Sixty Lectures of Dynamical Systems

Will J. Merry

ETH Zürich

Lecture notes for the course Dynamical Systems given in the academic year 2019-2020. The Problem Sheet solutions were written by Patricia Dietzsch and Alessio Pellegrini. Please let me know by email at merry@math.ethz.ch if you spot any typos!

LECTURE 1

The Definition of a Dynamical System

Dynamical systems studies the long-term behaviour of evolving systems.

What does this mean? As a motivating example, consider a family $\mathcal{P}$ of pigs. See Figure 1.1. Suppose at time $k = 0, 1, 2, \dots$ there are p_k pigs in $\mathcal{P}$. Let us assume that at time $k+1$ the number of pigs depends only on the number at time k. This means that the population can be described by a law of the form

$$p_{k+1} = f(p_k), \qquad k = 0, 1, 2, \dots$$

where f is an appropriate map. Inductively, we see that

$$p_k = f^k(p_0),$$

and hence the behaviour of the population of $\mathcal{P}$ is completely determined by initial number p_0 of pigs and the map f.

Figure 1.1: The family $\mathcal{P}$ of pigs[1].

Moving on from the pigs, let us suppose that the set of possible "states" of a "system" (be it physical/chemical/ontological/whatever...) are given by a set X,

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]Credit: Annie Spratt.

and that the evolution of said system is described by a map $f: X \to X$. This means that if the system at time k is given by $x_k \in X$, then at time $k+1$ the system is given by $x_{k+1} := f(x_k)$.

In order to get a more mathematically interesting theory, one typically endows the set X with extra structure. The precise structure one chooses depends on the nature of the system in question. Throughout these lectures, we will be concerned with four such flavours of dynamics.

(i) The set X carries a metric d, and hence is a **metric space**. The study of this type of dynamics is called **topological dynamics**, and it will form the first half of this course.

(ii) The set X carries a measure μ, and hence is a **measure space**. The study of this type of dynamics is called **measure-theoretic dynamics**, or sometimes **ergodic theory**, and it will form the second half of this course.

(iii) The set X carries a **differentiable structure**, and hence is **manifold**[2]. The study of this type of dynamics is called **differentiable dynamics**. We will come back to this in the first half of Dynamical Systems II.

(iv) The set X is the complex plane $\mathbb{C}$, or more generally a **Riemann surface**. The study of this type of dynamics is called **complex dynamics**, and this will form the second half of Dynamical Systems II.

Once one has endowed the set X with extra structure, it makes sense to require the map $f: X \to X$ to preserve this structure. Thus in option (i) we require f to be *continuous*, in option (ii) we require f to be *measure-preserving*, in option (iii) we require f to be *differentiable*, and finally in option (iv) we require f to be *holomorphic* (or *meromorphic*).

We begin in the topological setting. Here is our first formal definition.

DEFINITION 1.1. A **topological discrete dynamical system** consists of a metric space X and a continuous map $f: X \to X$.

Since this is rather a mouthful, whenever possible we will omit both the adjectives *topological* and *discrete* and simply call f a **dynamical system**.

The word "discrete" in Definition 1.1 refers to the fact that time takes integer values $k = 0, 1, 2\ldots$ We will discuss the "continuous" time version later this lecture.

REMARK 1.2. We could also work on an arbitrary topological space, rather than restrict to metric spaces. This is more general (since not every topological space is metrisable). However for us this extra level of abstraction is unnecessary—all of our interesting examples of topological dynamical systems occur on metric spaces. Therefore to minimise the topological prerequisites of this course we will work solely with metric spaces.

[2]If you are not familiar with manifolds, don't worry! There will be no manifolds in Dynamical Systems I, and in Dynamical Systems II we will cover the basics of manifold theory from scratch.

Sometimes we wish to be able to let time run backwards, and be able to compute the value x_{k-1} from x_k. For this to make sense the map f needs to be invertible and have a continuous inverse.

DEFINITION 1.3. A dynamical system $f\colon X \to X$ is **reversible** if f is a homeomorphism.

REMARK 1.4. By definition, all the dynamical systems we consider in this course are **deterministic**. This means that for any state x_k there is a *unique* state x_{k+1} that the system can take at time $k+1$ (namely, $x_{k+1} = f(x_k)$). Of course not all systems that occur in "real life" are deterministic. Perhaps the simplest such example is a coin toss. Take

$$X := \{\text{heads, tails}\},$$

and declare that at time k the system is in the state "heads" if the kth coin toss resulted in heads, and likewise for tails. This system is clearly not deterministic, as it is not possible to predict what state the system will be in at time $k+1$ given knowledge of what state the system was at time k.

This is an example of a **stochastic dynamical system**, where instead of there being a unique state x_{k+1} that a state x_k can attain, there is a probability distribution that governs the possible values of x_{k+1}. Stochastic dynamical systems are extremely important in real-world applications. However they are more complicated to handle mathematically, and we will not even touch upon them.

Going back to our previous example of the family $\mathcal{P}$ of pigs, let us suppose that the change in p_k is proportional to its size, that is, there exists a constant $c > -1$ such that

$$\frac{p_{k+1} - p_k}{p_k} = c.$$

Thus $p_{k+1} = (1+c)p_k$ and the corresponding dynamical system can be described by

$$f\colon [0,\infty) \to [0,\infty), \qquad f(x) := (1+c)x.$$

This model isn't very realistic though. Pigs like to eat lots of food, and unfortunately there isn't an unlimited amount of food. So let us assume that the surroundings limits the maximum size of $\mathcal{P}$, say by a number $N > 0$. One way to implement this would be to assume that

$$\frac{p_{k+1} - p_k}{p_k} = c(N - p_k).$$

Rescaling by

$$x_k := \frac{1}{N + \frac{1}{c}}\, p_k, \qquad a := cN + 1,$$

we see that

$$x_{k+1} = a x_k (1 - x_k).$$

This leads us to our first example of a dynamical system.

Example 1.5. The **logistic map** with value a is the dynamical system[3]

$$\lambda_a \colon \mathbb{R} \to \mathbb{R}, \qquad \lambda_a(x) := ax(1-x).$$

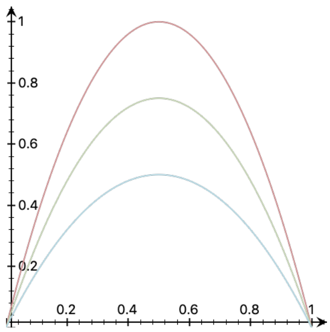

Figure 1.2: Plots of λ_2, λ_3, and λ_4.

Despite its simple form, for certain values of a the map λ_a can exhibit very complicated—or *chaotic*—dynamics. We will come back to this in Lecture 4 when we define chaos precisely. The name "logistic" comes from the fact that this is the discrete time version of the logistic population growth model discovered by the Belgium mathematician Verhulst, who was one of the pioneers of population dynamics.

Let us now give a few of the basic definitions.

Definition 1.6. Given a dynamical system $f \colon X \to X$, the **orbit** of a point $x \in X$, written as $\mathcal{O}_f(x)$ is the set

$$\mathcal{O}_f(x) := \{f^k(x) \mid k = 0, 1, 2, \dots\}.$$

If f is a reversible dynamical system then we can also look at the **negative orbit**

$$\mathcal{O}_f^-(x) := \mathcal{O}_{f^{-1}}(x) = \{f^{-k}(x) \mid k = 0, 1, 2, \dots\},$$

and the **total orbit**

$$\mathcal{O}_f^{\text{total}}(x) := \mathcal{O}_f(x) \cup \mathcal{O}_f^-(x) = \{f^k(x) \mid k \in \mathbb{Z}\}.$$

The simplest case is when the orbit consists of a single point.

Definition 1.7. A point x is called a **fixed point** of f if $f(x) = x$, so that $\mathcal{O}_f(x) = \{x\}$ We denote by $\mathsf{fix}(f)$ the set of fixed points of f.

Returning to the logistic map, one readily sees that provided $a \neq 0$, one has

$$\mathsf{fix}(\lambda_a) = \left\{0, 1 - \tfrac{1}{a}\right\}.$$

More generally, we can look at points that are fixed by some iterate of f.

[3]Note that we consider the domain and range of λ_a to be the entire real line $\mathbb{R}$. Mathematically this makes perfect sense, but as a model for population growth it loses its meaning for negative values of x. (One cannot speak of a family $\mathcal{P}$ of -5 pigs, for instance.)

Definition 1.8. A point $x \in X$ is called a **periodic point** of f if there exists $p \geq 1$ such that $f^p(x) = x$. We call such p a **period** of x. The minimal such p is called the **minimal period** of the periodic point x. We denote by $\mathsf{per}(f)$ the set of periodic points of f. Thus

$$\mathsf{per}(f) = \bigcup_{k=1}^{\infty} \mathsf{fix}(f^k).$$

Here is another example that we will come back to time and time again.

Example 1.9. Let $S^1 = \mathbb{R}/\mathbb{Z}$ denote the unit circle. Given $\theta \in [0, 1)$ the **circle rotation** with angle θ is the reversible dynamical system

$$\rho_\theta \colon S^1 \to S^1, \qquad \rho_\theta(x) := x + \theta \bmod 1.$$

The dynamics of ρ_θ depend on whether θ is a rational or an irrational number. We will see many examples of this of during the course. For instance, on Problem Sheet A you will prove:

Lemma 1.10. *The circle rotation ρ_θ satisfies:*

$$\mathsf{per}(\rho_\theta) = \begin{cases} S^1, & \theta \in \mathbb{Q}, \\ \emptyset, & \theta \notin \mathbb{Q}. \end{cases}$$

Moreover if $\theta \notin \mathbb{Q}$ then $\overline{\mathcal{O}_{\rho_\theta}(x)} = S^1$ for every $x \in S^1$.

Here is another example of a dynamical system on S^1. This one is not reversible.

Example 1.11. The **circle expansion** of order k is the dynamical system

$$e_k \colon S^1 \to S^1, \qquad e_k(x) := kx \bmod 1.$$

We usually call e_2 the **doubling map**.

Remark 1.12. Sometimes it is more convenient to view S^1 as the unit circle in $\mathbb{C}$. In this case the circle rotation ρ_θ is given by

$$\rho_\theta(z) = e^{2\pi i\theta} z,$$

and the circle expansion e_k is given by

$$e_k(z) = z^k.$$

Remark 1.13. Suppose $f \colon X \to X$ is a dynamical system. If $x \in \mathsf{per}(f)$ then clearly $\mathcal{O}_f(x)$ is a finite set, and when f is reversible the converse holds. However without the reversibility assumption it may fail, as the doubling map shows: $\mathcal{O}_{e_2}(\frac{1}{2})$ is the finite set $\{0, \frac{1}{2}\}$ but $\frac{1}{2}$ is not a periodic point of e_2.

An easy way to produce new dynamical systems from old ones is by restricting to invariant sets.

Definition 1.14. Suppose $f\colon X \to X$ is a dynamical system. A subset $A \subseteq X$ is called **f-invariant** or simply **invariant** if $f(A) \subseteq A$.

Thus if A is an invariant set then the restriction $f|_A\colon A \to A$ is also a dynamical system. Two examples of this are:

- If $x \in \mathsf{fix}(f)$ then $\{x\}$ is an invariant set for f. In this case the dynamical system $f|_{\{x\}}\colon \{x\} \to \{x\}$ is not very interesting.
- If $0 \le a \le 4$ then $[0,1]$ is an invariant set for $\lambda_a\colon \mathbb{R} \to \mathbb{R}$. (*Exercise:* Check this.)

Remark 1.15. Suppose $f\colon X \to X$ is a dynamical system and $A \subset X$ is an invariant set. Then by continuity of f, the closure $\overline{A}$ of A is also an invariant subset. Thus any invariant set can be replaced by a closed invariant set.

Definition 1.16. Suppose $f\colon X \to X$ is a dynamical system. An invariant subset A is called **completely invariant** if $f(A) = A$.

Remark 1.17. If f is reversible and A is completely invariant then $f|_A$ is also reversible. Moreover for a reversible system f a subset A is completely invariant if and only if it is both f-invariant and f^{-1}-invariant.

Every mathematical theory has its own notion of isomorphism.

Definition 1.18. Suppose $f\colon X \to X$ and $g\colon Y \to Y$ are dynamical systems. We say that f and g are **topologically conjugate**, or just **conjugate** if there exists a *homeomorphism* $H\colon X \to Y$ such that $g \circ H = H \circ f$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "H"'] & X \arrow[d, "H"] \\ Y \arrow[r, "g"'] & Y \end{tikzcd}$$

We write $f \simeq g$ to indicate that f and g are conjugate. If we want to explicitly mention the conjugacy H we write $f \simeq_H g$.

Lemma 1.19. *Conjugacy is an equivalence relation on the class of dynamical systems.*

Proof. Clearly $f \simeq f$ for any f. Moreover if $f \simeq_H g$ then $g \simeq_{H^{-1}} f$. Finally if $f \simeq_H g$ and $g \simeq_K h$ then $f \simeq_{K \circ H} h$. ∎

Here is another example of a dynamical system.

Example 1.20. The **tent map** is the dynamical system on $[0,1]$ defined by

$$\tau\colon [0,1] \to [0,1], \qquad \tau(x) := 2\min\{x, 1-x\}.$$

Equivalently,

$$\tau(x) := \begin{cases} 2x, & x \in [0, 1/2], \\ 2 - 2x, & x \in [1/2, 1], \end{cases}$$

which explains the name "tent". See Figure 1.3.

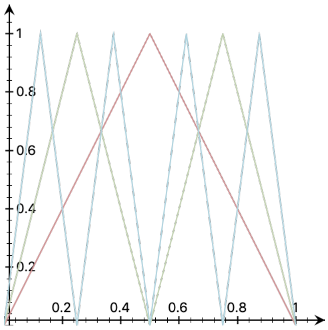

Figure 1.3: Plots of τ, τ^2, and τ^3.

Comparing the two red lines in Figures 1.2 and 1.3 suggests that the tent map might be conjugate to λ_4. This is indeed the case. On Problem Sheet A you will prove:

Lemma 1.21. *The tent map τ is conjugate to $\lambda_4|_{[0,1]}$.*

A weaker notion than conjugacy is the following:

Definition 1.22. Suppose $f\colon X \to X$ and $g\colon Y \to Y$ are dynamical systems. We say that g is a **(topological) factor** of f if there exists a continuous map $F\colon X \to Y$ with dense range such that $F \circ f = g \circ F$.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "F"'] & X \arrow[d, "F"] \\ Y \arrow[r, "g"'] & Y \end{tikzcd}$$

One calls F a **semiconjugacy**.

Remark 1.23. If X is compact and $F\colon X \to Y$ is a continuous map with dense range then F is necessarily surjective.

The simplest example of a factor comes from products:

Example 1.24. Suppose $f\colon X \to X$ and $g\colon Y \to Y$ are dynamical systems. Then the **product** dynamical system is

$$f \times g\colon X \times Y \to X \times Y, \qquad (x, y) \mapsto (f(x), g(y)).$$

This dynamical system has both f and g as factors, where the semiconjugacy is the projection onto X and Y respectively.

In all the examples so far, "time" has taken discrete values $k = 0, 1, 2, \dots$. In many physical systems of interest however, it is desirable to let time be continuous. It is quite easy to adapt the definitions to suit this case, as we now explain. In order to keep the discussion concise, we jump immediately to the topological category, and focus only on the reversible case.

Definition 1.25. Let X be a metric space. A **topological flow** (or just **flow**) on X is a map $\Phi\colon \mathbb{R} \times X \to X$ which is continuous with respect to the product topology such that

$$\Phi(0, x) = x, \qquad \forall\, x \in X, \tag{1.1}$$

and

$$\Phi(s, \Phi(t, x)) = \Phi(s + t, x), \qquad \forall\, x \in X,\ s, t \in \mathbb{R}. \tag{1.2}$$

Associated to a flow Φ is a family of maps $\varphi_t\colon X \to X$ for $t \in \mathbb{R}$ defined by

$$\varphi_t(x) := \Phi(t, x).$$

From this point of view the two properties (1.1) and (1.2) are rather more natural:

$$\varphi_0 = \mathrm{id}, \qquad \varphi_s \circ \varphi_t = \varphi_{s+t}. \tag{1.3}$$

This shows that the maps φ_t are necessarily homeomorphisms, since taking $s = -t$ in (1.3) tells us that $\varphi_t^{-1} = \varphi_{-t}$.

A flow on X can therefore be thought of as describing the evolution of a deterministic system where time is a continuous parameter. The condition $\varphi_0 = \mathrm{id}$ tells us that at time $t = 0$ the system is at rest.

Remark 1.26. Flows are always reversible (by definition). There is an analogous notion of a **semiflow**, which is a continuous map $\Phi\colon [0, \infty) \times X \to X$ satisfying (1.1) and (1.2), and this plays the role of a "non-reversible" flow. Nevertheless, for us all the continuous-time dynamical systems we will have cause to study are reversible, and so we will work exclusively with flows.

Convention: We will typically use capital Greek letters Φ, Ψ to denote flows. Moreover we will without comment always use the corresponding lowercase Greek letter to denote the associated family of maps (thus φ_t corresponds to Φ and ψ_t corresponds to Ψ, etc).

Before stating the next result, let us introduce a convention that will hold throughout the entire course: *anything marked with a (♣) is non-examinable.* There are various reasons for marking something with a (♣):

- it is only tangentially related to the course,
- it is rather technical or difficult,
- it is just a sketch,
- it requires more background knowledge (eg. differential geometry, functional analysis, etc) than the rest of the course assumes.

This following statement allows us construct many examples of flows.

PROPOSITION 1.27. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz continuous map satisfying $F(0) = 0$. Then for any $x \in \mathbb{R}^n$, the initial value problem*

$$\begin{cases} u'(t) = F(u(t)), \\ u(0) = x, \end{cases} \tag{1.4}$$

has a unique solution $u_x(t)$ which is defined for all $t \in \mathbb{R}$. Moreover the map

$$\Phi(t, x) \coloneqq u_x(t) \tag{1.5}$$

is a flow on $\mathbb{R}^n$.

The proof is not particularly hard, but since it involves ideas better suited to a class in analysis, it is non-examinable.

(♣) *Proof.* It follows from standard ODE theory that the initial value problem (1.4) has a unique solution $u_x(t)$ which is defined for t in some interval containing the origin. Moreover $(t, x) \mapsto u_x(t)$ depends continuously on both t and x.

Fix now a point $x \in \mathbb{R}^n$. Let us show that u_x is defined for all $t \in \mathbb{R}$. Since F is Lipschitz, there exists a constant $C > 0$ such that $\|F(y) - F(z)\| \le C\|y - z\|$ for all $y, z \in \mathbb{R}^n$. Then since

$$u_x(t) = x + \int_0^t F(u_x(s))\, ds, \tag{1.6}$$

we obtain

$$\|u_x(t)\| \le \|x\| + C \left| \int_0^t \|u_x(s)\|\, ds \right|,$$

and hence[4]

$$\|u_x(t)\| \le \|x\| e^{C|t|}.$$

From this it follows that u_x is defined for all $t \in \mathbb{R}$.

Next, fix $s \in \mathbb{R}$ and consider the map $v\colon \mathbb{R} \to \mathbb{R}^n$ defined by $v(t) \coloneqq u_x(s + t)$. Then $v(0) = u_x(s)$ and

$$v'(t) = u_x'(s + t) = F(u_x(s + t)) = F(v(t)).$$

Thus v is also a solution on the initial value problem (1.4) with initial condition $v(0) = u_x(s)$. Thus by the uniqueness of solutions with prescribed initial conditions it follows that

$$v(t) = u_{u_x(s)}(t).$$

This tells us that the map Φ from (1.5) satisfies

$$\Phi(s + t, x) = \Phi(s, \Phi(t, x)),$$

and hence is a flow. This completes the proof. ∎

[4] This is an application of the Gronwall Lemma.

Example 1.28. Apply Proposition 1.27 with F equal to the logistic map λ_a from Example 1.5. The resulting flow is the continuous-time version of the logistic dynamical system. Taking $a = 1$ and restricting to $[0, 1]$ for simplicity, one can solve the ordinary differential equation

$$u'(t) = u(t)(1 - u(t))$$

by separating variables and integrating by partial fractions to see that the flow $\varphi_t\colon [0, 1] \to [0, 1]$ is given by

$$\varphi_t(x) = \frac{xe^t}{1 + x(e^t - 1)}.$$

Many of the basic definitions are formally identical for flows—one just substitutes "for all $k \in \mathbb{Z}$" with "for all $t \in \mathbb{R}$". In order to avoid duplicating material unnecessarily, in later lectures we will often leave you to fill in the details. However since today is the first lecture, we will be friendly and write everything out.

Definition 1.29. Given a flow Φ on X, the **orbit** $\mathcal{O}_\Phi(x)$, the **negative orbit** $\mathcal{O}^-_\Phi(x)$, and the **total orbit** $\mathcal{O}^{\text{total}}_\Phi(x)$ are defined as you expect:

$$\mathcal{O}_\Phi(x) \coloneqq \{\varphi_t(x) \mid t \geq 0\}, \qquad \mathcal{O}^-_\Phi(x) \coloneqq \{\varphi_t(x) \mid t \leq 0\},$$

and

$$\mathcal{O}^{\text{total}}_\Phi(x) \coloneqq \mathcal{O}_\Phi(x) \cup \mathcal{O}^-_\Phi(x) = \{\varphi_t(x) \mid t \in \mathbb{R}\}.$$

Definition 1.30. For a flow Φ we define

$$\mathsf{fix}(\Phi) \coloneqq \{x \in X \mid \varphi_t(x) = x \text{ for all } t \in \mathbb{R}\} = \bigcap_{t>0} \mathsf{fix}(\varphi_t)$$

Similarly we say a point $x \in X$ is **periodic** if there exists $T > 0$ such that $\varphi_T(x) = x$, and we call such T a **period** of x. (Note it is important that we require T to be strictly positive!) The infimum of such T is said to be the **minimal period**[5] of the periodic point x, and we set

$$\mathsf{per}(\Phi) \coloneqq \{x \in X \mid x \text{ is a periodic point of } \Phi\}.$$

Thus

$$\mathsf{per}(\Phi) = \bigcup_{t>0} \mathsf{fix}(\varphi_t).$$

Next, the notion of invariant sets is defined similarly:

Definition 1.31. Let Φ be a flow on X. A subset $A \subseteq X$ is **invariant** if $\varphi_t(A) \subseteq A$ for all $t \geq 0$, and **completely invariant** if $\varphi_t(A) = A$ for all $t \in \mathbb{R}$. If A is completely invariant then $\Phi|_{\mathbb{R}\times A}$ is a flow on A.

The notion of conjugacy in the continuous case is slightly more subtle though.

[5] This implies that the minimal period of a fixed point is 0. For any periodic point which is not a fixed point, the minimal period is strictly positive (and the infimum is a minimum).

Definition 1.32. Let Φ be a flow on X and Ψ be a flow on Y. We say that Φ and Ψ are **topologically conjugate**, or just **conjugate** if there exists a homeomorphism $H\colon X \to Y$ such that $\psi_t \circ H = H \circ \varphi_t$ for all $t \in \mathbb{R}$:

$$\begin{tikzcd} X \arrow[r, "\varphi_t"] \arrow[d, "H"'] & X \arrow[d, "H"] \\ Y \arrow[r, "\psi_t"'] & Y \end{tikzcd}$$

Similarly we say that Ψ is a **(topological) factor** of Φ if there exists a continuous map $F\colon X \to Y$ with dense range such that $F \circ \varphi_t = \psi_t \circ F$ for all $t \in \mathbb{R}$.

Actually for flows this notion of conjugacy is often too restrictive to be useful (i.e. there exist flows that we would like to consider to be "isomorphic" yet they are not conjugate). We therefore conclude this lecture by introducing a weaker notion—*orbit equivalence*—which is typically more useful.

Definition 1.33. Let Φ and Ψ be two flows on the same space X. We say that Ψ is a **time change** of Φ if for every $x \in X$ both the orbit and the negative orbit agree,

$$\mathcal{O}_\Phi(x) = \mathcal{O}_\Psi(x) \qquad \text{and} \qquad \mathcal{O}^-_\Phi(x) = \mathcal{O}^-_\Psi(x), \qquad \forall\, x \in X. \tag{1.7}$$

Remark 1.34. Equation (1.7) implies that the total orbits are also preserved:

$$\mathcal{O}^{\text{total}}_\Phi(x) = \mathcal{O}^{\text{total}}_\Psi(x), \qquad \forall\, x \in X. \tag{1.8}$$

However (1.7) is stronger than (1.8). Indeed, if we set $\psi_t \coloneqq \varphi_{-t}$ then (1.8) is satisfied but (apart from in trivial cases) (1.7) is not. Thus time changes also require the "direction" of time to be preserved.

The next lemma clarifies the nature of time changes.

Lemma 1.35. *If Ψ is a time change of Φ then $\mathsf{fix}(\Psi) = \mathsf{fix}(\Phi)$. Moreover we can write*

$$\psi_t(x) = \varphi_{\alpha(t,x)}(x),$$

where $\alpha\colon \mathbb{R} \times X \to \mathbb{R}$ is a map such that

(i) *$\alpha(\cdot, x)\colon \mathbb{R} \to \mathbb{R}$ is surjective and strictly increasing for each $x \in X$.*

(ii) *For all $s, t \in \mathbb{R}$ and $x \in X$ one has*

$$\alpha(s + t, x) = \alpha(t, x) + \alpha(s, \psi_t(x)). \tag{1.9}$$

Conversely, if such a map α exists then Ψ is a time change of Φ.

Proof. The existence of α satisfying (i) is clear. Equation (1.9) is just (1.3) applied to ψ_t. ∎

Definition 1.36. Let Φ be a flow on X and Ψ be a flow on Y. We say that Φ and Ψ are **orbit equivalent** if Ψ is conjugate to a time change of Φ (in the sense of Definition 1.32). Similarly we say that Ψ is an **orbit factor** of Φ if Ψ is a factor of a time change of Φ.

Remark 1.37. The relationship of being orbit equivalent is another equivalence relation on the set of flows. It is clear that if Φ and Ψ are conjugate then they are also orbit equivalent, but the converse is not true. (*Exercise:* Find an example of this.)

We will not study flows much in this course. This is because most of the interesting ideas are already contained in the discrete setting, and as the discussion above shows, flows are often more complicated to handle. Nevertheless, flows will crop up now and again (mainly in the Problem Sheets), and we encourage the interested reader to correctly restate (and reprove) all the results from the course for flows.

LECTURE 2

Transitivity and Minimality

In the previous lecture we looked at the case where the orbit $\mathcal{O}_f(x)$ of a point x was as small as possible (i.e. fixed points and periodic points). Now we investigate the other extreme, when the orbit is as *large* as possible. Since an orbit $\mathcal{O}_f(x)$ is certainly at most countable (and most interesting metric spaces are not countable), it doesn't make sense to investigate points whose orbit is the entire space. Therefore we look at the next best thing: points $x \in X$ whose orbit $\mathcal{O}_f(x)$ is dense in X.

The following definition is the most important one of this lecture. As we will see below, under mild hypotheses on the metric space this definition guarantees (many) dense orbits.

DEFINITION 2.1. A dynamical system $f\colon X \to X$ is called **topologically transitive**, or just **transitive**, if for any pair U, V of non-empty open subsets of X, there exists some $k \geq 0$ such that $f^k(U) \cap V \neq \emptyset$.

REMARK 2.2. The transitivity condition is stronger than it appears at first sight. Indeed, as you will prove on Problem Sheet A, if $f\colon X \to X$ is transitive then for any pair U, V of non-empty open subsets there are actually *infinitely* many distinct $k \geq 0$ such that $f^k(U) \cap V \neq \emptyset$. This remarkable "recurrence" feature of transitive maps is the main reason why their dynamics are interesting.

A first easy lemma about transitive systems is:

LEMMA 2.3. *A transitive dynamical system has dense range.*

Proof. Let $f\colon X \to X$ be a dynamical system whose range is not dense. This means that there exists an open set V such that $f(X) \cap V = \emptyset$. After possibly shrinking V, there exists[1] an open set U such that $U \cap V = \emptyset$. Since $f^k(U) \subseteq f^k(X) \subseteq f(X)$ for all $k \geq 1$ we also have $f^k(U) \cap V = \emptyset$ for all $k \geq 0$, and hence f is not transitive. ■

We will shortly investigate which of the examples from the last lecture are transitive, but before doing so let us give several equivalent reformulations.

PROPOSITION 2.4. *Let $f\colon X \to X$ denote a dynamical system. Then following four conditions are equivalent.*

(i) *f is transitive.*

(ii) *X cannot be written as a disjoint union $X = A \cup B$ where both A and B have non-empty interior and A is f-invariant.*

(iii) *For any non-empty open subset $U \subseteq X$, the set $\bigcup_{k=0}^{\infty} f^k(U)$ is dense in X.*

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]The existence of such an open set U is clear, since X is a metric space (and thus Hausdorff).

(iv) *For any non-empty open subset $U \subseteq X$, the set $\bigcup_{k=0}^{\infty} f^{-k}(U)$ is dense in X.*

Proof. We first prove that (i) implies (ii). Suppose that $X = A \cup B$ with $A \cap B = \emptyset$, and suppose that $f(A) \subseteq A$. Then A° (the interior of A) and B° are open sets such that $f^k(A^\circ) \cap B^\circ = \emptyset$ for all $k \geq 0$. By (i) at least one of A° and B° must be empty.

Now let us prove that (ii) implies (iii). For this, suppose U is a non-empty open subset, and set $A := \bigcup_{k=0}^{\infty} f^k(U)$ and set $B := X \setminus A$. Then clearly A is f-invariant and has non-empty interior (since A contains the open set U). Thus by (ii), B must have empty interior, which is means that A is dense in X.

Next, (i) and (iii) are obviously equivalent, and to see that (iii) and (iv) are equivalent, note that $f^k(U) \cap V$ is non-empty if and only if $U \cap f^{-k}(V)$ is non-empty. This completes the proof. ■

Corollary 2.5. *If f is reversible then f is transitive if and only if f^{-1} is.*

Proof. This is immediate from the equivalence of (iii) and (iv) of Proposition 2.4. ■

Armed with Proposition 2.4, we can prove:

Lemma 2.6. *The tent map τ is transitive.*

Proof. To see this, note that τ^k is the piecewise linear map with $\tau^k(2i/2^k) = 0$ for $i = 0, 1, \dots, 2^{k-1}$ and $\tau^k((2i-1)/2^k) = 1$ for $i = 1, 2, \dots 2^{k-1}$. Thus τ^k has 2^{k-1} "tents". See Figure 1.3 again. Now suppose $U \subseteq [0,1]$ is open and non-empty. Then U contains an interval I of the form $I = [i/2^k, (i+1)/2^k]$ for some k and some i. Then $\tau^k(I) = [0,1]$, and hence also $\tau^k(U) = [0,1]$. ■

On Problem Sheet A you will prove:

Lemma 2.7. *The circle rotation ρ_θ is transitive if and only if θ is irrational.*

As mentioned above, under reasonable assumptions on X, a transitive dynamical system has (many) dense orbits. Here are the details.

Notation. If X is a metric space then we write $B(x,r)$ for the *open* ball of radius r about $x \in X$, and $\overline{B}(x,r)$ for the *closed* ball. If it is important to specify the metric d then we write $B_d(x,r)$ and $\overline{B}_d(x,r)$.

Recall a metric space is **complete** if every Cauchy sequence converges. You are hopefully already familiar with the next foundational result, but in case not, we supply a proof anyway.

Theorem 2.8 (The Baire Category Theorem). *If X is a complete metric space and $\{U_k\}_{k\geq 1}$ is a collection of open dense subsets of X then $\bigcap_{k=1}^{\infty} U_k$ is also dense in X.*

This proof is non-examinable, since it belongs to a course on point-set topology.

(♣) *Proof.* Let V be a non-empty open subset. We need to show there exists a point $x \in V$ such that $x \in U_k$ for all k. Since U_1 is open and dense, there exists $x_1 \in V \cap U_1$ and $0 < r_1 < 1$ such that $B(x_1, r_1) \subseteq V \cap U_1$. Since each U_k is open and dense, we can continue recursively to find sequences x_k and $0 < r_k < 1/k$ such that $\overline{B}(x_k, r_k) \subseteq B(x_{k-1}, r_{k-1}) \cap U_k$. Since $x_n \in B(x_k, r_k)$ whenever $n > k$, the sequence (x_k) is a Cauchy sequence. Since X is complete, it converges to a point x which satisfies $x \in \overline{B}(x_k, r_k)$ for all k. Thus $x \in V \cap U_k$ for all k. This completes the proof. ■

A metric space is **separable** if there exists a countable dense subset. Here is our promised result:

PROPOSITION 2.9. *Let $f\colon X \to X$ be a transitive dynamical system on a separable complete metric space. Let*

$$\mathcal{D}(f) := \{x \in X \mid \mathcal{O}_f(x) \text{ is dense in } X\}.$$

Then $\mathcal{D}(f)$ is itself dense in X (and thus in particular, non-empty).

Proof. Since X is separable there exists a countable collection of open sets $\{U_k\}_{k\geq 1}$ which form a basis for the topology on X (take a countable dense set, and then take all open balls of rational radii about those points). A point $x \in X$ has a dense orbit if and only if for every k there exists a non-negative integer n such that $f^n(x) \in U_k$. This means that

$$\mathcal{D}(f) = \bigcap_{k=1}^{\infty} \bigcup_{n=0}^{\infty} f^{-n}(U_k).$$

Since f is transitive, each set $\bigcup_{n=0}^{\infty} f^{-n}(U_k)$ is itself dense (and open, since f is continuous), by the equivalence of parts (i) and (iv) of Proposition 2.4. Thus by the Baire Category Theorem 2.8, $\mathcal{D}(f)$ is a dense subset of X. In particular it is non-empty. This completes the proof. ∎

The converse to Proposition 2.9 is (almost) true. The only difference is that we require different hypotheses on X. Recall that an **isolated point** in a metric space X is a point $x \in X$ with the property that $\{x\}$ is an open subset of X. For example, if $X := \{0\} \cup [1, 2]$ (considered as a subset of the real line, with the induced topology), then there is precisely one isolated point, namely 0.

