac{1}{N(\omega^{(1)}, \omega^{(2)})}$$

where $N(a, b)$ is the **separation time** – the first position where a and b are different.

Two metrics $d_1(x, y)$ and $d_2(x, y)$ are **strongly equivalent** if there exists $C > 0$ such that

$$C^{-1} d_1(x, y) \le d_2(x, y) \le C d_1(x, y).$$

The metrics given earlier are strongly equivalent.

Example 4.2. We consider an example of a metric on $\mathbb{R}^p$ that is not strongly equivalent to the previous ones. Consider

$$d(x,y) = \begin{cases} |x-y| & \text{if } x \text{ and } y \text{ lie on the same line} \\ |x|+|y| & \text{otherwise.} \end{cases}$$

5. Hausdorff Dimension

5.1. **Definition.** Take any set $Z \subset \mathbb{R}^n$, and a number $\alpha \geq 0$. Fix $\varepsilon > 0$. We want to cover Z with countably many open balls of different radii. This collection of balls is

$$\mathcal{U} = \{B(x_i, r_i) : r_i < \varepsilon, i = 1, 2, 3, \dots\}.$$

These are called ε**-covers**. Note that this depends on which metric we use; the metric determines the notion of the ball. This won't change significantly if we change between strongly equivalent metrics. We're interested in finding some sort of "optimal" cover, using the "smallest" number of balls. This is given when

$$M(Z,\alpha,\varepsilon) := \inf_{\mathcal{U}} \left\{ \sum_{i=1}^{\infty} r_i^{\alpha} : \bigcup_i B(x_i, r_i) \supset Z, r_i \leq \varepsilon \right\}$$

is achieved, where $M(Z,\alpha,\varepsilon) \in [0,\infty]$.

Definition 5.1. Let the **Hausdorff function** be

$$m(Z,\alpha) := \lim_{\varepsilon \to 0} M(Z,\alpha,\varepsilon)$$

This limit exists because $M(Z,\alpha,\varepsilon)$ is monotonically increasing. This construction was first done by Caratheodory in 1914. It was independently rediscovered by Hausdorff in 1919. The name shows that history is not always fair.

5.1.1. *$m(Z,\alpha)$ as a function of Z.*

Proposition 5.1. *First, we fix α and consider $m(Z,\alpha)$ as a function of Z. There are some properties:*

(1) $m(\emptyset,\alpha) = 0$ (normalization)
(2) $m(Z_1,\alpha) \leq m(Z_2,\alpha)$ when $Z_1 \subset Z_2$ (monotonicity)
(3) $m(\bigcup_{i=1}^{\infty} Z_i, a) \leq \sum_{i=1}^{\infty} m(Z_i,\alpha)$ (sub-additivity)

Proof. The first two properties are trivial. We sketch the proof of the third fact. For any $\delta > 0$, there exists $\varepsilon > 0$ such that

$$|m(Z,\alpha) - M(Z,\alpha,\varepsilon)| \leq \frac{\delta}{2}.$$

There then exists $\mathcal{U} = \{B(x_i, r_i)\}$ such that

$$|M(Z, \alpha, \varepsilon) - \sum r_i^\alpha| \leq \delta.$$

Denote $Z = \bigcup Z_j$. We now see that there exists $\varepsilon > 0$ such that $\mathcal{U}_j = \{B(t_{ji}, r_{ji})\}$ where

$$|m(Z_j, \alpha) - \sum r_{ji}^\alpha| \leq \frac{\delta}{2^j}.$$

Note that $\mathcal{U} = \{B(x_{ji}, r_{ji})\}$ is an ε-cover of Z. Therefore, we see that

$$\begin{aligned} M(Z, \alpha, \varepsilon) &\leq \sum_{j,i} r_{ij}^\alpha = \sum_j \sum_i r_{ji}^\alpha \\ &\leq \sum_j \left(m(Z_j, \alpha) + \frac{\delta}{2^j} \right) = \sum_j m(Z_i, \alpha) + \delta. \end{aligned}$$

Since δ was arbitrary, the result now follows. □

Definition 5.2. A set function (a function that depends on a set) is called a **measure** if it satisfies these three properties. Our $m(Z, \alpha)$ is called the **Hausdorff measure**.

Sometimes this is infinite. It is most interesting when it is finite.

5.1.2. *$m(Z, \alpha)$ as a function of α.* Now we fix Z and consider $m(Z, \alpha)$ as a function of α. We can draw its graph. This graph is constant except possibly at one point; before this point it is infinite and after this point it is 0.

Proposition 5.2. *If $m(Z, \alpha)$ is finite then for every $\beta > \alpha$, $m(Z, \beta) = 0$. If $m(Z, \alpha)$ is finite and positive then for every $\beta < \alpha$, $m(Z, \beta) = \infty$.*

Proof. We prove the first statement; the second is a simple exercise.
Consider

$$M(Z, \beta, \varepsilon) = \inf_{\mathcal{U}} \left\{ \sum r_i^\beta : \bigcup B(x_i, r_i) \supset Z_i, r_i \leq \varepsilon \right\}.$$

This can be easily estimated by writing $r_i^\beta = r_i^\alpha r_i^{\beta - \alpha}$. □

Note that $m(Z, \alpha)$ can only be finite and nonzero for at most one value α.

Definition 5.3. α is called the **Hausdorff dimension** of Z. We denote $\alpha = \dim_H Z$.

We are interested in finding the sets where this can be computed. This is true for fractal sets, and the Hausdorff dimension is what we call "fractal dimension".

5.2. **Properties of Hausdorff Dimension.**

Proposition 5.3. *Some properties of the Hausdorff dimension*[3]

(1) $\dim_H \emptyset = 0$
(2) $\dim_H Z_1 \leq \dim_H Z_2$ *when* $Z_1 \subset Z_2$
(3) $\dim_H \bigcup Z_j = \sup \dim_H Z_j$

Proof. These follow from the properties in 5.1 above. In particular, the third property follows from sub-additivity:

Choose α so that

$$\alpha > \sup \dim_H Z_j.$$

Then

$$m(Z_j, \alpha) = 0 \implies m\left(\bigcup Z_j, \alpha\right) = 0 \implies \alpha > \dim_H \bigcup Z_j.$$

□

Corollary 5.4. *Note in addition that* $\dim_H\{x\} = 0$ *and* $\dim_H Z = 0$ *for any countable set* Z*.*

Example 5.1. $\dim_H \mathbb{R}^1 = 1$ and $\dim_H \mathbb{R}^p = p$.

The first fact follows from the fact that $\dim_H[0,1] = 1$, and the second fact follows that $\dim_H B^p = p$. We consider these bounded cases instead.

Consider the unit square S^2. We wish to show that the Hausdorff dimension is $\dim_H S^2 = 2$. This can be done directly from the definition, but there is another way. Note that $S^2 = [0,1] \times [0,1]$. So we ask: is it true that

$$\dim_H([0,1] \times [0,1]) = \dim_H([0,1]) + \dim_H([0,1]) = 2?$$

Question. Is it true that

$$\dim_H Z_1 \times Z_2 = \dim_H Z_1 + \dim_H Z_2?$$

This is a difficult question. This was posed by Besicovich. He posed this question during a seminar at Cambridge, and they tried to prove it. They failed because there is a counterexample. Instead, it is only true that

$$\dim_H Z_1 \times Z_2 \geq \dim_H Z_1 + \dim_H Z_2.$$

Can we set some conditions on when equality holds? We'll answer this question later. The point now is that dimensions are tricky.

[3] 21 June 2010

5.2.1. *Hausdorff Dimension of the Cantor Set.* We now consider the Hausdorff dimension of the Cantor set.

Besicovich posed this question, and Moran obtained the result. The most puzzling question is whether the Hausdorff dimension depends on the locations of the balls in the Cantor set. Everyone in the seminar believed that it should. Moran found a proof that shows this, and he presented this in the seminar. Besicovich found a mistake, and the next day Moran found a proof that doesn't depend on locations. This was correct. It is amazing that the locations of the balls do not matter – it only depends on the ratio coefficients.

His result was:

Theorem 5.5. *The Hausdorff dimension of the Cantor set is*

$$\dim_H \mathcal{C} = \alpha,$$

where α *satisfies the equation*

$$\sum_{i=1}^{r} \lambda_i^\alpha = 1.$$

Note that α is not an integer. It also doesn't depend on dimension; the dimension of the construction in higher dimensions is the same as a linear case.

Define $F(x) = \sum_{i=1}^{r} \lambda_i^x$. We plot this function and see that there is a unique root. Why is this root equal to the dimension of the Cantor set?

6. Box Dimensions

6.1. **Definitions.** We modify the notion of the Hausdorff dimension. Consider a set $Z \subset \mathbb{R}^p$ and a number α. Define

$$R(Z, \alpha, \varepsilon) = \inf_{U=\{B(x_i,r_i)\}} \left\{ \sum r_i^\alpha : \bigcup_i B(x_i, r_i) \supset Z, r_i = \varepsilon \right\}$$
$$r(Z, \alpha) = \lim_{\varepsilon \to 0} R(Z, \alpha, \varepsilon)$$

The only difference with the previous definition is that all balls are now the same size. Note that it is not clear that the limit exists. Instead, we consider

$$\overline{r}(Z, \alpha) = \overline{\lim}_{\varepsilon \to 0} R(Z, \alpha, \varepsilon)$$
$$\underline{r}(Z, \alpha) = \underline{\lim}_{\varepsilon \to 0} R(Z, \alpha, \varepsilon).$$

Note that we do not have subadditivity in this case. We still have the properties that $\overline{r}(\emptyset, \alpha) = 0$ and $\overline{r}(Z_1, \alpha) \leq \overline{r}(Z_2, \alpha)$ for $Z_1 \subset Z_2$. Similarly, $\underline{r}(\emptyset, \alpha) = 0$ and $\underline{r}(Z_1, \alpha) \leq \underline{r}(Z_2, \alpha)$.

As in the case of Hausdorff dimension, this yields two critical values $\overline{\alpha}$ and $\underline{\alpha}$ by an analogous argument as Proposition 5.2.

Definition 6.1. $\overline{\alpha}$ and $\underline{\alpha}$ are the upper and lower **box dimensions** of Z, denoted $\overline{\dim}_B$ and $\underline{\dim}_B$. Note that $\underline{\alpha} \leq \overline{\alpha}$.

6.2. **Properties of Box Dimensions.** We now list several properties of box dimensions without proof.

Proposition 6.1. *The box dimensions satisfy*

(1) $\dim_H Z \leq \underline{\dim}_B Z \leq \overline{\dim}_B Z$. *When these three values agree, the common value is called the fractal dimension.*
(2) $\overline{\dim}_B\{x\} = 0$
(3) $\overline{\dim}_B Z = 0$ *for any finite set* Z.
(4) $\overline{\dim}_B Z = \overline{\dim}_B \text{Closure}(Z)$.
(5) $\overline{\dim}_B[0,1] = 1$.

Remark. Note that we can rewrite $R(Z, \alpha, \varepsilon)$ as $\inf \varepsilon^\alpha N(Z, \varepsilon)$, where $N(Z, \varepsilon)$ is the smallest number of balls of radius ε needed to cover Z.

Then

$$\overline{\underline{\dim}}_B Z = \overline{\underline{\lim}}_{\varepsilon \to 0} \frac{\log N(Z, \varepsilon)}{\log \frac{1}{\varepsilon}}.$$

Another interesting result is that

$$\overline{\underline{\dim}}_B Z = \overline{\underline{\lim}}_{\varepsilon \to 0} \frac{\log L(Z, \varepsilon)}{\log \frac{1}{\varepsilon}}.$$

where $L(Z, \varepsilon)$ is the largest number of disjoint balls of radius ε centered at Z. It is easy to relate $L(Z, \varepsilon)$ and $N(Z, \varepsilon)$. These formulas are often useful, but we will not need them. We omit the proof.

Problem 6.1. We do a Cantor-like construction with the sets

$$\Delta_1, \ldots, \Delta_r,$$

where the ratio coefficients $\lambda_1^{(n)}, \ldots \lambda_r^{(n)}$ depend on n, and $\sum \lambda_i^{(n)} = 1$, where $\lambda_i^{(n)} \to \lambda_i$ when $n \to \infty$. Is Moran's statement that

$$\dim_H \mathcal{C} = \underline{\dim}_B \mathcal{C} = \overline{\dim}_B \mathcal{C}$$

true in this case?

6.3. **Dimension of the Cantor Set.** We want to refine Moran's result of his Theorem 5.5, and we will obtain:

Theorem 6.2.

$$\dim_H \mathcal{C} = \underline{\dim}_B \mathcal{C} = \overline{\dim}_B \mathcal{C}.$$

6.3.1. *Two Technical Lemmas.* To prove this theorem, we need two technical lemmas. These allows us to handle the Hausdorff dimension. They estimate $\dim_H$ from above and from below.

Lemma 6.3. *Assume that there exists $C > 0$ such that for every $\varepsilon > 0$ there exists an ε-cover $U = \{B(x_i, r_i)\}$ with $\sum r_i^\alpha \leq C$. Then $\dim_H Z \leq \alpha$.*

Proof. Observe that $M(Z, \alpha, \varepsilon) \leq \sum r_i^\alpha \leq C$ for every $\varepsilon > 0$. Therefore, $m(Z, \alpha) \leq C$ is finite, so $\dim_H Z \leq \alpha$, as desired. ☐

Lemma 6.4. *Assume that there exists $C > 0$, $\varepsilon > 0$ such that for every ε-cover $U = \{B(x_i, r_i)\}$ where $\sum r_i^\alpha \geq C$. Then $\dim_H Z \geq \alpha$.*

Proof. We have $M(Z, \alpha, \varepsilon) \geq C$ for some $\varepsilon > 0$. So $m(Z, \alpha) \geq M(Z, \alpha, \varepsilon) \geq C$. As before, this is means that $\dim_H Z \geq \alpha$. ☐

Remark. In the first lemma, we need to build a cover for any given ε. In the second lemma, we need to work with all ε-covers. Therefore the second lemma is harder to work with. So estimates from below for Hausdorff dimension are usually much harder to establish than those from above.