LEMMA 2.10. *Let $f\colon X \to X$ be a dynamical system on a metric space with no isolated points. If $x \in X$ has a dense orbit then so does $f^k(x)$ for any $k \geq 1$.*

Proof. Note that $\mathcal{O}_f(f^k(x))$ contains the set $\mathcal{O}_f(x) \setminus \{x, f(x), f^2(x), \dots f^{k-1}(x)\}$. In a metric space without isolated points, a dense set remains dense after removing finitely many points. Thus if $\mathcal{O}_f(x)$ is dense then so is $\mathcal{O}_f(f^k(x))$. ∎

COROLLARY 2.11. *Let $f\colon X \to X$ be a dynamical system on a metric space with no isolated points. If there exists a point $x \in X$ with dense orbit then f is transitive.*

Proof. Suppose x has a dense orbit and U and V are non-empty open sets. Then there exists $k \geq 0$ such that $f^k(x) \in U$. Since $f^k(x)$ has a dense orbit by Lemma 2.10, there exists $n \geq k$ such that $f^n(x) \in V$. Thus $f^{n-k}(U) \cap V \neq \emptyset$. This completes the proof. ∎

Let us now discuss a strengthening of the notion of transitivity.

DEFINITION 2.12. A dynamical system $f\colon X \to X$ is called **minimal** if for *every* point $x \in X$, the orbit $\mathcal{O}_f(x)$ is dense in X.

Lemma 1.10 shows that irrational rotations are minimal. The same argument as in the proof of Corollary 2.11 shows:

Corollary 2.13. *A minimal dynamical system is transitive.*

If a dynamical system is minimal then it cannot have fixed or periodic points (apart from trivial cases when the metric space is finite). Thus the tent map is an example of a dynamical system that is transitive but not minimal.

There is an analogous version of Proposition 2.4 for minimal dynamical systems, which goes as follows:

Proposition 2.14. *Let $f\colon X \to X$ be a dynamical system. The following are equivalent:*

(i) *f is minimal.*

(ii) *The only closed invariant sets of X are X itself and the empty set.*

(iii) *For any non-empty open subset $U \subseteq X$, one has $\bigcup_{k=0}^{\infty} f^{-k}(U) = X$.*

Proof. To see that (i) implies (ii), suppose that $A \subseteq X$ is a non-empty closed invariant set. Let $x \in A$. Then since A is invariant, $\mathcal{O}_f(x) \subseteq A$. Since A is closed we have $\overline{\mathcal{O}_f(x)} \subseteq A$. But since f is minimal, $\overline{\mathcal{O}_f(x)} = X$, and thus $A = X$.

To see that (ii) implies (iii), let $U \subseteq X$ be a non-empty open set. Then $A := X \setminus \bigcup_{k=0}^{\infty} f^{-k}(U)$ is closed and invariant. Since $A \neq X$, by (ii) we must have $A = \emptyset$.

Finally to see that (iii) implies (i), let $x \in X$ and let U be an arbitrary non-empty open subset. Then by (iii), $x \in f^{-k}(U)$ for some $k \geq 0$. Thus $f^k(x) \in U$, and hence $\mathcal{O}_f(x) \cap U \neq \emptyset$. Since U was arbitrary, $\mathcal{O}_f(x)$ is dense. This completes the proof. ■

In general minimality is a less useful condition than transitivity, since it is too restrictive. Nevertheless, sometimes the restriction of a dynamical system to an invariant set can be minimal, even if the entire dynamical system is not.

Definition 2.15. Let $f\colon X \to X$ be a dynamical system. A non-empty closed invariant set $A \subseteq X$ is called **minimal** if the restriction $f|_A\colon A \to A$ is minimal.

Before proving the next result, we recall Zorn's Lemma.

Axiom 2.16 (Zorn's Lemma). Let S be a partially ordered set. If every totally ordered subset of S has an upper bound, then S contains a maximal element.

Zorn's Lemma is equivalent to the Axiom of Choice (note: we used the Axiom of Choice in the proof of the Baire Category Theorem!). Now recall the following elementary bit of point-set topology.

Definition 2.17. Let X be a set and $\mathcal{A} := \{A_j \mid j \in J\}$ be a collection of subsets of X. We say that $\mathcal{A}$ has the **finite intersection property** if given any finite subset $J_0 \subseteq J$, one has $\bigcap_{j \in J_0} A_j \neq \emptyset$.

The finite intersection property can be used to characterise compactness:

Proposition 2.18. *A topological space X is compact if and only if every collection of closed sets having the finite intersection property has non-empty intersection.*

(♣) *Proof.* Let $\{A_j \mid j \in J\}$ be a collection of closed sets having the finite intersection property. Set $U_j := X \setminus A_j$. Then $\{U_j \mid j \in J\}$ is a collection of open sets with the property that for any finite set $J_0 \subseteq J$, the collection $\{U_j \mid j \in J_0\}$ does *not* cover X. Moreover if $\bigcap_{j \in J} A_j = \emptyset$ then $\{U_j \mid j \in J\}$ is an open cover of X with no finite subcover. This shows that the hypotheses of the proposition are merely a restatement of the definition of compactness. ∎

This allows us to prove that dynamical systems on compact metric spaces always have minimal sets.

Proposition 2.19. *Let $f \colon X \to X$ be a dynamical system on a compact metric space X. Then f has a minimal set*[2].

Proof. Let $\mathcal{F}$ denote the collection of all closed non-empty f-invariant sets. Then $\mathcal{F} \neq \emptyset$ since $X \in \mathcal{F}$. We define a partial ordering on $\mathcal{F}$ by saying that $A \preceq B$ if $B \subseteq A$ (i.e. *reverse* inclusion). Suppose $\mathcal{A} \subseteq \mathcal{F}$ is a totally ordered subset. We claim that $\mathcal{A}$ has an upper bound. Indeed, it is clear that $\mathcal{A}$ has the finite intersection property. Thus by Proposition 2.18, $\bigcap_{A \in \mathcal{A}} A \neq \emptyset$, and since we are using reverse inclusion, this is an upper bound for $\mathcal{A}$ under $\preceq$.

Thus by Zorn's Lemma 2.16, $\mathcal{F}$ has a maximal element. By the equivalence of (i) and (ii) in Proposition 2.14, this maximal element is a minimal set for f. This completes the proof. ∎

Finally, let us conclude by briefly mentioning how today's material translates to flows. There are a few important differences in the statements.

Definition 2.20. Let Φ be a flow on X. We say that Φ is **(topologically) transitive** if for any pair U, V of non-empty open subsets of X, there exists some $t \geq 0$ such that $\varphi_t(U) \cap V \neq \emptyset$.

The next result is also on Problem Sheet A. This result is *not* true for discrete dynamical systems.

Lemma 2.21. *Suppose Φ is a transitive flow on X. Then X is connected.*

The next result is proved in the same fashion as Proposition 2.9. The conclusion is slightly stronger since flows are always reversible by definition (compare Lemma 2.5).

Proposition 2.22. *Suppose X is complete and separable, and that Φ is a transitive flow on X. Let*

$$\mathcal{D}(\Phi) := \left\{ x \in X \mid \mathcal{O}_\Phi(x) \text{ and } \mathcal{O}_\Phi^-(x) \text{ are dense in } X \right\}.$$

Then $\mathcal{D}(\Phi)$ is dense in X.

We move onto the flow version of Corollary 2.11. This time we do not need to assume that X has no isolated points in the statement. This is because the orbits of flows are connected subsets.

[2]This is arguably the single most "abstract" result in the entire course. It is the only time we will explicitly appeal to Zorn's Lemma.

Proposition 2.23. *Suppose there exists a point $x \in X$ such that both $\mathcal{O}_\Phi(x)$ and $\mathcal{O}^-_\Phi(x)$ are dense in X. Then Φ is transitive.*

We leave it up to you to formulate the minimality property for flows, and prove the analogue of Proposition 2.19.

LECTURE 3

The Non-Wandering Set and Its Friends

In this lecture we will define an entire menagerie of invariant sets associated to a dynamical system $f\colon X \to X$. So far we have met two: the fixed points, and the periodic points, which satisfy

$$\mathsf{fix}(f) \subseteq \mathsf{per}(f).$$

DEFINITION 3.1. Let $f\colon X \to X$ denote a dynamical system, and let $x \in X$. The **ω-limit set** of f at x, written $\omega_f(x)$, is the set of all points $y \in X$ for which there exists a sequence $k_n \to \infty$ such that $f^{k_n}(x) \to y$.

Equivalently,

$$\omega_f(x) = \bigcap_{k=1}^{\infty} \overline{\bigcup_{n \geq k} f^n(x)} = \bigcap_{k=1}^{\infty} \overline{\mathcal{O}_f(f^k(x))}. \tag{3.1}$$

We have the following result.

PROPOSITION 3.2. *The set $\omega_f(x)$ is a closed invariant subset of X. If X is compact then $\omega_f(x)$ is non-empty and completely invariant.*

Proof. Suppose $y_k \in \omega_f(x)$ converges to y. We want to show that $y \in \omega_f(x)$. For each $j \geq 1$, choose k_j such that $d(y_{k_j}, y) < 2^{-j}$. Then choose numbers n_j such that $d(f^{n_j}(x), y_{k_j}) < 2^{-j}$, and such that $n_j < n_{j+1}$. Then by the triangle inequality, $d(f^{n_j}(x), y) < 2^{1-j}$, and hence $y \in \omega_f(x)$. This shows that $\omega_f(x)$ is closed. It is clear that $f(\omega_f(x)) \subseteq \omega_f(x)$.

Now suppose that X is compact. Then $\omega_f(x)$ is certainly non-empty. It remains to show complete invariance. Let $y \in \omega_f(x)$. Choose $k_n \to \infty$ such that $f^{k_n}(x) \to y$. Compactness tells us that, after passing to a subsequence if necessary, $f^{k_n - 1}(x)$ converges to some point $z \in X$. Thus $f^{k_n}(x) \to f(z)$, and hence $f(z) = y$. Since $z \in \omega_f(x)$ we thus $\omega_f(x) \subseteq f(\omega_f(x))$. This completes the proof. ■

COROLLARY 3.3. *Suppose A is a minimal set for f. Then $\omega_f(x) = A$ for all $x \in A$. Conversely if A is any non-empty compact subset of X with the property that $\omega_f(x) = A$ for all $x \in A$ then A is minimal.*

Proof. By assumption A is closed, invariant and non-empty. If $x \in A$ then

$$\begin{aligned}
\omega_f(x) &= \bigcap_{k=1}^{\infty} \overline{\mathcal{O}_f(f^k(x))} \\
&= \bigcap_{k=1}^{\infty} \overline{\mathcal{O}_{f|_A}(f^k(x))} \\
&= \bigcap_{k=1}^{\infty} A \\
&= A.
\end{aligned}$$

To prove the converse, the assumption implies in particular that A is invariant. Thus by Proposition 2.19 there exists a minimal set $B \subseteq A$ for $f|_A$. The argument above shows that $\omega_f(x) = B$ for all $x \in B$, whence $A = B$. This completes the proof. ∎

COROLLARY 3.4. *Let $f\colon X \to X$ be a dynamical system on a metric space without isolated points. Then a point x has dense orbit if and only if $\omega_f(x) = X$*

Proof. If $\omega_f(x) = X$ then the orbit of x is certainly dense (this doesn't require X to have no isolated points). Conversely, Lemma 2.10 and (3.1) shows that if x has dense orbit then $\omega_f(x) = X$. ∎

DEFINITION 3.5. Assume that f is reversible, and fix $x \in X$. The **α-limit set** of f at x, written $\alpha_f(x)$, is the ω-limit set for f^{-1} at x:

$$\alpha_f(x) := \omega_{f^{-1}}(x).$$

Thus $y \in \alpha_f(x)$ if and only there exists a sequence $k_n \to \infty$ such that $f^{-k_n}(x) \to y$.

$$\alpha_f(x) = \bigcap_{k=1}^{\infty} \overline{\bigcup_{n \geq k} f^{-n}(x)}.$$

A periodic point x has the property that $f^k(x) = x$ for infinitely many $k \geq 0$. Generalising this, we say a point x is *recurrent* if $f^k(x)$ is arbitrarily close to x for infinitely many $k \geq 0$. This can be expressed concisely as follows:

DEFINITION 3.6. Let $f\colon X \to X$ be a dynamical system. A point $x \in X$ is called a **recurrent** point if $x \in \omega_f(x)$. The set of recurrent points is denoted by $\mathsf{rec}(f)$.

Obviously one has $\mathsf{per}(f) \subseteq \mathsf{rec}(f)$.

LEMMA 3.7. *The set $\mathsf{rec}(f)$ is invariant.*

Proof. If $x \in \mathsf{rec}(f)$ then there exists $k_n \to \infty$ such that $f^{k_n}(x) \to x$. Thus also $f^{k_n}(f(x)) = f(f^{k_n}(x)) \to f(x)$ and hence $f(x) \in \mathsf{rec}(f)$. This completes the proof. ∎

REMARK 3.8. The set $\mathsf{fix}(f)$ is always closed in X (by continuity). In contrast, neither $\mathsf{per}(f)$ nor $\mathsf{rec}(f)$ are necessarily closed subsets of X.

Next we define the notion of a *non-wandering* point, which is a further generalisation of a recurrent point.

DEFINITION 3.9. A point $x \in X$ is called a **non-wandering** point if for any neighbourhood[1] U of x there exists $k \geq 1$ such that $f^k(U) \cap U \neq \emptyset$. The set of non-wandering points is denoted[2] by $\mathsf{nw}(f)$.

[1]We use the convention that in a topological space, a **neighbourhood** of a point is an *open* set containing that point.

[2]Many books use the notation $\Omega(f)$ to denote the non-wandering set. We prefer $\mathsf{nw}(f)$ as it is (a) more descriptive and (b) less likely to be confused with the ω-limit sets.

REMARK 3.10. If f is transitive then $\mathsf{nw}(f) = X$. This does not quite follow immediately from the definitions, since the definition of transitivity requires $k \geq 0$ and here we require $k \geq 1$. However by Problem A.5 (see also Remark 2.2) if f is transitive then for any non-empty open U there exist infinitely many $k \geq 0$ such that $f^k(U) \cap U \neq \emptyset$.

PROPOSITION 3.11. *Let $f\colon X \to X$ be a dynamical system.*

(i) *The non-wandering set is closed and invariant.*

(ii) *One has $\omega_f(x) \subseteq \mathsf{nw}(f)$ for all $x \in X$.*

(iii) *One has $\overline{\mathsf{rec}(f)} \subseteq \mathsf{nw}(f)$.*

(iv) *If f is reversible then $\mathsf{nw}(f) = \mathsf{nw}(f^{-1})$ and $\mathsf{nw}(f)$ is completely invariant.*

Proof. To see that $\mathsf{nw}(f)$ is closed, we show its complement is open. If $x \notin \mathsf{nw}(f)$ then there exists a neighbourhood U of x such that $f^k(U) \cap U = \emptyset$ for all $k \geq 1$, and hence all points $y \in U$ also do not belong to $\mathsf{nw}(f)$. Thus $X \setminus \mathsf{nw}(f)$ is open.

Now we show $\mathsf{nw}(f)$ is invariant. Let $x \in \mathsf{nw}(f)$, and let V denote a neighbourhood of $f(x)$. Then $U := f^{-1}(V)$ is a neighbourhood of x, and hence there exists some $k \geq 1$ such that $f^k(U) \cap U \neq \emptyset$. The image of this intersection under f is contained in $f^k(V) \cap V$, and hence the latter is non-empty. Thus $f(x) \in \mathsf{nw}(f)$. This proves (i).

Next, suppose $x \in X$ and $y \in \omega_f(x)$. Let U be a neighbourhood of y. We want to find some $k \geq 1$ such that $f^k(U) \cap U$ is non-empty. In other words, we want to find $z \in U$ and $k \geq 1$ such that $f^k(z) \in U$. Since $y \in \omega_f(x)$ there exists $k_n \to \infty$ such that $f^{k_n}(x) \to y$. Thus there exists $k_{n_0} < k_{n_1}$ such that both $f^{k_{n_0}}(x) \in U$ and $f^{k_{n_1}}(x) \in U$. Set $z = f^{k_{n_0}}(x) \in U$ and $k = k_{n_1} - k_{n_0}$. Then $f^k(z) \in U$ as required. This proves (ii).

By definition $\mathsf{rec}(f) \subseteq \bigcup_{x \in X} \omega_f(x)$, and thus by (ii) we have $\mathsf{rec}(f) \subseteq \mathsf{nw}(f)$. Since $\mathsf{nw}(f)$ is closed by (i), we also have $\overline{\mathsf{ref}(f)} \subseteq \mathsf{nw}(f)$. This proves (iii).

Finally, if f is reversible and $x \in \mathsf{nw}(f)$, then for every neighbourhood U of x there exists $k \geq 1$ such that $f^k(U) \cap U \neq \emptyset$. The f^{-k}-image of this intersection is contained in $U \cap f^{-k}(U)$, which is non-empty, and hence $x \in \mathsf{nw}(f^{-1})$. Thus $\mathsf{nw}(f) \subseteq \mathsf{nw}(f^{-1})$, and by symmetry, the two sets are equal. Thus

$$f^{-1}(\mathsf{nw}(f)) = f^{-1}(\mathsf{nw}(f^{-1})) \subseteq \mathsf{nw}(f^{-1}) = \mathsf{nw}(f),$$

so $\mathsf{nw}(f)$ is completely invariant. This proves (iv) and thus completes the proof. ■

REMARK 3.12. On Problem Sheet B sheet you will show that if $x \in \mathsf{nw}(f)$ then for any neighbourhood U of x there exist infinitely many $k \geq 1$ such that $f^k(U) \cap U \neq \emptyset$.

We now move onto the final invariant set. For this set the choice of metric d on X is important, so we include it in our notation.

DEFINITION 3.13. Let $f\colon X \to X$ be a dynamical system on metric space (X, d). A tuple $(y_1, \ldots, y_k)$ is called an **ε-chain** if

$$d(f(y_i), y_{i+1}) < \varepsilon, \qquad \forall\, 1 \leq i \leq k-1.$$

We say that x is **ε-pseudo-periodic** if there exists an ε-chain that starts and ends at x. Finally, we say that x is **chain recurrent** if x is ε-pseudo-periodic for all $\varepsilon > 0$. The set of all chain recurrent points is denoted by $\mathsf{cha}_d(f)$.

REMARK 3.14. Here is an informal way to think about chain recurrent points. Suppose our "measuring device" that we use to "observe" points in X is accurate only to the nearest ε. All measurements of real-world dynamical systems have this defect to some extent. Then as far as our measuring device is concerned, a periodic orbit is indistinguishable from an ε-pseudo-periodic orbit. Thus a chain recurrent point is a point which is indistinguishable from a periodic point for an arbitrarily precise measuring device.

The chain recurrent set is another closed invariant subset. It is the "largest" of the various sets we have defined so far.

PROPOSITION 3.15. *Let (X, d) be a metric space and $f\colon X \to X$ a dynamical system. Then the chain recurrent set $\mathsf{cha}_d(f)$ is a closed invariant subset which contains $\mathsf{nw}(f)$.*

Proof. We first prove that $\mathsf{cha}_d(f)$ is closed. Fix $x \in \overline{\mathsf{cha}_d(f)}$ and $\varepsilon > 0$. Since f is continuous at x, there exists $0 < \delta < \varepsilon$ such that

$$d(x, y) < \delta \qquad \Rightarrow \qquad d(f(x), f(y)) < \varepsilon. \tag{3.2}$$

Now choose $y \in \mathsf{cha}_d(f)$ such that $d(x, y) < \delta$, and choose an ε-chain $(y, z_1, \dots, z_k, y)$. Then we claim that $(x, z_1, \dots z_k, x)$ is a 2ε-chain. Indeed, we need only check the start and end points, and

$$\begin{aligned} d(f(x), z_1) &\le d(f(x), f(y)) + d(f(y), z_1) \\ &< \varepsilon + \varepsilon \\ &= 2\varepsilon, \end{aligned}$$

and similarly $d(f(z_k), x)) < 2\varepsilon$. Since ε was arbitrary, this shows that $x \in \mathsf{cha}_d(f)$ as required.

We now prove that $\mathsf{cha}_d(f)$ is invariant. Let $x \in \mathsf{cha}_d(f)$ and fix $\varepsilon > 0$. We will produce an ε-chain from $f(x)$ to itself. This time choose $0 < \gamma < \varepsilon$ such that

$$d(f(x), y) < \gamma \qquad \Rightarrow \qquad d(f^2(x), f(y)) < \varepsilon$$

(such γ exists as f is continuous at $f(x)$). Let $(x, y_1, \dots, y_k, x)$ denote a γ-chain. Then since $d(f(x), y_1) < \gamma$ we have $d(f(y_1), f^2(x)) < \varepsilon$ and thus

$$\begin{aligned} d(f^2(x), y_2) &\le d(f^2(x), f(y_1)) + d(f(y_1), y_2) \\ &< \varepsilon + \gamma \\ &< 2\varepsilon. \end{aligned}$$

Thus $(f(x), y_2, \dots, y_k, x, f(x))$ is a 2ε-chain. See Figure 3.1.

Since ε was arbitrary, this shows that $f(x) \in \mathsf{cha}_d(f)$, and since x was an arbitrary point of $\mathsf{cha}_d(f)$, this shows that $\mathsf{cha}_d(f)$ is invariant.

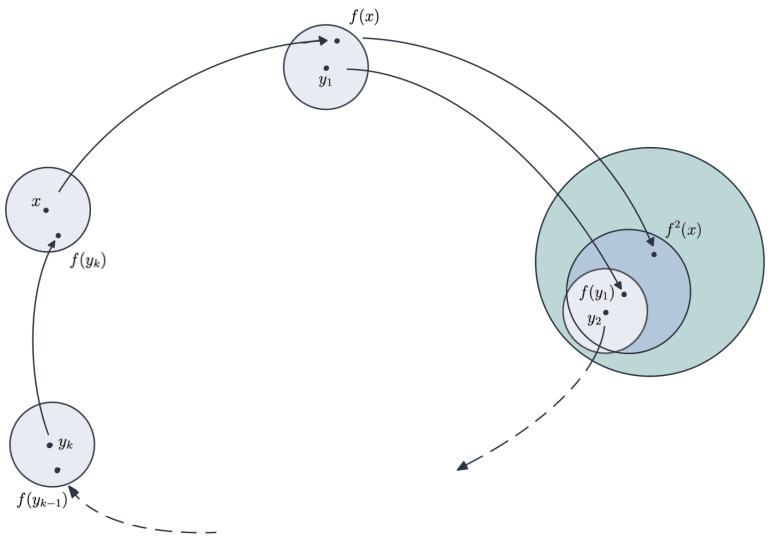

Figure 3.1: Proving $\mathsf{cha}_d(f)$ is invariant.

Finally let us prove $\mathsf{nw}(f) \subseteq \mathsf{cha}_d(f)$. Let $x \in \mathsf{nw}(f)$ and $\varepsilon > 0$. Take δ as in (3.2), and let U be a neighbourhood of x contained in the ball of radius δ about x. Since $x \in \mathsf{nw}(f)$, there exists $k \geq 1$ such that $f^k(U) \cap U \neq \emptyset$. If $k = 1$ then (x, x) is a 2ε-chain. If $k > 1$ then there exists $y \in U$ such that $f^k(y) \in U$. Then $(x, f(y), f^2(y), \ldots, f^{k-1}(y), x)$ is an ε-chain. Since ε was arbitrary, we have $x \in \mathsf{cha}_d(f)$ as required. ■

The next proposition tells us that on a compact metric space the chain recurrent set is actually independent of the choice of metric. The proof is on Problem Sheet B.

Proposition 3.16. *Let (X, d_1) be a compact metric space and $f \colon X \to X$ be a dynamical system. If d_2 is any other metric on X defining the same topology then*

$$\mathsf{cha}_{d_1}(f) = \mathsf{cha}_{d_2}(f).$$

Summary: Let $f \colon X \to X$ be a dynamical system on a metric space (X, d). Then we have invariant subsets

$$\mathsf{fix}(f) \subset \mathsf{per}(f) \subset \mathsf{rec}(f) \subset \mathsf{nw}(f) \subset \mathsf{cha}_d(f).$$

In general all of these inclusions can be strict. (*Exercise:* Find examples of this!)

We conclude this lecture by briefly discussing how these definitions work for flows. If Φ is a flow on X then the definitions of the ω-limit set $\omega_\Phi(x)$, the α-limit set $\alpha_\Phi(x)$, the recurrent set $\mathsf{rec}(\Phi)$, and the non-wandering set $\mathsf{nw}(\Phi)$ are all formally identical to the discrete case. One just replaces "k" with "t" where appropriate. Thus $\omega_\Phi(x)$ consists of all points $y \in X$ such that there exists a sequence $t_n \to \infty$ such that $\varphi_{t_n}(x) \to y$, and a point x belongs to $\mathsf{nw}(\Phi)$ if for every neighbourhood U of x there exists $t > 0$ such that $\varphi_t(U) \cap U \neq \emptyset$.

There is, however, a subtlety in the definition of the chain recurrent set of a flow, so we will go over the details here in more depth.

Definition 3.17. Let Φ be a flow on a metric space (X, d). Given $\varepsilon > 0$ and $T > 0$, a tuple $(y_1, \dots, y_k; t_1, \dots, t_{k-1})$ is called an (ε, T) **chain** if for each $1 \leq i \leq k-1$ one has

$$d(\varphi_{t_i}(y_i), y_{i+1}) < \varepsilon, \qquad \text{and} \qquad 0 \leq t_i \leq T.$$

We say that x is (ε, T)**-pseudo-periodic** if there exists an (ε, T)-chain that starts and ends at x. Finally, we say that x is **chain recurrent** if x is (ε, T)-pseudo-periodic for all $\varepsilon > 0$ and all $T > 0$. The set of all chain recurrent points is denoted by $\mathsf{cha}_d(\Phi)$.

It would appear at first glance that this definition is stronger than the corresponding definition for discrete dynamical systems, since we require chains to exist for all $T > 0$. Nevertheless, one has:

Theorem 3.18. *Let (X, d) be a compact metric space and let Φ be a flow on X. Let $f := \varphi_1$ so that $f\colon X \to X$ is a reversible dynamical system. Then*

$$\mathsf{cha}_d(\Phi) = \mathsf{cha}_d(f).$$

It is clear that $\mathsf{cha}_d(\Phi) \subseteq \mathsf{cha}_d(f)$ (as one can just take $T \equiv 1$). The converse is much harder and goes beyond the scope of this course[3].

[3] The ambitious reader is encouraged to try and prove it!

LECTURE 4

What Is Chaos?

Most people think of chaos as the so-called "butterfly effect", namely that a butterfly flapping its wings in Beijing can set off a cascading chain of atmospheric events that two weeks later leads to the formation of a catastrophic tornado that obliterates central Zürich.

More mathematically, this is *sensitive dependence on initial conditions*: in this case the dynamical system in question is the weather, and the small change (the butterfly) leads to a large change (the tornado) later down the road. A more pedestrian example of a chaotic dynamical system[1] that displays this "sensitive dependence" is the double pendulum. This is defined exactly as you'd guess: take a pendulum and then hang another pendulum on the end of it.

Unfortunately the mathematical definition of chaos is rather less glamorous than the popular science one. This is actually true of most things in life: adding ε's and δ's rarely make things exciting.

DEFINITION 4.1. Let (X, d) be a metric space. A dynamical system $f : X \to X$ has **sensitive dependence on initial conditions** if there exists a constant $\delta > 0$ such that for all $x \in X$ and all $\varepsilon > 0$, there exists a point $y \in X$ such that

$$d(x, y) < \varepsilon$$

but that x and y move far apart under sufficiently many applications of f, that is,

$$\text{there exists } k \geq 0 \qquad \text{such that} \qquad d(f^k(x), f^k(y)) > \delta.$$

The number δ is called a **sensitivity constant** for f.

REMARK 4.2. If X has isolated points then no dynamical system on X can have sensitive dependence on initial conditions. Indeed, if x is isolated then for sufficiently small ε the only point y satisfying $d(x, y) < \varepsilon$ is $y = x$.

REMARK 4.3. As we have defined it, the sensitivity constant δ is *not* unique (since if δ is a sensitivity constant then so is δ' for any $0 < \delta' < \delta$). This could be rectified by taking the supremum of all such δ's. However in practice this supremum is usually hard to compute, and since it is only important that *some* δ exists (rather than any precise value of δ), we will not do so.

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]Formally the double pendulum is actually flow on cotangent bundle of the torus $\mathbb{T}^2 := S^1 \times S^1$. To see this, note that the position of the pendulum is entirely specified by the two angles θ and ω that the two pendulums make with the vertical, and hence the motion may be described by recording these positions, together with their respective momenta. Newton's Second Law gives a system of ordinary differential equations for the motion. These equations cannot be solved analytically, but it is possible to use numerical methods. A MATLAB implementation of this can be found here.

On Problem Sheet B, you will show that both the tent map and the doubling map have sensitive dependence on initial conditions. Meanwhile circle rotations do not have sensitive dependence on initial conditions, as the following remark shows.

REMARK 4.4. Recall that a **contraction** of a metric space (X, d) is a continuous map $f \colon X \to X$ such that $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in X$. A dynamical system which is a contraction cannot have sensitive dependence on initial conditions. In particular, an **isometry** (i.e. a map such that $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$) cannot have sensitive dependence on initial conditions.

When the metric space is compact, the property of having sensitive dependence on initial conditions does not depend on the metric. (Compare this to Proposition 3.16).

LEMMA 4.5. *Let (X, d_1) be a compact metric space, and suppose $f \colon X \to X$ is a dynamical system which has sensitive dependence on initial conditions with respect to d_1. If d_2 is any other metric on X defining the same topology then f has sensitive dependence with respect to d_2 as well.*

Proof. Let

$$\eta(r) := \sup\{d_2(x, y) \mid d_1(x, y) \leq r\}.$$

and

$$\zeta(r) := \inf\{d_2(x, y) \mid d_1(x, y) \geq r\}.$$

By compactness, $\eta(r) \to 0$ as $r \to 0$ and $\zeta(r)$ is strictly positive for all $r > 0$. Suppose δ is a sensitivity constant for f with respect to d_1. We claim that $\zeta(\delta)$ is a sensitivity constant for f with respect to d_2. Since $\eta(r) \to 0$ as $r \to 0$, it suffices to show that for any $x \in X$ and $r > 0$ we can find $y \in X$ and $k \geq 0$ such that

$$d_2(x, y) \leq \eta(r), \qquad \text{and} \qquad d_2(f^k(x), f^k(y)) > \zeta(\delta). \tag{4.1}$$

Since f has sensitive dependence on initial conditions with respect to d_1 we find $y \in X$ and $k \geq 0$ such that $d_1(x, y) < r$ and $d_1(f^k(x), f^k(y)) > \delta$. It follows from the definition of η and ζ that this same y satisfies (4.1). This completes the proof. ∎

Lemma 4.5 is *not* true for non-compact metric spaces. An explicit example of this is on Problem Sheet B. This means that on non-compact metric spaces, sensitive dependence on initial conditions is a "bad" condition to study. Before stating this precisely, let us formalise the notion of "good" and "bad" properties.

DEFINITION 4.6. We say that a property $\mathcal{P}$ is a **(topological) dynamical invariant**[2] of a dynamical system if it is preserved under conjugacy, i.e., if f satisfies property $\mathcal{P}$ and g is conjugate to f then g also satisfies property $\mathcal{P}$.

There is also the stronger notion of an *inheritable* property.

DEFINITION 4.7. A property $\mathcal{P}$ is said to be **(topologically) inheritable** if it is preserved under passing to factors, i.e., if f satisfies property $\mathcal{P}$ and g is a factor of f then g also satisfies property $\mathcal{P}$.

[2]Do note confuse this with an invariant set. These are two different meanings of the word "invariant". (Don't blame me, I didn't come up with the terminology...)

Here is an example of an inheritable property:

LEMMA 4.8. *Transitivity is an inheritable property.*

Proof. Let $f: X \to X$ be a transitive dynamical system. Suppose $g: Y \to Y$ is a factor of f, and let $F: X \to Y$ denote a semiconjugacy. We show that g is also transitive. Thus suppose that U and V are non-empty open subsets of Y. Then $F^{-1}(U)$ and $F^{-1}(V)$ are non-empty open subsets of X, since F is continuous and has dense range. Thus there exists $x \in F^{-1}(U)$ and $k \geq 0$ such that $f^k(x) \in F^{-1}(V)$. Thus if $y := F(x)$ one has $y \in U$ and $g^k(y) = F(f^k(x)) \in V$. This completes the proof. ■

We can now give meaning to our claim that sensitive dependence on initial conditions is a "bad" property.

PROPOSITION 4.9. *Sensitive dependence on initial conditions is not a dynamical invariant. However it is a dynamical invariant when restricted to dynamical systems on compact spaces.*

Proof. To prove the first claim we need only exhibit a *single* example of a pair of conjugate dynamical systems with the property that one of them has sensitive dependence on initial conditions and the other does not. By Problem B.5 there exists a topological space[3] X with two metrics d_1 and d_2 that define the same topology, together with a dynamical system $f: X \to X$ such that f has sensitive dependence on initial conditions with respect to d_1 but not with respect to d_2. Since d_1 and d_2 define the same topology, the identity map $\mathrm{id}: X \to X$ is a homeomorphism from (X, d_1) to (X, d_2), and hence a conjugacy from f to itself:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\mathrm{id}"'] & X \arrow[d, "\mathrm{id}"] \\ X \arrow[r, "f"'] & X \end{tikzcd}$$

This proves the first claim. To prove the second claim, suppose X and Y are compact and $f: X \to X$ and $g: Y \to Y$ are conjugate dynamical systems on via a homeomorphism $H: X \to Y$.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "H"'] & X \arrow[d, "H"] \\ Y \arrow[r, "g"'] & Y \end{tikzcd}$$

Let d_X be an arbitrary metric defining the topology on X and let d_Y be an arbitrary metric defining the topology on Y. Assume f has sensitive dependence on initial conditions with respect to d_X. We must prove that g has sensitive dependence on

[3]In fact, $X = (0, \infty)$ with d_1 the standard metric works.

initial conditions with respect to d_Y. For this we argue in two steps. First, observe that

$$d'_X(x_1, x_2) := d_Y(H(x_1), H(x_2))$$

is another metric on X such that $H\colon (X, d'_X) \to (Y, d_Y)$ is an isometry. By Lemma 4.5, f has sensitive dependence on initial conditions with respect to d'. Suppose δ is a sensitivity constant for f with respect to d'_X. Then since H is an isometry it then follows easily that g has sensitive dependence on initial conditions with respect to d_Y, with the same sensitivity constant. Indeed, fix $y \in Y$ and $\varepsilon > 0$. Set $x := H^{-1}(y)$. Then there exists $z \in X$ and $k \geq 0$ such that $d'(x, z) < \varepsilon$ and $d'_X(f^k(x), f^k(z)) > \delta$. Set $w := H(z)$. Then

$$d_Y(y, z) = d'_X(x, z) < \varepsilon,$$

and

$$d_Y(g^k(y), g^k(w)) = d'_X(f^k(x), f^k(z)) > \delta.$$

This completes the proof. ■

We now move onto the mathematical definition of chaos. In contrast to the popular science definition, chaos has three ingredients, of which sensitive dependence on initial conditions is only one. To begin with we define chaos only on metric spaces without isolated points (we will shortly rectify this).

Definition 4.10 (Preliminary Version). A dynamical system $f\colon X \to X$ on a metric space (X, d) without isolated points is said to be **chaotic** if it satisfies the following three conditions:

(i) f has sensitive dependence on initial conditions.

(ii) f is transitive.

(iii) The set of periodic points of f is dense in X.

Even though chaos is defined using the notion of sensitive dependence on initial conditions, the next result shows that chaos is an invariant (and actually inheritable) property.

Theorem 4.11. *Let X be a metric space without isolated points, and $f\colon X \to X$ a dynamical system on X which is topologically transitive and for which the set of periodic points is dense in X. Then f has sensitive dependence on initial conditions with respect to any metric defining the topology on X.*

Proof. Fix a metric d defining the topology on X. Since X is necessarily an infinite set as it has no isolated points, we can choose two points $y_1, y_2 \in \mathsf{per}(f)$ such that $\mathcal{O}_f(y_1) \cap \mathcal{O}_f(y_2) = \emptyset$. Let

$$\delta := \frac{1}{8} d(\mathcal{O}_f(y_1), \mathcal{O}_f(y_2)).$$

We claim that δ is a sensitivity constant for f. Fix $x \in X$ and $\epsilon > 0$. The triangle inequality implies that at least one of $d(x, \mathcal{O}_f(y_1))$ and $d(x, \mathcal{O}_f(y_2))$ is at least 4δ. Without loss of generality assume the former:

$$d(x, f^k(y_1)) \geq 4\delta, \qquad \forall\, k \geq 0. \tag{4.2}$$

Next, since the periodic points are dense in X, we can choose $z \in \mathsf{per}(f)$ with

$$d(x, z) < \min\{\varepsilon, \delta\}. \tag{4.3}$$

Suppose z has period $p \geq 0$. Since f is continuous, there exists a neighbourhood U of y_1 such that

$$d(f^k(y_1), f^k(w)) < \delta, \qquad \forall\, k = 0, 1, 2, \dots, p \qquad \text{and} \quad w \in U. \tag{4.4}$$

Since f is transitive, there exists a point $x_1 \in X$ and $n \geq 0$ such that $d(x, x_1) < \epsilon$ and $f^n(x_1) \in U$. We now claim that one of z or x_1 is the point we are looking for. Since both of them are at within ε of x, we need only show that at least one of them gets mapped at least δ far away from x after some number of iterations of f. Let $q \geq 0$ denote the unique integer such that $n \leq qp < n + p$. Then since z has period p,

$$d\big(f^{qp}(z), f^{qp}(x_1)\big) = d\big(z, f^{qp-n}(f^n(x_1)\big),$$

and by the triangle inequality the right-hand side is at least

$$\begin{aligned} d\big(z, f^{qp-n}(f^n(x_1))\big) &\geq d\big(x, f^{qp-n}(y_1)\big) - d\big(f^{qp-n}(y_1), f^{qp-n}(f^n(x_1))\big) - d(x, z) \\ &> 4\delta - \delta - \delta \\ &= 2\delta, \end{aligned}$$

where we used (4.3),(4.2), and (4.4). Thus by the triangle inequality one last time, at least one of $d\big(f^{qp}(x), f^{qp}(z)\big)$ and $d\big(f^{qp}(x), f^{qp}(x_1)\big)$ is at least δ. This completes the proof. ■

We can now extend Definition 4.10 to all metric spaces, and show that it is a dynamical invariant.

Definition 4.12 (Final version). A dynamical system $f \colon X \to X$ is said to be **chaotic** if it is topologically transitive and the set of periodic points of f is dense in X.

Theorem 4.11 tells us that Definition 4.12 is equivalent to Definition 4.10 for metric spaces without isolated points. To prove that chaos is inheritable, we first prove:

Lemma 4.13. *The property of having a dense set of periodic points is inheritable.*

Proof. Let $g \colon Y \to Y$ be a factor of $f \colon X \to X$ with semiconjugacy $F \colon X \to Y$. Let $U \subset Y$ be an open non-empty set. We want to find a periodic point $y \in U$ for g. Since F is continuous and has dense range, $F^{-1}(U)$ is an open non-empty set of X, and hence by assumption there exists a periodic point $x \in F^{-1}(U)$ of S. If $f^p(x) = x$ then $g^p(F(x)) = F(x)$, and hence $y := F(x)$ is a periodic point for g. ■

Corollary 4.14. *Chaos is an inheritable property.*

Proof. Immediate from Lemma 4.8 and Lemma 4.13. ■

We now introduce another example of a dynamical system that is rather different to any that we have seen before. This dynamical system will turn out to be chaotic.

DEFINITION 4.15. Let Σ_2 denote[4] the space of all sequence $\mathtt{x} = (x_0, x_1, x_2, \dots)$ where each $x_k \in \{0,1\}$. We define a metric on d on Σ_2 by setting

$$d(\mathtt{x},\mathtt{y}) := \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{2^k}.$$

Recall that a topological space X is totally disconnected if the only connected subsets are the empty set and the singletons $\{x\}$ for $x \in X$.

PROPOSITION 4.16. *The space (Σ_2, d) is a compact metric space without isolated points which is totally disconnected.*

(♣) REMARK 4.17. A theorem[5] from point-set topology tells us that: *any two compact totally disconnected metric spaces without isolated points are homeomorphic.* A metric space with these properties that you are probably already familiar with is the **Cantor Set** (see Proposition 11.16 if you have forgotten the definition). Thus Σ_2 is homeomorphic to the Cantor Set. We will not use nor need this result in the course, however.

Proof. We begin by noting the following trivial statements about the metric d: given $\mathtt{x},\mathtt{y} \in \Sigma_2$, one has

$$x_k = y_k \qquad \forall \qquad k = 0,\dots n \qquad \Rightarrow \qquad d(\mathtt{x},\mathtt{y}) \le \frac{1}{2^n}, \tag{4.5}$$

and

$$d(\mathtt{x},\mathtt{y}) < \frac{1}{2^n} \qquad \Rightarrow \qquad x_k = y_k \qquad \forall \qquad k = 0,\dots n. \tag{4.6}$$

We now show that Σ_2 has no isolated points. Assume for contradiction that there exists $\mathtt{x} \in \Sigma_2$ and $\varepsilon > 0$ such that $B(\mathtt{x},\varepsilon) = \{\mathtt{x}\}$. Choose n so large, that $2^{-n} < \epsilon$. Let $\mathtt{y} \in \Sigma_2$ denote any element of Σ_2 such that $x_k = y_k$ for $0 \le k \le n$ and $y_{n+1} \ne x_{n+1}$. For instance if $x_{n+1} = 0$ then

$$\mathtt{y} = (x_0, x_1, \dots, x_n, 1, 1, 1, \dots)$$

works. Then by (4.5) we have

$$\begin{aligned} d(\mathtt{x},\mathtt{y}) &= \sum_{k \ge n+1} \frac{|x_k - y_k|}{2^k} \\ &\le \frac{1}{2^n} \\ &< \varepsilon. \end{aligned}$$

Since $\mathtt{y} \ne \mathtt{x}$ this contradicts the assumption that $B(\mathtt{x},\varepsilon) = \{\mathtt{x}\}$.

Now we show that Σ_2 is compact. Let $(\mathtt{x}^n) = ((x_0^n, x_1^n, x_2^n, \dots)) \subset \Sigma_2$ be a sequence. Either there exist infinitely many $n \ge 1$ for which $x_0^n = 0$ or there exist infinitely many $n \ge 1$ for which $x_0^n = 1$. Hence, passing to a subsequence we obtain

[4]The "2" in Σ_2 refers to the fact that each entry has exactly two choices: 0 or 1. There are similar spaces Σ_k for any $k \ge 1$, where each entry is allowed to take one of k designated values.

[5]See for instance Corollary 30.4 in Willard's book General Topology for a proof.

that x_0^n is constant in n. Similarly by induction we pass to subsequences such that x_k^n is constant in n for every fixed $k \geq 0$. A diagonal argument now gives a subsequence which is convergent by (4.5).

Finally to see that Σ_2 is totally disconnected it suffices to show that given any $\mathtt{x} \neq \mathtt{y}$ we can find open disjoint sets U and V such that $U \cup V = \Sigma_2$ and $\mathtt{x} \in U$ and $y \in V$. If $\mathtt{x} \neq \mathtt{y}$, then there is $n > 0$ such that $x_n \neq y_n$. Without loss of generality assume that $x_n = 0$ and $y_n = 1$. Define

$$U := \{\mathtt{z} \in \Sigma_2 \mid z_n = 0\}, \qquad V := \{\mathtt{z} \in \Sigma_2 \mid z_n = 1\}.$$

These sets have the desired properties. This completes the proof. ■

Here is an example of a dynamical system on Σ_2.

EXAMPLE 4.18. The **shift map** $\sigma \colon \Sigma_2 \to \Sigma_2$ is the map

$$\sigma(x_0, x_1, x_2, \dots) := (x_1, x_2, x_3, \dots).$$

This map is obviously continuous. We have:

LEMMA 4.19. *A point $\mathtt{x} \in \Sigma_2$ is periodic under σ if and only if the sequence (x_k) is periodic. Moreover a point $\mathtt{x}$ has dense orbit under σ if and only if every finite $0, 1$ sequence appears as a block in $\mathtt{x}$.*

Proof. The first statement is obvious. For the second, suppose $\mathtt{x}$ has a dense orbit under σ and suppose $(y_0, y_1, \dots y_n)$ is a finite $(0,1)$-sequence. Let

$$\mathtt{y} := (y_0, y_1, \dots, y_n, 0, 0, \dots).$$

Then there exists $k \geq 0$ such that $d(\sigma^k(\mathtt{x}), \mathtt{y}) < 2^{-n}$. It follows from (4.6) that $(x_k, \dots, x_{k+n}) = (y_0, \dots, y_n)$. The converse follows similarly, using (4.5). ■

Lemma 4.19 allows us to prove the shift map is chaotic.

PROPOSITION 4.20. *The shift map $\sigma : \Sigma_2 \to \Sigma_2$ is chaotic.*

Proof. Since the set of all finite $(0,1)$-sequences is countable, we can construct a $(0,1)$-sequence that contains each finite $(0,1)$-sequence as a block. Thus by Lemma 4.19, σ has a point with dense orbit. Since Σ_2 has no isolated points by Proposition 4.16, Corollary 2.11 implies that σ is transitive. Now let $\mathtt{y} \in \Sigma_2$ and $\epsilon > 0$. We will find a periodic point $\mathtt{x}$ for σ which satisfies $d(\mathtt{x}, \mathtt{y}) < \epsilon$. Choose n large enough so that $2^{-n} < \epsilon$. Then define

$$\mathtt{x} = (y_0, y_1, \dots, y_n, y_0, y_1, \dots, y_n, y_0, y_1, \dots).$$

By Lemma 4.19, $\mathtt{x}$ is periodic under σ, and by (4.5) one has $d(\mathtt{x}, \mathtt{y}) < 2^{-n}$. This completes the proof. ■

On Problem Sheet B you will show that the doubling map is a factor of the shift map. Combining this with Corollary 4.14 we obtain:

COROLLARY 4.21. *The doubling map is chaotic.*

We conclude this lecture with a remark on the terminology.

(♣) Remark 4.22. **Warning:** Now that we have defined chaos precisely, it is only prescient to warn you that the definition we have adopted is in fact but one of several possible different mathematical formulations of chaos. This should not surprise you: since chaos is a natural phenomenon, any mathematical definition is at best a "model". So why should there be only one?

The definition of chaos given in Definitions 4.10 and 4.12 is due to Devaney, and hence is commonly referred to as **Devaney chaos**. Another popular definition of chaos is called **Li-Yorke chaos** (which is due to Li and Yorke), and there are many more[6]. In general none of the definitions are equivalent, although there are various implications.

Nevertheless, in these notes we will only ever be concerned with Devaney chaos (i.e. Definition 4.12), and thus we will simply refer to it as "chaos".

[6] To mention a few by name: Block-Cappel chaos, Wiggins chaos, Martelli chaos,...

Mixing and Weakly Mixing Dynamical Systems

In this lecture we define a stronger version of transitivity, which is called *mixing*, and the intermediate notion of being *weakly mixing*, and explore various different characterisations of these properties.

DEFINITION 5.1. A dynamical system $f\colon X \to X$ is called **(topologically) mixing** if for any pair U, V of non-empty open subsets of X, there exists $n \geq 0$ such that for all $k \geq n$,

$$f^k(U) \cap V \neq \emptyset.$$

Clearly any mixing system is also transitive, but the converse is false, as we will shortly explain. Let us see an example:

LEMMA 5.2. *The tent map $\tau\colon [0,1] \to [0,1]$ is mixing.*

Proof. Let U and V be a pair of non-empty open subsets. As the proof of Lemma 2.6 showed, for any open set $U \subset [0,1]$ there exists a finite n such that $\tau^n(U) = [0,1]$. Thus also $\tau^k(U) = [0,1]$ for all $k \geq n$, and so in particular $\tau^k(U) \cap V \neq \emptyset$ for all $k \geq n$. ∎

Meanwhile a circle rotation is never mixing (cf. Example 5.7 below). Thus an irrational circle rotation is an example of a dynamical system that is transitive but not mixing. The next result is proved in the same way as Lemma 4.8.

LEMMA 5.3. *Mixing is an inheritable property.*

Proof. Let $f\colon X \to X$ and $g\colon Y \to Y$ be dynamical systems, and suppose that g is a factor of f with semiconjugacy $F\colon X \to Y$. Let U and V be a pair of non-empty open subsets of Y. Then since F is continuous and has dense range, $F^{-1}(U)$ and $F^{-1}(V)$ are non-empty open subsets of X. Thus there exists $n \geq 0$ and points $x_k \in F^{-1}(U)$ for each $k \geq n$ such that $f^k(x_k) \in F^{-1}(V)$. Then if $y_k := F(x_k)$ one has $y_k \in U$ and $g^k(y_k) = F(f^k(x_k)) \in V$. Thus $g^k(U) \cap V \neq \emptyset$ for all $k \geq n$. ∎

Recall from Example 1.24 that given two dynamical systems $f\colon X \to X$ and $g\colon Y \to Y$, we denote by $f \times g\colon X \times Y \to X \times Y$ the product system $(x,y) \mapsto (f(x), g(y))$.

PROPOSITION 5.4. *Suppose $f\colon X \to X$ and $g\colon Y \to Y$ are dynamical systems. Then:*

(i) *If $f \times g$ has a dense orbit then so do both f and g.*

(ii) *If $f \times g$ is transitive then so are both f and g.*

(iii) *If $f \times g$ is chaotic then so are both f and g.*

(iv) *If f and g are both transitive and at least one of them is mixing then $f \times g$ is topologically transitive.*

(v) *The system $f \times g$ is mixing if and only if both f and g are.*

Proof. Both f and g are a factor of $f \times g$ (cf. Example 1.24). Thus (ii) and (iii) and the "$\Rightarrow$" direction of (v) follow from Lemma 4.8, Corollary 4.14, and Lemma 5.3 respectively. A similar argument to the proof of Corollary 4.14 shows that the property of having a dense orbit is also inheritable, which proves (i).

To prove (iv), without loss of generality suppose that f is mixing and g is transitive. Let $O, Q \subseteq X \times Y$ be open non-empty sets. By definition of the product topology we can find open non-empty sets $U, V \subset X$ and $W, Z \subset Y$ such that $U \times W \subset O$ and $V \times Z \subset Q$. Now note that

$$(f \times g)^k(U \times W) \cap (V \times Z) = (f^k(U) \cap V) \times (g^k(W) \cap Z),$$

Choose $n \geq 0$ such that for all $k \geq n$ one has $f^k(U) \cap V \neq \emptyset$. Now using Problem A.5 we find a $k \geq n$ such that $g^k(W) \cap Z \neq \emptyset$. For this k, one thus has $(f \times g)^k(U \times W) \cap (V \times Z) \neq \emptyset$, and hence also $(f \times g)^k(O) \cap Q \neq \emptyset$.

Finally the proof of the "$\Leftarrow$" of (v) is a very similar argument. ∎

Corollary 5.5. *If f is mixing then $f \times f$ is transitive.*

Proof. Take $f = g$ and apply part (iv) of Proposition 5.4. ∎

Corollary 5.5 is not an if and only if statement: there exist dynamical systems f which are not mixing but for which $f \times f$ is transitive. Such dynamical systems are not easy to construct though—at the end of this lecture we will give an example. Nevertheless, this condition is important in its own right, and it gets its own name:

Definition 5.6. A dynamical system $f \colon X \to X$ is said to be **weakly (topologically) mixing** if $f \times f$ is transitive.

Lemma 5.7. *No circle rotation is weakly mixing.*

Proof. This follows immediately from Problem B.7, since the set $\{\theta, \theta, 1\}$ is never rationally independent. ∎

Therefore an irrational rotation is an example of a dynamical system that is transitive but not weakly mixing, which proves that weakly mixing is a strictly stronger property than transitivity. It is considerably harder to construct an example of a dynamical system that is weakly mixing but not mixing—we will construct such a system at the end of the next lecture (see Proposition 6.16).

Lemma 5.8. *A dynamical system $f \colon X \to X$ is weakly mixing if and only if for any quadruple U, V, W, Z of non-empty open subsets of X, there exists $k \geq 0$ such that*

$$f^k(U) \cap V \neq \emptyset \qquad \text{and} \qquad f^k(W) \cap Z \neq \emptyset.$$

See Figure 5.1.

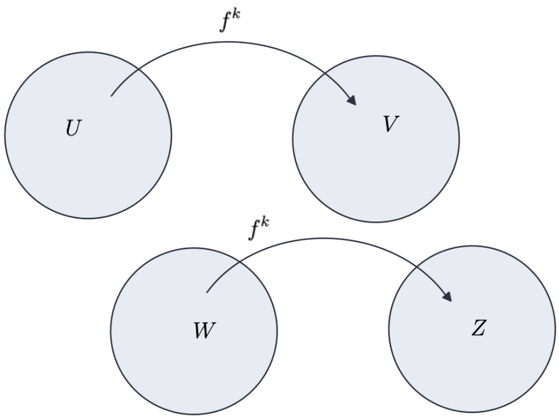

Figure 5.1: A weakly mixing dynamical system.

Proof. This is immediate from the fact that—as remarked in the proof of part (iv) of Proposition 5.4—any non-empty open set $O \subset X \times X$ contains a set of the form $U \times W$, where U, W are non-empty open subsets of X. ∎

Weakly mixing is again inheritable:

LEMMA 5.9. *Weakly mixing is an inheritable property.*

Proof. Let $f \colon X \to X$ and $g \colon Y \to Y$ be dynamical systems, and suppose that g is a factor of f. with semiconjugacy $F \colon X \to Y$. Then $F \times F$ is a semiconjugacy from $f \times f$ to $g \times g$, and hence $g \times g$ is a factor of $f \times f$. The claim now follows from Lemma 4.8. ∎

COROLLARY 5.10. *Let $f \colon X \to X$ and $g \colon Y \to Y$ be dynamical systems. If $f \times g$ is weakly mixing then so are both f and g.*

We can unify the concepts of transitivity, mixing and weakly mixing via the notion of *return times*.

DEFINITION 5.11. Let $f \colon X \to X$ denote a dynamical system. Given non-empty open subsets $U, V \subset X$ we define the set of **return times** for f as

$$\operatorname{ret}_f(U, V) := \{k \geq 0 \mid f^k(U) \cap V \neq \emptyset\} \subseteq \{0, 1, 2, \dots\}.$$

The next result expresses the three concepts in terms of return times.

COROLLARY 5.12. *A dynamical system $f \colon X \to X$ is:*

(i) *transitive if and only if given any non-empty open subsets $U, V \subset X$ one has $\operatorname{ret}_f(U, V) \neq \emptyset$;*

(ii) *mixing if and only if given any non-empty open subsets $U, V \subset X$, the subset $\operatorname{ret}_f(U, V)$ is cofinite in $\{0, 1, 2 \dots\}$;*

(iii) *weakly mixing if and only if given any four non-empty open subsets $U, V, W, Z \subset X$ one has*

$$\operatorname{ret}_f(U, V) \cap \operatorname{ret}_f(W, Z) \neq \emptyset.$$

Proof. Parts (i) and (ii) are immediate. Part (iii) is just a rephrasing of Lemma 5.8. ∎

In fact, we can reduce the number of sets needed to check the weakly mixing condition from four to two:

PROPOSITION 5.13. *Let $f\colon X \to X$ denote a dynamical system. Then the following three statements are equivalent.*

(i) *f is weakly mixing.*

(ii) *For any three non-empty open subsets $U, V, W \subseteq X$, one has*

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(U, W) \neq \emptyset.$$

(iii) *For any two non-empty open subsets $U, V \subset X$, one has*

$$\mathrm{ret}_f(U, U) \cap \mathrm{ret}_f(U, V) \neq \emptyset.$$

See Figure 5.2.

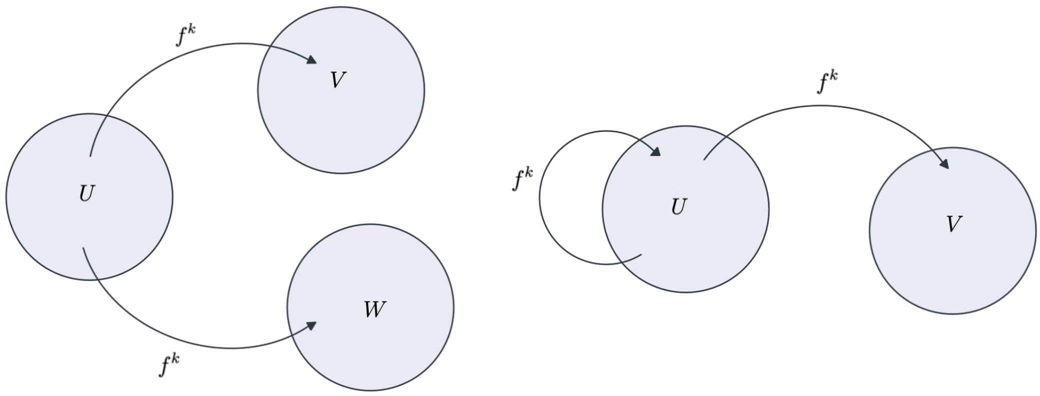

Figure 5.2: Weakly mixing with three and two sets.

Proof. It is obvious that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii).

Let us prove that (ii) $\Rightarrow$ (i). Let $U, V, W, Z \subseteq X$ be non-empty open sets. By part (iii) of Corollary 5.12 it suffices to show that

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z) \neq \emptyset.$$

We apply (ii) to the triple U, W, Z to find an element $k \in \mathrm{ret}_f(U, W) \cap \mathrm{ret}_f(U, Z)$. In particular, this tells us that both[1]

$$U_0 \coloneqq U \cap f^{-k}(W) \neq \emptyset \qquad \text{and} \qquad f^{-k}(Z) \neq \emptyset.$$

We now apply (ii) again, this time to the triple $U_0, V, f^{-k}(Z)$, to find an element $n \in \mathrm{ret}_f(U_0, V) \cap \mathrm{ret}_f(U_0, f^{-k}(Z))$.

[1]Actually $f^{-k}(Z)$ is always non-empty; see the proof of (iii) $\Rightarrow$ (ii) below.

Since $n \in \mathrm{ret}_f(U_0, f^{-k}(Z))$ there exists $x \in U_0$ with $f^n(x) \in f^{-k}(Z)$. Let $y := f^k(x)$. Since $x \in U_0$ we have $y \in W$. Moreover $f^n(y) = f^{n+k}(x) = f^k(f^n(x)) \in Z$. Thus $f^n(W) \cap Z \neq \emptyset$, and hence $n \in \mathrm{ret}_f(W, Z)$. Finally since $\mathrm{ret}_f(U_0, V) \subseteq \mathrm{ret}_f(U, V)$ as $U_0 \subseteq U$, it follows that

$$n \in \mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z)$$

and thus this intersection is non-empty, as desired.

Now we prove that (iii) $\Rightarrow$ (ii). Let $U, V, W \subseteq X$ be non-empty open sets. This time we want to show that $\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(U, W) \neq \emptyset$. Since f is necessarily transitive by (iii), there exists some $k \geq 0$ such that

$$U_0 := U \cap f^{-k}(V)$$

is a non-empty open set. Since transitive maps have dense range by Lemma 2.3, $f^{-k}(W)$ is a non-empty open set. Then by (iii) applied to the pair $U_0, f^{-k}(W)$, there exists some

$$n \in \mathrm{ret}_f(U_0, U_0) \cap \mathrm{ret}_f(U_0, f^{-k}(W)).$$

This means there exists $x, y \in U_0$ with $f^n(x) \in U_0$ and $f^n(y) \in f^{-k}(W)$. Since $x \in U$ and $f^{n+k}(x) \in V$, we see that $n + k \in \mathrm{ret}_f(U, V)$. Similarly $f^{n+k}(y) \in W$ and hence $n + k \in \mathrm{ret}_f(U, W)$. Since $U_0 \subseteq U$, one also has $n + k \in \mathrm{ret}_f(U, W)$, and thus

$$n + k \in \mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(U, W) \neq \emptyset.$$

This completes the proof. ∎

We conclude by restating how the various properties fit together.

Summary: Let $f \colon X \to X$ be a dynamical system. Then

$$\text{mixing} \quad \Rightarrow \quad \text{weakly mixing} \quad \Rightarrow \quad \text{transitive},$$

and neither of the implications can be reversed.

Note that we have not yet established that mixing really is a stronger property than weakly mixing; as mentioned earlier this is fairly tricky and will be done next lecture in Proposition 6.16.

LECTURE 6

Furstenberg's Theorem

In this lecture we continue our discussion of weakly mixing dynamical systems, starting with an important theorem of Furstenberg. We then take a brief excursion into *linear* dynamical systems on Banach spaces and construct an example of a dynamical system which is weakly mixing but not mixing, thus fulfilling a promise from the last lecture.

DEFINITION 6.1. Two dynamical systems $f\colon X \to X$ and $g\colon X \to X$ are said to **commute** if

$$f \circ g = g \circ f.$$

We begin with the following useful trick.

LEMMA 6.2. *Let $f\colon X \to X$ be a transitive dynamical system, and let U, V, W, Z be non-empty open subsets of X. Suppose there exists a dynamical system $g\colon X \to X$ which commutes with f and satisfies*

$$g(U) \cap W \neq \emptyset \qquad \text{and} \qquad g(V) \cap Z \neq \emptyset. \tag{6.1}$$

Then

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z) \neq \emptyset. \tag{6.2}$$

Informally, Lemma 6.2 can be thought of as saying that the existence of such a system g implies that f is "weakly mixing for this particular quadruple of sets U, V, W, and Z".

Proof. Since g is continuous, (6.1) implies there exist non-empty open subsets $U_0 \subset U$ and $V_0 \subset V$ such that $g(U_0) \subset W$ and $g(V_0) \subset Z$. See Figure 6.1. Suppose $k \in \mathrm{ret}_f(U_0, V_0)$. Then there exists $x \in U_0$ such that $f^k(x) \in V_0$. Then since $g(x) \in W$ and $f^k(g(x)) = g(f^k(x)) \in Z$, one sees that $k \in \mathrm{ret}_f(W, Z)$, which shows that

$$\mathrm{ret}_f(U_0, V_0) \subseteq \mathrm{ret}_f(W, Z).$$

Since clearly $\mathrm{ret}_f(U_0, V_0) \subseteq \mathrm{ret}_f(U, V)$, we therefore always have

$$\mathrm{ret}_f(U_0, V_0) \subseteq \mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z). \tag{6.3}$$

Since f is transitive the left-hand side of (6.3) is non-empty, and thus the right-hand side of (6.3) is also non-empty. This completes the proof. ∎

Our first application of Lemma 6.2 is the following curiosity.

COROLLARY 6.3. *Let $f\colon X \to X$ be a topologically transitive dynamical system, and suppose U, V, W, Z are four open non-empty subsets of X. Then*

$$\mathrm{ret}_f(U, W) \cap \mathrm{ret}_f(V, Z) \neq \emptyset \qquad \Rightarrow \qquad \mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z) \neq \emptyset.$$

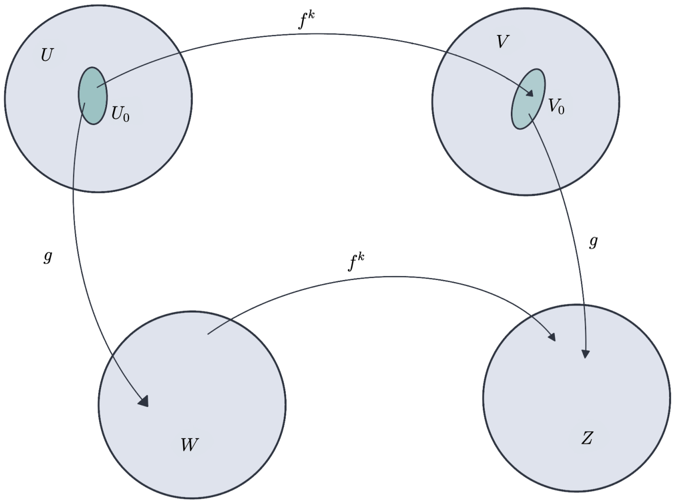

Figure 6.1: Commuting systems f and g.

Proof. Suppose $k \in \mathrm{ret}_f(U, W) \cap \mathrm{ret}_f(V, Z)$. Set $g := f^k$. Then g commutes with f and satisfies (6.1). Thus Lemma 6.2 implies that $\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, Z) \neq \emptyset$. ∎

We could attempt to strengthen the weakly mixing condition by requiring higher products to be transitive.

Definition 6.4. Let $f: X \to X$ be a dynamical system. We say that f is ***n*-fold transitive** if the n-fold product

$$\underbrace{f \times \cdots \times f}_{n \text{ times}} : \underbrace{X \times \cdots \times X}_{n \text{ times}} \to \underbrace{X \times \cdots \times X}_{n \text{ times}}$$

is transitive.

With this terminology, a transitive system is a 1-fold transitive system and a weakly mixing system is a 2-fold transitive system. Moreover if $k \leq n$ then the k-fold product is a factor of the n-fold factor, and hence an n-fold transitive system is also k-fold transitive for all $k \leq n$.

Actually Definition 6.4 turns out to be redundant. This is the content of the next result, which is due to the Israeli mathematician Furstenberg. The proof makes use of Lemma 6.2 again.

Theorem 6.5 (Furstenberg's Theorem). *Let $n \geq 2$. A dynamical system is n-fold transitive if and only if it is weakly mixing.*

Proof. We prove by induction on n that if f is n-fold transitive then f is also $(n+1)$-fold transitive. By arguing as in Lemma 5.8, to show that f is $(n+1)$-fold transitive it suffices to show that given any $2n+2$ non-empty open subsets $U_k, V_k \subseteq X$ for $k = 1, \ldots, n+1$, one has

$$\bigcap_{k=1}^{n+1} \mathrm{ret}_f(U_k, V_k) \neq \emptyset. \tag{6.4}$$

Since f is weakly mixing, applying Lemma 5.8 to the quadruple U_1, U_2, V_1 and V_2 we find $k \geq 0$ such that

$$f^k(U_1) \cap U_2 \neq \emptyset, \qquad \text{and} \qquad f^k(V_1) \cap V_2 \neq \emptyset.$$

Then by Lemma 6.2, applied with $g = f^k$, we find non-empty open subsets $U_0 \subset U_1$ and $V_0 \subset V_1$ such that

$$\mathrm{ret}_f(U_0, V_0) \subseteq \mathrm{ret}_f(U_1, V_1) \cap \mathrm{ret}_f(U_2, V_2) \tag{6.5}$$

(see (6.3)). Now by the induction hypothesis applied to the $2n$ non-empty open sets $U_0, U_3, \dots, U_{n+1}$ and $V_0, V_3, \dots, V_{n+1}$, we obtain that

$$\mathrm{ret}_f(U_0, V_0) \cap \left(\bigcap_{k=3}^{n+1} \mathrm{ret}_f(U_k, V_k) \right) \neq \emptyset. \tag{6.6}$$

Then (6.5) and (6.6) imply that (6.4) holds, which thus completes the proof. ■

Here is an application of Furstenberg's Theorem, which gives us another way to characterise weakly mixing maps.

Proposition 6.6. *Let $f\colon X \to X$ be a dynamical system. Then f is weakly mixing if and only if for any two non-empty open subsets $U, V \subset X$, the set $\mathrm{ret}_f(U, V)$ contains arbitrarily long intervals.*

Proof. First assume that f is weakly mixing. Let $U, V \subset X$ be non-empty open sets and let $n \geq 2$. We will show that $\mathrm{ret}_f(U, V)$ contains an interval of length n. By Furstenberg's Theorem 6.5, f is n-fold transitive, and hence there exists some $i \geq 0$ such that

$$i \in \bigcap_{k=1}^{n} \mathrm{ret}_f(U, f^{-k}(V)),$$

that is,

$$f^i(U) \cap f^{-k}(V) \neq \emptyset \qquad \forall\, 1 \leq k \leq n.$$

This implies that $i + k \in \mathrm{ret}_f(U, V)$ for each $1 \leq k \leq n$.