Proposition 6.5.

$$\dim_H[0, 1] = 1$$

Proof. To prove this, we'll first show $\dim_H[0, 1] \leq 1$. We want to find a good cover of intervals of size ε. To do this, divide the interval into strips of length ε. We can do this with $\frac{2}{\varepsilon}$ intervals of length ε, which has finite total length. Then $N([0, 1], \varepsilon) = \frac{2}{\varepsilon}$, so that

$$\lim_{\varepsilon \to 0} \frac{\log \frac{2}{\varepsilon}}{\log \frac{1}{\varepsilon}} = 1.$$

Therefore $\overline{\dim}_B[0, 1] \leq 1$.

Now, we want to show that $\dim_H[0, 1] \geq \alpha$ for any $\alpha < 1$. We use Lemma 6.4 with $C = 1$. We take any ε-cover. For this cover, we compute $\sum r_i^\alpha = \sum r_i r_i^{\alpha-1}$. Since $r_i < \varepsilon$, we have $r_i^{\alpha-1} \geq \frac{1}{\varepsilon^{1-\alpha}}$. Therefore

$$\sum r_i^\alpha \geq \frac{1}{2\varepsilon^{1-\alpha}} \sum 2r_i \geq \frac{1}{2\varepsilon^{1-\alpha}} \geq 1$$

for sufficiently small ε. Hence, $\dim_H[0, 1] \geq \alpha$ for any $\alpha < 1$, and we're done. ☐

6.3.2. *Proof of Dimension of $\mathcal{C}$.* We can now prove Moran's theorem.

Proof of Moran Theorem 6.2. We want to find a good cover for which we can use Lemma 6.3. Note that

$$|\Delta_{i_1,\dots,i_n}| = |\Delta_{i_1}|\lambda_{i_2},\dots\lambda_{i_n} \le |\Delta_i|\lambda^{n-1} \le a\lambda^{n-1}$$

where $0 < \lambda := \max_{1\le i\le r}\lambda_i \le 1$ and $a = \max|\Delta_i|$. This means that for any $\varepsilon > 0$, there exists n such that $|\Delta_{i_1,\dots,i_n}| \le \varepsilon$. These sets $\Delta_{i_1,\dots,i_n}$ give a cover for our problem. They are called **basic sets** of the construction.

We now need to show that

$$\sum |\Delta_{i_1,\dots,i_n}|^\alpha \le C.$$

To do this, observe that

$$\begin{aligned}
&\sum |\Delta_{i_1,\dots,i_n}|^\alpha = \sum_{i_1,\dots,i_n} |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \dots \lambda_{i_n}^\alpha \\
&= \sum_{i_1,\dots,i_{n-1}} |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \dots \lambda_{i_{n-1}}^\alpha \sum_{i_n} \lambda_{i_n}^\alpha = \sum_{i_1,\dots,i_{n-1}} |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \dots \lambda_{i_{n-1}}^\alpha \\
&= \dots = \sum_j |\Delta_j|^\alpha = C
\end{aligned}$$

because $\sum_{j=1}^r \lambda_j^\alpha = 1$. That gives us our estimate from above by Lemma 6.3. We have shown that $\dim_H \mathcal{C} \le \alpha$.

Now we need to estimate from below, which is more challenging.[4] We shall show that $\dim_H \mathcal{C} \ge \alpha$ using Lemma 6.4.

Consider a finite cover $U = \{\Delta_{i_1,\dots,i_n}\}$ of basic sets, where we allow basic intervals of different n's to be chosen. We require that $\bigcup \Delta_{i_1,\dots,i_n} \supset \mathcal{C}$. For any $\varepsilon > 0$, there exist m such that $|\Delta_{i_1,\dots,i_m}| < \varepsilon$. Therefore, U is an ε-cover.

Now, consider

$$\sum |\Delta_{i_1,\dots,i_n}|^\alpha$$

Pick $|\Delta_{i_1,\dots,i_m}|^m$ with the smallest m. We then have

$$\begin{aligned}
|\Delta_{i_1,\dots,i_m}|^m &= |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \cdots \lambda_{i_m}^\alpha = |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \cdots \lambda_{i_m}^\alpha \cdot (\lambda_1^\alpha + \lambda_2^\alpha) \\
&= |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \cdots \lambda_{i_m}^\alpha \cdot \lambda_1^\alpha + |\Delta_{i_1}|^\alpha \lambda_{i_2}^\alpha \cdots \lambda_{i_m}^\alpha \cdot \lambda_2^\alpha.
\end{aligned}$$

This means that that we can replace $\Delta_{i_1,\dots,i_m}$ with two elements of the next level in our sum. We can therefore replace it in our cover

[4] 22 July 2010

with the two new elements. Repeating this procedure, we obtain every subinterval at the l-th level of the Cantor construction. Therefore,

$$\sum |\Delta_{i_1,\dots,i_n}|^\alpha = \sum_{i_1,\dots,i_l} |\Delta_{i_1,\dots,i_l}|^\alpha = \sum_{j=1}^{r} |\Delta_j|^\alpha,$$

which is a fixed number. This is a remarkable property of covers of basic sets. We shall use this to bound the box dimension from above.

We can produce a cover by intervals with size of exactly ε, and it should have the smallest possible number of elements. Here is a clever trick.

Take a point $x \in \mathcal{C}$. This point is coded:

$$x = (i_1, i_2, \dots i_n, \dots) = \bigcap \Delta_{i_1,\dots,i_n}.$$

Along this sequence, we write a corresponding sequence of ratio coefficients: $(\lambda_{i_1}, \dots \lambda_{i_n})$. We can choose a sufficiently small ε such that $\lambda_{i_1} > \varepsilon$. If $\lambda_{i_1}\lambda_{i_2} < \varepsilon$, we stop here. Otherwise, we consider $\lambda_{i_1}\lambda_{i_2}\lambda_{i_3}$. Continuing, we reach a point at which

$$\begin{aligned} \lambda_{i_1} \cdots \lambda_{i_{m-1}} &> \varepsilon \\ \lambda_{i_1} \cdots \lambda_{i_m} &\le \varepsilon \end{aligned}$$

This gives us a cutoff time, and we get

$$\varepsilon \min_{1 \le j \le r} \lambda_j \le |\Delta_{i_1,\dots,i_m}| \le \varepsilon$$

is close to ε. Declare this to be an element of our cover. Now, consider any point not in this set and repeat, yielding more intervals $\Delta_{j_1,\dots,j_k}$. This gives finitely many disjoint basic sets, each of which has length almost ε. (Finite follows from compactness.)

Remark. Are the $\Delta_{i_1,\dots,i_n}$ open or closed? Consider the set $\Sigma_2^+ = \{(i_1, \dots, i_n, \dots)\}$ and the coding map $h : \Sigma_2^+ \to \mathcal{C}$. We want to choose sets such that

$$h(C_{i_1,\dots,i_n}) = \Delta_{i_1,\dots,i_n}.$$

These sets $C_{i_1,\dots,i_n}$ are

$$C_{i_1,\dots,i_n} = (j_1, \dots, j_n, \dots)$$

where $j_1 = i_1, \dots, j_n = i_n$. This is called a **cylinder**. We have a metric in the space of cylinders:

$$d(\omega^{(1)}, \omega^{(2)}) = \sum \frac{|i_j^{(1)} - i_j^{(2)}|}{2^j}$$

So our question becomes: Is $C_{i_1,\ldots,i_n}$ open or closed? Actually, these cylinders are clopen; this is a peculiar property of Σ_2^+. This is because we have a discrete topology.

So

$$\sum |\Delta_{i_1,\ldots,i_n}|^\alpha = \sum_{j=1}^{r} |\Delta_j|^\alpha$$

is finite. The only problem is that the sizes of the intervals are not precisely ε; they are a bit less. We fix this by expanding each interval in the cover by a little bit to make each have length precisely ε. After expansion, we have

$$\sum |I_{i_1,\ldots,i_n}|^\alpha = \frac{1}{\min \lambda_j} \sum_{j=1}^{r} |\Delta_j|^\alpha,$$

and by analogous results to Lemmas 6.3 and 6.4, we get our result for box dimensions.

We still need to prove that $\dim_H \mathcal{C} \geq \alpha$. We will use Lemma 6.4. Choose any $\varepsilon > 0$ and consider a cover $U = \{B(x_i, r_i)\} = \{I_i\}$. For any point $x = (i_1, i_2, \ldots, i_n, \ldots) \in \mathcal{C}$, we have an interval $I_i \in U$ that contains x. Define $r = |I_i|$. We have a cutoff time m such that

$$\begin{aligned} \lambda_{i_1} \cdots \lambda_{i_{m-1}} &> r \\ \lambda_{i_1} \cdots \lambda_{i_m} &\leq r. \end{aligned}$$

Therefore, we have the basic set $\Delta_{i_1,\ldots,i_m}$, where $m = m(I_i) = m(|I_i|)$ depends on the length of I_i. The idea is to replace the I_i with appropriate basic sets to obtain a cover by basic sets. Then, as before, the sum will become a constant, and we will be able to estimate $\sum r_i^\alpha = \frac{1}{2} \sum |I_i|^\alpha$.

We know two things about the basic sets:

$$\begin{aligned} &x \in \Delta_{i_1,\ldots,i_m} \\ &r \min \lambda_j \leq |\Delta_{i_1,\ldots,i_m}| \leq r. \end{aligned}$$

Instead of using I_j to cover, we need to choose intervals A_j that is around twice as big as I_j that is associated with a basic interval $\Delta_{i_1,\ldots,i_m}$. These A_j's form a cover. Then,

$$\sum r_i^\alpha = \frac{1}{2} \sum |I_i|^\alpha \sim \sum |\Delta_i|^\alpha \sim |\Delta_{i_1,\ldots,i_m}|^\alpha.$$

Therefore, we are done by Lemma 6.4. □

7. Further Ideas in Dynamical Systems

We want to understand the meaning of the formula

$$\sum_{i=1}^{r} \lambda_i^\alpha = 1.$$

In particular, what is the dynamical interpretation of the number α? Our proof of Moran's theorem used fractal geometry; we want to consider its implications in dynamics.

7.1. Markov Processes.

We now want to consider the linear map that we considered a few days ago in Section 4.2. Here, the invariant set is a Cantor set $\mathcal{C}$. Note that $f(\mathcal{C}) = \mathcal{C}$ is a repellor for f. Recall that $\dim_H \mathcal{C} = \alpha$ such that $\lambda_1^\alpha + \lambda_2^\alpha = 1$. This is too simple; we want to have something more complicated. We put this on hold.

Definition 7.1. Define

$$\Sigma_2^+ = \{\omega = (i_1, i_2, \dots, i_n, \dots)\}, i = 1, 2, \dots.$$

Let $A = (a_{ij})$ be an $r \times r$ matrix with all entries equal to 0 or 1. This is called a **transition matrix**.

Example 7.1. For example, consider

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

This means that the transition from 1 to 1 is not allowed, but the other transitions are allowed. This produces a graph with r vertices.

Let Σ_A^+ be the sequences allowed by this matrix; this is a subset $\Sigma_A^+ \subset S_2^+$:

$$\Sigma_A^+ = \{\omega = (i_1, i_2, \dots)\}, a_{i_n i_{n+1}} = 1.$$

Note that S_A^+ is an invariant subset under the shift map. It is also closed.

Note that this yields a map σ that is called a **subshift of finite type**.

$$\begin{array}{ccc} \Sigma_A^+ & \xrightarrow{\sigma} & \Sigma_A^+ \\ \downarrow h & & \downarrow h \\ C & \longrightarrow & C \end{array}$$

Apply these transition rules to the Cantor set construction, and erase all $\Delta_{i_1,\dots,i_n}$ that have sequences $(i_1, \dots, i_n)$ that are not allowable.

Definition 7.2. Define a new type of Cantor set by

$$\mathcal{C} = \bigcap_{n\geq 0} \bigcup_{\substack{(i_1,\ldots,i_n)\\ \text{allowable}}} \Delta_{i_1,\ldots,i_n)}$$

This is called a **Markov Cantor set construction**. Note that what happens at one step does not depend what happened before; the events are independent and the process has no memory. In our case, we have one step of memory; it only depends on the previous step. This is a **Markov process**.

Is there a good formula for the Hausdorff dimension of this type of construction? This should depend on the matrix chosen.

Here's some motivation for considering these new constructions. Our dynamical system before was modelled by Σ_2^+. We can consider much bigger classes of systems if we allow them to be modelled by S_A^+.

Definition 7.3. The map $f : \Delta_1 \cup \Delta_2 \to [0,1]$ is a **Markov map** if the following holds: If $f(\Delta_i) \cup \Delta_j \neq \emptyset$ then $f(\Delta_i) \supset \Delta_j$. This is called the Markov property. The simpler example we considered earlier is called the **full-branched** Markov map.

Example 7.2. As before, we use a linear map on Δ_2. Our map on Δ_1, however, doesn't have range in all of $[0,1]$. Instead, it has range Δ_2.