To prove the other direction, it is convenient to use part (ii) of Proposition 5.13. Let $U, V, W \subset X$ be non-empty open sets. We will prove that

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(U, W) \neq \emptyset.$$

Since $\mathrm{ret}_f(V, W)$ is certainly non-empty, by continuity there exists $k \geq 0$ and a non-empty open subset $V_0 \subset V$ such that $f^k(V_0) \subset W$. Then by assumption there exists $n \geq 0$ such that the interval $[n, n+k] \subseteq \mathrm{ret}_f(U, V_0)$. In particular

$$n + k \in \mathrm{ret}_f(U, V_0) \subseteq \mathrm{ret}_f(U, V).$$

To complete the proof we show that $n + k$ also belongs to $\mathrm{ret}_f(U, W)$. But this follows from the fact that $f^{n+k}(U) \cap W$ contains the set $f^{n+k}(U) \cap f^k(V_0)$, which itself contains the set $f^k(f^n(U) \cap V_0)$. Since also $n \in \mathrm{ret}_f(U, V_0)$ by assumption, the latter set is non-empty. This completes the proof. ■

On Problem Sheet C you will find yet another characterisation of weakly mixing.

There remains a serious defect of our definition of weakly mixing: we have yet to exhibit a single example of a dynamical system that is weakly mixing but not mixing! We now rectify this. The example we construct (Proposition 6.16) is not the simplest example of such a system, but it has the virtue of being easy to understand and gives us an excuse to introduce linear dynamical systems. Our construction will use a little bit of elementary functional analysis[1].

Definition 6.7. A **Banach space** is a vector space E equipped with a norm $\|\cdot\|$ such that the associated metric $d(v,w) := \|v-w\|$ is complete.

Any finite-dimensional normed vector space[2] is a Banach space. Here is an easy to understand example of an infinite-dimensional Banach space:

Definition 6.8. Let

$$\ell^\infty(\mathbb{R}) := \{\text{sequences } \mathtt{x} = (x_k)_{k\geq 0} \text{ of real numbers} \mid \sup |x_k| < \infty\}$$

denote the space of all **bounded sequences**. This is a Banach space under the norm

$$\|\mathtt{x}\|_\infty := \sup_{k\geq 0} |x_k|.$$

We denote by e_k the vector with a 1 in the kth position and 0 in all the other entries. Thus $(\mathsf{e}_k)_{k\geq 0}$ is a basis of $\ell^\infty(\mathbb{R})$.

Definition 6.9. Let

$$c_0(\mathbb{R}) := \left\{\mathtt{x} \in \ell^\infty(\mathbb{R}) \mid \lim_{k\to\infty} x_k = 0\right\}$$

denote the space of all **null sequences**. This is a closed subspace of $\ell^\infty(\mathbb{R})$ and hence is a Banach space under the same norm $\|\cdot\|_\infty$.

Remark 6.10. The space $c_0(\mathbb{R})$ is typically better behaved than $\ell^\infty(\mathbb{R})$. For instance, $c_0(\mathbb{R})$ is separable as a metric space, whereas $\ell^\infty(\mathbb{R})$ is not. We will not need nor use this fact however.

Definition 6.11. Let

$$c_{00}(\mathbb{R}) := \{\mathtt{x} \in c_0(\mathbb{R}) \mid \text{ there exists } n \geq 1 \text{ such that } x_k = 0 \text{ for all } k \geq n\}$$

denote the space of all **finite sequences**.

It is easy to see that $c_{00}(\mathbb{R})$ is dense in $c_0(\mathbb{R})$. However $c_{00}(\mathbb{R})$ is *not* closed in $c_0(\mathbb{R})$, and hence is *not* a Banach space with respect to $\|\cdot\|_\infty$.

Definition 6.12. Suppose $(E, \|\cdot\|)$ is a Banach space. A continuous linear map $L\colon E \to E$ is called a **linear dynamical system**.

[1] Do not worry if you are not familiar with functional analysis and Banach spaces—the only Banach spaces we will meet in Dynamical Systems I are Euclidean spaces $\mathbb{R}^n$, or the space of bounded sequences defined below.

[2] By convention, *all* vector spaces in this course are assumed to be vector spaces over $\mathbb{R}$ or $\mathbb{C}$.

Remark 6.13. A linear map $L\colon E \to E$ is continuous if and only if it is *bounded* in the sense that there exists $c > 0$ such that[3]

$$\|Lv\| \leq c\|v\|, \qquad \forall\, v \in E.$$

If E is finite-dimensional then every linear map is continuous (and hence bounded).

Just as in Example 4.18, one can consider shift operators on the spaces $\ell^\infty(\mathbb{R})$, $c_0(\mathbb{R})$, and $c_{00}(\mathbb{R})$.

Example 6.14. The **shift operator**[4]

$$\sigma(x_0, x_1, x_2, \dots) \coloneqq (x_1, x_2, \dots)$$

is a linear dynamical system on both $\ell^\infty(\mathbb{R})$ and $c_0(\mathbb{R})$.

With these preliminaries out of the way, we present a useful (albeit somewhat contrived) criterion for a linear dynamical system to be weakly mixing.

Theorem 6.15. *Let $L\colon E \to E$ be a linear dynamical system on a Banach space. Assume there exists a dense subset*[5] *$X \subseteq E$ and a dynamical system $f\colon X \to X$ such that:*

$$Lf(v) = v, \qquad \forall\, v \in X. \tag{6.7}$$

Assume moreover that there exists a strictly increasing sequence (k_n) of numbers such that for all $v \in X$,

$$\lim_{n\to\infty} L^{k_n} v = 0, \qquad \text{and} \qquad \lim_{n\to\infty} f^{k_n}(v) = 0. \tag{6.8}$$

Then L is weakly mixing.

Theorem 6.15 is a special case of a result due to Gethner and Shapiro. In our case however the proof is almost easier than the statement.

Proof. Let U, V, W, Z be non-empty open sets in E. By part (iii) of Corollary 5.12 it suffices to show that

$$\mathrm{ret}_L(U, V) \cap \mathrm{ret}_L(W, Z) \neq \emptyset.$$

Since X is dense in E, we can find vectors

$$u \in U \cap X, \qquad v \in V \cap X, \qquad w \in W \cap X, \qquad z \in Z \cap X.$$

For n sufficiently large it follows from (6.8) that

$$u + f^{k_n}(v) \in U, \qquad v + L^{k_n} u \in V, \qquad w + f^{k_n}(z) \in W, \qquad z + L^{k_n} w \in Z. \tag{6.9}$$

Next, using (6.7) we have

$$L^{k_n}(u + f^{k_n}(v)) = L^{k_n} u + v, \qquad L^{k_n}(w + f^{k_n}(z)) = L^{k_n} w + z. \tag{6.10}$$

Combining (6.9) and (6.10) shows that $\mathrm{ret}_L(U, V) \cap \mathrm{ret}_L(W, Z) \neq \emptyset$. ■

[3] By convention, for a linear map where possible we omit the brackets and write Lv instead of $L(v)$.

[4] We use the word “operator” instead of “map” to help distinguish this from the shift map in Example 4.18.

[5] We do *not* assume that X is a linear subspace, and the map f does *not* have to be linear.

We will use Theorem 6.15 to produce an example of a linear dynamical system on $c_0(\mathbb{R})$ that is weakly mixing but not mixing. For this let $\mathtt{w}, \mathtt{w}' \in \ell^\infty(\mathbb{R})$ denote the vectors

$$\mathtt{w} = (w_k) := \left(2, \tfrac{1}{2}, 2, 2, \tfrac{1}{2}, \tfrac{1}{2}, 2, 2, 2, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 2, 2, 2, 2, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \dots\right),$$

and

$$\mathtt{w}' = (w'_k) := \left(\tfrac{1}{2}, 2, \tfrac{1}{2}, \tfrac{1}{2}, 2, 2, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 2, 2, 2, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 2, 2, 2, 2, \dots\right).$$

Define an operator

$$L \colon c_0(\mathbb{R}) \to c_0(\mathbb{R}), \qquad L(x_0, x_1, x_2, \dots) := (w_0 x_1, w_1 x_2, w_2 x_3, \dots).$$

One can think of L as a "weighted" version of the shift operator from Example 6.14. Note L is continuous by Remark 6.13.

Proposition 6.16. *The operator L is weakly mixing but not mixing.*

Proof. To show that L is weakly mixing we apply Theorem 6.15. Take $X = c_{00}(\mathbb{R})$, and define

$$f \colon c_{00}(\mathbb{R}) \to c_{00}(\mathbb{R}), \qquad f(x_0, x_1, x_2, \dots) := \left(0, w'_0 x_0, w'_1 x_1, w'_2 x_2, \dots\right).$$

Then clearly

$$Lf(\mathtt{x}) = \mathtt{x}, \qquad \forall\, \mathtt{x} \in c_{00}(\mathbb{R}),$$

and moreover

$$\lim_{k\to\infty} L^k(\mathtt{x}) = \mathtt{0}, \qquad \forall\, \mathtt{x} \in c_{00}(\mathbb{R}),$$

where $\mathtt{0} := (0, 0, 0, \dots)$. Thus to apply Theorem 6.15 we need only find a sequence (k_n) of increasing numbers such that

$$\lim_{n\to\infty} f^{k_n}(\mathtt{x}) = \mathtt{0}, \qquad \forall\, \mathtt{x} \in c_{00}(\mathbb{R}).$$

We take $k_n := n^2$. Observe that

$$\prod_{i=1}^{n^2} w'_{i-1} = \frac{1}{2^n},$$

and hence

$$f^{n^2}(\mathtt{e}_1) = \Big(0, \dots, 0, \underbrace{\tfrac{1}{2^n}}_{n\text{th position}}, 0, 0, \dots\Big),$$

and thus $\lim_{n\to\infty} f^{n^2}(\mathtt{e}_1) = \mathtt{0}$. Similarly $f^{n^2}(\mathtt{e}_k) \to 0$ for any basis vector $\mathtt{e}_k$. Since any element of $c_{00}(\mathbb{R})$ can be written as a finite sum of the $\mathtt{e}_k$, it follows that

$$\lim_{n\to\infty} f^{n^2}(\mathtt{x}) = \mathtt{0}, \qquad \forall\, \mathtt{x} \in c_{00}(\mathbb{R}).$$

Theorem 6.15 therefore implies that L is weakly mixing.

Finally we show that L is not mixing. For this observe that if

$$m_n := n^2 + n$$

then

$$\prod_{i=1}^{m_n} w_{i-1} = 1,$$

$$L^{m_n}(x_0, x_1, x_2, \dots) = (x_{m_n}, *, *, \dots),$$

where the value of the $*$'s are unimportant. This means that if

$$U := \{\mathtt{x} \in c_0(\mathbb{R}) \mid \|\mathtt{x}\| < 1\}$$

and

$$V := \{\mathtt{x} \in c_0(\mathbb{R}) \mid |x_0| > 1\}$$

then U and V are open sets[6] in $c_0(\mathbb{R})$ for which

$$L^{m_n}(U) \cap V = \emptyset, \qquad \forall\, n \geq 1.$$

Thus L is not mixing. This completes the proof. ■

We will return to linear dynamical systems next semester when we discuss hyperbolicity.

[6] To see that V is open, note that since $|x_k| \leq \|\mathtt{x}\|$ for any $\mathtt{x} \in c_0(\mathbb{R})$ and $k \geq 0$, the projection operators $P_k \colon c_0(\mathbb{R}) \to \mathbb{R}$ defined by $P_k(\mathtt{x}) = x_k$ are all continuous.

LECTURE 7

Topological Entropy

Let $f \colon X \to X$ be a dynamical system on a compact metric space. In this lecture we introduce the **topological entropy** $\mathsf{h}_{\mathrm{top}}(f)$ of f. This is a non-negative real number (or ∞), which attempts to give a quantitative measure of how "complex" the dynamics of f are. As one might expect, trying to reduce the entire dynamics of f to a single number is only partially successful. Nevertheless, it is remarkable quite how much information can be captured by the topological entropy[1].

The definition of $\mathsf{h}_{\mathrm{top}}(f)$ is rather complicated, but roughly speaking it measures the rate at which orbits of a dynamical system move apart as time increases. Before getting started with the formal definition, let us give a more heuristic outline in a similar vein to our informal discussion of the chain recurrent set from Remark 3.14.

Suppose we start with a dynamical system $f \colon X \to X$. As a first measure of complexity of f, we could try and "count" the number of orbit segments of f up to time k:

$$\{x, f(x), f^2(x), \dots, f^k(x)\}.$$

Suppose that our measuring device is only accurate up to the nearest ε, and hence it is unable to distinguish between two orbits segments $\{x, f(x), f^2(x), \dots f^k(x)\}$ and $\{y, f(y), f^2(y), \dots f^k(y)\}$ if at every stage the distance between $f^i(x)$ and $f^i(y)$ is less than ε. However as time goes on one might hope that the orbits of distinct points diverge. This would mean that eventually the measuring device could distinguish them. See Figure 7.1.

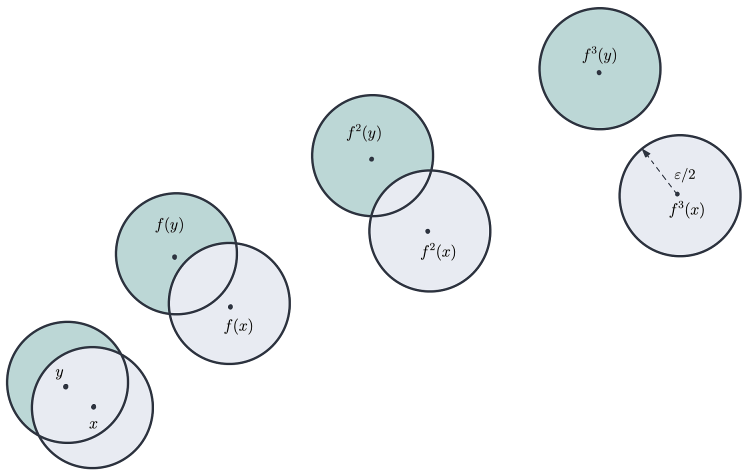

Figure 7.1: The orbit segments of x and y cannot be distinguished until time $k = 3$.

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]See Problem E.5.

Let us temporarily denote by $n(k, \varepsilon)$ the total number of orbits segments our measuring device finds at time k. We then look at the growth rate of the function $k \mapsto n(k, \varepsilon)$, that is, the growth rate of orbits over time, as seen by our measuring device. If our dynamical system has "simple" dynamics then given two nearby states x and y, it may take a very long time before our detector can tell that the orbits of x and y are different, and hence the growth rate could be low. If however our dynamical system has sensitive dependence on initial conditions (i.e. is chaotic), then the orbits of nearby states will diverge very rapidly. This will therefore give rise to a large growth rate.

Next, note that with a better measuring device (i.e. one with a smaller error value), less time will be required to tell distinct orbits apart. Thus the growth rate of the function $k \mapsto n(k, \varepsilon)$ increases as ε decreases. Since we are mathematicians (and don't need to do any actual experiments!), we can quite happily pretend that we have access to measuring devices of unlimited precision and let $\varepsilon \to 0$. The resulting quantity can therefore be thought of as the growth rate of orbits over time, as seen through the eyes of an arbitrarily precise measuring device. This is what we call the **topological entropy** of the system.

Now onto the formal definitions:

Throughout our discussion of topological entropy, we will always[2] assume that the underlying metric spaces are **compact**. This will considerably simplify the construction, and is sufficient for all of our examples.

Definition 7.1. Let (X, d) be a compact metric space and $f \colon X \to X$ a dynamical system. For each $k \geq 1$, we can define a new metric d_k^f on X by setting

$$d_k^f(x, y) \coloneqq \max_{0 \leq i \leq k-1} d\big(f^i(x), f^i(y)\big).$$

Thus $d_1^f = d$ and $d_k^f \leq d_{k+1}^f$ for all $k \geq 1$. Since f is continuous and X is compact, one easily sees that all metrics d_k^f are strongly equivalent. In the new metric d_k^f, points are ε close if they remain ε close for k iterations of f.

Given a set A, we denote by $\#A$ its cardinality. We now introduce three related notions that "count" the number of orbits segments of f of length k that are distinguishable at the scale ε.

Definition 7.2. Let $f \colon X \to X$ denote a dynamical system on a compact metric space. A subset $A \subseteq X$ is called a (k, ε)**-spanning set** for f if for every $x \in X$ there exists $y \in A$ such that $d_k^f(x, y) < \varepsilon$. Compactness of X implies there exist finite (k, ε)-spanning sets, and we set

$$\operatorname{span}(f, k, \varepsilon) \coloneqq \min\{\#A \mid A \text{ is a } (k, \varepsilon)\text{-spanning set for } f\}.$$

[2]The exception to this is the (non-examinable) Remark 7.17 at the end of the lecture, where we briefly outline how to go about extending the definition of entropy to certain dynamical systems on non-compact spaces.

Remark 7.3. We should really include the metric d in our notation and write $\mathrm{span}_d(f,k,\varepsilon)$, since this quantity *does* depend on the choice of the metric. Nevertheless, we have elected not to, since the notation is already pretty horrendous.

Next, we have:

Definition 7.4. Let $f\colon X\to X$ denote a dynamical system on a compact metric space. A subset $B\subseteq X$ is called a (k,ε)**-separated set** for f if given any two distinct points $x,y\in B$ one has $d_k^f(x,y)\geq\varepsilon$. Compactness of X implies that any such set is finite, and we set

$$\mathrm{sep}(f,k,\varepsilon):=\sup\{\#A\mid A\text{ is a }(k,\varepsilon)\text{-separated set for }f\}.$$

The proof of Proposition 7.6 below will show this is a finite number.

Recall given a set A in a metric space (X,d), we define the diameter of A as

$$\mathrm{diam}_d(A):=\sup\{d(x,y)\mid x,y,\in A\}.$$

Finally, we have:

Definition 7.5. Let $f\colon X\to X$ denote a dynamical system on a compact metric space. We denote by $\mathrm{cov}(f,k,\varepsilon)$ the minimum cardinality of an open covering of X by sets whose d_k^f diameter is less than ε. By compactness, $\mathrm{cov}(f,k,\varepsilon)$ is finite.

The three quantities are related as follows:

Proposition 7.6. *For each $k\geq 1$ and $\varepsilon>0$, one has*

$$\mathrm{cov}(f,k,2\varepsilon)\overset{(7.1)}{\leq}\mathrm{span}(f,k,\varepsilon)\overset{(7.2)}{\leq}\mathrm{sep}(f,k,\varepsilon)\overset{(7.3)}{\leq}\mathrm{span}(f,k,\varepsilon/2)\overset{(7.4)}{\leq}\mathrm{cov}(f,k,\varepsilon/2).$$

Proof. We first prove inequality (7.1). Suppose A is a (k,ε)-spanning set for f of minimum cardinality. Then the open balls of radius ε in the d_k^f metric centred at points of A must cover X. By compactness the same is true for some $0<\delta<\varepsilon$. The diameter of such a set is $2\delta<2\varepsilon$, and hence

$$\mathrm{cov}(f,k,2\varepsilon)\leq\mathrm{span}(f,k,\varepsilon).$$

We next[3] prove (7.3). Let A be a minimal $(k,\varepsilon/2)$-spanning set and let B be an arbitrary (k,ε)-separated set. We define an injective map $\phi\colon B\to A$ as follows: If $x\in B$ there exists at least one $y\in A$ such that $d_k^f(x,y)<\varepsilon/2$. Pick one such y and call it $\phi(x)$. If $\phi(x_1)=\phi(x_2)$ then by the triangle inequality we have

$$d_k^f(x_1,x_2)\leq d_k^f(x_1,\phi(x_1))+d_k^f(\phi(x_2),x_2)<\varepsilon.$$

Since B is a (k,ε)-separated set, this implies that $x_1=x_2$. Therefore ϕ is injective, and hence $\#B\leq\#A$. Since B was an arbitrary (k,ε)-separated set and A was a minimal $(k,\varepsilon/2)$-spanning set, this shows that

$$\mathrm{sep}(f,k,\varepsilon)\leq\mathrm{span}(k,f,\varepsilon/2).$$

[3]The reason for proving (7.3) before (7.2) is that until we proved (7.3), we do not know that $\mathrm{sep}(f,k,\varepsilon)$ is finite. Note also that (7.3) implies that we can replace the “sup” in Definition 7.4 with “max”.

Now we prove (7.2). let B be a (k,ε)-separated set for f with maximal cardinality. Then $d_k^f(x,y) \geq \varepsilon$ for all $x, y \in B$. We claim that B is also a (k,ε)-spanning for f. For this let $x \in X$. We must produce some $y \in B$ such that $d_k^f(x,y) < \varepsilon$. If $x \in B$ there is nothing to prove. If $x \in X \setminus B$ and no such y existed then $B \cup \{x\}$ would also be a (k,ε)-separated set for f. This contradicts the maximality of B. Thus B is also a (k,ε)-spanning set for f, and hence in particular we have.

$$\operatorname{span}(f,k,\varepsilon) \leq \operatorname{sep}(f,k,\varepsilon).$$

Finally let us prove (7.4). Suppose $\{U_1,\dots,U_n\}$ is any cover of X consisting of sets of d_k^f-diameter less than ε. Pick any point $x_i \in U_i$. Then the set $\{x_1,\dots,x_n\}$ forms a (k,ε)-spanning set for f. This shows that

$$\operatorname{span}(f,k,\varepsilon) \leq \operatorname{cov}(f,k,\varepsilon),$$

and so the proof is complete. ■

Before going any further, let us recall the following elementary lemma from calculus:

Lemma 7.7 (Fekete's Lemma). *Let $\alpha\colon \mathbb{N} \to \mathbb{R}$ denote a subadditive function, i.e.*

$$\alpha(k+n) \leq \alpha(k) + \alpha(n), \qquad \forall\, k,n \in \mathbb{N}. \tag{7.5}$$

Assume that $\inf_k \frac{\alpha(k)}{k} > -\infty$. Then the limit of $\frac{\alpha(k)}{k}$ exists as $k \to \infty$, and moreover

$$\lim_{k\to\infty} \frac{\alpha(k)}{k} = \inf_{k\in\mathbb{N}} \frac{\alpha(k)}{k}.$$

This proof is non-examinable, since it belongs to a course on real analysis.

(♣) *Proof.* Let $a := \inf_k \frac{\alpha(k)}{k}$. Fix $\varepsilon > 0$. By definition of the infimum there exists $n \geq 1$ such that

$$\left| \frac{\alpha(n)}{n} - a \right| < \frac{\varepsilon}{2}. \tag{7.6}$$

Choose m large enough such that

$$\frac{\alpha(i)}{mn} < \frac{\varepsilon}{2}, \qquad \forall\, 1 \leq i < n. \tag{7.7}$$

Now choose $k \geq mn$. Then there exist integers q, r such that $k = qn + r$ where $q \geq m$ and $0 \leq r < n$. Then

$$\begin{aligned}
\frac{\alpha(k)}{k} &\overset{(7.5)}{\leq} \frac{\alpha(qn)}{qn+r} + \frac{\alpha(r)}{qn+r} \\
&\overset{(7.5)}{\leq} \frac{q\alpha(n)}{qn} + \frac{\alpha(r)}{mn} \\
&\overset{(7.6)}{<} a + \frac{\varepsilon}{2} + \frac{\alpha(r)}{mn} \\
&\overset{(7.7)}{<} a + \varepsilon.
\end{aligned}$$

This shows that

$$\left|\frac{\alpha(k)}{k} - a\right| < \varepsilon, \qquad \text{for} \qquad k \geq mn,$$

and hence the limit of $\frac{\alpha(k)}{k}$ as $k \to \infty$ exists and is equal to a. This completes the proof. ■

We can now prove:

Proposition 7.8. *Let $f\colon X \to X$ be a dynamical system on a compact metric space and let $\varepsilon > 0$. Then the limit*

$$\mathsf{h}_\varepsilon^{\mathrm{cov}}(f) := \lim_{k\to\infty} \frac{1}{k} \log \operatorname{cov}(f, k, \varepsilon)$$

exists and is finite.

Proof. We will show that

$$\alpha(k) := \log \operatorname{cov}(f, k, \varepsilon) \tag{7.8}$$

is a subadditive function.

Suppose U has d_k^f-diameter less than ε and V has d_n^f-diameter less than ε. We claim that if $U \cap f^{-k}(V)$ is non-empty[4] then the d_{k+n}^f diameter of $U \cap f^{-k}(V)$ is also less than ε. Indeed, if $x, y \in U \cap f^{-k}(V)$ then since $x, y \in U$ one has

$$\max_{0\leq i\leq k-1} d\big(f^i(x), f^i(y)\big) < \varepsilon.$$

But also $f^k(x)$ and $f^k(y) \in V$, and hence

$$\max_{0\leq i\leq n-1} d\big(f^i(f^k(x)), f^i(f^k(y))\big) < \varepsilon.$$

Therefore

$$d_{k+n}^f(x, y) = \max_{0\leq i\leq k+n-1} d\big(f^i(x), f^i(y)\big) < \varepsilon,$$

as claimed.

Thus if $\{U_i\}_{i\in I}$ is a covering of X of sets with d_k^f-diameter at most ε and $\{V_j\}_{j\in J}$ is a covering of $X$of sets with d_n^f-diameter at most ε then

$$\{U_i \cap f^{-k}(V_j) \mid (i, j) \in I \times J\}$$

is a covering of X of sets with d_{k+n}^f-diameter at most ε. The cardinality of this covering is at most $\#I \cdot \#J$ (equality holds when each intersection $U_i \cap f^{-k}(V_j)$ is non-empty). This shows that

$$\operatorname{cov}(f, k+n, \varepsilon) \leq \operatorname{cov}(f, k, \varepsilon) \cdot \operatorname{cov}(f, n, \varepsilon).$$

Taking log of both sides shows that (7.8) satisfies (7.5). The result now follows from Fekete's Lemma 7.7. ■

[4] If we define the diameter of the empty set to be zero then we don't need to make a case distinction here.

COROLLARY 7.9. *Let $f\colon X \to X$ be a dynamical system on a compact metric space and let $\varepsilon > 0$. Then all four of the following quantities are finite:*

$$\begin{aligned}
\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{+} &:= \limsup_{k\to\infty} \frac{1}{k} \log \mathrm{span}(f,k,\varepsilon) \\
\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{-} &:= \liminf_{k\to\infty} \frac{1}{k} \log \mathrm{span}(f,k,\varepsilon) \\
\mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{+} &:= \limsup_{k\to\infty} \frac{1}{k} \log \mathrm{sep}(f,k,\varepsilon) \\
\mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{-} &:= \liminf_{k\to\infty} \frac{1}{k} \log \mathrm{sep}(f,k,\varepsilon)
\end{aligned}$$

REMARK 7.10. The quantities $\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{\pm}$ and $\mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{\pm}$ are less well-behaved that $\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$. In general it may happen that

$$\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{+} \neq \mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{-}$$

and

$$\mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{+} \neq \mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{-},$$

and thus unlike $\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$ the limits need not exist. Indeed, there are examples[5] of dynamical systems for which $\lim_{k\to\infty} \frac{1}{k} \log \mathrm{span}(f,k,\varepsilon)$ diverges for arbitrarily small values of ε.

Proof. By Proposition 7.6 we have

$$\mathsf{h}^{\mathrm{cov}}_{2\varepsilon}(f) \leq \mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{-} \leq \mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{-} \leq \mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{+} \leq \mathsf{h}^{\mathrm{span}}_{\varepsilon/2}(f)^{+} \leq \mathsf{h}^{\mathrm{cov}}_{\varepsilon/2}(f), \tag{7.9}$$

and the result follows by Proposition 7.8. ∎

Returning briefly to the heuristic discussion at the start of the lecture, we see that the quantities $\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^{+}$ and $\mathsf{h}^{\mathrm{sep}}_{\varepsilon}(f)^{+}$ are two (in general, different) ways of formalising the notion of the growth rate of orbits over time as seen through the eyes of a measuring device with error ε. The quantity $\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$ is slightly harder to understand heuristically, but it is mathematically better behaved (i.e. the limit exists). Moreover in the limit $\varepsilon \to 0$ it doesn't matter which quantity we use.

The quantity $\mathrm{cov}(f,k,\varepsilon)$ is obviously a monotonically increasing function as ε decreases, and hence $\varepsilon \mapsto \mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$ does as well. Thus the limit as $\varepsilon \to 0$ exists. We finally arrive at our main definition.

DEFINITION 7.11. Let $f\colon X \to X$ be a dynamical system on a compact metric space. The limit

$$\mathsf{h}_{\mathrm{top}}(f) := \lim_{\varepsilon\to 0} \mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f) \in [0,\infty]$$

exists and is called the **topological entropy** of f.

We can alternatively define $\mathsf{h}_{\mathrm{top}}(f)$ using the other two quantities:

[5] If I manage to think of a sufficiently simple one it will appear on a forthcoming Problem Sheet.

PROPOSITION 7.12. *Let $f\colon X \to X$ be a dynamical system on a compact metric space. Then*

$$\mathsf{h}_{\mathrm{top}}(f) = \lim_{\varepsilon\to 0} \mathsf{h}_{\varepsilon}^{\mathrm{span}}(f)^{-} = \lim_{\varepsilon\to 0} \mathsf{h}_{\varepsilon}^{\mathrm{span}}(f)^{+} = \lim_{\varepsilon\to 0} \mathsf{h}_{\varepsilon}^{\mathrm{sep}}(f)^{-} = \lim_{\varepsilon\to 0} \mathsf{h}_{\varepsilon}^{\mathrm{sep}}(f)^{+}.$$

Proof. Immediate from (7.9). ■

REMARK 7.13. As Remark 7.10 shows, one *cannot* compute the topological entropy using spanning or separating sets and just taking a limit—one must use either the limsup or the liminf. This subtle fact is occasionally stated incorrectly in textbooks.

The exact value of $\mathsf{h}_{\mathrm{top}}(f)$ is usually very hard to compute. However as we will see, what typically matters most is whether $\mathsf{h}_{\mathrm{top}}(f) = 0$ or $\mathsf{h}_{\mathrm{top}}(f) > 0$, and this is usually much easier to discover.

A priori, the quantity $\mathsf{h}_{\mathrm{top}}(f)$ might depend on the choice of metric d, since the quantities $\mathrm{span}(f,k,\varepsilon)$, $\mathrm{sep}(f,k,\varepsilon)$ and $\mathrm{cov}(f,k,\varepsilon)$ certainly do (cf. Remark 7.3). However luckily this is not the case, as we now prove.

PROPOSITION 7.14. *Let $f\colon X \to X$ be a dynamical system on a compact metric space. Then $\mathsf{h}_{\mathrm{top}}(f)$ does not depend on the choice of metric defining the topology on X.*

Proof. Suppose d_1 and d_2 are two different metrics defining the same topology on X. As in the proof of Lemma 4.5, we consider the function

$$\eta(r) \coloneqq \sup\{d_2(x,y) \mid d_1(x,y) \le r\}.$$

Since X is compact, $\eta(r) \to 0$ as $r \to 0$, and hence with the obvious notation we can alternatively compute

$$\mathsf{h}_{\mathrm{top}}(f,d_2) = \lim_{r\to 0}\lim_{k\to\infty} \frac{1}{k}\log \mathrm{cov}_{d_2}(f,k,\eta(r))$$

(this only works because we already know the limit exists).

If U has $(d_1)^f_k$-diameter at most r then U has $(d_2)^f_k$-diameter at most $\eta(r)$. Thus

$$\mathrm{cov}_{d_2}(f,k,\eta(r)) \le \mathrm{cov}_{d_1}(f,k,r),$$

and hence

$$\begin{aligned}\mathsf{h}_{\mathrm{top}}(f,d_2) &= \lim_{r\to 0}\lim_{k\to\infty} \frac{1}{k}\log \mathrm{cov}_{d_2}(f,k,\eta(r)) \\ &\le \lim_{r\to 0}\lim_{k\to\infty} \frac{1}{k}\log \mathrm{cov}_{d_1}(f,k,r) \\ &= \mathsf{h}_{\mathrm{top}}(f,d_1).\end{aligned}$$

Interchanging the roles of d_1 and d_2 shows that $\mathsf{h}_{\mathrm{top}}(f,d_1) \le \mathsf{h}_{\mathrm{top}}(f,d_2)$, and the result follows. ■

COROLLARY 7.15. *Topological entropy is a dynamical invariant.*

Proof. Suppose $f\colon X \to X$ is conjugate to $g\colon Y \to Y$ via a homeomorphism $H\colon X \to Y$:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "H"'] & X \arrow[d, "H"] \\ Y \arrow[r, "g"'] & Y \end{tikzcd}$$

Let d be a metric on Y. By Proposition 7.14 we are free to choose a convenient metric on X. For this we choose

$$\tilde{d}(x_1, x_2) := d\big(H(x_1), H(x_2)\big),$$

which is a metric on X generating the topology on X for which H is an isometry. For this choice of metric, we have

$$\begin{aligned} \tilde{d}^f_k(x_1, x_2) &= \max_{0 \le i \le k-1} \tilde{d}\big(f^i(x_1), f^i(x_2)\big) \\ &= \max_{0 \le i \le k-1} d\big(H(f^i(x_1)), H(f^i(x_2))\big) \\ &= \max_{0 \le i \le k-1} d\big(g^i(H(x_1)), g^i(H(x_2))\big), \\ &= d^g_k\big(H(x_1), H(x_2)\big). \end{aligned}$$

Since H is a bijection, it is clear that coverings as well as spanning and separated sets must have the same cardinality for both f and g. The result follows. ∎

On Problem Sheet D you will prove the following stronger version of Corollary 7.15:

Corollary 7.16. *If g is a factor of f then $\mathsf{h}_{\mathrm{top}}(g) \le \mathsf{h}_{\mathrm{top}}(f)$. In particular, having zero topological entropy is an inheritable property.*

(♣) Remark 7.17. Let us briefly outline how to (partially) remove our standing assumption that the underlying metric spaces are compact. Recall that a dynamical system $f\colon X \to X$ on a metric space (X, d) is said to be **uniformly continuous** with respect to d if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(x, y) < \delta \qquad \Rightarrow \qquad d(f(x), f(y)) < \varepsilon.$$

If X is compact then any dynamical system is uniformly continuous. However in the non-compact case this ceases to be case.

We say that two metrics d_1 and d_2 are **uniformly equivalent** if there exists a constant $c > 0$ such that

$$\frac{1}{c} d_1(x, y) \le d_2(x, y) \le c d_1(x, y), \qquad \forall\, x, y \in X.$$

Two uniformly equivalent metrics define the same topology, but the converse is false (even when the underlying space is compact). If d_2 is a uniformly equivalent metric to d_1 and $f\colon X \to X$ is uniformly continuous with respect to d_1 then f is also uniformly continuous with respect to d_2.

Suppose now that (X, d) is an arbitrary metric space and $f\colon X \to X$ is uniformly continuous with respect to d. If $K \subseteq X$ is compact then by mimicking the construction above one can define the topological entropy of f with respect to the compact set K, denoted by $\mathsf{h}_{\mathrm{top}}(f, d, K)$. We then define the **topological entropy** of f as

$$\mathsf{h}_{\mathrm{top}}(f, d) \coloneqq \sup \left\{ \mathsf{h}_{\mathrm{top}}(f, d, K) \mid K \subseteq X \text{ is compact} \right\}.$$

One can show that $\mathsf{h}_{\mathrm{top}}(f, d_1) = \mathsf{h}_{\mathrm{top}}(f, d_2)$ if d_1 and d_2 are uniformly equivalent. Moreover this new definition of topological entropy includes the original as a special case (since if X is compact we can take $K = X$).

LECTURE 8

Hyperbolic Toral Automorphisms

Over the next few lectures we will compute the topological entropy of most of our model dynamical systems. Along the way we will also introduce another important class of dynamical systems, called **hyperbolic toral automorphisms**. The results of our various computations are summarised in Table 8.1 below.

Dynamical system	Topological Entropy	Proved in
The circle rotation ρ_θ	$\mathsf{h}_{\mathrm{top}}(\rho_\theta) = 0$	Corollary 8.2
The circle expansion e_k	$\mathsf{h}_{\mathrm{top}}(e_k) = \log k$	Corollary 8.7
Any reversible system $f\colon S^1 \to S^1$	$\mathsf{h}_{\mathrm{top}}(f) = 0$	Proposition 8.8
Any reversible system $f\colon [0,1] \to [0,1]$	$\mathsf{h}_{\mathrm{top}}(f) = 0$	Problem D.3
A hyperbolic toral automorphism f_L	$\mathsf{h}_{\mathrm{top}}(f_L) = \log \lambda$	Theorem 8.18
Any expansive f	$\mathsf{h}_{\mathrm{top}}(f) < \infty$	Theorem 9.10
The tent map τ	$\mathsf{h}_{\mathrm{top}}(\tau) = \log 2$	Corollary 11.8
The logistic map $\lambda_4\vert_{[0,1]}$	$\mathsf{h}_{\mathrm{top}}(\lambda_4\vert_{[0,1]}) = \log 2.$	Corollary 11.9
The shift map σ	$\mathsf{h}_{\mathrm{top}}(\sigma) = \log 2$	Problem F.1

Table 8.1: The topological entropy of some of our model dynamical systems.

We begin with the following trivial computation.

Lemma 8.1. *Suppose $f\colon X \to X$ is a dynamical system on a compact metric space. Suppose there exists a metric d defining the topology on X for which f is an isometry. Then $\mathsf{h}_{\mathrm{top}}(f) = 0$.*

Proof. For such a metric d, one has $d_k^f = d$ for all k, and hence $\operatorname{cov}(f, k, \varepsilon) = \operatorname{cov}(f, 1, \varepsilon)$ for all k. Thus for any $\varepsilon > 0$ one has

$$\mathsf{h}_\varepsilon^{\mathrm{cov}}(f) = \lim_{k\to\infty} \frac{1}{k} \log \operatorname{cov}(f, k, \varepsilon) = 0,$$

and hence $\mathsf{h}_{\mathrm{top}}(f) = \lim_{\varepsilon\to 0} \mathsf{h}_\varepsilon^{\mathrm{cov}}(f) = 0$. ∎

Corollary 8.2. *Any circle rotation has zero topological entropy.*

Proof. A circle rotation is an isometry with respect to the usual metric on S^1 (cf. the proof of Problem A.2). ∎

The next result shows how entropy behaves with respect to products.

PROPOSITION 8.3. *Let $f\colon X \to X$ and $g\colon Y \to Y$ be dynamical systems on compact metric spaces. Then $\mathsf{h}_{\mathrm{top}}(f \times g) = \mathsf{h}_{\mathrm{top}}(f) + \mathsf{h}_{\mathrm{top}}(g)$.*

Proof. Consider the metric d on $X \times Y$ given by

$$d\big((x_1, y_1), (x_2, y_2)\big) \coloneqq \max\{d(x_1, x_2), d(y_1, y_2)\}$$

(where by a slight abuse of notation d also denotes both the metrics on X and Y). With this metric one has

$$d_k^{f\times g}\big((x_1, y_1), (x_2, y_2)\big) = \max\left\{d_k^f(x_1, x_2), d_k^g(y_1, y_2)\right\}.$$

Thus if $U \subseteq X$ has d_k^f-diameter at most ε and $V \subseteq Y$ has d_k^g diameter at most ε, then $U \times V$ has $d_k^{f\times g}$ diameter at most ε, whence it follows that

$$\mathrm{cov}(f \times g, k, \varepsilon) \le \mathrm{cov}(f, k, \varepsilon) \cdot \mathrm{cov}(g, k, \varepsilon),$$

and hence taking logs of both sides tells us that

$$\mathsf{h}_{\mathrm{top}}(f \times g) \le \mathsf{h}_{\mathrm{top}}(f) + \mathsf{h}_{\mathrm{top}}(g).$$

To see the other direction, with the same metric we observe that if $A \subseteq X$ is (k, ε)-separated with respect to f and $B \subseteq Y$ is (k, ε)-separated with respect to g then $A \times B$ is also (k, ε)-separated with respect to $f \times g$. Thus

$$\mathrm{sep}(f \times g, k, \varepsilon) \ge \mathrm{sep}(f, k, \varepsilon) \cdot \mathrm{sep}(g, k, \varepsilon),$$

and hence taking logs of both sides tells us that

$$\mathsf{h}_{\mathrm{top}}(f \times g) \ge \mathsf{h}_{\mathrm{top}}(f) + \mathsf{h}_{\mathrm{top}}(g).$$

Thus we must have $\mathsf{h}_{\mathrm{top}}(f \times g) = \mathsf{h}_{\mathrm{top}}(f) + \mathsf{h}_{\mathrm{top}}(g)$. This completes the proof. ∎

COROLLARY 8.4. *Any product of rotations*

$$\rho_{\theta_1} \times \cdots \times \rho_{\theta_n} \colon \mathbb{T}^n \to \mathbb{T}^n$$

has zero topological entropy.

Now we investigate how topological entropy behaves with respect to invariant sets.

PROPOSITION 8.5. *Suppose $f\colon X \to X$ is a dynamical system on a compact metric space. If $A \subseteq X$ is a closed invariant set then $\mathsf{h}_{\mathrm{top}}(f|_A) \le \mathsf{h}_{\mathrm{top}}(f)$. Moreover if $A_1, \dots, A_n$ are closed (not necessarily disjoint) invariant sets whose union is all of X then*

$$\mathsf{h}_{\mathrm{top}}(f) = \max_{1\le i\le n} \mathsf{h}_{\mathrm{top}}(f|_{A_i}).$$

Proof. Suppose that A is a closed invariant set. A subset $B \subseteq A$ that is (k,ε)-separated for $f|_A$ is also clearly (k,ε)-separated for f itself, and hence

$$\operatorname{sep}(f|_A, k, \varepsilon) \le \operatorname{sep}(f, k, \varepsilon).$$

Thus $\mathsf{h}_{\mathrm{top}}(f|_A) \le \mathsf{h}_{\mathrm{top}}(f)$. To see the second statement, note that if $B_i \subseteq A_i$ is a (k,ε)-spanning set for $f|_{A_i}$ then $\bigcup_{i=1}^n B_i$ is a (k,ε)-spanning set for f itself. Thus

$$\operatorname{span}(f,k,\varepsilon) \le \sum_{i=1}^n \operatorname{span}(f|_{A_i},k,\varepsilon) \le n \cdot \max_{1\le i\le n} \operatorname{span}(f|_{A_i},k,\varepsilon).$$

Thus

$$\begin{aligned}
\mathsf{h}^{\mathrm{span}}_{\varepsilon}(f)^- &= \liminf_{k\to\infty} \frac{1}{k}\log \operatorname{span}(f,k,\varepsilon) \\
&\le \lim_{k\to\infty} \frac{1}{k}\log n + \liminf_{k\to\infty} \max_{1\le i\le n} \frac{1}{k}\log\operatorname{span}(f|_{A_i},k,\varepsilon) \\
&= 0 + \max_{1\le i\le n} \liminf_{k\to\infty} \frac{1}{k}\log\operatorname{span}(f|_{A_i},k,\varepsilon).
\end{aligned}$$

Letting $\varepsilon \to 0$ shows that $\mathsf{h}_{\mathrm{top}}(f) \le \max_{1\le i\le m} \mathsf{h}_{\mathrm{top}}(f|_{A_i})$. Since $\mathsf{h}_{\mathrm{top}}(f|_{A_i}) \le \mathsf{h}_{\mathrm{top}}(f)$ for each i by the first part, we must have equality. This completes the proof. ∎

Here is a more involved explicit computation.

PROPOSITION 8.6. *The doubling map $e_2\colon S^1 \to S^1$ satisfies $\mathsf{h}_{\mathrm{top}}(e_2) = \log 2$.*

Proof. We endow S^1 with the usual metric d, given explicitly by

$$d(x,y) := \min\{|x-y|, 1-|x-y|\}, \tag{8.1}$$

where we think of S^1 as the unit interval $[0,1]$ with 0 and 1 identified. Observe that:

$$d(x,y) \le \frac{1}{4} \qquad \Rightarrow \qquad d(e_2(x), e_2(y)) = 2d(x,y). \tag{8.2}$$

Now let A_k denote the set of numbers of the form $\frac{i}{2^k}$ for $i = 0, 1, \dots, 2^k - 1$. Note A_k has cardinality 2^k. Fix $0 < \varepsilon < 1/4$ and choose $n \ge 2$ such that

$$\frac{1}{2^{n+1}} < \varepsilon \le \frac{1}{2^n}.$$

We claim that for any $k \ge 1$, A_{n+k} is a (k,ε)-spanning set for e_2. Indeed, if $x \in S^1$ then there exists $0 \le i \le 2^{n+k} - 1$ such that

$$\frac{i}{2^{n+k}} \le x < \frac{i+1}{2^{n+k}}.$$

Take

$$y := \frac{i+1}{2^{n+k}},$$

so that $y \in A_{n+k}$. Then $d(x,y) \le 2^{-n-k}$, and for any $0 \le i \le k-1$, one has using (8.2) repeatedly that

$$d\big(e_2^i(x), e_2^i(y)\big) \le 2^i d(x,y) < \frac{1}{2^{n+1}} < \varepsilon.$$

Next, we claim that A_{n+k-1} is a (k,ε)-separated set for e_2. For this take two distinct points $x, y \in A_{n+k-1}$. We must show that there exists $0 \le i \le k-1$ for which $d\big(e_2^i(x), e_2^i(y)\big) \ge \varepsilon$. Since $\varepsilon < 1/4$ by assumption we may assume (for if not, we are done) that $d\big(e_2^i(x), e_2^i(y)\big) \le 1/4$ for all $0 \le i \le k-1$, and hence by (8.2) we have

$$d\big(e_2^{k-1}(x), e_2^{k-1}(y)\big) = 2^{k-1} d(x,y).$$

Since $x \neq y$ both belong to A_{n+k-1}, we have $d(x,y) \ge 2^{1-k-n}$, and thus

$$d\big(e_2^{k-1}(x), e_2^{k-1}(y)\big) \ge \frac{2^{k-1}}{2^{n+k-1}} = \frac{1}{2^n} \ge \varepsilon.$$

We have therefore shown that

$$\operatorname{span}(e_2, k, \varepsilon) \le 2^{n+k} \qquad \text{and} \qquad \operatorname{sep}(e_2, k, \varepsilon) \ge 2^{n+k-1}.$$

Thus

$$\begin{aligned} \mathsf{h}_\varepsilon^{\mathrm{span}}(e_2)^+ &= \limsup_{k\to\infty} \frac{1}{k} \log \operatorname{span}(e_2, k, \varepsilon) \\ &\le \limsup_{k\to\infty} \frac{(n+k)\log 2}{k} \\ &= \log 2, \end{aligned}$$

and similarly

$$\begin{aligned} \mathsf{h}_\varepsilon^{\mathrm{sep}}(e_2)^+ &= \limsup_{k\to\infty} \frac{1}{k} \log \operatorname{sep}(e_2, k, \varepsilon) \\ &\ge \limsup_{k\to\infty} \frac{(n+k-1)\log 2}{k} \\ &= \log 2. \end{aligned}$$

Letting $\varepsilon \to 0$ shows that

$$\log 2 \le \mathsf{h}_{\mathrm{top}}(e_2) \le \log 2,$$

and thus $\mathsf{h}_{\mathrm{top}}(e_2) = \log 2$. This completes the proof. ■

Corollary 8.7. *The circle expansion $e_k \colon S^1 \to S^1$ satisfies $\mathsf{h}_{\mathrm{top}}(e_k) = \log k$.*

Proof. The proof is very similar to that of Proposition 8.6—one just uses fractions with denominators k^i instead of 2^i. ■

The dynamical system e_k is not reversible for $k \ge 2$, since it is not injective. We next prove that any reversible system on S^1 necessarily has zero topological entropy.

Proposition 8.8. *Suppose $f \colon S^1 \to S^1$ is a reversible dynamical system. Then $\mathsf{h}_{\mathrm{top}}(f) = 0$.*

Proof. We know that f maps intervals to intervals because the intervals are the connected subsets of S^1. Fix $\varepsilon > 0$ and let B be a (k, ε)-separated set for f of cardinality $\operatorname{sep}(f, k, \varepsilon)$. Let n denote the integer part of $1/\varepsilon$ and choose a subset $Y \subset S^1$ of $n+1$ evenly spaced points. Then any consecutive points $z, w \in Y$ satisfy $d(z, w) < \varepsilon$. Set

$$Y_k := \bigcup_{i=0}^{k-1} f^{-i}(Y),$$

then $\#Y_k \leq k(n+1)$. Now let I be an interval of $S^1 \setminus B$ having endpoints $x, y \in B$. Then $d\big(f^i(x), f^i(y)\big) \geq \varepsilon$ for some $0 \leq i \leq k-1$. The distance $d\big(f^i(x), f^i(y)\big)$ is the smallest of the lengths of the interval $f^i(I)$ and its complementary interval $S^1 \setminus f^i(I)$. Thus $f^i(I)$ must have length at least ε and therefore contains a point of Y. Thus I itself must contain a point of Y_k.

This shows that the cardinality of B, which is the same as the number of components of $S^1 \setminus B$, is less than or equal to the cardinality of Y_k. Hence

$$\operatorname{sep}(f, k, \varepsilon) \leq k(n+1),$$

and hence

$$\begin{aligned} \mathsf{h}_\varepsilon^{\mathrm{sep}}(f)^+ &= \limsup_{k \to \infty} \frac{1}{k} \log \operatorname{sep}(f, k, \varepsilon) \\ &\leq \limsup_{k \to \infty} \frac{1}{k} \log(k(n+1)) = 0. \end{aligned}$$

Letting $\varepsilon \to 0$ shows that $\mathsf{h}_{\mathrm{top}}(f) \leq 0$, and hence $\mathsf{h}_{\mathrm{top}}(f) = 0$. This completes the proof. ■

Remark 8.9. On Problem Sheet D you will prove that Proposition 8.8 is also true for any reversible dynamical system on $[0, 1]$. However as Theorem 8.18 shows, this result is not true for reversible dynamical systems on $\mathbb{T}^2$.

We now introduce another important class of dynamical systems on tori. We regard the n-dimensional torus $\mathbb{T}^n$ as $\mathbb{R}^n/\mathbb{Z}^n$ (equipped with quotient topology), and we let

$$\pi \colon \mathbb{R}^n \to \mathbb{Z}^n \tag{8.3}$$

denote the quotient map.

Definition 8.10. Let $L \colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear map which can be represented by a matrix (also denoted by L) whose entries are all integers. Then L induces a map

$$f_L \colon \mathbb{T}^n \to \mathbb{T}^n$$

such that the following diagram commutes:

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "L"] \arrow[d, "\pi"'] & \mathbb{R}^n \arrow[d, "\pi"] \\ \mathbb{T}^n \arrow[r, "f_L"'] & \mathbb{T}^n \end{tikzcd}$$

This map is continuous by definition of the quotient topology, and hence is a dynamical system on $\mathbb{T}^n$. We call f_L the **toral endomorphism induced by** L. If f_L is reversible we say that f is a **toral automorphism**.

The following easy lemma is left as an exercise.

LEMMA 8.11. *The dynamical system f_L is reversible if and only if $|\det L| = 1$.*

DEFINITION 8.12. A matrix L is said to be **hyperbolic** if $|\lambda| \neq 1$ for every eigenvalue λ of L.

See Remark 8.16 below for a partial explanation of the word "hyperbolic".

DEFINITION 8.13. Let $f_L \colon \mathbb{T}^n \to \mathbb{T}^n$ denote a toral endomorphism. We say that f_L is a **hyperbolic toral endomorphism** if L is a hyperbolic matrix, and that f_L is a **hyperbolic toral automorphism** if f_L is reversible and L is a hyperbolic matrix.

EXAMPLE 8.14. Take $n = 1$. Then the circle expansions $e_k \colon S^1 \to S^1$ are hyperbolic toral endomorphisms for $k \geq 2$.

To focus the ideas, let us now concentrate on the case $n = 2$. We come back to the general n in Remark 8.19 at the end of the lecture. For $n = 2$, an easy way to check hyperbolicity is to look at the trace:

LEMMA 8.15. *Let $L \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map represented by a matrix with integer entries. Suppose that $|\det L| = 1$ and that $|\operatorname{tr} L| > 2$. Then L is hyperbolic. Moreover the two eigenvalues λ_1 and λ_2 of L are real and irrational and there exists a unique $\lambda > 1$ such that $\{|\lambda_1|, |\lambda_2|\} = \{\lambda, \lambda^{-1}\}$.*

Proof. Let $t = \operatorname{tr} L$. Then the eigenvalues of L are given by

$$\lambda^{\pm} := \frac{t \pm \sqrt{t^2 - 4 \det L}}{2}.$$

If $|t| > 2$ then these are both real. Moreover these numbers are rational if and only if there exists a $k \in \mathbb{N}$ such that $t^2 \pm 4 = k^2$. However if such a k exists then $(t - k)(t + k) = \pm 4$, which implies that either

$$t + k = 4, \qquad t - k = 1$$

or

$$t + k = -1, \qquad t - k = -4.$$

No such integer k exists. ■

From now on to avoid needing to make case distinctions, let us assume that $|\operatorname{tr} L| > 2$ and that $\det L = 1$. Then the eigenvalues of L can be written as $\{\lambda, \lambda^{-1}\}$ for some $\lambda > 1$. Let v_1, v_2 be unit length eigenvectors of L such that

$$Lv_1 = \lambda v_1, \qquad Lv_2 = \lambda^{-1} v_2.$$

If $\hat{d}$ is a metric on $\mathbb{R}^n$ that is translation-invariant then $\hat{d}$ induces a metric d on $\mathbb{T}^n$ for which the map π from (8.3) becomes a local isometry. In this case

$$d\big(\pi(v), \pi(w)\big) = \hat{d}\big(\pi^{-1}([v]), \pi^{-1}([w])\big).$$

In the following it will be convenient to work with a metric on $\mathbb{R}^2$ that is nicely adapted to the eigendirections of L. Since $\{v_1, v_2\}$ is basis of $\mathbb{R}^2$, given any two vectors $v, w \in \mathbb{R}^2$ we can write

$$v - w = a_1 v_1 + a_2 v_2$$

for $a_1, a_2 \in \mathbb{R}$. We then define a metric $\hat{d}$ on $\mathbb{R}^2$ by

$$\hat{d}(v, w) := \max\{|a_1|, |a_2|\}. \tag{8.4}$$

This is translation-invariant metric on $\mathbb{R}^2$, and hence defines a metric d on $\mathbb{T}^2$. A "ball" in this metric is a parallelogram whose sides are parallel to v_1 and v_2. See Figure 8.1 below. If ℓ_i is a sufficiently short line segment in $\mathbb{T}^2$ parallel to v_i then $f_L(\ell_i)$ is another line segment in $\mathbb{T}^2$ whose length (with respect to the induced metric d on $\mathbb{T}^2$) is multiplied by λ or λ^{-1}. Since $\lambda > 1$, we see that f_L acts an **expansion** in the direction v_1 and as a **contraction** in the direction v_2.

Remark 8.16. Hyperbolic toral automorphisms will play a much greater role in the first half of Dynamical Systems II, when we study differentiable dynamics, and in particular, **hyperbolic dynamics**. The precise definition of "hyperbolic" is a little complicated, and so we defer it until next semester, but roughly speaking it means that the dynamical system expands in some directions and contracts in others. The argument above shows that f_L is indeed a hyperbolic dynamical system on $\mathbb{T}^2$.

As we will see in Dynamical Systems II, hyperbolicity will turn out to be the main mechanism for producing positive topological entropy for differentiable dynamical systems. Theorem 8.18 below should be thought of as an example of this mechanism at work.

On Problem Sheet D you will prove:

Proposition 8.17. *Any hyperbolic toral automorphism $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ is mixing.*

Here we compute topological entropy of a hyperbolic toral automorphism.

Theorem 8.18. *Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic toral automorphism as above. Then $\mathsf{h}_{\mathrm{top}}(f_L) = \log \lambda$.*

Proof. Let $\hat{d}$ denote the translation-invariant metric from (8.4). Abbreviate[1] by $\hat{d}_k = d_k^L$ and $d_k = d_k^{f_L}$. Since π is a local isometry for $\varepsilon > 0$ sufficiently small one has

$$B_{d_k}\big(\pi(v), r\big) = \pi\Big(B_{\hat{d}_k}(v, r)\Big) \tag{8.5}$$

A "ball" of radius r in the $\hat{d}_k$ metric of radius is a parallelogram whose sides are parallel to v_1 and v_2 and whose (Euclidean) lengths are $2\varepsilon\lambda^{-k}$ and 2ε respectively.

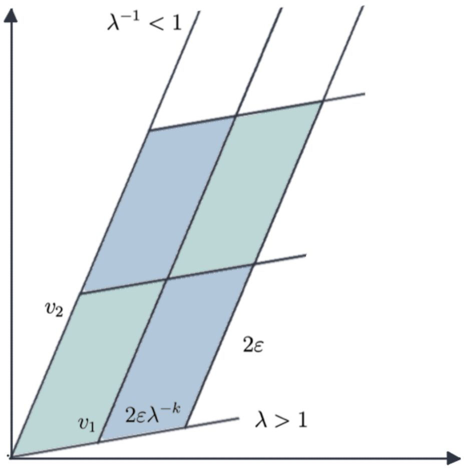

Figure 8.1: The $\hat{d}_k$ ball of radius ε.

See Figure 8.1 . The Euclidean area of such a ball is $4c\varepsilon^2\lambda^{-k}$, where $0 < c \leq 1$ is a constant that depends on the angle between v_1 and v_2 (if v_1 and v_2 are orthogonal then $c = 1$).

This means that to cover the unit square $[0,1] \times [0,1]$ (which has Euclidean area 1) we need at least $\frac{\lambda^k}{4c\varepsilon^2}$ such balls. Since any set with d_k-diameter less than ε is contained in a d_k-ball of radius ε, using (8.5) we see that

$$\operatorname{cov}(f_L, k, \varepsilon) \geq \frac{\lambda^k}{4c\varepsilon^2}, \qquad \text{for } \varepsilon \text{ sufficiently small,}$$

and hence

$$\begin{aligned} \mathsf{h}_{\text{top}}(f_L) &\geq \lim_{\varepsilon \to 0} \liminf_{k \to \infty} \frac{1}{k} \log \frac{\lambda^k}{4c\varepsilon^2} \\ &= \lim_{\varepsilon \to 0} \liminf_{k \to \infty} \frac{1}{k} \big(\log \lambda^k - \log 4c\varepsilon^2 \big) \\ &= \log \lambda. \end{aligned} \tag{8.6}$$

For the converse direction, we tile $\mathbb{R}^2$ by closed $\hat{d}_k$-balls (i.e. parallelograms). If ε is small enough then any such ball that intersects the unit square $[0,1] \times [0,1]$ is entirely contained in the larger square $[-1,2] \times [-1,2]$. This larger square has Euclidean area 9, and hence this tiling of $\mathbb{R}^2$ has at most $\frac{9\lambda^k}{c\varepsilon^2}$ balls that intersect the unit square. The image under π of all the open balls that intersect the unit square form an open cover of $\mathbb{T}^2$, and using (8.5) again we therefore obtain

$$\operatorname{cov}(f_L, k, \varepsilon) \leq \frac{9\lambda^k}{c\varepsilon^2}, \qquad \text{for } \varepsilon \text{ sufficiently small.}$$

[1]Strictly speaking, $\hat{d}_k^L$ does not fit into the framework of Definition 7.1 as $(\mathbb{R}^n, d)$ is not a compact metric space. This does not matter as far as the forthcoming proof is concerned.

Then

$$
\begin{aligned}
\mathsf{h}_{\mathrm{top}}(f_L) &\leq \limsup_{\varepsilon \to 0} \liminf_{k \to \infty} \frac{1}{k} \log \frac{9\lambda^k}{c\varepsilon^2} \\
&= \lim_{\varepsilon \to 0} \liminf_{k \to \infty} \frac{1}{k} \left(\log \lambda^k - \log \frac{1}{9} c\varepsilon^2 \right) \\
&= \log \lambda. \qquad (8.7)
\end{aligned}
$$

Comparing (8.6) and (8.7) we see that $\mathsf{h}_{\mathrm{top}}(f_L) = \log \lambda$. This completes the proof. ■

(♣) REMARK 8.19. Theorem 8.18 generalises to tori of arbitrary dimension. Suppose $f_L \colon \mathbb{T}^n \to \mathbb{T}^n$ has is a hyperbolic toral automorphism such that $\det L = 1$. Order the eigenvalues of L as

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_j| > 1 > |\lambda_{j+1}| > \ldots |\lambda_n|.$$

Then an argument similar to that of Theorem 8.18 shows that

$$\mathsf{h}_{\mathrm{top}}(f_L) = \sum_{i=1}^{j} \log |\lambda_i|. \qquad (8.8)$$

Note that (8.8) includes both Corollary 8.7 and Theorem 8.18 as a special case.

Expansive Dynamical Systems

In this lecture we introduce two related notions of expansiveness, and prove that expansive dynamical systems always have finite topological entropy. This lecture contains two reasonably tricky proofs, Theorem 9.9 and Theorem 9.10, which—at least in my opinion—are rather more involved than anything else we have done so far[1].

DEFINITION 9.1. Suppose $f\colon X \to X$ is a dynamical system on a metric space (X, d). We say that f is **expansive** if there exists a constant $\delta > 0$ such that

$$d\big(f^k(x), f^k(y)\big) \le \delta, \quad \forall\, k \ge 0 \qquad \Rightarrow \qquad x = y.$$

Any constant $\delta > 0$ with this property is called an **expansivity constant** for f.

REMARK 9.2. An expansivity constant is not unique, since if δ is an expansivity constant then so is δ' for any $0 < \delta' < \delta$. As in Remark 4.3 (which was concerned with the similar situation of sensitivity constants), one could take the supremum of all expansivity constants to obtain something unique. However this is rarely helpful, and typically hard to compute.

The doubling map is our prototypical example of a expansive map.

EXAMPLE 9.3. Consider S^1 with the standard metric. We claim that the doubling map $e_2\colon S^1 \to S^1$ is expansive with expansivity constant $\delta = 1/4$. Indeed, this follows immediately from equation (8.2). If $x, y \in S^1$ satisfy $d\big(e_2^k(x), e_2^k(y)\big) < 1/4$ then

$$d\big(e_2^{k+1}(x), e_2^{k+1}(y)\big) = 2d\big(e_2^k(x), e_2^k(y)\big).$$

Thus if $d\big(e_2^k(x), e_2^k(y)\big) < 1/4$ for all $k \ge 0$ then $d(x, y) = 0$.

Here is a slightly more exotic example.

EXAMPLE 9.4. Consider the logistic map λ_a for $a > 2 + \sqrt{5}$. Set

$$X \coloneqq \bigcap_{k=0}^{\infty} \lambda_a^{-k}([0, 1])$$

Then X is a compact λ_a-invariant subset of $[0, 1]$. Equip X with the metric inherited from the standard metric on $[0, 1]$. We claim that $\lambda_a|_X$ is expansive, and that moreover an expansivity constant can be taken as $0 < \delta < b \coloneqq \sqrt{1 - 4/a}$.

To see this, set

$$I \coloneqq [0, (1 - b)/2], \qquad J \coloneqq [(1 + b)/2, 1].$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]Hopefully I won't mess things up too badly in lecture...

Since $\lambda_a(x) = 1$ for $x = (1 \pm b)/2$ we have $I \cup J = \lambda_a^{-1}([0,1])$ and moreover that

$$|\lambda_a'(x)| = a|1-2x| \geq ab, \qquad \forall\, x \in I \cup J. \tag{9.1}$$

Now suppose $x, y \in X$ satisfy $|\lambda_a^k(x) - \lambda_a^k(y)| \leq \delta$ for all $k \geq 0$. Then for each $k \geq 1$, either $\lambda_a^k(x)$ and $\lambda_a^k(y)$ both belong to I or they both belong to J. Thus by the mean value theorem, one has from (9.1) that

$$\delta \geq |\lambda_a^k(x) - \lambda_a^k(y)| \geq (ab)^k |x-y|$$

for all $k \geq 1$. Since $a > 2 + \sqrt{5}$ one has $ab > 1$, and hence $x = y$.

Example 9.5. Consider Σ_2 with its natural metric d (Definition 4.15). In Problem F.1 you will prove that $\sigma \colon \Sigma_2 \to \Sigma_2$ is expansive.

The next result is similar to Lemma 4.5, and the proof is left as an exercise.

Lemma 9.6. *Let (X, d_1) be a compact metric space, and suppose $f \colon X \to X$ is a dynamical system which is expansive with respect to d_1. If d_2 is any other metric on X defining the same topology then f is expansive with respect to d_2.*

We emphasise however that the precise value of an expansivity constant *does* depend on the metric.

Remark 9.7. An expansive dynamical system automatically has sensitive dependence on initial conditions. However expansive systems are not necessarily chaotic, since the periodic points need not be dense.

In the reversible case another definition is possible.

Definition 9.8. Suppose $f \colon X \to X$ is a reversible dynamical system on a metric space (X, d). We say that f is **weakly expansive**[2] if if there exists a constant $\delta > 0$ such that

$$d\big(f^k(x), f^k(y)\big) \leq \delta, \quad \forall\, k \in \mathbb{Z} \qquad \Rightarrow \qquad x = y.$$

Again, any such constant $\delta > 0$ with this property is called an **weak expansivity constant** for f.

It is obvious that an expansive reversible dynamical system is also weakly expansive, but the converse is not true. In fact, provided the metric space is not a finite set, no expansive dynamical system is ever reversible:

Theorem 9.9. *Suppose X is an infinite compact metric space. Then any expansive dynamical system $f \colon X \to X$ is not injective.*

[2]**Warning:** The terminology "weakly expansive" is not standard. Many textbooks continue to use the name "expansive" for what we are calling "weakly expansive". This can be misleading, since—as Theorem 9.9 shows—expansive and weakly expansive reversible systems are *not* the same thing!

Proof. Assume[3] for contradiction that f is both injective and expansive, and let $\delta > 0$ be an expansivity constant for f. We prove the result in three steps.

1. We first claim there exists $p \geq 1$ such that

$$d\big(f^i(x), f^i(y)\big) \leq \delta \qquad \forall\, i = 1, \dots, p \qquad \Rightarrow \qquad d(x, y) \leq \delta. \tag{9.2}$$

Indeed, if (9.2) is not true then for every $k \geq 1$ we can find points x_k, y_k such that

$$d\big(f^i(x_k), f^i(y_k)\big) \leq \delta \qquad \forall\, i = 1, \dots, k \qquad \text{and} \qquad d(x_k, y_k) \geq \delta.$$

By compactness, we can choose convergent subsequences $x_{k_n} \to x$ and $y_{k_n} \to y$. Then $d(x, y) \geq \delta$; in particular $x \neq y$. However we claim that $f(x) = f(y)$. Indeed, by continuity for any $i \geq 0$ we have

$$d\big(f^i(x), f^i(y)\big) = \lim_{n \to \infty} d\big(f^i(x_{k_n}), f^i(y_{k_n})\big). \tag{9.3}$$

Since $d\big(f^i(x_{k_n}), f^i(y_{k_n})\big) \leq \delta$ for all $1 \leq i \leq k_n$ we deduce from (9.3) that

$$d\big(f^i(x), f^i(y)\big) \leq \delta, \qquad \forall\, i \in \mathbb{N}.$$

Then expansivity (applied to the points $f(x)$ and $f(y)$) tells us that $f(x) = f(y)$, which contradicts f being injective. This shows that (9.2) holds, and finishes the proof of Step 1.