In this example, $f(\Delta_1) \cap \Delta_1 = \emptyset$ and

$$f(\Delta_1) \cap \Delta_2 \neq \emptyset \implies f(\Delta_1) \supset \Delta_2.$$

Similarly, $f(\Delta_2) \supset \Delta_1$ and $f(\Delta_2) \supset \Delta_2$. It is easy to see that this map is Markov.

We consider the repellor for this map. It turns out that this is precisely the Markov Cantor set construction that we did in the previous example 7.1.

7.2. **Nonlinear Case.** Now, what if we replace the linear pieces in our simple linear example by nonlinear pieces? We set the condition that $|f'(x)| > 1$. This is to ensure that our limit set is a repellor, which is the core of chaotic behavior. We study $f : \Delta_1 \cup \Delta_2 \to [0,1]$. Here, $|f'(x)| > a > 1$. By a similar construction that we did earlier, we obtain a Cantor set $\mathcal{C}$ that is invariant and it is a repellor. We still have our coding map $h : \Sigma_2^+ \to \mathcal{C}$. This means that the number of periodic orbits of order n is 2^n, and the total number of periodic orbits is infinite countable. This is identical to the linear case.

Question. What is the Hausdorff dimension for this Cantor set? Does it do anything to the dynamical characteristics of this system?

7.3. **Nonrigorous Introduction to Lyapunov Exponents.** Define $\lambda_1 = |\Delta_1|$ and $\lambda_2 = |\Delta_2|$. Then

$$f'(x) = \begin{cases} \lambda_1^{-1} & x \in \Delta_1 \\ \lambda_2^{-1} & x \in \Delta_2. \end{cases}$$

We want to have that

$$|\Delta_{i_1,\dots,i_n}| = \prod_{j=1}^{n} \lambda_{i_j} \sim \lambda^n.$$

This number λ could be computed as

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log \prod_{j=1}^{n} \lambda_{i_j} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \log \lambda_{i_j} = -\lim_{n\to\infty} \sum_{j=1}^{n} \log f'(f^j(x))$$
$$= -\lim_{n\to\infty} \frac{1}{n} \log \prod_{j=1}^{\infty} f'(f^j(x))$$

if the limit existed. We'll consider the basic idea now and fill in rigor later.

Now, starting from a point x and a basic interval $\Delta_{i_1,\dots,i_n}$ containing x, we would have

$$|\Delta_{i_1,\dots,i_n}| \sim \prod_{j=1}^{\infty} f'(f^j(x)).$$

Definition 7.4.

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log \left| \prod_{j=0}^{n-1} f'(f^j(x)) \right|$$

is the **Lyapunov exponent** of x.

For a physics paper, this would be perfect. For a mathematical paper, we have a problem: How do we know that this limit exists? It doesn't have to. We're going to consider only points where it does exist (good points), and ignore points where it does not (bad points). Most points will be good.

7.4. **Measure.** Recall that we have considered three examples of maps – the full-branched Markov map, the general Markov map, and non-linear Markov maps. We are interested in their Hausdorff dimensions. We work over Σ_2^+; it is easy to replace 2 with other numbers.[5]

[5] 26 July 2010

Let $0 < p < 1$ and $q = 1 - p$. From a probabilistic point of view, we have a sequence of independent events. This random probabilistic process can be interpreted by putting a measure on this space.

Example 7.3. Say we want to put a measure on the interval $[0, 1]$; this can be thought of as some notion of length. For example, we can write that the length of interval $[a, b]$ is $l([a, b]) = |b - a|$. We can extend to countable collections of intervals. In order to define length, we need to have some basic sets; in this case, we can use intervals. We should have the property

$$l\left(\bigcup [a_i, b_i]\right) \leq \sum |b_i - a_i|$$

with equality if they are all disjoint. This is a notion of subadditivity. We now do something analogous to the Caratheodory construction that we considered earlier.

Here, we use $\lambda(Z) = m_H(Z, 1)$. This gives a measure that is an extension of the notion of length.

If we define length in the whole line, there is a transformation $x \mapsto x + a$. Then $l(A) = l(A + a)$. This is the only measure on the line with this property.

We give a very general way of building measures. Our basic sets will be cylinders

$$C_{i_1,\ldots,i_n} = \{\omega \in \Sigma_2^+ : \omega = (j_1, \ldots, j_n), j_1 = i_1, \ldots, j_n = i_n\}.$$

Definition 7.5. Define a measure

$$m_p(C_{i_1,\ldots,i_n}) = p^{a_n} q^{n-a_n},$$

where a_n is the number of 1's in the n-tuple $(i_1, \ldots, i_n)$.

Since any two cylinders are disjoint, declare the measure of a collection of cylinders as a sum; this then satisfies additivity and subadditivity. This is known as the **Bernoulli measure**.

This measure has an interesting property. Note that

$$\mu_p(E) = \mu_p(\sigma^{-1}(E)).$$

This is because

$$\sigma^{-1}(C_{i_1,\ldots,i_n}) = \bigcup_{i_0=1}^{r} C_{i_0,i_1,\ldots,i_n}.$$

Then

$$\mu_p(\sigma^{-1}(C_{i_1,\ldots,i_n})) = \mu_p\left(\bigcup_{i_0=1}^{r} C_{i_0,i_1,\ldots,i_n}\right) = \sum_{i_0=1}^{r} \mu_p(C_{i_0,\ldots,i_n}).$$

Our commutative diagram then gives measures for our basic sets. So

$$\mu_p(\Delta_{i_1,\dots,i_n}) = p^{a_n} q^{n-a_n}.$$

This gives a measure on Cantor sets. We put a measure on the repellor, which had measure zero. This repellor has full measure under our new measure. We study some characteristics of the measure.

8. Computing Dimension

8.1. **Entropy.** This is one of the major characteristics of invariant measures. For each such measure, we can define the notion of entropy. For μ_p, we'll make an idea of entropy that is convenient; this won't work for a general measure. There are whole courses about entropy of dynamical systems, and we don't have time to be too general.

Pick $\omega \in \Sigma_2^+$, $\omega = (i_1, \dots, i_n, \dots)$ and a cylinder $C_{i_1,\dots,i_n}$. Then as $n \to \infty$, we want

$$\mu_p(C_{i_1,\dots,i_n}) \sim e^{-hn}$$

for some $h > 0$. If this is the case, h is the entropy.

Definition 8.1.

$$h(\omega) = h_{\mu_p}(\omega) = -\lim_{n\to\infty} \frac{1}{n} \log \mu_p(C_{i_1,\dots,i_n})$$

is the entropy.

8.1.1. *Entropy of Bernoulli measure.* We want to compute this for our Bernoulli measure. Then

$$\begin{aligned} h(\omega) &= -\lim_{n\to\infty} \frac{1}{n} \log\left(p^{a_n} q^{n-a_n}\right) = -\lim_{n\to\infty} \frac{1}{n}\left(a_n \log p + (n-a_n)\log q\right) \\ &= -\left(\lim_{n\to\infty} \frac{a_n}{n} \log p + \left(1 - \lim_{n\to\infty}\frac{a_n}{n}\right)\log q\right). \end{aligned}$$

The existence of this limit boils down to the existence of $\lim_{n\to\infty} \frac{a_n}{n}$. Simple probability suggests that this limit is p. If this were true, we would get that

$$h(\omega) = -(p \log p + q \log q).$$

This is the **entropy of** μ_p; this does not depend on ω.

There is a subtle question: For which ω does this limit exist? We postpone discussion of this for now; however, this is true for the "majority of points".

Consider the function

$$\phi(p) = -(p \log p + (1-p)\log(1-p)).$$

We can draw its graph. It is zero at $p = 0$ and $p = 1$; it looks like a bump. There is a maximum at $p = \frac{1}{2}$. For this p, the entropy is maximal. So $\mu_{1/2}$ is called the **the measure of maximal entropy**. This matches physical common sense; a fair coin is most chaotic.

8.2. **Lyapunov exponent.** Consider this notion of entropy on our Cantor set. Recall the Lyapunov exponent

$$\begin{aligned}
\lambda(x) &= \lim_{n\to\infty}\frac{1}{n}\log\left|\prod_{j=0}^{n-1} f'(f^j(x))\right| = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\log|f'(f^j(x))| \\
&= -\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\log\lambda_{i_j} = -\lim_{n\to\infty}\frac{1}{n}\log\prod_{j=0}^{n-1}\lambda_{i_j} \\
&= -\lim_{n\to\infty}\frac{1}{n}\log\lambda_1^{a_n}\lambda_2^{n-a_n} \\
&= -\left(\lim_{n\to\infty}\frac{1}{n}a_n\log\lambda_1 + \left(1-\lim_{n\to\infty}\frac{1}{n}a_n\right)\log\lambda_2\right) \\
&= -(p\log\lambda_1 + q\log l_2).
\end{aligned}$$

This exists whenever the previous limit exists. The formulas for entropy and the Lyapunov exponent are very similar.

8.3. **Pointwise dimension.**

Definition 8.2. Now, pick a point $x\in\mathcal{C}$, $x=(i_1,\iota_2,\dots)$. Define

$$d_{\mu_p}(x) = \lim_{n\to\infty}\frac{\log\mu_p(\Delta_{i_1,\dots,i_n})}{\log|\Delta_{i_1,\dots,i_n}|}$$

This is the **pointwise dimension of μ_p at x.**

Once again, there is the question of whether the limit exists; we postpone this yet again. We compute this:

$$\begin{aligned}
d_{\mu_p}(x) = \lim_{n\to\infty}\frac{\frac{1}{n}\log\mu(\Delta_{i_1,\dots,i_n})}{\frac{1}{n}|\Delta_{i_1,\dots,i_n}|} &= \frac{p\log p + q\log q}{p\log\lambda_1 + q\log\lambda_2} \\
&= \frac{p\log p + (1-p)\log(1-p)}{p\log\lambda_1 + (1-p)\log\lambda_2} =: \psi(p).
\end{aligned}$$

This again looks like a bump that is zero at $p=0$ and $p=1$. There is a maximum at λ_1^α, where $\psi(\lambda_1^\alpha)=\alpha$. The highest value of the pointwise dimension is the Hausdorff dimension. This is key. Let $\mu_{\lambda_1^\alpha}$ be the **measure of maximal dimension**.

8.4. **Computing the dimension.** To compute the dimension of the repellor, find the measure of maximal dimension. This gives a good way of computing the Hausdorff dimension.

We can interpret this differently; the dimension is the quotient of entropy over Lyapunov exponent:

$$\dim_H \mathcal{C} = \frac{h_{\mu_p}}{\lambda_{\mu_p}}.$$

This is good for computation. The only problem is to find the measure of maximal dimension. This formula ties together three important quantities in the subject; the formula relates fractal sets, chaos, randomness, and instability. This is beautiful.

Remark. We still need to see why our limits exist, and we still need to construct measures of maximal dimension.

The limit $\lim_{n\to\infty} \frac{1}{n} a_n$ is an asymptotic frequency of ones. We believe from probability that this should exist; the frequency of independent events is an mathematical expectation. This is a corollary of probability. This limit must exist, but not necessarily for all orbits; it exists for almost all orbits; bad orbits have measure zero. This is a fact from probability theory that we will not prove.

Joke. This was first told by Littlewood; the precise source is unclear. There was a big party, and there was a probabilist. He suggested that they bet: Count 100 people, and there will be between 45 and 55 men. So they bet, and counted 100 people. At that moment, 100 soldiers went by, and the probabilist was ashamed.

The moral is that if you make a bet on a large number, there is a small chance that things will be go wrong. It's just that the number 100 is too small.

The limit exists for a set of full measure. Therefore, this is true for the entropy and the Lyapunov exponent. Hence, it is true for the pointwise dimension as well.

8.5. **Markov measure.** Consider Markov constructions. We wish to define Markov measure. Everything called "Markov" is named after a Russian mathematician who pioneered these notions.

Definition 8.3. Define $P = (p_{ij})$ to be a **stochastic matrix** if $0 \leq p_{ij} \leq 1$, $\sum_{i=1}^n p_{ij} = 1$.

The transition matrix $A = (a_{ij})$ has $a_{ij} = 1$ when $p_{ij} > 1$, and 0 otherwise; it is defined by the stochastic matrix.

Recall that we have the set

$$\Sigma_A^+ = \{\omega = (i_1, \dots) : p_{i_n i_{n+1}} > 0\}$$

We also need the probability vector $\bar{p} = (p_1, \dots, p_r)$. It is sufficient to define a measure on a cylinder.

Definition 8.4. The Markov measure is

$$\mu_{P,\bar{p}}(C_{i_1,\dots,i_n}) = p_{i_1} P_{i_1 i_2} P_{i_2 i_3} \cdots P_{i_{n-1} i_n}.$$

This is used to carry out the calculation of the Hausdorff dimension. This calculation is straightforward. In nonlinear cases, the calculation of entropy is the same, but the Lyapunov exponent is much messier.

Example 8.1. In the case of the linear full-branched Markov map. If we have $\lambda := \lambda_1 = \lambda_2$, then it is easy to show that $\lambda(x) = -\log \lambda$. This is true for every x, without exceptions. In the case $p = \frac{1}{2}$, we also have $h_{\mu_{1/2}} = \log 2$. The Hausdorff dimension is then

$$\dim_H \mathcal{C} = -\frac{\log 2}{\log \lambda}.$$

This is an exceptional case where there are no exceptions. However, when we allow different slopes and $\lambda_1 \neq \lambda_2$, we only have it is true almost everywhere.

Remark. Given any point $x \in \mathcal{C}$, $\mu(\Delta_{i_1,\dots,i_n}) \sim |\Delta_{i_1,\dots,i_n}|^{d_\mu(x)}$. We would like to look at the points in $\mathcal{C}$ where $d_\mu(x) = a$ is constant. This is a better fractal than $\mathcal{C}$ because it has better self-similarity; the scale of self-similarity is constant. Each value of a gives a **real fractal**, and there are many real fractals packed in the Cantor set. Therefore, the Cantor set is a **multifractal**. In the case of $\lambda_1 = \lambda_2$; the slope is the same and the Cantor set is a pure fractal. This is unbelieveably complicated structure generated from a piecewise linear function.

9. Two Dimensional Dynamical Systems

So far, we've only worked in a one-dimensional world. For the remaining three lectures, we will discuss the two-dimensional case. This is more complicated, but the basic idea is the same: Use the measure of maximal dimension. There are simple classes of dynamical systems where we don't know if such a measure exists. That is the frontier of current research. We will start with a real model of propagation of voltage through a neural system. We will show that everything is governed by a simple two-dimensional dynamical system, and we'll use it to discover things in general. Going further up in dimension produce more complicated phenomenon, though many ideas still hold. We are only considered low-dimensional dynamical systems; there are many open questions in higher dimensions – we only opened the door to this subject.

9.1. **Fitzhugh-Nagumo Model.**

We discuss the FitzHugh-Nagumo Model of an electrical signal through a neuron.[6]

Each neuron is composed of several parts: the dendrids going into the soma, which has the axon as the only output. The signals go in through the dendrids, excites the soma, and sends output to the axon. The axon has length an order or two longer than the dendrids; it is a very thin tube. The signal propagates through the tube. The question is: What is the voltage in this tube at any time?

Let the length of the tube be parametrized by x, and let time be given by t. We want to study the voltage $U(x,t)$. There are several different ways to model this. There is a very complicated Hutchinson model that deals with all physical aspects, but that's hard to use. We'll consider the FitzHugh-Nagumo model, which treats this as a circuit.

In biology, the models are called *phenomenological* models; they are not based on fundamental laws. This is different from physics, where they have fundamental laws (e.g. Maxwell's Law). In biology, using fundamental laws is hopelessly complicated and unrealistic. So they forget about fundamental laws and find a simple model. They build models by looking at physical characteristics, based on things like diffusion or viscosity.

The FitzHugh-Nagumo model is

$$\begin{aligned}\frac{\partial u}{\partial t} &= -au(u-\theta)(u-1) - bv + k\frac{\partial^2 u}{\partial t^2}\\ \frac{\partial v}{\partial t} &= cu - dv\end{aligned}$$

The term $k\frac{\partial^2 u}{\partial t^2}$ is the diffusion term. It is very small. This model has parameters, and they have biological meanings.

9.2. **Interpretation of the Model.**

Joke. A teacher was explaining to some students that the Riemann hypothesis, and the students said to tell the computer to do it.

[6] 27 July 2010

One way to study this equation is to ask the computer to do it for us. This is done by approximations for the derivatives:

$$\frac{\partial u}{\partial t} \mapsto \frac{u(x,t+\Delta)-u(x,t)}{\Delta}$$
$$\frac{\partial^2 u}{\partial x^2} \mapsto \frac{u(x+\Delta,t)-2u(x,t)+u(x-\Delta,t)}{\Delta^2}.$$

We now replace continuous space with discrete space by defining

$$u_k(n) = u(k\Delta, u\Delta)$$
$$v_k(n) = v(k\Delta, u\Delta).$$

This yields

$$\begin{aligned} u_k(n+1) &= u_k(n) - a\Delta u_k(n)(u_k(n)-\theta)(u_k(n)-1) - bv_k(u) \\ &\quad + k\frac{u_{k+1}(n)-2u_k(n)+u_{k-1}(n)}{\Delta} \\ v_k(n+1) &= v_k(n) - d\Delta v_k(n) + c\Delta u_k(n). \end{aligned}$$

Simplifying gives

$$\begin{aligned} u_k(n+1) &= u_k(n) - A(u_k(n)-\theta)(u_k(n)-1) - \alpha v_k(n) + kgu_k(n) \\ v_k(n+1) &= v_k(n) - \beta u_k(n) + \delta v_k(n). \end{aligned}$$

The interpretation of this model is to produce a discrete lattice on the line. Due to the length of the axon, we consider the whole line. At each point on this lattice, we have a plane. On each such plane, we have a two-dimensional map, namely

$$\begin{aligned} f(u,v) &= (f_1(u,v), f_2(u,v)) \\ f_1(u,v) &= u - Au(u-\theta)(u-1) - \alpha v \\ f_2(u,v) &= \beta u + \gamma v. \end{aligned}$$

So moving along n in the above formulas is equivalent to studying the trajectories of this two-dimensional system. The computer approximates this solution to the PDE by writing this dynamical system and iterating.

What happens for a given k does not influence the behavior at $k+1$ except for the behavior of the diffusion term. The diffusion term forces neighbors to interact; this is called an **interaction term**. Without this term, this is an independent dynamical system. Note that the diffusion term is very small. Therefore, for the first approximation of what happens, we can drop it. We'll see what happens without interaction, and add the interaction later. With the interaction term, this is called a **coupled map lattice**.

So now, well just study this two-dimensional map and forget the little interactions.

Remark. There are several parameters here. There are physical or biological explanations for these parameters, and we want to have $\theta \approx \frac{1}{2}$. In addition, $\gamma = 1 - d\Delta < 1$, and α, β are very small (or even extremely small). A is called a **leading parameter**, which means that this is a parameter that we will vary. We want to see how the dynamics change when we change A. This is the only parameter that we will change. This is responsible for concentrations of molecules of the axon, controlling the strength of the signal.

What happens in the discrete model may not be the behavior of the PDE. When A is small, the solutions of the discrete are close to the solution of PDE. The only way to prove this is through computations and empirical evidence. As A gets bigger, the solutions of the discrete system do not produce solutions of the PDE system. Why should we study them?

This is an idea that first came in the work of Kaneko, a great Japanese scientist, in around 1983. He published a paper asking why the discrete system is any worse as a phenomenological model as the PDE; it could be just as good. They compared each model against real experiment. It turns out that for sufficiently large A and for certain other values of parameters, the discrete system describes real neurons better than the PDE. (This depends on the type of neuron.) So the discrete system has its own legitimacy outside of its relationship to the PDE, and in any case, it is mathematically interesting. As mathematicians (unlike real scientists), we like the system and ignore everything else; we are not responsible for interpretation of results.

9.3. **Studying the Discrete System.** We leave the world of neurons and PDEs, and focus on the discrete system. This two-dimensional system is called the **local map** – the behavior at each local site.

The system is

$$f(u, v) = (u - Au(u - \theta)(u - 1) - \alpha v, \beta u + \gamma v)$$

with $\theta \approx \frac{1}{2}$, $0 < \gamma < 1$, $\alpha, \beta \approx 0$.

First, we study the fixed points of the map, where $f(u, v) = (u, v)$. To do this, we should solve

$$\begin{aligned} u &= u - Au(u - \theta)(u - 1) - \alpha v \\ v &= \beta u + \gamma v. \end{aligned}$$

The three fixed points are then

$$u_0 = 0, \quad v_0 = 0$$

$$u_{1,2} = \frac{1}{2}\left(\theta + 1 \pm \sqrt{(\theta-1)^2 - \frac{4\alpha\beta}{A(1-\gamma)}}\right), \quad v_{1,2} = \frac{\beta u_{1,2}}{1-\gamma}.$$

Of course, if the quantity inside the radical is negative, $(0, 0)$ is the only fixed point. So:

$$0 < A < A_0 = \frac{4\alpha\beta}{(1-\gamma)(1-\theta)^2} \implies (0,\,0) \text{ is the only fixed point}$$

$$A > A_0 \implies \text{there are 3 fixed points}$$

We are now interested in the stability of these fixed points: attracting, repelling, or hyperbolic? To do this, we study the Jacobian matrix

$$Df(x) = \begin{pmatrix} \frac{\partial f_1(u,v)}{\partial u} & \frac{\partial f_1(u,v)}{\partial v} \\ \frac{\partial f_2(u,v)}{\partial u} & \frac{\partial f_2(u,v)}{\partial v}. \end{pmatrix}$$

We compute this matrix at a fixed point, and we look at the eigenvalues. The eigenvalues tell us which type of stability we have. In this particular case, we see that this matrix is

$$Df(x) = \begin{pmatrix} 1 - A\theta + 2A(1+\theta)u - 3Au^2 & -\alpha \\ \beta & \gamma \end{pmatrix}.$$

Substituting $u = 0$, we obtain

$$Df(0) = \begin{pmatrix} 1 - A\theta & -\alpha \\ \beta & \gamma \end{pmatrix},$$

and we can work out the eigenvalues. We are insufficiently courageous to do this for the other eigenvalues. We try a better idea: Using the fact that α, β are very small; three orders less than the other numbers. So the Jacobian matrix is approximately

$$Df(x) \approx \begin{pmatrix} 1 - A\theta + 2A(1+\theta)u - 3Au^2 & 0 \\ 0 & \gamma \end{pmatrix}.$$

We want to be mathematically rigorous. First, we study this case. As long as u is not too big, everything depends continuously on α and β, so dropping those terms won't change the type of behavior of the eigenvalues, and we can judge the type of behavior without α and β; that's a rigorous mathematical statement.

By this simplification, we obtain

$$Df(0) \approx \begin{pmatrix} 1 - A\theta & 0 \\ 0 & \gamma \end{pmatrix}$$

so that γ is always an eigenvalue. So there is always a contracting direction. Hence, we have an attracting point or a hyperbolic point; it can't be a repellor. If $|1 - A\theta| < 1$, we have an attracting point; if $|1 - A\theta| > 1$, we have a hyperbolic point.

Before A_0, the behavior is very simple; everything contracts to 0. After A_0, we must consider the other fixed points. At $P_1 = (1, 0)$ we obtain that

$$Df(P_1) \approx \begin{pmatrix} 1 + A\theta - A\theta^2 & 0 \\ 0 & \gamma \end{pmatrix},$$

which is always hyperbolic. At P_2,

$$Df(P_2) \approx \begin{pmatrix} 1 + A\theta - A & 0 \\ 0 & \gamma \end{pmatrix},$$

so if $0 < A < A_1' \approx \frac{2}{1-\theta}$ then P_2 is attracting. If $A > A_1'$ then P_2 is hyperbolic.

The final outcome is that there are several possibilities:

(1) Everything contracts to the origin.
(2) 0 and P_2 are attracting while P_1 is hyperbolic. (Draw a phase portrait.) There are no periodic points and everything is simple.
(3) All three points are hyperbolic. The problem is that it seems like there should be something attracting. In fact, there are attracting periodic orbits of period 2.

In this last case, our system is too complicated. We study a simpler example and come back to this later.

9.4. **A Great Example.** We give one of the greatest examples of mathematics. There is more about this than any other dynamical system. Consider

$$f(x) = x^2 + c.$$

What if we choose c to be really big? By drawing a picture, it is clear that $f^n(x) \to \infty$ for every x. Every trajectory goes to $+\infty$.

This isn't very interesting. We consider a smaller value of c. When $c = \frac{1}{4}$, $f(x) = x^2 + \frac{1}{4}$ is tangent to $y = x$, so the point of tangency is a fixed point. Points after the fixed point go to ∞, while some points in an interval before the fixed point converge to the fixed point. Draw a picture! Any sudden change in behavior is called a **bifurcation**. Then $c = \frac{1}{4}$ is the first bifurcation.

When c is a little bit less than $\frac{1}{4}$, then the graphs intersect twice and we have two fixed points. We want to consider stability of these fixed points P_1 and P_2. Then

$$f'(P_i) = 1 \mp \sqrt{1 - 4c}.$$

So the lower fixed point attracts and the upper fixed point repels. We have an attracting interval and everything else goes to ∞.

Do we have any periodic orbits of period 2? We need to solve $f^2(x) = x$; we must solve

$$(x^2 + c)^2 + c = x \implies x^4 + 2cx^2 - x + c^2 = 0.$$

Factoring out the two fixed points, we get two periodic points

$$q_{1,2} = -\frac{1}{2} \mp \sqrt{-\frac{3}{4} - c}.$$

So if $-\frac{3}{4} < c < \frac{1}{4}$, there are no periodic orbits of period 2. But if $c < -\frac{3}{4}$, we suddenly get two more periodic orbits of period 2. What happens there? This is when both fixed points become repelling. There must be an attractor between two repellors, so we actually have an attracting orbit of period 2.

As c decreases further, we get another bifurcation at $c = -\frac{5}{4}$. This is when our periodic orbit becomes repelling. We therefore must have another attracting periodic orbit, and we actually get two periodic orbits of period 4.

This process goes on, and new periodic orbits are born while old periodic orbits become unstable. This is a fantastic picture – so much complexity for such a simple map. There are infinitely bifurcation points:

$$c_1 = \frac{1}{4}, \quad c_2 = -\frac{3}{4}, \quad c_3 = -\frac{5}{4}, \quad \to \quad c_\infty.$$

The behavior at each bifurcation points is different. The first bifurcation is called a **tangent (or saddle-node) bifurcation**. The second bifurcation is called a **period doubling bifurcation** because we double the period of the periodic points. This yields a cascade of bifurcations in an infinite process.