2. In Step 2 we will show that this p actually has a stronger property: for any $x, y \in X$ and $m \geq 0$ if

$$d\big(f^i(x), f^i(y)\big) \leq \delta, \qquad \text{for} \qquad m + 1 \leq i \leq m + p, \tag{9.4}$$

then in fact

$$d\big(f^i(x), f^i(y)\big) \leq \delta \qquad 0 \leq i \leq m. \tag{9.5}$$

We prove this by induction on m; equation (9.2) was the case $m = 0$. For the inductive step, let us suppose that x, y satisfy

$$d\big(f^i(x), f^i(y)\big) \leq \delta, \qquad \text{for} \qquad m + 2 \leq i \leq m + 1 + p.$$

Then $d\big(f^i(f(x)), f^i(f(y))\big) \leq \delta$ for $m+1 \leq i \leq m+p$ so by applying the inductive hypothesis to $f(x)$ and $f(y)$ we have $d\big(f^i(f(x)), f^i(f(y))\big) \leq \delta$ for $i = 0, \dots, m$, and hence we have

$$d\big(f^i(x), f^i(y)\big) \leq \delta, \qquad \text{for} \qquad i = 1, \dots, m + 1 + p. \tag{9.6}$$

Applying (9.2) to (9.6) we also obtain $d(x, y) \leq \delta$. This finishes the inductive step.

3. We are now ready to prove the theorem. Since X is compact, there exist finitely many points $z_1, \dots, z_q$ such that the balls of radius $\frac{\delta}{2}$ in the d^f_{p+1} metric cover X. Since X is assumed to be an infinite metric space, in particular there exists a subset $A \subset X$ containing $q + 1$ points.

[3] This elegant proof is due to Coven and Keane.

For every $k \geq 0$ by the Pigeonhole Principle there exists $x_k \neq y_k$ in A such that $f^k(x_k)$ and $f^k(y_k)$ lie in the same ball $B_{d^f_{p+1}}\left(z_i, \frac{\delta}{2}\right)$. Thus from (9.4) and (9.5) we have that $d^f_{k+p+1}(x_k, y_k) \leq \delta$.

Since there are only finitely many pairs of these $q+1$ points, by the Pigeonhole Principle again there exists $x \neq y \in A$ such that $x_k = x$ and $y_k = y$ for infinitely many k. Thus $d\big(f^i(x), f^i(y)\big) \leq \delta$ for all $i \geq 0$. This violates expansivity. We have therefore obtained a contradiction to our assumption that f was both injective and expansive; the proof is complete. ∎

So far all the dynamical systems whose topological entropy we've computed[4] have had finite topological entropy. In general there is no reason for this to be the case, and on Problem Sheet F there is an explicit example of a dynamical system whose topological entropy is infinite. However, as the next pair of results show, any (weakly) expansive dynamical system necessarily has finite topological entropy.

Theorem 9.10. *Let $f \colon X \to X$ be an expansive dynamical system on a compact metric space. Then $\mathsf{h}_{\mathrm{top}}(f) < \infty$.*

Proof. Let δ denote a (weak) expansivity constant for f, and choose

$$0 < 4\gamma < \varepsilon < \delta. \tag{9.7}$$

We will prove that there exists $p = p(\gamma, \varepsilon) \geq 1$ such that

$$\mathrm{sep}(f, k, \gamma) \leq \mathrm{sep}(f, k + 2p, \varepsilon), \qquad \forall\, k \geq p. \tag{9.8}$$

Let us assume (9.8) for the moment and show how the result follows. Using Proposition 7.6 we then have

$$\mathrm{cov}(f, k, 2\gamma) \leq \mathrm{cov}\left(f, k + 2p, \tfrac{\varepsilon}{2}\right).$$

Thus

$$\begin{aligned}
\mathsf{h}^{\mathrm{cov}}_{2\gamma}(f) &= \lim_{k \to \infty} \frac{1}{k} \log \mathrm{cov}(f, k, \gamma) \\
&\leq \lim_{k \to \infty} \frac{1}{k} \log \mathrm{cov}\left(f, k + 2p, \tfrac{\varepsilon}{2}\right) \\
&= \lim_{n \to \infty} \frac{1}{n} \log \mathrm{cov}\left(f, n, \tfrac{\varepsilon}{2}\right) \\
&= \mathsf{h}^{\mathrm{cov}}_{\frac{\varepsilon}{2}}(f).
\end{aligned}$$

Since $\gamma < \frac{\varepsilon}{4}$ we always have $\mathsf{h}^{\mathrm{cov}}_{2\gamma}(f) \geq \mathsf{h}^{\mathrm{cov}}_{\frac{\varepsilon}{2}}(f)$, and thus

$$\mathsf{h}^{\mathrm{cov}}_{\gamma}(f) = \mathsf{h}^{\mathrm{cov}}_{\frac{\varepsilon}{2}}(f).$$

Since $0 < 4\gamma < \varepsilon < \delta$ were arbitrary, this shows that the quantity $\mathsf{h}^{\mathrm{cov}}_{r}(f)$ is independent of r for all $r < \frac{\delta}{2}$. Since for any fixed value of r the quantity $\mathsf{h}^{\mathrm{cov}}_{r}(f)$ is finite (by Proposition 7.8), we see that $\mathsf{h}_{\mathrm{top}}(f)$ is also finite.

[4] Admittedly, this is a rather short list.

It remains therefore to prove (9.8). We claim that there exists $p = p(\gamma, \varepsilon)$ such that:

$$d(x, y) \geq \gamma \qquad \Rightarrow \qquad d\big(f^i(x), f^i(y)\big) \geq \varepsilon \text{ for some } 0 \leq i \leq p. \tag{9.9}$$

The proof of (9.9) is similar to that of (9.2): if no such p exists then we find sequences x_k, y_k such that $d(x_k, y_k) \geq \gamma$ for all k but $d\big(f^i(x_k), f^i(y_k)\big) < \varepsilon$ for all $0 \leq i \leq k$. Passing to a subsequence we may assume that $x_k \to x$ and $y_k \to y$. Then $d(x, y) \geq \gamma$ but $d\big(f^i(x), f^i(y)\big) < \varepsilon$ for all $i \geq 0$. Since $\varepsilon < \delta$ this violates expansivity.

Finally[5] let us see how (9.9) implies (9.8). Let $k \geq p$ and suppose now that B is a maximal (k, γ)-separated set for f. Then for $z \neq w$ belonging to B, one has $d^f_k(z, w) \geq \gamma$. Thus there exists $0 \leq j \leq k - 1$ such that $d(f^j(z), f^j(w)) \geq \gamma$. Applying (9.9) with $x = f^j(z)$ and $y = f^j(w)$, we see there exists $0 \leq i \leq p$ such that

$$d\big(f^{i+j}(z), f^{i+j}(w)\big) \geq \varepsilon.$$

Since certainly $0 \leq i + j \leq k + p - 1$, this shows that $d^f_{k+p}(z, w) \geq \varepsilon$. This shows that B is $(k + p, \varepsilon)$-separated set for f. Since B was a maximal (k, γ)-separated set, we obtain

$$\operatorname{sep}(f, k + p, \varepsilon) \geq \#B = \operatorname{sep}(f, k, \gamma).$$

The proof is complete. ∎

Corollary 9.11. *Let X be compact metric space and let $f \colon X \to X$ be a reversible weakly expansive dynamical system. Then $\mathsf{h}_{\mathrm{top}}(f) < \infty$.*

Proof. The argument is mostly the same as in the proof of Theorem 9.10. If δ is a weak expansivity constant for f, then with γ, ε as in (9.7), we again claim there exists $p = p(\gamma, \varepsilon)$ such that (9.8) holds. Using weak expansivity, we first argue as in (9.9) that there exists $p = p(\gamma, \varepsilon)$ such that

$$d(x, y) \geq \gamma \qquad \Rightarrow \qquad d\big(f^i(x), f^i(y)\big) > \varepsilon \text{ for some } -p \leq i \leq p. \tag{9.10}$$

It then readily follows that if B is a (k, γ)-separated set then $f^{-p}(B)$ is a $(k+2p, \varepsilon)$-separated set. In the reversible case $\#B = \#f^{-p}(B)$, and thus $\operatorname{sep}(f, k + 2p, \varepsilon) \geq \operatorname{sep}(f, k, \gamma)$. The proof concludes as before. ∎

Remark 9.12. In practice, perhaps the most useful consequence of the *proof* of Theorem 9.10 and Corollary 9.11 is that for (weakly) expansive systems f, it is not necessary to take the limit as $\varepsilon \to 0$ of $\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$ in order to compute the topological entropy. This makes computing the topological entropy much easier. We will see an example of this in Lecture 11.

Remark 9.13. We will see another proof of Theorem 9.10 and Corollary 9.11 which is slightly slicker.

[5] Thank you to J. Hächl for simplifying this proof.

Topological Entropy via Open Covers

In this lecture we will first give another related way of defining expansivity. We will then build on this to give a new way of defining topological entropy. This new definition will be the main ingredient we use to find the link between topological entropy and measure-theoretic entropy in Lecture 28. Finally, we use our new definition of expansiveness and our new definition of entropy to give a second proof of Theorem 9.10 (and Corollary 9.11).

DEFINITION 10.1. Let X be a metric space and let $f\colon X \to X$ be a dynamical system. Let $\mathcal{U} = \{U_1, \dots, U_q\}$ be a finite open cover of X. We say that $\mathcal{U}$ is a **generator** for f if given any sequence $(i_k)_{k\geq 0}$ where $i_k \in \{1, \dots, q\}$ for each $k \geq 0$ the intersection

$$\bigcap_{k=0}^{\infty} f^{-k}(\overline{U}_{i_k})$$

is either empty or contains exactly one point.

In the reversible case we can also define the notion of a weak generator.

DEFINITION 10.2. Let X be a metric space and let $f\colon X \to X$ be a reversible dynamical system. Let $\mathcal{U} = \{U_1, \dots, U_q\}$ be a finite open cover of X. We say that $\mathcal{U}$ is a **weak generator** for f if given any sequence $(i_k)_{k\in\mathbb{Z}}$ where $i_k \in \{1, \dots, q\}$ for each $k \in \mathbb{Z}$ the intersection

$$\bigcap_{k=-\infty}^{\infty} f^{-k}(\overline{U}_{i_k})$$

is either empty or contains exactly one point.

If $\mathcal{U}$ is a generator for a reversible system f then clearly $\mathcal{U}$ is also a weak generator. The converse is not necessarily true though. In fact, it follows from Proposition 10.5 and Theorem 9.9 that reversible maps on infinite compact metric spaces *never* admit generators.

Let us now recall the following basic piece of point-set topology.

LEMMA 10.3. *Let X be a compact metric space and let $\mathcal{U}$ denote a (not necessarily finite) open cover of X. Then there exists $\delta > 0$ such that any set $A \subset X$ of diameter less than δ is contained in some element of $\mathcal{U}$.*

REMARK 10.4. In general a **Lebesgue number** for an open cover $\mathcal{U}$ of a metric space (X, d) is a positive number $\delta > 0$ with the property that any set $A \subset X$ of diameter less than δ is contained in some element of $\mathcal{U}$. Thus Lemma 10.3 tells us that on a compact metric space, every open cover admits a Lebesgue number.

The following proof is non-examinable, since it belongs to a course on real-analysis.

(♣) *Proof of Lemma 10.3.* Since X is compact, by passing to a finite subcover, we may assume $\mathfrak{U}$ is finite, say $\mathfrak{U} = \{U_1, \dots, U_q\}$. We may also assume that $U_i \neq X$ for each i, otherwise there is nothing to prove. Consider the function $a \colon X \to [0, \infty)$ given by

$$a(x) := \frac{1}{q} \sum_{i=1}^{q} d(x, X \setminus U_i).$$

This is a continuous positive function on a compact set, and hence $\delta \colon = \inf a > 0$. Then if A has diameter less that δ and $x \in A$ then $A \subset B(x, \delta)$. Since $a(x) \geq \delta$ there must exist at least one U_i such that $d(x, X \setminus U_i) \geq \delta$. Thus $B(x, \delta) \subset U_i$ for some i, and hence also $A \subset U_i$. This completes the proof. ■

We can now prove:

PROPOSITION 10.5. *Let $f \colon X \to X$ be a dynamical system on a compact metric space. Then:*

(i) *f is expansive if and only if f admits a generator.*

(ii) *If f is reversible then f is weakly expansive if and only if f admits a weak generator.*

Proof. We prove (ii) only; the proof of (i) is almost identical. Assume that f is weakly expansive, and let δ denote a weak expansivity constant for f. Since X is compact, there is a finite cover $\mathfrak{U} = \{U_1, \dots, U_q\}$ of X by open balls of radius $\frac{\delta}{2}$. Suppose we are given a sequence $(i_k)_{k \geq 0}$ where $i_k \in \{1, \dots, q\}$ for each $k \geq 0$. Suppose

$$x, y \in \bigcap_{k=-\infty}^{\infty} f^{-k}\left(\overline{U}_{i_k}\right)$$

Then in particular $d\big(f^k(x), f^k(y)\big) \leq \delta$ for all $k \in \mathbb{Z}$, whence by weak expansivity we have $x = y$. Thus $\mathfrak{U}$ is a weak generator.

For the converse, let $\mathfrak{U} = \{U_1, \dots, U_q\}$ be a weak generator for f and let δ be a Lebesgue number for $\mathfrak{U}$. Suppose that $d\big(f^k(x), f^k(y)\big) \leq \frac{\delta}{2}$ for all $k \in \mathbb{Z}$. Then for each k we can choose an element $U_{i_k} \in \mathfrak{U}$ such that $f^k(x), f^k(y) \in U_{i_k}$. Then both x and y belong to $\bigcap_{k \in \mathbb{Z}} f^{-k}\big(\overline{U}_{i_k}\big)$, whence $x = y$. Thus f is weakly expansive with weak expansivity constant δ. This completes the proof. ■

COROLLARY 10.6. *Let $f \colon X \to X$ be a dynamical system on a compact metric space. Then*

(i) *f is expansive if and only if f^k is expansive for all $k \geq 1$.*

(ii) *If f is reversible then f is weakly expansive if and only if f^k is weakly expansive for all $k \neq 0$.*

Proof. This easiest way to prove this is to use (weak) generators. Again, we prove (ii) only. If $\mathcal{U}$ is a weak generator for f then the collection of sets of the form

$$U_0 \cap f^{-1}(U_1) \cap f^{-2}(U_2) \cap \cdots \cap f^{-(k-1)}(U_{k-1}),$$

where $(U_0, \dots, U_{k-1})$ runs over all possible $|k|$-tuples of elements of $\mathcal{U}$, is a weak generator for f^k (this is true for any $k \neq 0$). Conversely any weak generator for f^k is also a weak generator for f. This completes the proof. ∎

Having seen that we can characterise (weakly) expansive maps via the existence of special open covers, we now proceed to study open covers in general.

Definition 10.7. Let X be a metric space. Suppose $\mathcal{U}$ and $\mathcal{V}$ are two (not necessarily finite) open covers of X. We define their **join** to be the open cover

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V \mid U \in \mathcal{U}, V \in \mathcal{V}\}.$$

Similarly given finitely[1] many open covers $\{\mathcal{U}_i\}$, $i = 1, \dots, k$, we define their join $\bigvee_{i=1}^k \mathcal{U}_i$ to be the open cover whose sets are all intersections of the form $\bigcap_{i=1}^k U_i$, where $U_i \in \mathcal{U}_i$.

Definition 10.8. Let X be a metric space. Suppose $\mathcal{U}$ and $\mathcal{V}$ are two open covers of X. We say that $\mathcal{V}$ is a **refinement** of $\mathcal{U}$ if for each $V \in \mathcal{V}$ there exists $U \in \mathcal{U}$ such that $V \subseteq U$. We write $\mathcal{U} \preceq \mathcal{V}$ to indicate that $\mathcal{V}$ is a refinement of $\mathcal{U}$.

Two examples of this are:

- If $\mathcal{V}$ is a subcover (i.e. a collection of elements of $\mathcal{U}$ that also covers X) of $\mathcal{U}$ then $\mathcal{U} \preceq \mathcal{V}$.
- If $\mathcal{U}$ and $\mathcal{V}$ are any two covers then

$$\mathcal{U} \preceq \mathcal{U} \vee \mathcal{V}, \qquad \text{and} \qquad \mathcal{V} \preceq \mathcal{U} \vee \mathcal{V}. \tag{10.1}$$

Definition 10.9. If $\mathcal{U}$ is an open cover of X and $f \colon X \to X$ is a dynamical system, we denote by $f^{-1}(\mathcal{U})$ the open cover

$$f^{-1}(\mathcal{U}) := \{f^{-1}(U) \mid U \in \mathcal{U}\}.$$

We define $f^{-k}(\mathcal{U})$ inductively.

Definition 10.10. Let $f \colon X \to X$ be a dynamical system on a metric space, and suppose $\mathcal{U}$ is an open cover of X. For $k \in \mathbb{N}$ we denote by $\mathcal{U}_f^k$ the open cover

$$\mathcal{U}_f^k := \bigvee_{i=0}^{k-1} f^{-i}(\mathcal{U}).$$

[1] Note we cannot form the join of infinitely many open covers, since in this case the intersections may fail to be open in X.

Thus

$$\mathcal{U}_f^1 = \mathcal{U}, \qquad \mathcal{U}_f^2 = \mathcal{U} \vee f^{-1}(\mathcal{U}),$$

and in a general a typical element of $\mathcal{U}_f^k$ is an intersection of the form

$$U_0 \cap f^{-1}(U_1) \cap \cdots \cap f^{-(k-1)}(U_{k-1}),$$

where $U_i \in \mathcal{U}$. Moreover for any $k, n \geq 0$ one has

$$\mathcal{U}_f^{k+n} = \mathcal{U}_f^k \vee f^{-k}(\mathcal{U}_f^n). \tag{10.2}$$

REMARK 10.11. With this language the proof of Corollary 10.6 can be shortened: if $\mathcal{U}$ is a generator for f then $\mathcal{U}_f^k$ is a generator for f^k.

The next lemma is immediate.

LEMMA 10.12. *Let $f\colon X \to X$ be a dynamical system on a metric space, and suppose $\mathcal{U}$ and $\mathcal{V}$ are two open covers of X. Then*

$$f^{-1}(\mathcal{U} \vee \mathcal{V}) = f^{-1}(\mathcal{U}) \vee f^{-1}(\mathcal{V}), \tag{10.3}$$

and

$$\mathcal{U} \preceq \mathcal{V} \qquad \Rightarrow \qquad f^{-1}(\mathcal{U}) \preceq f^{-1}(\mathcal{V}). \tag{10.4}$$

We will now work towards an alternative definition of topological entropy that uses open covers. This generalises the definition of entropy using $\mathsf{h}_\varepsilon^{\mathrm{cov}}$, and has the advantage that it gives us more freedom in the choice of open covers we use. This will be crucial in our proof of the Variational Principle in Lecture 28. The first step is to define the entropy of a cover.

DEFINITION 10.13. Let X be a compact metric space. If $\mathcal{U}$ is an open cover of X, we denote by $\min \mathcal{U}$ the number of sets in a finite subcover of $\mathcal{U}$ of minimal cardinality. We define the **entropy** of a cover $\mathcal{U}$ to be

$$\mathsf{H}(\mathcal{U}) := \log \min \mathcal{U}.$$

PROPOSITION 10.14. *Let $\mathcal{D}(\varepsilon)$ denote the cover consisting of all open sets with diameter strictly less than ε. Then*

$$\min \mathcal{D}(\varepsilon)_f^k = \operatorname{cov}(f, k, \varepsilon)$$

(♣) *Proof.* An element of $\mathcal{D}(\varepsilon)_k^f$ is an intersection of the form

$$V = U_0 \cap f^{-1}(U_1) \cap \cdots \cap f^{-(k-1)}(U_{k-1})$$

where each U_i has diameter less than ε. Let $x, y \in V$. Then for each $0 \leq i \leq k-1$ one has

$$f^j(x), f^j(y) \in U_j$$

and thus as U_j has diameter less than ε we have $d(f^j(x), f^j(y)) < \varepsilon$. This shows that

$$x, y \in V \qquad \Rightarrow \qquad d_k^f(x, y) < \varepsilon,$$

and hence

$$\operatorname{cov}(f,k,\varepsilon) \le \min \mathcal{D}(\varepsilon)^f_k.$$

The converse is rather trickier[2]. We will first show that if W is any set with d^f_k-diameter less than ε then there exists $V \in \mathcal{D}(\varepsilon)^f_k$ such that $W \subseteq V$.

Indeed, if W has d^f_k-diameter less than ε then there exists $0 < \delta < \varepsilon$ such that $\operatorname{diam}_{d^f_k} W \le \delta$. Now set

$$V := \bigcap_{i=0}^{k-1} f^{-i}\left(\bigcup_{x\in W} B_d\Big(f^i(x), \frac{\varepsilon-\delta}{3}\Big)\right).$$

Then clearly $W \subseteq V$. We claim that $V \in \mathcal{D}(\varepsilon)^f_k$. To see this we must show that

$$\operatorname{diam}_d\left(\bigcup_{x\in W} B_d\Big(f^i(x), \frac{\varepsilon-\delta}{3}\Big)\right) < \varepsilon.$$

Suppose

$$y_1, y_2 \in \bigcup_{x\in W} B_d\Big(f^i(x), \frac{\varepsilon-\delta}{3}\Big).$$

Then there exist $x_1, x_2 \in W$ such that

$$y_1 \in B_d\Big(f^i(x_1), \frac{\varepsilon-\delta}{3}\Big), \qquad y_2 \in B_d\Big(f^i(x_2), \frac{\varepsilon-\delta}{3}\Big).$$

Then

$$\begin{aligned} d(y_1,y_2) &\le d\big(y_1, f^i(x_1)\big) + d\big(f^i(x_1), f^i(x_2)\big) + d\big(f^i(x_2), y_2\big) \\ &\le \frac{\varepsilon-\delta}{3} + d^k_f(x_1,x_2) + \frac{\varepsilon-\delta}{3} \\ &< \varepsilon. \end{aligned}$$

Thus $V \in \mathcal{D}(\varepsilon)^f_k$ as claimed.

We now complete the proof of the converse direction. Suppose $\operatorname{cov}(f,k,\varepsilon) = n$. Then there exists an open covering $W_1, \dots, W_n$ of X of sets with d^f_k diameter less than ε. By the above procedure we can find sets $V_1, \dots, V_n$ that belong to $\mathcal{D}(\varepsilon)^f_k$ and satisfy $W_n \subseteq V_n$. Thus $V_1, \dots, V_n$ also covers X, and hence $\min \mathcal{D}(\varepsilon)^f_k \ge n$. This completes the proof. ■

Proposition 10.15. *Let $f\colon X \to X$ be a dynamical system on a metric space, and suppose $\mathcal{U}$ and $\mathcal{V}$ are two open cover of X. Then:*

(i) $\mathsf{H}(\mathcal{U}) \ge 0$, *and* $\mathsf{H}(\mathcal{U}) = 0$ *if and only if* $X \in \mathcal{U}$.

(ii) *If* $\mathcal{U} \preceq \mathcal{V}$ *then* $\mathsf{H}(\mathcal{U}) \le \mathsf{H}(\mathcal{V})$.

(iii) $\mathsf{H}(\mathcal{U} \vee \mathcal{V}) \le \mathsf{H}(\mathcal{U}) + \mathsf{H}(\mathcal{V})$.

[2] Thanks to J. Hächl for this argument.

(iv) $\mathsf{H}\big(f^{-1}(\mathcal{U})\big) \le \mathsf{H}(\mathcal{U})$. *If f is surjective then* $\mathsf{H}\big(f^{-1}(\mathcal{U})\big) = \mathsf{H}(\mathcal{U})$.

Proof. The proof of (i) is immediate. To prove (ii), suppose $\{V_1, \dots V_q\}$ is a subcover of $\mathcal{V}$ with minimum cardinality. For each $1 \le i \le q$ there exists $U_i \in \mathcal{U}$ such that $V_i \subseteq U_i$. Thus $\{U_1, \dots, U_q\}$ also covers X, and hence $\min \mathcal{U} \le q$.

To prove (iii), if $\{U_1, \dots, U_p\}$ and $\{V_1, \dots, V_q\}$ are subcovers of minimal cardinality, then $U_i \cap V_j$ for $1 \le i \le p$ and $1 \le j \le q$ is a subcover of $\mathcal{U} \vee \mathcal{V}$, and thus

$$\min \mathcal{U} \vee \mathcal{V} \le \min \mathcal{U} \cdot \min \mathcal{V}.$$

To prove (iv), if $\{U_1, \dots U_q\}$ is a subcover of $\mathcal{U}$ of minimum cardinality then the collection $\{f^{-1}(U_1,), \dots, f^{-1}(U_q)\}$ is a subcover of $f^{-1}\mathcal{U}$, so

$$\min f^{-1}(\mathcal{U}) \le \min \mathcal{U}.$$

If f is surjective and $\{f^{-1}(V_1), \dots, f^{-1}(V_p)\}$ is a subcover of $f^{-1}(\mathcal{U})$ of minimum cardinality then $\{V_1, \dots, V_p\}$ also covers X, and hence $\min \mathcal{U} \le \min f^{-1}(\mathcal{U})$. This completes the proof. ■

The next result generalises Proposition 7.8 to arbitrary covers.

Proposition 10.16. *Let $f\colon X \to X$ be a dynamical system on a compact metric space. If $\mathcal{U}$ is an open cover of X then the limit* $\lim_{k\to\infty} \frac{1}{k}\mathsf{H}(\mathcal{U}_f^k)$ *exists.*

Proof. This is another application of Fekete's Lemma 7.7. Define $\alpha\colon \mathbb{N} \to [0, \infty)$ by

$$\alpha(k) := \mathsf{H}\big(\mathcal{U}_f^k\big).$$

We will show that α is subadditive. Indeed,

$$\begin{aligned} \alpha(k+n) &= \mathsf{H}\big(\mathcal{U}_f^{k+n}\big) \\ &\le \mathsf{H}\big(\mathcal{U}_f^k\big) + \mathsf{H}\big(f^{-k}(\mathcal{U}_f^n)\big) \\ &\le \alpha(k) + \alpha(n) \end{aligned}$$

where the first inequality used (10.2) and part (iii) of Proposition 10.15, and the second inequality used part (iv) of Proposition 10.15. Now the result follows from Fekete's Lemma 7.7. ■

We can therefore make the following definition.

Definition 10.17. Let $f\colon X \to X$ denote a dynamical system on a compact metric space. Let $\mathcal{U}$ denote an open cover of X. We define the **entropy of f relative to** $\mathcal{U}$, written $\mathsf{h}^*(f, \mathcal{U})$ to be

$$\mathsf{h}^*(f, \mathcal{U}) = \lim_{k\to\infty} \frac{1}{k}\mathsf{H}\big(\mathcal{U}_f^k\big).$$

By taking the supremum over all open covers we get another quantity.

Definition 10.18. Let $f\colon X \to X$ be a dynamical system on a compact metric space. We define

$$\mathsf{h}_{\mathrm{top}}^*(f) := \sup_{\mathcal{U}} \mathsf{h}^*(f, \mathcal{U}),$$

where the supremum is over all open covers of X.

LEMMA 10.19. *Let $f\colon X \to X$ be a dynamical system on a compact metric space. Then*

$$\mathsf{h}_{\mathrm{top}}(f) \le \mathsf{h}^*_{\mathrm{top}}(f).$$

Proof. By Proposition 10.14 for any $\varepsilon > 0$ we have

$$\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f) = \mathsf{h}^*(f, \mathcal{D}(\varepsilon)),$$

and thus

$$\mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f) \le \mathsf{h}^*_{\mathrm{top}}(f), \qquad \forall\, \varepsilon > 0.$$

Letting $\varepsilon \to 0$ yields the result. ■

In fact, we always have equality:

$$\mathsf{h}_{\mathrm{top}}(f) = \mathsf{h}^*_{\mathrm{top}}(f). \tag{10.5}$$

We will prove this in Corollary 10.23 after a couple of preliminary statements.

REMARK 10.20. If we take (10.5) as for given for a moment, then at first sight, it may appear that we have taken the already rather complicated definition of topological entropy from Lecture 7 and made it. . . worse. Indeed, to compute $\mathsf{h}^*_{\mathrm{top}}(f)$, one is faced with with an ungainly supremum over all open covers, which certainly does *not* appear to be a pleasant thing to compute. Nevertheless, this new definition has a number of advantages over the old one. We will come back to this in Remark 10.24 below.

LEMMA 10.21. *If $\mathcal{U} \preceq \mathcal{V}$ then $\mathsf{h}^*(f, \mathcal{U}) \le \mathsf{h}^*(f, \mathcal{V})$.*

Proof. If $\mathcal{U} \preceq \mathcal{V}$ then by applying (10.4) repeatedly we see that

$$\mathcal{U}^k_f \preceq \mathcal{V}^k_f$$

for all $k \in \mathbb{N}$, and thus by part (ii) of Proposition 10.15 we obtain $\mathsf{h}^*(f, \mathcal{U}) \le \mathsf{h}^*(f, \mathcal{V})$. ■

Let us define the **diameter** of a cover $\mathcal{U}$ to be

$$\operatorname{diam} \mathcal{U} := \sup_{U \in \mathcal{U}} \operatorname{diam} U.$$

PROPOSITION 10.22. *Let X be a compact metric space and let $\{\mathcal{U}_n\}$ be a sequence of covers such that $\operatorname{diam} \mathcal{U}_n \to 0$. Then if $f\colon X \to X$ is any dynamical system on X one has*

$$\mathsf{h}^*_{\mathrm{top}}(f) = \lim_{n \to \infty} \mathsf{h}^*(f, \mathcal{U}_n)$$

(where both sides could be equal to ∞).

Proof. First suppose that $\mathsf{h}^*_{\mathrm{top}}(f) < \infty$. Fix $\epsilon > 0$ and let $\mathcal{V}$ denote an open cover with $\mathsf{h}^*(f, \mathcal{V}) > \mathsf{h}^*_{\mathrm{top}}(f) - \epsilon$. Let $\delta > 0$ denote Lebesgue number for $\mathcal{V}$. Choose $p \ge 1$ large enough such that for all $n \ge p$, one has $\operatorname{diam} \mathcal{U}_n < \delta$. Then $\mathcal{V} \preceq \mathcal{U}_n$ for all $n \ge p$, and hence $\mathsf{h}^*(f, \mathcal{V}) \le \mathsf{h}^*(f, \mathcal{U}_n)$ for all $n \ge p$.

Thus for $n \geq p$ one has

$$\mathsf{h}^*_{\mathrm{top}}(f) \geq \mathsf{h}^*(f, \mathcal{U}_n) > \mathsf{h}^*_{\mathrm{top}}(f) - \epsilon.$$

Since ϵ was arbitrary we deduce that $\mathsf{h}^*_{\mathrm{top}}(f) = \lim_{n\to\infty} \mathsf{h}^*(f, \mathcal{U}_n)$.

If $\mathsf{h}^*_{\mathrm{top}}(f) = \infty$ then for any $c > 0$ we can find an open cover $\mathcal{V}$ with $\mathsf{h}^*(f, \mathcal{V}) \geq c$. The same argument as above then shows that $\lim_{n\to\infty} \mathsf{h}^*(f, \mathcal{U}_n) = \infty$. This completes the proof. ■

We can now prove that the two definitions of entropy coincide.

Corollary 10.23. *Let $f \colon X \to X$ be a dynamical system on a compact metric space X. Then*

$$\mathsf{h}_{\mathrm{top}}(f) = \mathsf{h}^*_{\mathrm{top}}(f).$$

Proof. Apply Proposition 10.22 with $\mathcal{U}_k = \mathcal{D}_{1/k}$. ■

Remark 10.24. Let us now come back to Remark 10.20 and explain why computing $\mathsf{h}^*_{\mathrm{top}}(f)$ is often easier than computing $\mathsf{h}_{\mathrm{top}}(f)$ directly. Firstly, Lemma 10.21 tells us that in Definition 10.18 we need only take the supremum over *finite* covers $\mathcal{U}$ (since every open cover admits a finite subcover by compactness). This is already a major improvement over our definition of $\mathsf{h}_{\mathrm{top}}(f)$ as $\lim_{\varepsilon\to 0} \mathsf{h}^{\mathrm{cov}}_{\varepsilon}(f)$, since by Proposition 10.14 this is equivalent to working with the (very non-finite) cover $\mathcal{D}(\varepsilon)$. Secondly, Proposition 10.22 tells us that to compute $\mathsf{h}^*_{\mathrm{top}}(f)$ we are free to choose our favourite sequence of covers $(\mathcal{U}_n)$, provided their diameters go to zero. This freedom allows for major computational simplifications. Indeed, several of the proofs later on in the course will hinge upon choosing a "clever" such sequence $(\mathcal{U}_n)$.

To illustrate this idea, we conclude this lecture by giving a new proof of Theorem 9.10 that makes use of generators. We will see another example next lecture (cf. Step 3 of the proof of Theorem 11.7).

Theorem 10.25. *Let X be a compact metric space.*

(i) *Let $f \colon X \to X$ be an expansive dynamical system on X. Let $\mathcal{U}$ be a generator for f. Then*

$$\mathsf{h}^*_{\mathrm{top}}(f) = \mathsf{h}^*(f, \mathcal{U}),$$

and hence $\mathsf{h}^*_{\mathrm{top}}(f) < \infty$.

(ii) *Let $f \colon X \to X$ be a weakly expansive reversible dynamical system on X Let $\mathcal{U}$ be a weak generator for f. Then*

$$\mathsf{h}^*_{\mathrm{top}}(f) = \mathsf{h}^*(f, \mathcal{U}),$$

and hence $\mathsf{h}^*_{\mathrm{top}}(f) < \infty$.

Proof. We give the proof of (i) only. The modifications needed for (ii) are minor and can be safely left to you.

Suppose $\mathcal{U} = \{U_1, \dots, U_q\}$. We first show that for any $\varepsilon > 0$ there exists $p \in \mathbb{N}$ such that

$$\operatorname{diam} \mathcal{U}^p_f < \varepsilon. \tag{10.6}$$

The argument is very similar to the argument used to prove (9.9). If (10.6) is false, then we find sequences $(x_k), (y_k)$ of points, and tuples $(i_j^k)_{1\leq j\leq k}$ of integers such that $i_j^k \in \{1,\dots,q\}$ such that

$$d(x_k,y_k) \geq \varepsilon, \qquad \text{and} \qquad x_k, y_k \in \bigcap_{j=0}^{k} f^{-j}\big(U_{i_{j+1}^k}\big).$$

Since X is compact we may assume that $x_k \to x$ and $y_k \to y$. Then $d(x,y) \geq \varepsilon$, and hence in particular $x \neq y$. Next, by the Pigeonhole Principle there exists $i_0 \in \{1,\dots,q\}$ such that $i_0^k = i_0$ for infinitely many k. This implies that

$$x, y \in \overline{U}_{i_0}.$$

Similarly there exists $i_1 \in \{1,\dots,q\}$ such that $i_1^k = i_1$ for infinitely many k, and hence

$$x, y \in f^{-1}\big(\overline{U}_{i_1}\big).$$

Continuing inductively, we find a sequence (i_j) such that

$$x, y \in \bigcap_{j=0}^{\infty} f^{-j}\big(\overline{U}_{i_j}\big).$$

Since $x \neq y$, this contradicts $\mathcal{U}$ being a generator.