At this point, we discuss behavior of bifurcations without proof. The bifurcation points converge to a number c_∞, where the process stops. At any point before that, there are finitely many periodic orbits – this is a **Morse-Smale system**. Every trajectory must converge to an attracting fixed point or periodic orbit. This constitutes fairly simple dynamics. Each particular trajectory can be complicated, but the overall picture isn't too bad.

At this point, there are two questions. What happens after c_∞, and are there periodic orbits of other periods?

We have a jewel of one-dimensional dynamics called Sharkovski's Theorem from around 1960. He published it in an Ukrainian journal and it went unnoticed. A couple years later, Li and Yorke wrote another paper where they proved part of Sharkovski's Theorem. The two results were independent due to the Iron Curtain; there were no connections between the two sides. At some point many years later, this theorem became known in the West, and produced a revolution.

Li-Yorke's paper has a pretentious title: "Period 3 Implies Chaos".

Theorem 9.1 (Li-Yorke)**.** *Consider a one-dimensional dynamical system* $x \mapsto f(x)$*,* $x \in \mathbb{R}^1$*. Assume that* f *has a periodic orbit of period 3. Then for every* $n \geq 3$*,* f *has a periodic orbit of period* n*.*

Theorem 9.2 (Sharkovski)**.** *We list all natural numbers in the following way:*

$$\begin{array}{ccccccccc} 3, & 5, & 7, & 9, & \cdots & & & & \\ 2\cdot 3, & 2\cdot 5, & 2\cdot 7, & 2\cdot 9 & \cdots & & & & \\ 2^2\cdot 3, & 2^2\cdot 5, & 2^2\cdot 7, & 2^2\cdot 9 & \cdots & & & & \\ \vdots & & & & & & & & \\ \cdots & 2^n, & 2^{n-1}, & \cdots, & 2^2, & 2, & 1. \end{array}$$

Write $n < m$ *if* n *is before* m *in this list. This is a new ordering of the natural numbers. If* f *has a periodic orbit of period* n *then for every* $m > n$*, it has a periodic orbit of period* m*.*

In particular, the only way to have finitely many periodic orbits, we have to stop at periods of 2^n.

Remark. The FitzHugh-Nagumo model is effectively a one dimensional system represented by a cubic equation. The period doubling also occurs in this case as A changes. So there's a point A_∞. We'll spend the next two lectures talking about two phenomena that we can see near A_∞, and we'll do some mathematical speculation.

10. Nonlinear Two Dimensional Systems

10.1. **Approximation by the Linear Case.** Suppose we are on the plane, and we have a hyperbolic point with expansion μ in the y-direction and contraction λ in the x-direction.[7] This matrix of the map is therefore

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

[7] 28 July 2010

Now consider any map; in a small neighborhood of a hyperbolic fixed point, we can write it as

$$f(x,y) = (f_1(x,y), f_2(x,y))$$
$$f(0,0) = (0,0)$$
$$f(x,y) = f(0,0) + \left.\begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}\right|_{(0,0)} (x,y) + \varepsilon g(x,y).$$

Therefore, we can write any map as $f(x,y) = A(x,y) + \varepsilon g(x,y)$, so it becomes

$$f(x,y) = (\lambda x + \varepsilon g_1(x,y), \mu y + \varepsilon g_2(x,y)).$$

Here, the g_i are of higher order, so we can write

$$g_i(0,0) = 0$$
$$\frac{\partial g_i}{\partial x}(0,0) = \frac{\partial g_i}{\partial y}(0,0) = 0.$$

In addition, we need to have that

$$|dg(x,y)| \leq c$$

for all (x,y) in a small neighborhood. This is the Taylor decomposition in two variables. Therefore, for nonlinear maps, we need a small correction in the non-linear term. There is no longer any guarantee that points on the x axis stay on the axis. This suggests a question: Is there a curve such that points on the curve stay on the curve? This is a substitute for the x-axis. Assume that we can. Then a similar argument could be made for the vertical axis – points on that curve would move toward 0 under the inverse map. Note that here the behavior is very similar to the linear case with some distortions. The difference between the linear and nonlinear case is that lines become curves and we only know things about behavior in a small neighborhood.

Definition 10.1. Such a stable curve is called a **stable curve (separatrix)** and the analogous unstable curve is called an **unstable curve (separatrix)**.

The existence of such curves is called the Grobman-Hartman Theorem. An earlier version of the theorem is called Hadamard-Perron Theorem.

10.2. **Attractors.** We will now consider some consequences of these results.

Suppose we take a point on the stable curve. If we move it forward, it moves toward the origin. What if we move it back? It will eventually leave the neighborhood where we can do our local analysis, and we lose control of it. If might wander back to our neighborhood eventually.

To see how these preimages behave, we can consider preimages of a separatrix, and we move it further and further back. It might behave strangely, but we can do this. We can do this around every hyperbolic point, in which case we get more separatrices. This quickly becomes complicated. Can two stable separatrices intersect? No: One point cannot go to two different points, so the stable separatrices never cross. However, stable and unstable separatrices can intersect.

Definition 10.2. The point of intersection of a stable and an unstable separatrix is called a **heteroclinic point**.

At heteroclinic points, the positive and negative iterates lie on different separatrices. This was first discovered by Poincare.

Recall that we have period doubling behavior. Our periodic points eventually become hyperbolic points, which generates more separatrices. So we get an unbelievably complicated structure with lots of packed separatrices. It's too complicated to even draw a picture.

Now, we let A grow until $A \to A_\infty$, where there are countably many periodic orbits, each of which is hyperbolic. As we pass A_∞ and let A get large, we can draw a rectangle R containing the three fixed points $0, P_1, P_2$ such that $\overline{f(R)} \subset R$.

Theorem 10.1. *For every $0 < \gamma < 1$ and every $\beta > 0$ small enough, there exists $A' > 0$ such that for each $0 < A < A'$, there exist $R \ni 0, P_1, P_2$ and $\overline{f(R)} \subset R$.*

This means that once a point gets inside of R, it cannot escape.

Definition 10.3. Such a rectangle R is called **trapping region**.

We can now look at the images of R:

$$R \supset f(R) \supset f^2(R) \supset f^3(R) \supset \cdots .$$

Consider

$$\Lambda = \bigcap_{n \geq 0} f^n(R).$$

Note that by definition, Λ is an **attractor**: it is an intersection of trapping regions. So Λ attracts all trajectories that sit in R; if $x \in R$ then $f^n(x) \to \Lambda$.

Theorem 10.2. *Λ is the largest invariant set. This means that $f(\Lambda) = \Lambda$.*

Proof. Consider

$$\begin{aligned} f(\Lambda) = f\left(\bigcap_{n\geq 0} f^n(R)\right) &= \bigcap_{n\geq 0} f(f^n(R)) \\ = \bigcap_{n\geq 0} f^{n+1}(R) &= \bigcap_{k\geq 1} f^k(R) = \Lambda. \end{aligned}$$

Therefore, Λ is an invariant set. Now suppose there were an invariant set $Z \supset \Lambda$. Then $f(Z) = Z$, and it must be Λ. □

In the case where there are three fixed points, the attractor contains those points. As structure gets more complicated as A increases, all fixed points and all periodic points must be in the attractor. As A passes A_∞, the structure of this attractor becomes much more complicated.

Theorem 10.3. *Take a hyperbolic fixed point $x \in \Lambda$. Let $\gamma^s(x)$ be the stable separatrix and let $\gamma^u(x)$ be the unstable separatrix. We claim that $\gamma^u(x) \subset \Lambda$.*

Proof. We prove this by contradiction. Suppose $y \in \gamma^u(x)$ but $y \notin \Lambda$. Say that $y \in R$ is in the trapping region; otherwise there is nothing to do. Consider its positive trajectory $f^n(y)$. This also lies in the trapping region; for every $n > 0$, $f^n(y) \in R$. Now consider the negative trajectory; then for every $n < 0$, $f^n(y)$ converges to x, and hence $f^n(y) \in R$. Therefore, $f^n(y) \in R$ for every n.

Let $Z = \Lambda \cup \{f^n(y)\} \supset \Lambda$, which is a contradiction to the previous theorem 10.2. □

So the attractor is bigger than simply the fixed points; it includes unstable separatrices.

Remark. We see from numerical calculations that the previous results of Sharkovski's Theorem 9.2 seem to hold for the FitzHugh-Nagumo map. However, Sharkovski's Theorem only holds in one dimension. There is no rigorous reason for this to be the case; it's an open problem. There is overwhelming numerical evidence, however.

What do the attractors look like? There are infinitely many periodic orbits, and is it very complicated. We can compute it numerically, and the pictures exist in Google. Note that this is only numerical evidence. There is no rigorous theory yet.

We want to analyze two things: How do we know that there exists a stable or unstable separatrix? What is the structure of the attractor? We postpone the proof of the existence of separatrices. It is important, but we don't have time.

10.3. **Construction of an Attractor.** We consider the construction of an attractor. We consider a model that is simpler so that we can actually prove things. The model is three-dimensional – it is more complicated in dimension, but in all other aspects it is simpler.

Consider a map $f : P = D^2 \times S^1 \to P$. Here, P is a solid torus. We make a coordinate system on P. Note that $D^2 = \{(x, y) : x^2 + y^2 = 1\}$, so coordinates on D^2 are given by x and y. Coordinates on S^1 are given by a parameter θ. Then let

$$f(x, y, \theta) = (\lambda x + r \cos\theta, \mu y + r \sin\theta, 2\theta),$$

where $0 < \lambda < \mu < 1$ and $r > 0$. Choose r appropriately so that this map actually takes P into P. We want to see what $f(P)$ looks like.

We can cut the solid torus P and unfold to obtain a solid cylinder. Multiply the x-coordinate by λ and the y-direction by μ, and double the length by $\theta \mapsto 2\theta$, so that it looks like an elliptical sausage.

We fold it back into the solid torus P. Note that because it's twice as big, we have to curl it around P twice. A picture would really help here. That's precisely $f(P)$. Here, we need to make sure that $f(P)$ should fit inside P, so r needs to be small enough. Since P is an attracting region, we have $\overline{f(P)} \subset P$.

Now consider the attractor

$$\Lambda = \bigcap f^n(P).$$

Consider the cross section of the solid torus by a plane. Locally, this looks like $D^2 \times [-\varepsilon, \varepsilon]$. We end up getting lots of pairs of nested ellipses; a picture is needed. I should figure out how to draw pictures.

Exercise 10.1. In one set of nested ellipses, the ellipses will be vertically positioned; in the other, the ellipse will be horizontally positioned. This is an effect of sin and cos is left as an exercise.

Clearly, this is a Cantor-like procedure, and we end up with a Cantor set in each cross section.

In our slice, we therefore have that the attractor is

$$\Lambda \cap D^2 \times [-\varepsilon, \varepsilon] = \mathcal{C} \times [-\varepsilon, \varepsilon].$$

We want to see the global picture. From a point in the Cantor set, we go around the torus and come back to a different point on the Cantor set, and repeat forever. This gets a countably many number of points. Repeating with other points, we obtain the whole attractor. It is called the **Smale-Williams Solenoid**, and it is a fractal-like set, kind of like $\mathcal{C} \times S^1$, but not really.

Consider any point $x \in \Lambda$. Then $df(x)$ is clearly a 3×3 triangular matrix:

$$df(x) = \begin{pmatrix} \lambda & 0 & -\sin\theta \\ 0 & \mu & \cos\theta \\ 0 & 0 & 2 \end{pmatrix}.$$

The eigenvalues are therefore $\lambda, \mu, 2$. Therefore it is hyperbolic: It has two directions of contraction and one direction of expansion; there is a two-dimensional stable set (a disc) and a one-dimensional unstable set that lies in the attractor. So every point here is hyperbolic. Therefore, Λ is a **hyperbolic attractor**. It is the first example of a hyperbolic example ever known in dynamical systems.

We are interested in the Hausdorff dimension of our attractor. We make things simpler, we consider the case where $\lambda = \mu$. In this case, we have circles and we can apply Moran's formula 5.5. From $\lambda^\alpha + \lambda^\alpha = 1$, this yields

$$\dim_H \mathcal{C} = \frac{\log 2}{-\log \lambda},$$

which is the entropy divided by the Lyapunov exponent. Now, the Hausdorff dimension of Λ should be

$$\dim_H \Lambda = \frac{\log 2}{-\log \lambda} + 1.$$

However, the Hausdorff dimension of the product is not in general the sum of the Hausdorff dimensions. However, we have a theorem

Theorem 10.4 (Besicovich). *$\dim_H A \times B = \dim_H A + \dim_H B$ if $\underline{\dim}_B A = \overline{\dim}_B B = \dim_H A$.*

We omit the technical proof. This gives us our formula for $\dim_H \Lambda$.

Now, what if $\lambda < \mu < \frac{1}{2}$? The result is still the same; μ does not affect the Hausdorff dimension. This is very difficult; it was an open problem for around twenty years, and it was solved around ten years ago.

10.4. **Hadamard-Perron Theorem.** We now prove the Hadamard-Perron Theorem about the existence of stable curves.[8]

Proof. Consider a map on the plane given by

$$f(x, y) = (\lambda x + \varepsilon g_1(x, y), \mu x + \varepsilon g_2(x, y)).$$

As we discussed earlier, we can write this as $f(\bar{x}) = A\bar{x} + \varepsilon\bar{g}(x)$, where

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} = df(0).$$

[8] 29 July 2010

We would like to find a stable curve $\gamma^s = (x, \phi(s))$ that is the graph of $y = \phi(x)$. We require that

$$\phi(0) = 0$$
$$\frac{d\phi}{dx}(0) = 0$$
$$\left|\frac{d\phi}{dx}(x)\right| \leq L.$$

The invariance of this curve under the map f is given by the condition that $f(x, \phi(x))$ is the point of the graph of f. We now have

$$f(x, \phi(x)) = (\lambda x + \varepsilon g_1(x, \phi(x)), \mu y + \varepsilon g_2(x, \phi(x))).$$

We want this to lie on the graph, which means that

$$\phi(\lambda x + \varepsilon g_1(x, \phi(x))) = \mu\phi(x) + \varepsilon g_2(x, \phi(x)).$$

This allows us to solve for ϕ, yielding

$$\phi(x) = \mu^{-1}\phi(\lambda x + \varepsilon g_1(x, \phi(x))) - \varepsilon\mu^{-1} g_2(x, \phi(x)).$$

This is a functional equation for the function ϕ. How do we solve this? We need to prove that this equation has a unique solution ϕ satisfying our conditions.

Hadamard proved the existence of solutions. His method was to pick any function ψ satisfying our conditions that

$$\psi(0) = 0$$
$$\frac{d\psi}{dx}(0) = 0$$
$$\left|\frac{d\psi}{dx}(x)\right| \leq L.$$

We generate a new function via

$$\psi(x) \mapsto \psi_1(x) = \mu^{-1}\psi(\lambda x + \varepsilon g_1(x, \psi(x))) - \varepsilon\mu^{-1} g_2(x, y).$$

We note that the same properties hold for the new function ψ_1. Iterating this transformation, we obtain a sequence of functions $\psi_1, \psi_2, \cdots$. These functions $\psi_n(x)$ converge to a function $\phi(x)$ as $n \to \infty$. This follows from the condition that

$$\phi(x) = \mu^{-1}\phi(\lambda x + \varepsilon g_1(x, \phi(x))) - \varepsilon\mu^{-1} g_2(x, \phi(x)).$$

For this to work, we need to verify that the conditions on the derivative do not change after iteration; we stay within the same class of functions. We also need to verify that the sequence converges, and we need to see what type of convergence we have. We won't do all of the details; instead, we'll only give an idea of the proof.

10.4.1. *Conditions on the Derivatives.* It is easy to see that $\psi_1(0) = 0$. To see the properties on its derivatives, we need to differentiate the expression for ψ_1 using the chain rule. This is a simple verification that is too messy to type. We also need to show that the derivative $\frac{d\psi_1}{dx}$ is bounded by the same constant as $\frac{d\psi}{dx}$. This is a bit trickier and requires manipulating some inequalities. The point is that we have estimates on each derivative by assumption, so we can use the triangle inequality and the fact that $0 < \lambda < 1 < \mu$. In order to get our bound, we have to shrink our neighborhood. If we choose a neighborhood sufficiently small, our ε will be sufficiently small and the bounds will work out.

10.4.2. *Convergence.* Now, we need to show that $\psi_n(x) \to \phi(x)$. By the Arzela-Ascoli theorem, we see that the $\psi_n(x)$ does converge to a continuous function. This is good, but not good enough; we don't know if $\phi(x)$ is differentiable. We need to go beyond the techniques of standard calculus. Consider the set of all differentiable fucntions on $[-r, r]$ satisfying our conditions

$$\begin{aligned} \psi(0) &= 0 \\ \frac{d\psi}{dx}(0) &= 0 \\ \left|\frac{d\psi}{dx}(x)\right| &\leq L. \end{aligned}$$

Call this space of functions $\mathcal{A}$.

Define a map $\mathcal{F} : \mathcal{A} \to \mathcal{A}$ satisfying $(\mathcal{F}\psi)(x) = \psi_1(x)$. As we just proved, this map moves $\mathcal{A}$ into itself. Hence, it is a well-defined operator.

We introduce a metric into this space $\mathcal{A}$ defined by

$$d(\psi_1(x), \psi_2(x)) = \max_{x\in[-r,r]} |\psi_1(x) - \psi_2(x)| + \max_{x\in[-r,r]} \left|\frac{d\psi_1(x)}{dx} - \frac{d\psi_2(x)}{dx}\right|.$$

It is easy to check that this satisfies the properties of a metric.

Lemma 10.5. *Now, take $\psi_1, \psi_2 \in \mathcal{A}$, then*

$$d(\mathcal{F}\psi_1, \mathcal{F}\psi_2) \leq \gamma d(\psi_1, \psi_2)$$

for some $0 < \gamma < 1$.

This lemma is the end of the story; we would have that

$$d(\psi_{n+1}, \psi_n) \leq \gamma d(\psi_n, \psi_{n-1}) \leq \cdots ,$$

so $\sum \psi_n$ is a geometric series and hence it converges to some function in the same space. Here, we used the fact that $\mathcal{A}$ is complete; this is the Arzela-Ascoli theorem. This operator is known as a contracting operator. We omit the proof of the lemma. □

10.5. Smale Horseshoe.

$$f(u,v) = (u - Au(u-\theta)(u-1) - \alpha v, \beta u + \gamma v)$$

We return to the FitzHugh-Nagumo map. We use the parameters $A \approx 7.0 - 7.5$, $\theta = \frac{1}{2}$, and α, β are small. We previously considered the attractor of this map. Given any point in the trapping region, we can get to the attractor by iterating f. If we start to change A in the range $0 < A < A'$, we obtain bifurcations. We can plot trajectories, and we'll eventually get some kind of periodic picture. It would help to draw something here. After we pass the bifurcations and get to the attractors, our plot produces completely chaotic behavior; we get forever random behavior because our attractor is chaotic. The cross section of the attractor is a Cantor-like set, and trajectories chaotically move around the attractor. If we cut the attractor into a number of pieces, the coding will be entirely chaotic. Here, nothing is proven and we only have numerical evidence and a conjecture. In the case of the Smale-Williams solenoid, this result has been proven.

There is something even more interesting. When $A' < A < A''$ (approximately $8 < A < 10$, though numerical evidence suggests this is true for all $A > 8$), we can repeat the argument that we did before. We no longer have an attracting region; some trajectories escape from the trapping region. What is the set of points that will not leave the rectangular trapping region? To answer this question, we consider a simpler model.

10.5.1. *A Simpler Model.* Consider a rectangle R, and fix two numbers $0 < \lambda < 1 < \mu$. We contract the rectangle in the vertical direction by λ and expand the rectangular in the horizontal direction by μ. The picture looks like a long thin rectangle. We fold this new rectangle in the shape of a horseshoe, and put it back into the original rectangle R. This folding is done such that the horseshoe is straight inside R and only bends outside of R. This is a map $f : R \to \mathbb{R}^2$. A picture would help here. Some points map out of R and can no longer be iterated. We are interested in the set of points that can be iterated twice; this set consists of two vertical strips of R; these strips are precisely $f^{-1}(R \cap f(R))$. Geometrically, the set of points that can be iterated three times consist of four strips – two substrips of each of the two previous strips. Continuing this process, we see that the set of points that can be iterated infinitely form the product of a Cantor set with an interval. This can be seen by a proof by picture. We know the lengths of the Cantor intervals because the ratio coefficient is μ and we can compute the Hausdorff dimension to be $\frac{\log 2}{\log \mu}$. This is a perfect world where we know everything.

Now, we instead consider the preimages instead of the forward images. It happens that our map is again a horseshoe, but placed vertically instead of horizontally. In this case, we again get a product of a Cantor set with an interval, and we obtain a set of Hausdorff dimension $\frac{\log 2}{-\log \lambda}$ because the ratio coefficient was λ.

The set of points which can be iterated infinitely forward and backward is the intersection of these two Cantor collection of vertical and horizontal strips. This is the biggest invariant set of points in the rectangle R.

Definition 10.4. This set

$$\Lambda = \bigcap_{-\infty<n<\infty} f^n(R).$$

is known as the **Smale horseshoe**.

Smale discovered this in 1959 or 1960. He discovered two important properties.

Firstly, every point in the set has a unique two-sided coding, so the symbolic space is

$$\Sigma_2 = \{\omega = (i_j), -\infty < j < \infty\}.$$

This gives a commutative diagram

$$\begin{array}{ccc} \Sigma_2 & \xrightarrow{\sigma} & \Sigma_2^+ \\ \downarrow h & & \downarrow h \\ C & \xrightarrow{f} & C \end{array}$$

where the σ is the full shift. Given any random sequence, there is a trajectory corresponding to this random sequence. This is a very chaotic system.

Secondly, there are 2^n periodic orbits of period n, and countably many periodic orbits overall; this is again by our topological conjugacy.

Also, every trajectory is hyperbolic, and our separatrices are precisely the horizontal and vertical lines in our horseshoe. This is highly chaotic.

We can choose two points x and y on a stable separatrix (vertical line) so that they lie on the same horseshoe shaped curve. The points on the stable separatrix move toward x, so the point y moves toward x. This means that the horseshoe curve folds around and crosses the stable separatrix countably many times. Reversing time, we have the symmetric situation with the horizontal line as stable. This deserves a picture.

Now, take a hyperbolic point. If the unstable separatrix folds back and intersects the stable separatrix again, their intersection is called a **homoclinic point**. Repeating this procedure, we again get a countable number of intersections. All of these points of parts of the invariant set of the horseshoe map. This is nontrivial, but it can be proven.

Remark. Homoclinic orbits were discovered by Poincare. He tried to solve the three-body problem. He discovered a homoclinic orbit while trying to solve this. This is one of his results that is now considered part of the foundation of our theory. He wrote that he could not even attempt to draw a picture – it is too complicated. This shows that the behavior of the three-body problem is very complicated. Numerically computing the equations of the three-body problem, we can get a graph. It starts with nice behavior, changes to chaotic behavior, and returns to nice periodic behavior, and this happens repeatedly. This is called **persistent intermittency** – intermittent chaotic behavior. This puzzled scientists for some time. The source of such behavior is a horseshoe. Since the horseshoe has zero measure, there is no chance to pick a horseshoe point as the initial point. Therefore, the trajectory cannot stay on the horseshoe, so it travels between horseshoes and hits nice periodic behavior in between. We can't see the horseshoes, but we can see their behavior.

So we've seen three types of behavior: There is the Morse-Smale system with finitely many chaotic points, so that we eventually get nice periodic behavior. There is persistent intermittent chaos that never dies out. The third case is permanent chaotic behavior. These are the only these three types of chaotic behavior, and fractal sets are in all of them. In these cases, the Lyapunov exponents and entropy characterize fractal sets and instability. This forms the intersection of fractal geometry and dynamical systems.

11. Homework 1

These notes were generated at the problem session after the end of the first half of the course. Much of the problem session was recorded verbatim, and as such, may be less coherent than the previous lecture notes.

Exercise 11.1. Consider the Hausdorff space (Σ_r^+, d_a) with $r \geq 2$. Compute its Hausdorff dimension $\alpha_r = \dim_H \Sigma_r^+$.

Solution.

$$\Sigma_r^+ = \{1, 2, \dots, r\}^{\mathbb{N}}$$

where the metric is

$$d_a(x, y) = \sum_{j \geq 1} \frac{|x_j - y_j|}{a^j}$$

where $a > 1$. We can ask what its Hausdorff dimension is. We have two questions: Guessing the right number and proving it. To find this dimension, we consider a one-parameter family of Hausdorff measures: Assume

$$0 < m(\Sigma_r^+, \alpha) < \infty.$$

We're looking for the critical value, where this is finite. Recall that

$$m(\Sigma_r^+, \alpha) = \lim_{\varepsilon \to 0} \inf_U \left\{ \sum_i (\operatorname{diam} U_i)^\alpha \right\}$$

We can turn the symbolic space into a Cantor set. Note that it is self similar. If a set Y is a rescaling of X by a factor of $\lambda > 0$, then the

$$m(Y, \alpha) = \lambda^\alpha m(X, \alpha).$$

This helps us compute the dimension. Here, our self similarity occurs in the symbolic space:

$$\Sigma_r^+ = [1] \cup [2] \cup \cdots \cup [r],$$

where

$$[i] = \{x \in \Sigma_R^+ \colon x_1 = i\}$$

are cylinders. Similar, we have

$$[a_1 a_2 \dots a_k] = \{x \in \Sigma_R^+ \colon x_i = a_i \forall i \leq k\}.$$

Now, consider the shift map $\sigma : [i] \to \Sigma_r^+$. This has the property that $d(\sigma(x_1), \sigma(x_2)) = \lambda d(x_1, x_2)$. Therefore,

$$m(\Sigma_r^+, \alpha) = \sum_{i=1}^r m([i], \alpha) = \left(\frac{1}{a}\right)^\alpha r m(\Sigma_r^+, \alpha).$$

due to the rescaling.

If $0 < m(S_r^+, \alpha) < \infty$, then

$$1 = \left(\frac{1}{a}\right)^\alpha r \implies \alpha = \frac{\log r}{\log a}.$$

The problem is that this is not a proof. We also don't know that the Hausdorff measure is nonzero.

Why this argument work? To show two things are equal, we check two inequalities. First, we'll show that

$$\dim_H \Sigma_r^+ \leq \frac{\log r}{\log a}.$$

To do this, we need to find a good cover. For every $\alpha > 0$, we need

$$\lim_{\varepsilon \to 0} \inf \sum_i (\operatorname{diam} U_i)^\alpha = 0$$

There is a nice ε-cover to use: We cover by cylinders by noting that

$$\operatorname{diam}([a_1, \ldots, a_k]) =: d_k$$

and

$$\sum_i (\operatorname{diam} U_i)^\alpha = r^k d_k^\alpha = r^k \left(\frac{1}{a}\right)^{k\alpha} d_0$$

which can be made arbitrarily small, and $m(\Sigma_r^+, \alpha) = 0$. This doesn't give a bound in the other direction, however. There are a number of ways to do this. We omit the technical details. The basic idea of a direct argument for a lower bound is that we can assume our cover is finite by compactness. We can then consider the smallest element in the cover.