Now suppose that $\mathcal{V}$ is an arbitrary open cover. Let $\delta > 0$ be a Lebesgue number for $\mathcal{V}$, and choose $p \in \mathbb{N}$ such that

$$\operatorname{diam} \mathcal{U}_f^p < \delta.$$

Then $\mathcal{V} \preceq \mathcal{U}_f^p$, and hence by Lemma 10.21,

$$\mathsf{h}^*(f,\mathcal{V}) \leq \mathsf{h}^*\big(f,\mathcal{U}_f^p\big).$$

Then using (10.2) we compute

$$\begin{aligned}
\mathsf{h}^*\big(f,\mathcal{U}_f^p\big) &= \lim_{k\to\infty} \frac{1}{k} \mathsf{H}\left(\bigvee_{i=0}^{k-1} f^{-i}\big(\mathcal{U}_f^p\big)\right) \\
&= \lim_{k\to\infty} \frac{1}{k} \mathsf{H}\big(\mathcal{U}_f^{k+p}\big) \\
&= \lim_{n\to\infty} \frac{1}{n} \mathsf{H}\big(\mathcal{U}_f^{n}\big) \\
&= \mathsf{h}^*(f,\mathcal{U}).
\end{aligned}$$

This completes the proof. ■

LECTURE 11

Piecewise Monotone Dynamical Systems

In this lecture, we start by investigating another situation where it is possible to compute the topological entropy on the nose. This will allow us to finally compute the entropy of the tent map τ, thus concluding the list of computations we announced in Table 8.1.

DEFINITION 11.1. Let $f\colon [0,1] \to [0,1]$ be a dynamical system. We say that f is **piecewise monotone** if there exists $m \geq 1$ and closed intervals $I_1, \dots, I_m$ whose interiors are pairwise disjoint such that $\bigcup_{k=1}^m I_k = [0,1]$ and such that $f\colon I_k \to [0,1]$ is strictly monotone for each $k = 1, \dots, m$. We denote by $\mathrm{mon}(f)$ the minimal such m.

EXAMPLE 11.2. The tent map τ is piecewise monotone with $\mathrm{mon}(\tau) = 2$. More generally, the kth iterate τ^k has 2^{k-1} tents (see Figure 1.3 and the proof of Lemma 2.6), and hence $\mathrm{mon}(\tau^k) = 2^k$.

LEMMA 11.3. *Let $f, g\colon [0,1] \to [0,1]$ be two piecewise monotone dynamical systems. Then $f \circ g$ is also piecewise monotone, and moreover*

$$\mathrm{mon}(g) \leq \mathrm{mon}(f \circ g) \leq \mathrm{mon}(f) \cdot \mathrm{mon}(g).$$

Proof. Assume that f has maximal intervals of monotonicity $\{I_i\}_{i=1}^m$ and g has maximal intervals of monotonicity $\{J_j\}_{j=1}^n$. If $g^{-1}(I_i) \cap J_j \neq \emptyset$ then $f \circ g$ is strictly monotone on $g^{-1}(I_i) \cap J_j$. Thus $f \circ g$ is piecewise monotone, and each interval of monotonicity of $f \circ g$ is contained in an interval of monotonicity of g. This implies that $\mathrm{mon}(f \circ g) \geq \mathrm{mon}(g)$. Moreover

$$\begin{aligned}\mathrm{mon}(f \circ g) &\leq \# \left\{ (i,j) \mid g^{-1}(I_i) \cap J_j \right) \neq \emptyset \right\} \\ &\leq \mathrm{mon}(f) \cdot \mathrm{mon}(g).\end{aligned}$$

This completes the proof. ∎

COROLLARY 11.4. *Let $f\colon [0,1] \to [0,1]$ denote a piecewise monotone dynamical system. Then*

$$\mathrm{mon}(f^{i+j}) \leq \mathrm{mon}(f^i) \cdot \mathrm{mon}(f^j), \qquad \forall\, i, j \geq 0, \tag{11.1}$$

and

$$\mathrm{mon}(f^k) \leq \mathrm{mon}(f^{k+1}), \qquad \forall\, k \in \mathbb{N}. \tag{11.2}$$

Proof. Take $f = g$ in Lemma 11.3. ∎

REMARK 11.5. The bound (11.1) is sharp, as the tent map illustrated in Example 11.2.

Corollary 11.6. *Let $f\colon [0,1] \to [0,1]$ denote a piecewise monotone dynamical system. Then the quantity*

$$\mathsf{m}(f) \coloneqq \lim_{k\to\infty} \frac{1}{k} \log \operatorname{mon}(f^k)$$

exists and is a finite number.

Proof. The map $k \mapsto \log \operatorname{mon}(f^k)$ is subadditive by (11.1). Thus by Fekete's Lemma 7.7 the limit exists and is equal to $\inf_{k\ge 1} \frac{1}{k} \log \operatorname{mon}(f^k)$. ■

We now prove that $\mathsf{m}(f)$ agrees with the topological entropy of f. This is a fairly involved result.

Theorem 11.7. *Let $f\colon [0,1] \to [0,1]$ denote a piecewise monotone dynamical system. Then $\mathsf{m}(f) = \mathsf{h}_{\mathrm{top}}(f)$.*

This proof is non-examinable.

(♣) *Proof.* We prove the result in three steps.

1. In this first step, we prove that $\mathsf{m}(f) \ge \mathsf{h}_{\mathrm{top}}(f)$. Fix $\varepsilon > 0$ and $k \ge 1$. Let $n \coloneqq \operatorname{sep}(f, k, \varepsilon)$ and let $A = \{x_1, \dots, x_n\}$ denote a maximal (k, ε)-separated set for f. Thus for each $1 \le i \le n-1$ there exists $0 \le r_i \le k-1$ such that

$$\left| f^{r_i}(x_i) - f^{r_i}(x_{i+1}) \right| \ge \varepsilon.$$

The Pigeonhole Principle then tells us that there exists:

- a positive integer $q \ge \frac{n}{k}$;
- an integer $0 \le r \le k-1$;
- a subset
$$\{x_{i_1} < \cdots < x_{i_q}\} \subseteq A$$
of cardinality q,

such that

$$\left| f^{r}(x_{i_j}) - f^{r}(x_{i_{j+1}}) \right| \ge \varepsilon, \qquad 1 \le j < q. \tag{11.3}$$

Set

$$c \coloneqq \sup_{x\in[0,1]} |f(x)|.$$

Then (11.3) tells us that f^r has at least $\frac{q\varepsilon}{c}$ intervals of monotonicity, and hence

$$\operatorname{mon}(f^r) \ge \frac{q\varepsilon}{c} \ge \frac{n\varepsilon}{ck}.$$

Since $r \le k$ we have by (11.2) that also

$$\operatorname{mon}(f^k) \ge \frac{\varepsilon \cdot \operatorname{sep}(f, k, \varepsilon)}{ck}.$$

Thus

$$\begin{aligned}
\mathsf{m}(f) &= \lim_{k\to\infty} \frac{1}{k} \log \operatorname{mon}(f^k) \\
&\geq \limsup_{k\to\infty} \frac{1}{k} \Big(\log(\operatorname{sep}(f,k,\varepsilon) - (\log c + \log k - \log \varepsilon) \Big) \\
&= \limsup_{k\to\infty} \frac{1}{k} \log \operatorname{sep}(f,k,\varepsilon) \\
&\geq \lim_{k\to\infty} \frac{1}{k} \log \operatorname{cov}(f,k,2\varepsilon) \\
&= \mathsf{h}^{\mathrm{cov}}_{2\varepsilon}(f),
\end{aligned}$$

where the last inequality used Proposition 7.6. Since ε was arbitrary, we obtain

$$\mathsf{m}(f) \geq \lim_{\varepsilon\to 0} \mathsf{h}^{\mathrm{cov}}_{2\varepsilon}(f) = \mathsf{h}_{\mathrm{top}}(f).$$

2. Before proving the converse direction, in this step we will show that for any $p \in \mathbb{N}$ one has

$$\mathsf{m}(f^p) = p \cdot \mathsf{m}(f). \tag{11.4}$$

To prove this fix $p \in \mathbb{N}$ and set $g := f^p$.

$$\begin{aligned}
p \cdot \mathsf{m}(f) &= \lim_{k\to\infty} \frac{p}{k} \log \operatorname{mon}(f^k) \\
&= \limsup_{k\geq 1} \frac{p}{k} \log \operatorname{mon}(f^k) \\
&\overset{(\heartsuit)}{\geq} \limsup_{n\to\infty} \frac{1}{n} \log \operatorname{mon}(g^n) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \operatorname{mon}(g^n) \\
&= \mathsf{m}(g),
\end{aligned}$$

where $(\heartsuit)$ used the fact $\left(\frac{\log \operatorname{mon}(g^n)}{n}\right)_{n\geq 1}$ is a subsequence of $\left(\frac{p \log \operatorname{mon}(f^k)}{k}\right)_{k\geq 1}$. On the other hand by Lemma 11.3 we have

$$\begin{aligned}
p \cdot \mathsf{m}(f) &= \lim_{k\to\infty} \frac{p}{k} \log \operatorname{mon}(f^k) \\
&\leq \lim_{k\to\infty} \frac{p}{k} \log \Big(\operatorname{mon}\left(g^{\lfloor \frac{k}{p} \rfloor}\right) \cdot \operatorname{mon}\left(f^{k - \lfloor \frac{k}{p} \rfloor p}\right) \Big) \\
&\leq \lim_{k\to\infty} \frac{p}{k} \Big(\log \operatorname{mon}\left(g^{\lfloor \frac{k}{p} \rfloor}\right) + \max_{0\leq i\leq p-1} \log \operatorname{mon}(f^i) \Big) \\
&= \lim_{n\to\infty} \frac{1}{n} \operatorname{mon}(g^n).
\end{aligned}$$

3. In this last step we prove that $\mathsf{m}(f) \leq \mathsf{h}_{\mathrm{top}}(f)$, which combined with Step 1 completes the proof. It is convenient to use Corollary 10.23. Let us denote by $\{I_i\}_{i=1}^m$ the maximal intervals of monotonicity of f, ordered so that $0 \in I_1$ and the neighbouring intervals are labelled consecutively. Let $\mathcal{U}$ denote the open cover of

$[0,1]$ whose elements are open intervals formed from neighbouring intervals of the $\{I_i\}$:

$$U_1 := I_1 \cup I_2^\circ, \qquad U_m := I_{m-1}^\circ \cup I_m,$$

and for $2 \le i \le m-1$,

$$U_i := (I_{i-1} \cup I_i \cup I_{i+1})^\circ$$

Now for each $k \ge 1$, every non-empty and non-trivial[1] interval of the form

$$I_{i_0} \cap f^{-1}\big(I_{i_1}\big) \cap \cdots \cap f^{-(k-1)}\big(I_{i_{k-1}}\big)$$

is an interval of monotonicity for f^k which intersects at most 3^k elements of the cover $\mathcal{U}_f^k$ (the 3 comes from that each U_i intersects up to three of I_j). This shows that

$$\operatorname{mon}(f^k) \le 3^k \min \mathcal{U}_f^k.$$

Thus

$$\begin{aligned}
\mathsf{h}_{\mathrm{top}}(f) &= \mathsf{h}^*_{\mathrm{top}}(f) \\
&\ge \mathsf{h}^*(f, \mathcal{U}) \\
&= \lim_{k\to\infty} \frac{1}{k} \log \min \mathcal{U}_f^k \\
&\ge \lim_{k\to\infty} \frac{1}{k} \log \frac{\operatorname{mon}(f^k)}{3^k} \\
&= \lim_{k\to\infty} \frac{1}{k} \log \operatorname{mon}(f^k) - \log 3 \\
&= \mathsf{m}(f) - \log 3.
\end{aligned}$$

At first glance this estimate is not so useful, due to the annoying $-\log 3$. This is where Step 2 comes in handy. Since $\mathsf{h}_{\mathrm{top}}(f^p) = p\,\mathsf{h}_{\mathrm{top}}(f)$ by Problem D.2, using (11.4) we obtain for any $p \in \mathbb{N}$ that

$$\begin{aligned}
\mathsf{h}_{\mathrm{top}}(f) &= \frac{1}{p}\mathsf{h}_{\mathrm{top}}(f^p) \\
&\ge \frac{1}{p}(\mathsf{m}(f^p) - \log 3) \\
&= \mathsf{m}(f) - \frac{\log 3}{p}.
\end{aligned}$$

Since $p \ge 1$ was arbitrary, it follows that $\mathsf{h}_{\mathrm{top}}(f) \ge \mathsf{m}(f)$, and this completes the proof. ■

Theorem 11.7 gives a very simple way to compute the topological entropy of the tent map.

Corollary 11.8. *The tent map τ has $\mathsf{h}_{\mathrm{top}}(\tau) = \log 2$.*

Proof. By Example 11.2 we have

$$\mathsf{m}(\tau) = \lim_{k\to\infty} \frac{1}{k} \log \operatorname{mon}(\tau^k) = \lim_{k\to\infty} \frac{1}{k} \log 2^k = \log 2.$$

Thus by Theorem 11.7, $\mathsf{h}_{\mathrm{top}}(\tau) = \log 2$ as well. ■

[1] A "trivial" interval is, by definition, a single point.

Corollary 11.9. *The restriction of the logistic map* λ_4 *to* $[0,1]$ *has* $\mathsf{h}_{\mathrm{top}}\big(\lambda_4|_{[0,1]}\big) = \log 2$.

Proof. Use Lemma 1.21, Corollary 7.15 and Corollary 11.8. ■

We can also now give a quick solution to Problem D.3.

Corollary 11.10. *Let* $f\colon [0,1] \to [0,1]$ *be a reversible dynamical system. Then* $\mathsf{h}_{\mathrm{top}}(f) = 0$.

Proof. If f is reversible then f is strictly monotone; thus $\mathrm{mon}(f^k) = 1$ for all k. ■

We conclude this lecture by exploring another criterion that guarantees finite topological entropy.

Definition 11.11. Let f denote a dynamical system on a compact metric space (X, d). Define the **Lipschitz constant** $\mathrm{lip}_d(f)$ by

$$\mathrm{lip}_d(f) := \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x,y)} \in [0, \infty]$$

We say that f is **Lipschitz continuous** if $\mathrm{lip}_d(f) < \infty$.

Remark 11.12. Since X is compact, if d_1 and d_2 are any two metrics defining the topology on X then $\mathrm{lip}_{d_1}(f) < \infty$ if and only if $\mathrm{lip}_{d_2}(f) < \infty$.

Under a mild assumption on the metric space, Lipschitz continuous dynamical systems always have finite topological entropy. Before proving this, we need another definition. Let us denote by $\mathcal{B}(\varepsilon)$ the open cover of all balls of radius ε, and abbreviate by

$$\beta(\varepsilon) := \min \mathcal{B}(\varepsilon).$$

Thus $\beta(\varepsilon)$ is the minimum cardinality of a cover of X by open balls of radius ε.

Definition 11.13. Let (X, d) be a compact metric space. We define the **ball dimension** of (X, d) to be the quantity

$$\text{ball-dim}_d(X) := \limsup_{\varepsilon \to 0} \frac{\mathsf{H}\big(\mathcal{B}(\varepsilon)\big)}{|\log \varepsilon|} = \limsup_{\varepsilon \to 0} \frac{\log \beta(\varepsilon)}{|\log \varepsilon|}.$$

Thus the ball dimension is either a non-negative real number or $+\infty$. It is not necessarily an integer.

To get a feeling for this definition, let us prove:

Proposition 11.14. *The ball dimension of* $[0,1]^n$ *with the standard metric* d *is* n.

Proof. Consider first the case $n = 1$. One clearly has $\beta(\varepsilon) \geq \frac{1}{2\varepsilon}$. Moreover one also has $\beta(\varepsilon) \leq \lfloor 2 + \frac{1}{2\varepsilon} \rfloor$, since the $\lfloor 2 + \frac{1}{2\varepsilon} \rfloor$ balls centred at the points $k\varepsilon(2 - \varepsilon)$ for $k = 0, 1, \dots, \lfloor 2 + \frac{1}{2\varepsilon} \rfloor$ cover $[0,1]$.

Thus

$$\begin{aligned}\text{ball-dim}_d([0,1]) &\geq \lim_{\varepsilon\to 0} \frac{\log(1/2\varepsilon)}{|\log \varepsilon|} \\ &= \lim_{\varepsilon\to 0} \frac{\log 2 + \log \varepsilon}{\log \varepsilon} \\ &= 1,\end{aligned}$$

and similarly

$$\text{ball-dim}_d([0,1]) \leq \lim_{\varepsilon\to 0} \frac{\log(\lfloor 2 + \frac{1}{2\varepsilon} \rfloor)}{|\log \varepsilon|} = 1.$$

A similar argument shows that it takes (ignoring constants that don't matter) about $(1 + 1/\varepsilon)^n$ balls to cover $[0,1]^n$. Thus (ignoring constants) we have

$$\begin{aligned}\text{ball-dim}_d([0,1]^n) &= \limsup_{\varepsilon\to 0} \frac{\log(1+1/\varepsilon)^n}{|\log \varepsilon|} \\ &= \limsup_{\varepsilon\to 0} \frac{n \log \varepsilon}{\log \varepsilon} \\ &= n.\end{aligned}$$

This completes the proof. ■

(♣) Remark 11.15. The ball dimension is a way of assigning a "dimension" to spaces that do not have a well-defined "dimension" in the usual sense. It follows from Proposition 11.14 that the ball dimension of an n-dimensional topological manifold is n. Thus for well-behaved spaces, the ball dimension agrees with the standard notion of dimension.

The next result, whose proof is on Problem Sheet F, shows that for badly behaved spaces the ball dimension is rather harder to grok. It is *not* a topological invariant (i.e. homeomorphic spaces can have different ball dimensions).

Proposition 11.16. *Let C denote the Cantor ternary set obtained by iteratively deleting the open middle third from subintervals of* $[0,1]$*:*

$$C := [0,1] \setminus \left(\bigcup_{k=0}^{\infty} \bigcup_{n=0}^{3^k-1} \left(\frac{3n+1}{3^{k+1}}, \frac{3n+2}{3^{k+1}} \right) \right).$$

Then if d denotes the metric inherited from $[0,1]$*, one has*

$$\text{ball-dim}_d(C) = \frac{\log 2}{\log 3}.$$

Here then is our final promised result.

Proposition 11.17. *Let f denote a dynamical system on a compact metric space (X,d). Then*

$$\mathsf{h}_{\text{top}}(f) \leq \text{ball-dim}_d(X) \cdot \max\{0, \log \text{lip}(f)\}.$$

In particular, provided X has finite ball dimension, any Lipschitz continuous map has finite topological entropy.

Proof. There is noting to prove if ball-dim$_d(X) = \infty$ or lip$_d(f) = \infty$, so we may assume that f is Lipschitz continuous and (X, d) has finite ball dimension. Let

$$c > \max\{1, \mathrm{lip}(f)\}$$

be arbitrary, and choose $k \in \mathbb{N}$ and $\varepsilon \in (0, 1)$. Then, for any $0 \le i \le k$, we have

$$f^i\big(B_d(x, c^{-k}\varepsilon)\big) \subset B_d(f^i(x), \varepsilon).$$

It follows that

$$B_d(x, c^{-k}\varepsilon) \subset B_{d_k^f}(x, \varepsilon),$$

and thus also

$$\mathrm{span}(f, k, \varepsilon) \le \beta(c^{-k}\varepsilon).$$

Since $|\log(c^{-k}\varepsilon)| = k \log c - \log \varepsilon$ we can write

$$k = \left|\frac{\log(c^{-k}\varepsilon)}{\log c}\right| \left(1 + \frac{\log \varepsilon}{|\log(c^{-k}\varepsilon)|}\right)$$

The fraction in the last bracket tends to 0 as $k \to \infty$, and hence using Proposition 7.6 we obtain

$$\begin{aligned}
\mathsf{h}^{\mathrm{cov}}_{2\varepsilon}(f) &= \lim_{k\to\infty} \frac{1}{k} \log \mathrm{cov}(f, k, 2\varepsilon) \\
&\le \limsup_{k\to\infty} \frac{1}{k} \log \mathrm{span}(f, k, \varepsilon) \\
&\le \limsup_{k\to\infty} \frac{\log(\beta(c^{-k}\varepsilon))}{k} \\
&\le \log c \limsup_{k\to\infty} \frac{\log(\beta(c^{-k}\varepsilon))}{|\log(c^{-k}\varepsilon)|} \\
&= \log c \cdot \text{ball-dim}_d(X)
\end{aligned}$$

Letting $\varepsilon \to 0$ we obtain

$$\mathsf{h}_{\mathrm{top}}(f) \le \log c \cdot \text{ball-dim}_d(X).$$

Finally since $c > \max\{1, \mathrm{lip}(f)\}$ was arbitrary the result follows. ■

LECTURE 12

Chaos Versus Positive Entropy

In this lecture we investigate the relationship between chaos[1] and positive topological entropy. This is a remarkably rich and interesting story, and many seemingly basic problems remain unsolved.

Both chaos and positive entropy are ways of measuring the amount of complexity (or instability) in the system, and thus one would naturally infer that the two concepts are related. However in fact they are quite different: chaos is a *global* property, in the sense that it makes an assumption on the dynamics of f across the entire space X. Meanwhile positive topological entropy is a *local* property. Indeed, Proposition 8.5 shows that if f has an invariant set A such that $\mathsf{h}_{\mathrm{top}}(f|_A) > 0$, then also $\mathsf{h}_{\mathrm{top}}(f) > 0$. This means that on many spaces X, it is possible to construct dynamical systems which are not chaotic (e.g. for which the periodic points are not dense), but which have arbitrarily large topological entropy.

It is perhaps more reasonable to hope that if f is chaotic then $\mathsf{h}_{\mathrm{top}}(f)$ is positive.

Question A. Let X be a compact metric space. If $f \colon X \to X$ is chaotic, does it necessarily follow that $\mathsf{h}_{\mathrm{top}}(f) > 0$?

The answer to Question A is definitely *no* for some spaces X: for example, there exists[2] a chaotic dynamical system $f \colon \Sigma_2 \to \Sigma_2$ which has zero topological entropy. Now perhaps this doesn't surprise you: after all, Σ_2 is about as badly behaved a space as it is possible to be (it is homeomorphic to the Cantor Set, cf. Remark 4.17). So what about more reasonable spaces? For $X = [0,1]$ or $X = S^1$ the answer to Question A is *yes*, although in both cases this is difficult to prove (we will deal with the case $X = [0,1]$ in this lecture). Amazingly however, if $X = [0,1]^n$, or $X = S^n$, or $X = \mathbb{T}^n$, then for $n \geq 2$ the answer to Question A is[3] not known!

A related question concerns the infimum of the topological entropies of all possible chaotic maps. To this end, we define:

$$\mathsf{h}_{\mathrm{top}}^{\inf}(X) := \inf \{ \mathsf{h}_{\mathrm{top}}(f) \mid f \colon X \to X \text{ is chaotic} \}.$$

Then one can ask:

Question B. Let X be a compact metric space. Is $\mathsf{h}_{\mathrm{top}}^{\inf}(X) > 0$?

For $X = [0,1]$ the answer to Question B is again *yes*. Indeed, as we will prove in Theorem 12.1 below, the infimum is $\log \sqrt{2}$. Meanwhile for $X = S^1$ however the

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1] We remind the reader (who may get confused when googling) that by "chaos" we mean *Devaney chaos* (see Remark 4.22).

[2] An explicit example is constructed by Weiss here.

[3] To the best of my knowledge.

answer to Question B is *no*: it is possible to construct chaotic dynamical systems on S^1 with arbitrarily small topological entropy. In general Question B is slightly easier to answer than Question A, although there remains many open problems.

The following table summarises some of what is known:

Space X	**Chaos** $\Rightarrow$ **positive entropy?**	**What is** $\mathsf{h}_{\mathrm{top}}^{\mathrm{inf}}(X)$**?**
$X = [0,1]$	Yes	$\log\sqrt{2}$
$X = S^1$	Yes	0
$X = \Sigma_2$	No	0
$X = [0,1]^n$ for $n \geq 2$	Open problem!	0
$X = S^n$ for $n \geq 2$	Open problem!	0
$X = \mathbb{T}^n$ for $n \geq 2$	Open problem!	0
Most other spaces X	Open problem!	Open problem!

Table 12.1: Chaos versus positive topological entropy.

The main result of the next two lectures is the following theorem, which justifies the first row of Table 12.1.

THEOREM 12.1. *Let* $f\colon [0,1] \to [0,1]$ *be a chaotic dynamical system. Then* $\mathsf{h}_{\mathrm{top}}(f) \geq \log\sqrt{2}$.

A partial converse to Theorem 12.1 is also true.

THEOREM 12.2. *Let* $f\colon [0,1] \to [0,1]$ *be a dynamical system such that* $\mathsf{h}_{\mathrm{top}}(f) > 0$. *Then there exists a closed invariant set* $I \subseteq [0,1]$ *such that* $f|_I$ *is chaotic.*

We will prove Theorem 12.1 next lecture (see Theorem 13.18). Theorem 12.2 is actually slightly easier than Theorem 12.1, but we don't have time for both and Theorem 12.1 is the more important and interesting of the two. In today's lecture we will establish several intermediate results that will be needed in the proof of Theorem 12.1.

REMARK 12.3. Theorems 12.1 and 12.2 are stated for the unit interval $[0,1]$. However we could instead consider dynamical systems $f\colon [a,b] \to [a,b]$ defined on an arbitrary compact interval $[a,b]$ and obtain the same result. Indeed, this is immediate from Corollary 7.15, since $[a,b]$ is homeomorphic to $[0,1]$. We will use this observation without further comment (including in the next remark).

REMARK 12.4. The bound in Theorem 12.1 is sharp. To see this consider the dynamical system $f\colon [-1,1] \to [-1,1]$ defined by

$$f(x) := \begin{cases} 2x+2, & -1 \leq x \leq -1/2, \\ -2x, & -1/2 \leq x \leq 0, \\ -x, & 0 \leq x \leq 1. \end{cases}$$

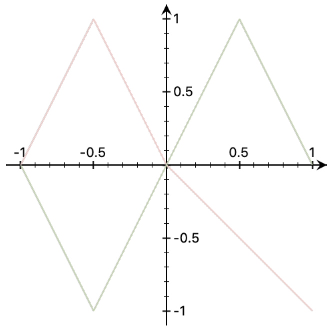

Figure 12.1: Plots of f and f^2.

See Figure 12.1.

By arguing as in the proof of Lemma 2.6, we see that f is transitive[4]. Thus by Proposition 12.6 below, f is chaotic, and therefore Theorem 12.1 tells us that $\mathsf{h}_{\mathrm{top}}(f) \geq \log\sqrt{2}$. On the other hand, from Figure 12.1 it is clear that the dynamical system f^2 is Lipschitz, with Lipschitz constant equal to 2. Thus[5] by Proposition 11.17 one has $\mathsf{h}_{\mathrm{top}}(f^2) \leq \log 2$. Since $\mathsf{h}_{\mathrm{top}}(f^2) = 2\,\mathsf{h}_{\mathrm{top}}(f)$ by Problem D.2, it follows that

$$\mathsf{h}_{\mathrm{top}}(f) = \log\sqrt{2}.$$

Remark 12.5. Combining Theorem 12.1 and Theorem 12.2 yields the following curiosity: *there exists no dynamical system $f\colon [0,1] \to [0,1]$ such that $0 < \mathsf{h}_{\mathrm{top}}(f) < \log\sqrt{2}$.* Thus the functional

$$\mathsf{h}_{\mathrm{top}}\colon \{\text{ dynamical systems on } [0,1] \} \to [0,\infty]$$

has a "gap".

The first step in the proof of Theorem 12.1 is the following result, which is interesting in its own right.

Proposition 12.6. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. Then $\mathsf{per}(f)$ is dense in $[0,1]$, and hence f is chaotic.*

Proposition 12.6 requires two preliminary lemmas.

Lemma 12.7. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous. If $f\big([a,b]\big) \subseteq [a,b]$ or $[a,b] \subseteq f\big([a,b]\big)$ then f has a fixed point.*

Proof. Let $g(x) := f(x) - x$. If $f\big([a,b]\big) \subseteq [a,b]$ then $g(a) \geq 0$ and $g(b) \leq 0$. By the Intermediate Value Theorem, there exists $c \in [a,b]$ such that $g(c) = 0$. If instead $[a,b] \subseteq f\big([a,b]\big)$ then there exists $a_1, b_1 \in [a,b]$ such that $f(a_1) \leq a$ and $f(b_1) \geq b$. Then $g(a_1) \leq 0$ and $g(b_1) \geq 0$, and thus as before we find c such that $g(c) = 0$. ■

The next result shows how Lemma 12.7 can be strengthened when f is transitive.

Lemma 12.8. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. Then:*

[4]The map f^2 is not transitive; thus this is an example of a dynamical system which is transitive but not weakly mixing (cf. Problem C.3).

[5]Alternatively, one could argue using Theorem 11.7 here.

(i) *There exists a fixed point a of f such that $0 < a < 1$.*

(ii) *The image of a non-trivial interval under f is another non-trivial interval.*

(iii) *f is surjective.*

The proof of Lemma 12.8 is deferred to Problem Sheet F. Instead we now prove a somewhat technical looking result, which is the lynchpin of the proof of Proposition 12.6.

LEMMA 12.9. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Let $(a,b) \subset [0,1]$ be an open interval such that $(a,b) \cap \mathsf{per}(f) = \emptyset$. Let $x, y \in (a,b)$ and suppose there exist $p, q \in \mathbb{N}$ such that $f^p(x) \in (a,b)$ and $f^q(y) \in (a,b)$. Then*

$$x < f^p(x) \qquad \Rightarrow \qquad y < f^q(y).$$

Proof. Assume that $f^p(x) > x$. We first prove by induction on k that

$$f^{kp}(x) > x, \qquad \forall\, k \in \mathbb{N}. \tag{12.1}$$

The case $k = 1$ is by assumption. For the inductive step, abbreviate by $g := f^p$ and let

$$z = g^j(x) := \min\left\{ g^i(x) \mid 1 \le i \le k-1 \right\},$$

so that $z > x$. Now assume for contradiction that $g^k(x) \le x$. Then by the Intermediate Value Theorem and connectedness,

$$[x,z] \subseteq \left[g^k(x), g^{k-j}(x)\right] \subset g^{k-j}\big([x,z]\big).$$

Lemma 12.7 tells us that g^{k-j} has a fixed point in $[x,z]$, and thus f has a periodic point in $[x,z]$. But since $z \le g(x)$ by definition, we have $[x,z] \subset (a,b)$, and this contradicts the assumption $(a,b) \cap \mathsf{per}(f) = \emptyset$.

Assume now that $f^q(y) \le y$. Then in fact $f^q(y) < y$ as y is not periodic. A similar induction argument as above (but in reverse order) tells us that

$$f^{kq}(y) < y, \qquad \forall\, k \in \mathbb{N}. \tag{12.2}$$

Combining (12.1) and (12.2) tells us that

$$x < f^{pq}(x), \qquad \text{and} \qquad y > f^{pq}(y). \tag{12.3}$$

f $x < y$ then $f^{pq}\big([x,y]\big) \subseteq [x,y]$. If $y < x$ then $[y,x] \subseteq f^{pq}\big([y,x]\big)$. In either case, Lemma 12.7 furnishes a point w lying between x and y such that $f^{pq}(w) = w$. Thus $w \in (a,b) \cap \mathsf{per}(f)$, which is a contradiction. The proof is complete. ∎

We can now prove Proposition 12.6.

Proof of Proposition 12.6. Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system, and suppose there exists $0 \le a < b \le 1$ such that $(a,b) \cap \mathsf{per}(f) = \emptyset$. By Proposition 2.9 there exists a point $x \in (a,b)$ with a dense orbit. Thus there exists $p, m, n \in \mathbb{N}$ with $m < n$ such that

$$x < f^p(x) < b, \qquad \text{and} \qquad a < f^n(x) < f^m(x) < x.$$

Set $y := f^m(x)$ and $q := n - m$. Then

$$a < f^q(y) < y < x < f^p(x) < b.$$

This contradicts Lemma 12.9. ∎

The eventual proof of Theorem 12.1 will make a case distinction depending whether the dynamical system is mixing or not. The rest of this lecture is devoted to understanding the difference between mixing and transitivity on the interval.

PROPOSITION 12.10. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Then f is mixing if and only if for every open interval $(a,b) \subset [0,1]$ and every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that*

$$[\varepsilon, 1-\varepsilon] \subset f^k\big((a,b)\big), \qquad \forall\, k \geq n. \tag{12.4}$$

Proof. Suppose first that f is mixing, and let $\varepsilon > 0$. Then there exists $n_1 \in \mathbb{N}$ such that

$$f^k\big((a,b)\big) \cap (0,\varepsilon) \neq \emptyset, \qquad \forall\, k \geq n_1,$$

and similarly there exists $n_2 \in \mathbb{N}$ such that

$$f^k\big((a,b)\big) \cap (1-\varepsilon,1) \neq \emptyset, \qquad \forall\, k \geq n_2,$$

Then if $n := \max\{n_1, n_2\}$, the set $f^k\big((a,b)\big)$ intersects both $(0,\varepsilon)$ and $(1-\varepsilon,1)$. Since $f^k\big((a,b)\big)$ is connected, this implies that $[\varepsilon, 1-\varepsilon] \subset f^k\big((a,b)\big)$ for all $k \geq n$.