We will do this by the mass distribution principle:

Theorem 11.1 (Mass Distribution Principle)**.** *Suppose μ is a measure such that $\mu(X) > 0$ and there exists $C > 0$ and $\alpha > 0$ such that $\mu(B(x,r)) \leq Cr^\alpha$.*

For every $x \in X$, $r > 0$, we have $\dim_H X \geq \alpha$.

Proof. We want to show that $m(X, \alpha) > 0$. This is by definition

$$m(X, \alpha) = \lim_{\varepsilon \to 0} \inf \left\{ \sum_i r_i^\alpha : \{(x_i, r_i)\} \right\} \geq \lim_{\varepsilon \to 0} \inf \left\{ \sum_i \frac{1}{C} \mu(B(x,r)) \right\}.$$

This isn't quite enough. We want to say that this is approximately $\frac{1}{C}\mu(X)$. The problem is that there are a lot of points that are covered multiple times and are overcounted. So therefore $m(X, \alpha) \geq \frac{1}{C}\mu(X)$. This gives a lower bound on Hausdorff dimension. This is a very useful tool – if you can find such measures. This is a good way of doing our problem. □

We want to consider Bernoulli measures. To do this, we first need the idea of Lebesgue measure. We omit the discussion of Lebesgue measure.

We present a general procedure to construct a measure. We start by defining X to be a set.

Definition 11.1. A collection of subsets $S \subset 2^X$ is a **semi-algebra** if

(1) $\emptyset, X \in S$
(2) $A, B \in S \implies A \cap B \in S$.
(3) $A, B \in S \implies A \setminus B = \cap_{i=1}^k C_i$ for $C_i \in S$ (where we take the disjoint union).

Now, let $l : S \to [0, \infty]$ be such that

(1) $l(\emptyset) = 0$
(2) $A, B \in S$ then $A \subset B \implies l(A) \leq l(B)$ (this follows from (3)).
(3) If $A_i \in S$ are disjoint and $A := \bigcup_i A_i \in S$ then $l(A) = \sum_i l(A_i)$.

This mimics the σ-additivity property.

Definition 11.2. We want to extend this to a function on a larger σ-algebra. There are two ways of doing this. One is to gradually make the collection of sets bigger. Suppose $\mathcal{A}$ is a semi-algebra satisfying a modified version of property (3):

$$A, B \in \mathcal{A} \implies A \setminus B \in \mathcal{A}.$$

Then A is an **algebra**. This property allows us to see that if $A, B \in \mathcal{A}$ then $A^c, B^c \in \mathcal{A}$. Then

$$A \cup B = ((A \cup B)^c)^c = (A^c \cup B^c)^c \in \mathcal{A}.$$

In addition, if $\mathcal{A}$ also satisfies $A_i \in \mathcal{A}$ then $\bigcup_i A_i \in \mathcal{A}$ then $\mathcal{A}$ is a **σ-algebra**.

To construct a measure, we start with a semi-algebra, we make an algebra, and then we produce a σ-algebra. One way to do this is to throw a powerful theorem at it. This is called Caratheodory's extension theorem. We omit the theorem; we are rushing through multiple lectures in measure theory. Just trust that it works.

Alternatively, once we have a set function defined on a semi-algebra, we can do a construction of an outer measure. Once we've done this, we can then restrict to a collection of measurable sets, and we end up at a σ-algebra of measurable sets.

Definition 11.3. A set is **measurable** if $m^*(E) = m^*(E \cap A) = m^*(E \cap A^c)$.

That is very briefly how we construct measures. This is what we want to use to build Bernoulli measures on symbolic space so that we can use the mass distribution principle.

We've done a lot of abstract nonsense. Why is this relevant? Apply the previous discussion in the case when $x = \Sigma_r^+$ and S is the set of cylinders. This means

$$S = \{[a_1, \ldots, a_k] : a_1, \ldots, a_k \in \{1, \ldots, r\}\}.$$

Recall the middle-thirds Cantor set construction is homeomorphic to Σ_2^+. At the k-th level of the Cantor set construction, there are 2^k intervals that correspond to the 2^k k-cylinders, so its natural to except cylinders to behave like intervals.

Define a set function $l : S \to [0, \infty]$. We'll want to find a measure so that each cylinder at the same level have the same weight. So the measure of any k-cylinder should be 2^{-k}. We check that this satisfies the additivity property. This gives us a measure called a **Bernoulli measure**. This is nice because we know exactly the measure of balls in symbolic space – they are just cylinders. (In general, for Bernoulli measure, we can split up the measure of the cylinders (intervals) differently.)

This means that $m(B(x, \varepsilon)) = 2^{-k(\varepsilon)}$ for some function $k(\varepsilon)$. Therefore, after some straightforward computations analogous to the generalized nonsense discussed earlier, we conclude that

$$\dim_H \Sigma_r^+ \geq \frac{\log r}{\log a},$$

and we're done.

We've finished one out of nine exercises and used three-fourths the time. We'll discuss some other exercises in much less detail.

Exercise 11.2. Construct an uncountable subset of $[0, 1]$ whose Hausdorff dimension is zero.

Solution. We do this using a Cantor-like construction. We let X_1 the middle third construction for the first step. We then break each interval into two intervals scaled down by a quarter. We continue with scaling of one-fifth, one-sixth, etc. Then let $X = \bigcap_n X_n$. This is uncountable because it is homeomorphic to symbolic space by coding.

If C_λ^r is the Cantor set we get from r intervals each scaling down by a factor λ, then Moran's theorem tells us that

$$\dim_H C_\lambda^r = \frac{\log r}{-\log \lambda}.$$

Note that this also answers problem 4.

Exercise 11.3. Compute the value of the Hausdorff function $m_h(C, \alpha)$ where C is the middle-third Cantor set and $a = \dim_H C$.

Solution. We know that

$$\dim_H \Sigma_2^+ = \frac{\log 2}{\log 3}.$$

Fix a number $\beta \in [0, 1]$, and let

$$X_\beta = \left\{ x \in S_2^+ : \lim_{n\to\infty} \frac{\#\text{ of 1's in } x_1, \dots x_n}{n} = \beta \right\}.$$

We can use the same argument for $m(X, \alpha) = r(\frac{1}{a})^\alpha m(X, \alpha)$ for each set, and we might see that they all have the same Hausdorff dimension. This is false and more complicated. This is a caution that this technique is dangerous.

Now, take the standard middle-third Cantor set and we have

$$m(C, \alpha) = \lim_{\varepsilon\to 0} \inf \sum_i |U_i|^\alpha.$$

We can take covers by basic intervals and compute that $m(C, \alpha) \leq 1$.

We're leaving a major portion of the solution to this problem as an exercise. The other direction is harder, as we've said many times before. Let U be an arbitrary ε-cover. If every $U_i \in U$ is a basic interval, we wave our hands and show that

$$\sum |U_i|^\alpha \geq 1.$$

The problem is that we may have covers that are not basic intervals.

Suppose $U_i \in U$ is *not* a basic interval. To deal with such an interval, we decompose it into three pieces. Let G be the largest gap in the interval, and let the left and right pieces be L and R for some strange reason. We want to relate $|U|^\alpha$ to $|L|^\alpha$ and $|R|^\alpha$.

By choice of G, note that $|G| \geq |L|$ and $|G| \geq |R|$. Therefore,

$$|G| \geq \frac{1}{2}(|L| + |R|)$$

This lets us see that

$$|U|^\alpha = (|L| + |R| + |G|)^\alpha \geq \left(\frac{3}{2}|L| + \frac{3}{2}|R|\right)^\alpha \geq \frac{1}{2}(3|L|^\alpha) + \frac{1}{2}(3|R|)^\alpha$$
$$= \frac{3^\alpha}{2}(|L|^\alpha + |R|^\alpha) = |L|^\alpha + |R|^\alpha.$$

We want to make G disappear; it doesn't stand for "good"; it stands for "gap". We needed to use a convexity argument:

$$\left(\frac{1}{2}(x+y)\right)^\alpha \geq \frac{1}{2}(x^\alpha + y^\alpha).$$

So we can break intervals into basic intervals, and hence $m(C, \alpha) = 1$.

Exercise 11.4. Given a number $\alpha \in [0, 1]$, construct a subset $Z \subset [0, 1]$ whose Hausdorff dimension is α.

Solution. Use the standard Moran construction to get any Hausdorff dimension $\alpha \in [0, 1]$. In fact, we can get $\alpha \in [0, \infty]$ by doing this in $\mathbb{R}^n$.

Remark. How do we get something with Hausdorff dimension ∞? Consider the Hilbert cube

$$[0, 1] \times [0, \frac{1}{2}] \times [0, \frac{1}{3}] \times \cdots$$

It has Hausdorff dimension ∞ because of the monotonicity property.

Exercise 11.5. Show that the lower and upper box dimensions of the set

$$A = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \cdots\right\}$$

are equal to $\frac{1}{2}$.

Solution. Clearly, the box dimension is in $[0, 1]$. Given $\varepsilon > 0$, let $N(\varepsilon)$ be the minimum cardinality of an ε-cover of A. Recall the definitions of the box dimensions. To find them, we need to estimate $N(\varepsilon)$. There's a part where A is pretty sparse and a part where A is pretty dense; they are separated by $\frac{1}{k(\varepsilon)}$. We have a region where each ball covers one point, and we have a region where we want to cover all of $[0, k(\varepsilon)]$. We can compute $k = k(\varepsilon)$ is such that

$$\frac{1}{k} - \frac{1}{k+1} < \varepsilon \leq \frac{1}{k-1} - \frac{1}{k}$$

so that

$$\frac{1}{k(k+1)} < \varepsilon \leq \frac{1}{k(k-1)}.$$

Then we need $k(\varepsilon)$ balls to cover the sparse region, and we need $\frac{1}{\varepsilon k(\varepsilon)}$ to cover the dense region.

$$k(\varepsilon) \leq N(\varepsilon) \leq k(\varepsilon) + \frac{1}{\varepsilon k(\varepsilon)}.$$

Observe that $\varepsilon \approx \frac{1}{k^2}$ so that $\frac{1}{\varepsilon k} \approx k$. Therefore,

$$k \leq N(\varepsilon) < 2k.$$

The box dimensions now follow from the definition.

Exercise 11.6. Compute the lower and upper box dimensions of the set

$$A = \left\{0, 1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \cdots\right\}.$$

Solution. Same approach as problem 5.

Exercise 11.7. Compute the Hausdorff dimension and lower and upper box dimensions of the subset $E \subset [0, 1]$ whose decimal expansions do not contain the digit 5.

Exercise 11.8. Compute the Hausdorff dimension of the "Cantor tartan", i.e. the set

$$E = \{(x, y) \in \mathbb{R}^2 \colon \text{either } x \in C \text{ or } y \in C\},$$

where C is the middle third Cantor set.

Exercise 11.9. Compute the Hausdorff dimension of the plane set given by

$$E = \{(x, y) \in \mathbb{R}^2 \colon x \in C \text{ and } 0 \leq y \leq x^2\},$$

where C is the middle third Cantor set.

12. Homework 2

These notes are from the second problem session at the end of the course.

Exercise 12.1. Estimate from below the Hausdorff dimension of the Markov geometric construction, which starts from three disjoint intervals $I_1, I_2, I_3 \subset [0,1]$. The ratio coefficients of the construction are $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ with $0 < \lambda < \frac{1}{3}$ and the transition matrix is

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Solution. We draw the directed graph for the transition matrix. Note that from the third interval it is only possible to go to the third interval, and from the second interval it is only possible to go to the first interval. From the first interval, there are two choices. Therefore, there are two subintervals I_{11}, I_{12} for I_1, and only subintervals I_{21} for I_2 and I_{33} for I_3.

Let X_n be the set of all intervals of the n-th level. We get a Cantor set

$$X = \bigcap_n X_n.$$

We are interested in estimating $\dim_H X$. As we saw in the previous problem set, we need to use either Bernoulli measures or Markov measures.

For a Markov measure, we need a probability vector $\vec{\pi}$ and a stochastic matrix P. The Markov measure of a cylinder is then

$$\mu([x_1, \ldots, x_n]) = \pi_{x_1} P_{x_1 x_2} \cdots P_{x_{n-1} x_n}.$$

It is important to check the properties of measures, but this is a simple verification. We will use this measure for our calculation.

It is natural to split the problem into two parts as a disjoint union; we either have strings of 1s and 2s or a string with only 3s. This second case doesn't change the dimension, so we've reduced the problem to one about the matrix

$$\tilde{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

We also need to have that $\mu(x) > 0$. It is known that if there is a zero transition probability but the corresponding entry in the stochastic matrix is nonzero, then $\mu(x) = 0$. Here, we assume that the matrix A is irreducible – for every i, j there exists n such that $(A^n)_{ij} \neq 0$; you can get to any vertex from any other vertex).

There's a more general statement that we need. It's best thought of as a generalization of the law of large numbers.

Proposition 12.1. *For μ-a.e. $x \in \Sigma_d^+$, we have*

$$\lim_{n\to\infty} \frac{1}{n}(\#\text{ of times that } i \text{ is followed by } j \text{ in } x_1, \dots, x_n) = \pi_i P_{ij}.$$

Note that this is the only reasonable thing the limit can be; π_i is the probability that we start with i and P_{ij} is the probability of going from i to j. This can be proved using the Birkhoff Ergodic Theorem, which we didn't discuss.

In our case, we have $\vec{\pi} = (p, 1-p)$ and the stochastic matrix is

$$P = \begin{pmatrix} 1-a & a \\ 1 & 0 \end{pmatrix}.$$

We need to have $\vec{\pi}P = \vec{\pi}$, so that

$$(p, 1-p)\begin{pmatrix} 1-a & a \\ 1 & 0 \end{pmatrix} = (p - ap + 1 - p, ap) = (p, 1-p),$$

so that

$$p = \frac{1}{1+a}.$$

This gives a one-parameter family μ_a of Markov measures. The support of the measure is X; $\mu_a(X) = 1$. This means that if we know something of the local dimension of the Markov measure, then we know a lot. We need to know the measure of a cylinder: What is $\mu_a([x_1, \cdots, x_n])$? More precisely, we want to know

$$h(\mu_a) = \lim_{n\to\infty} -\frac{1}{n}\log \mu_a([x_1, \cdots, x_n])$$

for μ-a.e. x. This would give that $\mu_a([x_1, \cdots, x_n]) \approx e^{-nh(\mu_a)}$.

For any Markov measure,

$$\mu([x_1, \dots, x_n]) = \pi_{x_1} P_{x_1 x_2} \cdots P_{x_{n-1}x_n}.$$

Then

$$\log \mu([x_1, \dots, x_n]) = \log \pi_{x_1} + \sum_{k=1}^{n-1} P_{x_k x_{k+1}}$$

$$= \log \pi_{x_i} + \sum_{i,j} (\log p_{ij})(\#\text{ of times } i \text{ is followed by } j \text{ in } x_1, \dots, x_n).$$

Now, for a.e. x, we have that

$$-\frac{1}{n}\log \mu([x_1, \dots, x_n]) \to \sum \pi_i P_{ij} \log P_{ij}.$$

This is the entropy of the general Markov measure.

For μ_a in our problem, we get

$$\begin{aligned} h(\mu_a) &= -\left[p\left((1-a)\log(1-a) + a\log a\right) + (1-p)\left(1\log 1 + 0\log 0\right)\right] \\ &= \frac{-((1-a)\log(1-a) + a\log a)}{1+a}. \end{aligned}$$

where we use the convention $0\log 0 = 0$.

We keep talking about entropy, but we want to estimate the dimension. We do this by using the very general relationship

$$\text{dimension} = \frac{\text{entropy}}{\text{Lyapunov exponent}},$$

so the local dimension of μ_a is

$$\frac{h(\mu_a)}{-\log\lambda}.$$

We need to maximize this function to get the best lower bound. This is an exercise in first-year calculus:

$$\frac{d}{da}h_{\mu_a} = 0 \implies a = \frac{3-\sqrt{5}}{2}.$$

For this value of a, we have

$$P = \begin{pmatrix} \frac{\sqrt{5}-1}{2} & \frac{3-\sqrt{5}}{2} \\ 1 & 0 \end{pmatrix}.$$

After a further calculation that we again omit, we have

$$h(\mu_a) = \log\left(\frac{1+\sqrt{5}}{2}\right).$$

We now have the definite statement that this is indeed a lower bound for the Markov construction;

$$\dim_H X \geq \frac{\log\left(\frac{1+\sqrt{5}}{2}\right)}{-\log\lambda}.$$

We conjecture that equality holds. Why is this true?

We've seen the Mass Distribution Principle 11.1. There is a similar statement for the upper bound:

Proposition 12.2. *If there exists a constant* C *such that* $\mu(B(x,r)) \geq C' r^\beta$ *for* every $x \in Z$, $\mu(Z) > 0$, *then* $\dim_H Z \leq \beta$.

The proof requires covering lemmas and is harder than the proof of 11.1 despite it being a statement about the upper bound; requiring everywhere instead of almost everywhere makes this harder to use.

For example, if we have a Bernoulli measure, then

$$\mu([x_1, \ldots, x_n]) = p^{\#1\text{s}} q^{\#2\text{s}}.$$

This works everywhere only where $p = q = \frac{1}{2}$, which is exactly where our lower bound was maximized; this is not a coincidence. So to show the upper bound, we just need to say that our limit exists everywhere and not just almost everywhere.

We'll do our best to present this construction without pulling too many rabbits out of a hat. Here's the general procedure for the construction of a **Parry measure.**

We want C, C' such that for a Markov measure $\mu = (\pi, P)$,

$$C' e^{-nh(\mu)} \leq \mu([x_1, \cdots, x_n]) \leq C e^{-nh(\mu)}.$$

We want the measure $\mu([x_1, \cdots, x_n]) = \pi_{x_1} P_{x_1 x_2} \cdots P_{x_{n-1} x_n}$ to not depend on the x_i. We can do this using a telescoping product.

The idea is to let

$$P_{ij} = \frac{a_{ij} v_j}{\chi v_i}$$

for some vector $\vec{v} \in \mathbb{R}^d$, for some χ that will soon be determined; the a_{ij} serves to set this to zero when the transition matrix has a zero entry. This is all happening in the context of one transition matrix; through all of this, a transition matrix A is fixed – we want to find $\dim_H \Sigma_A^+$.

There are two things that we need to check. Firstly, the rows of P must sum to 1. Indeed,

$$1 = \sum_j P_{ij} = \sum_j \frac{a_{ij}}{\chi} \frac{v_j}{v_i}.$$

Therefore,

$$\chi v_i = \sum_j a_{ij} v_j \implies \chi \vec{v} = A \vec{v}.$$

This is an eigenvector equation, so $\vec{v}$ is a right eigenvector of A and χ is the corresponding eigenvalue.

We also need a probability vector $\vec{\pi}$ such that it is a left eigenvector of the stochastic matrix: $\vec{\pi} P = \vec{\pi}$. This means that

$$\pi_j = \sum_i \pi_i P_{ij} = \sum_i \pi_i \frac{a_{ij} v_j}{\chi v_i},$$

which yields another eigenvalue equation

$$\chi \left(\frac{\pi_i}{v_j} \right) = \sum_i \left(\frac{\pi_i}{v_i} \right) a_{ij}.$$

Define $\vec{u}$ by $u_i = \frac{\pi_j}{v_j}$. Then $\chi \vec{u} = \vec{u} A$, so $\vec{u}$ is a left eigenvector of A.

All of this is entirely heuristic and somewhat roundabout; this won't be found in any textbook.

So now we're almost done. For a transition matrix A with $\vec{u}$ and $\vec{v}$ as left and right eigenvectors for the *largest* eigenvalue χ, define $\vec{\pi}$ be $\pi_i = u_i v_i$ (normalized) and $P_{ij} = \frac{a_{ij} v_j}{\chi v_i}$. This defines a Markov measure. Now, we have our telescoping product

$$\pi_{x_1} P_{x_1 x_2} \cdots P x_{n-1} x_n = \chi^{-n} (u_{x_1} v_{x_n} \chi),$$

which gives us the inequality that we claimed:

$$C' e^{-nh(\mu)} \leq \mu([x_1, \cdots, x_n]) \leq C e^{-nh(\mu)}.$$

Note that here we needed an eigenvector with all real and positive entries. This is true for irreducible transition matrices by the Perron-Frobenius Theorem; this is also why we need the largest eigenvalue. There is a nice geometric proof of this theorem.

That was an incredibly roundabout but fully complete answer to this problem. So we can now do this problem in thirty seconds by computing eigenvalues.

Exercise 12.2. Describe the behavior of all trajectories of the map $f : \mathbb{R} \to \mathbb{R}$ given by

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

for $c \geq 0$.

Solution. Note that this is really two one-dimensional maps that have met each other but not really mingled. One of them isn't even an interesting map. We only care about the x-coordinate with $g(x) = x^2 + c$. This map is very interesting, but for this parameter range, there isn't much going on.

We want to know about fixed points, periodic points, stability of each, and how trajectories behave. We quickly see that for large values of c, all trajectories go to ∞ (draw a picture). For some value of c, $x^2 + c$ intersects the line $y = x$, which is a fixed point. Anything starting to the left of the point horizontally across from it also goes to infinity, and in between we have a closed interval where trajectories converge to the fixed point. For $c < \frac{1}{4}$, there are two fixed points, one of which is stable. For $c > \frac{1}{4}$, there are no fixed points and all trajectories go to infinity.

Exercise 12.3. Find all periodic points of period 2 of the map

$$f(x, y) = (x^2 - (3/8)y^2, 2x)$$

and determine the type of stability (stable, unstable, or saddle).

Solution. This is a genuinely two-dimensional map. The brute force way is to compute the second iterate and get a polynomial of degree 4 and solve. We'll be more intelligent. We actually already know some roots of the equation

$$f(f(x, y)) = (x, y);$$

all fixed points satisfy it, for example. A little computation solving $f(x, y) = (x, y)$ shows that $(0, 0)$ and $(-2, -4)$ are the fixed points. This makes it easier to factor and solve the quartic polynomial for the period 2 orbits.

We now have that the period orbit contains the points

$$\left(\frac{2}{5}, -\frac{8}{5}\right), \quad \left(-\frac{4}{5}, \frac{4}{5}\right)$$

We are interested in stability, so we compute the Jacobian and find eigenvalues. The Jacobian is

$$Df = \begin{pmatrix} 2x & -\frac{3}{4}x \\ 2 & 0 \end{pmatrix}.$$

In particular, evaluating at the period 2 orbit and multiplying together, we have

$$\begin{pmatrix} \frac{4}{5} & \frac{6}{5} \\ 2 & 0 \end{pmatrix} \begin{pmatrix} -\frac{8}{5} & -\frac{3}{5} \\ 2 & 0 \end{pmatrix}.$$

The product is an exercise in linear algebra, and the eigenvalues of the product give the stability of the orbit. The same can be done at each of the fixed points.

Exercise 12.4. Describe the type of bifurcation at the given value of the parameter for the following maps:

(1) $f_\lambda(x) = x + x^2 + \lambda$ for $\lambda = 0$;
(2) $f_\lambda(x) = \lambda \sin x$ for $\lambda = -1$ or $\lambda = 1$;

Solution. For the first equation:

The easiest way is to look at the graphs. For $\lambda = 0$, this is the same picture as what we saw in problem 2. This is called a **tangent bifurcation** or a **saddle-node bifurcation**.

For the second equation: We look at this in a neighborhood of zero since otherwise x and $\lambda \sin x$ are far apart and therefore uninteresting. We plot the graph to see the behavior.

The fixed point whose behavior is changing is the one at 0. For $\lambda < 1$, 0 is a stable fixed point; for $\lambda = 1$, 0 is neutral, and for $\lambda > 1$, 0 is unstable and we have two more stable fixed points. This process is called a **pitchfork bifurcation**. This is because the bifurcation diagram looks like a pitchfork. (draw a picture!)

Near $\lambda = -1$, we again draw a picture. Again, the stability of the fixed point 0 changes but in a different way. For $\lambda > -1$, 0 is stable; for $\lambda = -1$, 0 is neutral; for $\lambda < -1$, 0 is unstable – but there are no new fixed points. Instead, we get a periodic orbit. Looking at the second iterate, we see a **period-doubling bifurcation**. By plotting the graph of $f^2(x) = \lambda \sin(\lambda \sin(x))$, we see a fixed point which corresponds to a period 2 orbit. This is an important phenonomenon, and it deserves a picture.

The period-doubling and tangent bifurcations are generic while the pitchfork bifurcation is not; it is "codimension two". Small changes to the pitchfork bifurcation lead to bifurcations of the other two types.

Exercise 12.5. Consider the map f given by

$$f(x) = \begin{cases} 3x & \text{if } 0 \le x \le \frac{1}{2} \\ 3 - 3x & \text{if } \frac{1}{2} < x \le 1. \end{cases}$$

Show that

(1) $\frac{3}{13}$ and $\frac{3}{28}$ lie on 3-cycles for f;
(2) the set

$$C = \{x \in [0,1] : f^n(x) \in [0,1] \text{ for all } n\}$$

is the middle-third Cantor set.

Solution. Plotting this, we see that this is a piecewise linear map. By plotting the graph of f^2, we see that this looks like the Cantor construction from the piecewise linear two-branched map considered earlier in the lectures. This is a very geometric argument that is fairly straightforward. One difference is that the conjugacy map with symbolic space turns out differently; instead of lexicographic ordering, we need a messier ordering.

Exercise 12.6. Consider the map f of the plane given by

$$f(x, y) = (x^2 + y^3 - 2a^2, x + y).$$

Find all fixed points of the map and determine the type of their stability depending on the value of the parameter a; sketch the phase portrait of the system.

Exercise 12.7. Show that the Smale horseshoe contains a point whose orbit is everywhere dense in the horseshoe.

Solution. This is invertible, so we code it by two-sided sequences $\Sigma_2 = \{0,1\}^{\mathbb{Z}}$. Then

$$\Lambda = \{x : f^n(x) \text{ defined for all } n \in \mathbb{Z}\}.$$

We can code this in terms of all two-sided sequences. Finding an orbit here is like finding a dense orbit in the two-sided shift space. This means that we need $\{\sigma^n y\}$ to enter every open cylinder in the shift space. Each cylidner looks like

$$[x_{-n} \cdots x_{-1} x_0 x_1 \cdots x_n].$$

To do this, list all words of length 1, and then all words of length 2, and continue forever. This defines a particular sequence such that every finite word occurs in it. That's what we need to have, since each of these finite words would appear after a certain shift. We're therefore done.

E-mail address: annoyingpi@gmail.com