Conversely, suppose that (12.4) holds. Let U, V be two nonempty open subsets of $[0,1]$. Choose open intervals $(a,b) \subset U$ and $(c,d) \subset V$ such that there exists $\varepsilon > 0$ such that $\varepsilon < c < d < 1-\varepsilon$. By (12.4) there exists $n \in \mathbb{N}$ such that

$$(c,d) \subset [\varepsilon, 1-\varepsilon] \subset f^k\big((a,b)\big), \qquad \forall\, k \geq n.$$

In particular, $f^k(U) \cap V \neq \emptyset$ for all $k \geq n$. Thus f is mixing, and the proof is complete. ■

COROLLARY 12.11. *Suppose $f\colon [0,1] \to [0,1]$ is mixing. Then there exists a periodic point whose minimal period is odd and at least three.*

Proof. Since f is transitive, the set $\mathsf{fix}(f)$ of fixed points has empty interior. Moreover $\mathsf{fix}(f)$ is closed. Thus there exists a non-empty open interval $(a,b) \subset [0,1]$ such that $[a,b] \cap \mathsf{fix}(f) = \emptyset$. By Proposition 12.10 there exists $n \in \mathbb{N}$ such that

$$[a,b] \subset f^k\big([a,b]\big), \qquad \forall\, k \geq n.$$

Let $q \geq n$ be odd. By Lemma 12.7 there exists a fixed point z of f^q in $[a,b]$. This point is not a fixed point of f by assumption. It it therefore a periodic point whose minimal period divides q. Since q is odd, the minimal period is not 2. ■

REMARK 12.12. We will see next lecture in Corollary 13.7 that the existence of a periodic point whose minimal period is *exactly* three implies the existence of periodic points with minimal periods of *all* orders.

The next result is on Problem Sheet F.

PROPOSITION 12.13. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. Suppose there exists a fixed point $a \in (0,1)$ such that either a lies in the interior of $f\big([0,a]\big)$, or a lies in the interior of $f\big([a,1]\big)$. Then f is mixing.*

So much for mixing dynamical systems. What about systems that are transitive but *not* mixing?

Proposition 12.14. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system that is not mixing. Then* $\mathsf{fix}(f) = \{a\}$ *for some $0 < a < 1$. Moreover*

$$f\big([0,a]\big) = [a,1], \qquad f\big([a,1]\big) = [0,a],$$

and both $f^2|_{[0,a]}\colon [0,a] \to [0,a]$ and $f^2|_{[a,1]}\colon [a,1] \to [a,1]$ are mixing dynamical systems.

Proof. By part (i) of Lemma 12.8 there exists a fixed point $a \in (0,1)$ of f. By Proposition 12.13 this fixed point does not lie in the interior of $f\big([0,a]\big)$ or in the interior of $f\big([a,1]\big)$. By part (ii) of Lemma 12.8 it follows that $f\big([0,a]\big) = [a,1]$ and $f\big([a,1]\big) = [0,a]$. Moreover $f^2|_{[0,a]}$ and $f^2|_{[a,1]}$ are transitive. By part (i) of Lemma 12.8 , $f^2|_{[0,a]}$ has a fixed point $b \in (0,a)$. This fixed point must satisfy the hypotheses of Proposition 12.13, as if not then arguing as above we would deduce that $f^2|_{[0,a]}$ permuted the two intervals $[0,b]$ and $[b,a]$, which is impossible as a is a fixed point. Then by Proposition 12.13, $f^2|_{[0,a]}$ is mixing. Similarly $f^2|_{[a,1]}$ is mixing. This completes the proof. ■

This allows us to extend Corollary 12.11 to an if and only if statement:

Corollary 12.15. *Suppose $f\colon [0,1] \to [0,1]$ is a transitive dynamical system. Then f is mixing if and only if there exists a periodic point whose minimal period is an odd number at least three.*

Proof. Proposition 12.14 tells us that if f is not mixing then every periodic point that is not a fixed point has even minimal period. ■

Another useful corollary is:

Corollary 12.16. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. If* $\#\mathsf{fix}(f) \geq 2$ *then f is mixing.*

LECTURE 13

Turbulent Dynamical Systems

In this lecture we introduce the notion of a **turbulent** dynamical system on the interval, and use this to prove Theorem 12.1. Along the way, we show that the existence of a point of period 3 implies the existence of periodic points of all order, which is a special case of the famous *Sharkovsky Theorem* that we will prove next lecture. We begin with some notation.

Definition 13.1. If $x, y \in \mathbb{R}$, we denote by $[\![x, y]\!]$ the interval whose endpoints are x and y. Thus $[\![x, y]\!] = [x, y]$ if $x \leq y$ and $[\![x, y]\!] = [y, x]$ if $y \leq x$. Similarly we denote by $(\!(x, y)\!)$ the interior of $[\![x, y]\!]$.

Here is the first key definition for today.

Definition 13.2. Let $I, J \subseteq [0,1]$ be closed[1] intervals, and let $f\colon [0,1] \to [0,1]$ be a dynamical system. We say that I **f-covers** J if $J \subseteq f(I)$. We write this symbolically as $I \to_f J$. If the dynamical system f is understood we will often say simply that I **covers** J and write $I \to J$.

The next lemma is merely a restatement of (half) of Lemma 12.7.

Lemma 13.3. *Suppose $I \to_f I$. Then f has a fixed point in I.*

Definition 13.4. Suppose $I_0, I_1, \dots, I_k$ are closed intervals such that $I_i \to_f I_{i+1}$ for $0 \leq i \leq k-1$. In this case we call $(I_0, \dots, I_k)$ a **chain of intervals** for f and write

$$I_0 \to_f I_1 \to_f \cdots \to_f I_k.$$

The next lemma is somewhat technical. The payoff is worth it, though.

Lemma 13.5. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Suppose $(I_0, \dots, I_p)$ is a chain of intervals for f. There exists a closed interval $J \subseteq I_0$ such that:*

(i) *$f^i(J) \subseteq I_i$ for $1 \leq i \leq p-1$;*

(ii) *$f^p(J^\circ) = I_p^\circ$ and $f^p(\partial J) = \partial I_p$.*

Proof. We will prove a stronger statement. We will show by induction on k (for $1 \leq k \leq p$) that there exist closed intervals

$$J_k \subseteq J_{k-1} \subseteq \cdots \subseteq J_1 \subseteq I_0 \tag{13.1}$$

such that

$$f^j(J_i) \subseteq I_j, \qquad \forall\, 1 \leq i \leq k \text{ and } 0 \leq j \leq i-1. \tag{13.2}$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1] Throughout this lecture we implicitly assume all our closed intervals are both non-empty and non-trivial: i.e., they are not just a single point.

and

$$f^i(J_i^\circ) = I_i^\circ \text{ and } f^i(\partial J_i) = \partial I_i, \qquad \forall\, 1 \le i \le k. \tag{13.3}$$

The case $k = p$ implies the lemma, with $J := J_p$.

First we do the base case $k = 1$. In this case (13.2) is implied by (13.1), so we need only construct $J_1 \subseteq I_0$ such that (13.3) holds for $i = k = 1$. Suppose $I_1 = [a, b]$. Then since $I_0 \to I_1$ there exist points $z_0, w_0 \in I_0$ such that $f(z_0) = a$ and $f(w_0) = b$. Then $f\big([\![z_0, w_0]\!]\big) = I_1$. This may not satisfy (13.3) however, since there could be a smaller closed subinterval of $[\![z_0, w_0]\!]$ which also maps onto I_1. To fix this suppose for definiteness that $z_0 \le w_0$ (the other case is symmetric). Then define

$$z := \min\{x \ge z_0 \mid f(x) = b\}, \qquad w := \max\{y \le z \mid f(y) = a\}.$$

Then if $J_1 := [w, z]$ then I_1 satisfies (13.3).

Now we prove the inductive step. Assume $1 \le k \le p-1$ and that we have already found closed intervals as in (13.1), (13.2), and (13.3). Since $f^{k+1}(J_k) = f(I_k)$ we have $J_k \to_{f^{k+1}} I_{k+1}$. Thus applying the case $k = 1$ to the map f^{k+1} and the two intervals J_k and I_{k+1}, we find a new closed interval $J_{k+1} \subseteq J_k$ such that

$$f^{k+1}(J_{k+1}^\circ) = I_{k+1}^\circ, \qquad \text{and} \qquad f^{k+1}(\partial J_{k+1}) = \partial I_{k+1}.$$

Thus (13.3) is satisfied. Moreover since $J_{k+1} \subseteq J_k$ we also have $f^j(J_{k+1}) \subseteq I_j$ for all $0 \le j \le k-1$, which shows that (13.2) holds too. This completes the proof. ∎

Our main use of Lemma 13.5 is the next corollary, which shows how the existence of a chain of intervals can be used to find periodic points.

Corollary 13.6. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Suppose $(I_0, \dots, I_p)$ is a chain of intervals for f such that $I_0 \subseteq I_p$. Then there exists a point $x \in \mathsf{per}(f) \cap I_0$ of (not necessarily minimal) period p such that*

$$f^i(x) \in I_i, \qquad \forall\, 1 \le i \le p. \tag{13.4}$$

Proof. Lemma 13.5 tells us there exists an interval $J \subseteq I_0$ such that $J \to_{f^p} J$. Lemma 13.3 tells us that f^p has a fixed point in J, and thus f has a periodic point $x \in J$ of period p. Finally, (13.4) is immediate from part (i) of Lemma 13.5. ∎

Corollary 13.6 is completely elementary—the proof was essentially nothing more than repeated applications of the Intermediate Value Theorem. Nevertheless, it has the following surprisingly deep consequence.

Corollary 13.7. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Suppose f has a point x of minimal period 3. Then for any $p \in \mathbb{N}$ there exists a periodic point of minimal period p.*

Proof. Replacing x with $f(x)$ or $f^2(x)$ if necessary, we may assume that

$$x < f(x) < f^2(x), \tag{13.5}$$

or

$$x < f^2(x) < f(x). \tag{13.6}$$

Assume to begin with that (13.5) holds, and set

$$I := [x, f(x)], \qquad J := [f(x), f^2(x)].$$

Then $I \to J$, $J \to J$, and $J \to I$. Now let $p \geq 2$, and consider the chain of intervals

$$I \to \underbrace{J \to \cdots \to J}_{p-1 \text{ times}} \to I$$

Corollary 13.6 implies that f has a point y of period p in I. Now suppose that $p \geq 4$ and that the minimal period m of y is less than p. Then $y = f^m(y) \in J$, and hence $y \in I \cap J$, i.e. $y = f(x)$. But then by (13.4) we have $f^i(f(x)) \in J$ for all $1 \leq i \leq p-1$. This is false for $i = 2$.

Finally, if (13.6) holds, the argument proceeds similarly, with $I = [x, f^2(x)]$ and $J = [f^2(x), f(x)]$. ∎

Corollary 13.7 should be thought of as a "baby" version of the Sharkovsky Theorem that we will state and prove next lecture. We next introduce the notion of a turbulent dynamical system, which will be helpful in the proof of Theorem 12.1.

Definition 13.8. Let $f\colon [0,1] \to [0,1]$ be a dynamical system. We say that f is **turbulent** if there exist two closed intervals $I, J \subset [0,1]$ such that $I \cap J$ contains at most one point and such that

$$I \cup J \subseteq f(I) \cap f(J).$$

Equivalently, this means that all four of the following holds:

$$I \to_f I, \qquad I \to_f J, \qquad J \to_f I, \qquad J \to_f J.$$

We call (I, J) a **turbulent pair** for f. If I and J can be chosen such that $I \cap J$ is empty then we say f is **strictly turbulent**, and that (I, J) is a **strictly turbulent pair**.

On Problem Sheet G you will show that if f is (strictly) turbulent then f^k is also (strictly) turbulent for all $k \geq 1$ (see part (i) of Problem G.5).

Example 13.9. The tent map $\tau\colon [0,1] \to [0,1]$ is turbulent. Indeed, one may take $I = [0, 1/2]$ and $J = [1/2, 1]$. However it is easy to see that the tent map is not strictly turbulent.

Remark 13.10. In the literature the name "horseshoe" is often used for what we have called a turbulent pair. Whilst this is certainly a visually appealing moniker, we will not use it. This is because there is another dynamical system commonly referred to as a horseshoe. This is a differentiable dynamical system on the square $[0,1]^2$. We will study it next semester in the context of hyperbolic dynamics (see Definition 47.11 in Lecture 47).

The next result gives an equivalent formulation of the turbulence condition. The proof is deferred to Problem Sheet G.

Lemma 13.11. *A dynamical system $f\colon [0,1] \to [0,1]$ is turbulent if and only if there exists $a, b, c \in [0,1]$ such that*

$$c \in (\!(a, b)\!), \qquad f(a) = f(b) = a, \qquad f(c) = b. \tag{13.7}$$

Turbulent maps have periodic points of all orders.

Proposition 13.12. *Let $f\colon [0,1] \to [0,1]$ be turbulent. Then f has a periodic point of minimal period 3.*

Proof. Assume to begin with that f is strictly turbulent, and let (I, J) be a strictly turbulent pair. Then we have a chain of intervals $I \to J \to J \to I$, and hence by Corollary 13.6, there is a point of period 3. Since $I \cap J = \emptyset$, this point is not a fixed point by (13.4).

Now assume f is just turbulent, and let (I, J) denote a turbulent pair. Then without loss of generality we may write $I = [a, b]$ and $J = [b, c]$ where $a < b < c$. If b is not a fixed point then we can argue as before. If instead b is a fixed point of f then we proceed as follows. Set

$$z := \min \{x \geq b \mid f(x) \in \{a, c\}\} .$$

Then the image of $[b, z)$ under f contains neither a or c. Thus $f\big([z, c]\big)$ contains both a and c since $[a, c] \subseteq f\big([b, c]\big)$. Therefore $[z, c] \subseteq f\big([z, c]\big)$, and hence $(I, [z, c])$ is a strictly turbulent pair for f. This completes the proof. ∎

The converse to Proposition 13.12 is not true: there exist dynamical systems which have a periodic point of minimal period 3 but are not turbulent. Let us now connect turbulence with the results from last lecture.

Proposition 13.13. *Let $f\colon [0,1] \to [0,1]$ be mixing. Then f^2 is strictly turbulent.*

The proof of Proposition 13.13 is deferred until next lecture (see Proposition 14.12), when we will prove it at the same time as the Sharkovsky Theorem. For now we concentrate on finishing the proof of Theorem 12.1. This requires a preliminary lemma, whose proof is deferred to Problem Sheet G.

Lemma 13.14. *Let $f\colon [0,1] \to [0,1]$ be a mixing dynamical system. Suppose that $0 \notin f\big((0,1]\big)$. Then there exists a sequence (x_k) of fixed points of f such that $x_k \to 0$.*

Proposition 13.15. *Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system which has at least two fixed points. Then f is turbulent.*

Proof. Suppose that f is transitive and has at least two fixed points. Then by Corollary 12.16, f is mixing. The closed set $\mathsf{fix}(f)$ has empty interior by transitivity, and hence we may assume that there exist fixed points $0 \leq a < b \leq 1$ such that $(a, b) \cap \mathsf{fix}(f) = \emptyset$. Thus either $f(x) > x$ for all $x \in (a, b)$, or $f(x) < x$ for all $x \in (a, b)$. We prove the first case only; the latter is similar.

If $f(x) > a$ for all $x \in (a, 1]$ then $[a, 1]$ is an invariant set. By transitivity, this implies $a = 0$. But then by Lemma 13.14 there exists a whole sequence of fixed points x_k of f tending to 0. This contradicts the hypothesis there are no fixed points in (a,b). Therefore there exists $z \in (a,1]$ such that $f(z) \leq a$. Since $f(x) > x$ for all $x \in (a,b)$, we must have $z \in (b,1]$. Moreover since $f(b) = b > a$, by the Intermediate Value Theorem there exists a point in $(b,z]$ which maps onto a. We can therefore define

$$d := \min\{x \in (a,1] \mid f(x) = a\}.$$

Suppose that $f(x) \neq d$ for all $x \in (a,d)$. Then $f(x) < d$ for all $x \in (a,d)$, and minimality of d implies that $f(x) > a$ for all $x \in (a,d)$. This means that $[a,d]$ is an invariant set. As before, this forces $a = 0$ and $d = 1$. But this then implies that 1 is not the image of f, and this contradicts the fact that f is surjective by part (iii) of Lemma 12.8.

We can therefore define

$$c := \min\{x \in (a,d) \mid f(x) = d\}.$$

Then

$$f\big([a,c]\big) = [a,d] = \qquad f\big([c,d]\big).$$

Thus $\big([a,c],[c,d]\big)$ is a turbulent pair for f. This completes the proof. ∎

Corollary 13.16. *Let $f\colon [0,1] \to [0,1]$ be transitive. Then f^2 is turbulent.*

Proof. If f is mixing then f^2 is turbulent by Proposition 13.13. If f is not mixing, then by Proposition 12.14, f has a unique fixed point a, which lies strictly between 0 and 1. Moreover

$$f\big([0,a]\big) = [a,1], \qquad f\big([a,1]\big) = [0,a],$$

and both $f^2|_{[0,a]}\colon [0,a] \to [0,a]$ and $f^2|_{[a,1]}\colon [a,1] \to [a,1]$ are mixing dynamical systems. The map $f^2|_{[0,a]}$ has at least two fixed points by part (i) of Lemma 12.8 (since a is a fixed point of $f^2||_{[0,a]}$). Thus by Proposition 13.15, $f^2|_{[0,a]}$ is turbulent, and hence also f^2 is turbulent. ∎

We now connect turbulence to entropy. It is the final step needed for Theorem 12.1.

Theorem 13.17. *Let $f\colon [0,1] \to [0,1]$ be a turbulent dynamical system. Then* $\mathsf{h}_{\mathsf{top}}(f) \geq \log 2$.

Proof. We prove only the case where f is strictly turbulent (which is sufficient for the proof of Theorem 12.1). The extension to the turbulent case is left for you on Problem Sheet G (see part (iii) of Problem G.5.)

So let (I_0, I_1) be a strictly turbulent pair. Choose disjoint open sets U_0, U_1 such that $I_0 \subset U_0$ and $I_1 \subset U_1$. Set $U_2 := [0,1] \setminus (I_0 \cup I_1)$. Then $\mathfrak{U} = \{U_0, U_1, U_2\}$ is an open cover of $[0,1]$. We shall prove that

$$\mathsf{h}^*(f, \mathfrak{U}) \geq \log 2.$$

Fix $k \in \mathbb{N}$ and let $(i_0, \dots, i_{k-1})$ be a k-tuple, where $i_j \in \{0,1\}$ for each j. Define

$$C(i_0, \dots, i_{k-1}) := \big\{x \in [0,1] \mid f^j(x) \in I_{i_j},\ \forall\, 0 \leq j \leq k-1\big\}.$$

Then by Lemma 13.5, each set $C(i_0, \dots, i_{k-1})$ is nonempty. Moreover it is contained in a unique element of the cover $\mathcal{U}(f,k)$, namely:

$$C(i_0, \dots, i_{k-1}) \subset U_{i_0} \cap f^{-1}\big(U_{i_1}\big) \cap \dots \cap f^{-(k-1)}\big(U_{i_{k-1}}\big).$$

Since there are 2^k different choices of tuples $(i_0, \dots, i_{k-1})$, this shows that

$$\min \mathcal{U}(f,k) \ge 2^k.$$

Thus

$$\mathsf{h}^*(f, \mathcal{U}) \ge \lim_{k \to \infty} \frac{1}{k} \log 2^k = \log 2.$$

This completes the proof. ■

Modulo the proof of Proposition 13.13, which will come next lecture, we can now finally prove Theorem 12.1, which for convenience we restate for transitive systems.

Theorem 13.18. *Let $f \colon [0,1] \to [0,1]$ be transitive. Then $\mathsf{h}_{\mathrm{top}}(f) \ge \log \sqrt{2}$.*

Proof. Corollary 13.16 tells us that f^2 is turbulent. Then $\mathsf{h}_{\mathrm{top}}(f^2) \ge \log 2$ by Theorem 13.17. Therefore $\mathsf{h}_{\mathrm{top}}(f) \ge \frac{1}{2} \log 2 = \log \sqrt{2}$ by Problem D.2. ■

Remark 13.19. With a bit more work, one can show that if $f \colon [0,1] \to [0,1]$ is transitive and has $\mathsf{h}_{\mathrm{top}}(f) = \log \sqrt{2}$ then f is conjugate to the system defined in Remark 12.4. Moreover if $f \colon [0,1] \to [0,1]$ has at least two fixed points and $\mathsf{h}_{\mathrm{top}}(f) = \log 2$ then f is conjugate to the tent map.

LECTURE 14

The Sharkovsky Theorem

In today's lecture we state and prove the famous *Sharkovsky Theorem*, which can be thought of as a massive generalisation of Corollary 13.7. Along the way we will also wrap up the one remaining result from the last two lectures that has yet to be proved (Proposition 13.13).

DEFINITION 14.1. We define a new ordering $\prec$ on $\mathbb{N}$ called the **Sharkovsky ordering** such that 1 is the smallest number and 3 is the largest number. Any element $n \in \mathbb{N}$ can be written uniquely as $n = 2^r p$ where p is odd and $r \in \mathbb{N} \cup \{0\}$. The ordering is given as follows:

$$\begin{aligned} 1 \prec 2 \prec 4 \prec \cdots \prec 2^r \prec 2^{r+1} \prec \cdots \\ \prec \cdots \\ \prec 2^r(2k+1) \prec 2^r(2k-1) \prec \cdots 2^r 7 \prec 2^r 5 \prec 2^r 3 \\ \prec 2^{r-1}(2k+1) \prec 2^{r-1}(2k-1) \prec \cdots 2^{r-1}7 \prec 2^{r-1}5 \prec 2^{r-1}3 \\ \prec \cdots \\ \prec 2k+1 \prec 2k-1 \prec \cdots 7 \prec 5 \prec 3. \end{aligned}$$

The symbols $\preceq$, $\succ$, and $\succeq$ are defined from $\prec$ as you would expect.

Here is the main result of today's lecture.

THEOREM 14.2 (The Sharkovsky Theorem). *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. If f has a periodic point with period $n > 1$ then f has a periodic point with period k for all $k \preceq n$.*

Corollary 13.7 is of course a baby case of Theorem 14.2, since by definition $k \preceq 3$ for all $k \in \mathbb{N}$. Although not strictly necessary, the proof of the Sharkovsky Theorem can be made more visual by introducing the notion of a *graph* of a periodic orbit.

DEFINITION 14.3. Let $f\colon [0,1] \to [0,1]$ be a dynamical system, and suppose $x \in [0,1]$ is a periodic point of minimal period $p \geq 2$. Enumerate the orbit as

$$\mathcal{O}_f(x) = \{x_1 < x_2 < \cdots < x_p\},$$

and let

$$I_i := [x_i, x_{i+1}], \qquad 1 \leq i \leq p-1.$$

The **graph of the periodic orbit** $\mathcal{O}_f(x)$ is the directed graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ with vertex set

$$\mathsf{V}\big(\mathcal{O}_f(x)\big) = \{I_i \mid 1 \leq i \leq p-1\},$$

and such that there is an edge from I_i to I_j if and only if

$$I_j \subseteq [\![f(x_i), f(x_{i+1})]\!] \tag{14.1}$$

Remark 14.4. **Warning:** Equation (14.1) is **not** quite the same as requiring that $I_j \subseteq f(I_i)$. We will therefore use a **bold** arrow to denote this relation:

$$I_i \boldsymbol{\to} I_j \qquad \Leftrightarrow \qquad (14.1) \text{ is satisfied.}$$

Since we always have (by the Intermediate Value Theorem) that

$$[\![f(x_i), f(x_{i+1})]\!] \subseteq f(I_i),$$

we see that (14.1) is a stronger condition than Definition 13.2, i.e.

$$I_i \boldsymbol{\to} I_j \qquad \Rightarrow \qquad I_i \to I_j.$$

See Figure 14.1 below for an example of such a graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$. The following lemma is (almost) immediate from the definition; see Problem G.4.

Lemma 14.5. *Let $J \in \mathsf{V}\big(\mathcal{O}_f(x)\big)$, and suppose $y \in \partial J$. Then there is a unique vertex $K \in \mathsf{V}\big(\mathcal{O}_f(x)\big)$ such that $J \boldsymbol{\to} K$ and such that $f(y) \in \partial K$.*

Definition 14.6. A **cycle of length** k is a cycle $J_0 \boldsymbol{\to} J_1 \boldsymbol{\to} \cdots \boldsymbol{\to} J_{k-1} \boldsymbol{\to} J_0$ in the graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$. A cycle is called **primitive** if it is not obtained by iterating a shorter cycle.

We will be interested in a special class of cycles.

Definition 14.7. Suppose x has minimal period p. A cycle $J_0 \boldsymbol{\to} J_1 \boldsymbol{\to} \cdots \boldsymbol{\to} J_{p-1} \boldsymbol{\to} J_0$ of length p in the graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ is called **fundamental** if there exists an endpoint[1] y of J_0 such that

$$f^k(y) \in \partial J_k \text{ for all } 0 \leq k \leq p-1.$$

It is not immediate from the definition that fundamental cycles exist. Let us prove that they do.

Proposition 14.8. *Let x be a periodic point of period p. There exists a unique (up to cyclic permutation) fundamental cycle in the graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$. Moreover in the fundamental cycle, no vertex appears more than twice, and at least one vertex appears exactly twice. The fundamental cycle can be decomposed into two shorter primitive cycles.*

Proof. Let $\mathcal{O}_f(x) = \{x_1 < x_2 < \cdots < x_p\}$ and set $I_i := [x_i, x_{i+1}]$ for $i = 1, \ldots, p-1$. Set $J_0 := I_1$. Lemma 14.5 tells us there exists a unique sequence $(J_k)_{k \geq 0} \subset \mathsf{V}\big(\mathcal{O}_f(x)\big)$ such that

$$f^k(x_1) \in \partial J_k, \qquad \text{and} \qquad J_k \boldsymbol{\to} J_{k+1}. \tag{14.2}$$

Since $f^p(x_1) = x_1$ and $x_1 < x_i$ for $2 \leq i \leq p$, the interval J_p must be equal to J_0. Thus $J_0 \boldsymbol{\to} J_1 \boldsymbol{\to} \cdots J_{p-1} \boldsymbol{\to} J_0$ is a fundamental cycle. This proves existence.

To prove uniqueness, suppose $K_0 \boldsymbol{\to} K_1 \boldsymbol{\to} \cdots \boldsymbol{\to} K_{p-1} \boldsymbol{\to} K_0$ is another fundamental cycle. Let $y \in \partial K_0$ satisfy $f^k(y) \in \partial K_k$ for each k. There exists a

[1] Note that if y is an endpoint of a vertex in $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ then y must belong to $\mathcal{O}_f(x)$.

unique $0 \leq j \leq p-1$ such that $y = f^j(x_1)$. Then $f^{p-j}(y) = f^p(x_1) = x_1$ is an endpoint of K_{p-j}. Thus $K_{p-j} = J_0$, and the uniqueness part of Lemma 14.5 tells us that

$$(K_0, K_1, \ldots, K_{p-1}, K_0) = (J_j, J_{j+1}, \ldots, J_{p-1}, J_0, J_1, \ldots, J_j).$$

Thus the fundamental cycle is unique up to a cyclic permutation.

Next, for each $1 \leq k \leq p-1$, there exist two distinct integers i, j such that $I_k = [f^i(x_1), f^j(x_1)]$. Therefore J_i and J_j are the only two vertices of the fundamental cycle that may be equal to I_k. Thus each vertex I_k appears at most twice in the fundamental cycle. Since the fundamental cycle has length p and there are only $p-1$ vertices in $\mathsf{G}\big(\mathcal{O}_f(x)\big)$, at least one vertex must appear exactly twice.

Finally, if we cut the fundamental cycle at the unique vertex that appears twice, we obtain two shorter cycles. Each of these cycles is necessarily primitive. ■

We now reinterpret Lemma 13.6 in terms of the graph of a periodic orbit. This lemma shows us how to find other periodic points of f provided we understand the structure of the graph of one periodic orbit.

Lemma 14.9. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system, and suppose x is a periodic point. Suppose the graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ contains a primitive cycle $J_0 \rightarrow J_1 \rightarrow \cdots \rightarrow J_{q-1} \rightarrow J_0$ of length q. Then there exists a periodic point y of minimal period q such that $f^i(y) \in J_i$ for each $1 \leq i \leq q-1$.*

Proof. By Corollary 13.6 there exists a periodic point y with (not necessarily) minimal period q such that $f^i(y) \in J_i$ for each $1 \leq i \leq q-1$.

Suppose y has minimal period m. Then m divides q. If $f^i(y) \notin \mathcal{O}_f(x)$ for each $0 \leq i \leq q-1$ then J_i is the unique vertex of $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ containing $f^i(y)$. Therefore $m = q$, as otherwise the cycle would not be primitive.

Now suppose there exists $0 \leq i \leq q-1$ such that $f^i(y) \in \mathcal{O}_f(x)$. Then $y = f^{q-i}(f^i(y)) \in \mathcal{O}_f(x)$, and thus x also has period q. Since $J_0 \rightarrow J_1 \rightarrow \cdots \rightarrow J_{q-1} \rightarrow J_0$ is primitive, it must be equal to the fundamental cycle of $\mathsf{G}\big(\mathcal{O}_f(x)\big)$. Thus x (and hence also y) has minimal period q. This completes the proof. ■

Corollary 14.10. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system. Assume f has a periodic point which is not a fixed point. Then f has a point of period 2.*

Proof. Let $p > 1$ denote the minimal period of a periodic point, and suppose $p \geq 3$. Suppose x has a period p. By Proposition 14.8 the fundamental cycle of the graph $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ can be decomposed into two shorter primitive cycles. Thus one of them has length $2 \leq q \leq p-1$. Then Lemma 14.9 tells us that f has a periodic point of minimal period q. This contradicts the choice of p, and thus completes the proof. ■

We now have three pieces of evidence for the Sharkovsky ordering:

- 1 should be the smallest number, since every dynamical system $f\colon [0,1] \to [0,1]$ has a fixed point (Lemma 12.7).
- 3 should be the largest number, since the existence of a point of period 3 implies the existence of all other orders (Corollary 13.7).

- 2 should be the second smallest number, since the existence of a point of period greater than 2 implies the existence of a point of period 2 (Corollary 14.10).

The proof of the Sharkovsky Theorem has one more crucial ingredient. The next result tells us exactly the structure of the graph of a periodic orbit of minimal odd period.

PROPOSITION 14.11. *Suppose $f\colon [0,1] \to [0,1]$ is a dynamical system admitting a periodic point x of odd period $p \geq 3$. Assume moreover that f has no periodic points of odd order $1 < q < p$. Let $z \in \mathcal{O}_f(x)$ denote the middle point (so that there are $\frac{1}{2}(p-1)$ points on the orbit smaller than z and $\frac{1}{2}(p-1)$ points on the orbit larger than z). If $z < f(z)$ then the points on the orbit of x are ordered as*[2]

$$f^{p-1}(z) < f^{p-3}(z) < \cdots < f^2(z) < z < f(z) < f^3(z) < \cdots < f^{p-2}(z).$$

If $z > f(z)$ then the reverse order holds.

This proof is non-examinable, as it is rather tricky.

(♣) *Proof.* The case $p = 3$ is easy (see the proof of Corollary 13.7), so we will assume $p \geq 5$. By Proposition 14.8 there is a (unique up to ordering) fundamental cycle of length p, which can be decomposed as two shorter primitive cycles. One of these cycles has odd length l, say (since their lengths sum to p). The hypotheses together with Lemma 14.9 imply that $l = 1$. Thus the fundamental cycle can be written as

$$J_1 \to J_1 \to J_2 \to \cdots \to J_{p-1} \to J_1,$$

where $J_i \neq J_1$ for each $2 \leq i \leq p-1$. In fact, this cycle has three more properties:

- The vertices $J_2, \dots, J_{p-1}$ are all distinct.
- For $2 \leq i \leq p-2$, there is no edge $J_i \to J_1$.
- If $1 \leq i, j \leq p-1$ are such that $j \geq i+2$ then there is no edge $J_i \to J_j$.

To prove the first bullet point, suppose $J_i = J_j$ for $2 \leq i < j \leq p-1$. Then there is a primitive cycle

$$J_1 \to J_2 \to \cdots J_i = J_j \to J_{j+1} \to \cdots \to J_{p-1} \to J_1$$

of length $p + i - j - 1$. If this number is odd, by Lemma 14.9 we find a periodic point with period $p + i - j - 1 < p$, which contradicts the choice of p. If instead $p + i - j - 1$ is even then we simply add an additional J_1 at the front:

$$J_1 \to J_1 \to J_2 \to \cdots J_i = J_j \to J_{j+1} \to \cdots \to J_{p-1} \to J_1$$

This is a primitive cycle of odd length $p + i - j < p$, and Lemma 14.9 again gives us a periodic point of minimal period $p + i - j$, which contradicts the choice of p. The proof of the second and third bullet points goes along similar lines.

[2] If $p = 3$ this should be read as $f^2(z) < z < f(z)$.

Now write $\mathcal{O}_f(x) = \{x_1 < \cdots < x_p\}$ and set $I_i = [x_i, x_{i+1}]$ for $i = 1, \dots, p-1$. Let $1 \leq k \leq p-1$ be the integer such that $J_1 = I_k$. Abbreviate by $z := x_k$. We have shown that the only two edges emanating from J_1 are $J_1 \to J_1$ and $J_1 \to J_2$. Thus J_2 and J_1 have a common endpoint, and hence J_2 is either equal to I_{k+1} or I_{k-1}. This means that we are now in one of two possible situations:

(i) $J_2 = I_{k+1}$, so that $x_{k+1} = f(z)$ and $x_{k-1} = f^2(z)$.

(ii) $J_2 = I_{k-1}$, so that $z = f(x_{k+1})$ and $x_{k+2} = f^2(x_{k+1})$.

Suppose to begin with that case (i) holds. We must have $f^3(z) > z$, otherwise there would be an arrow $J_2 \to J_1$. Thus $f^3(z) = x_i$ for some $i \geq k+2$. Since there is an arrow $J_2 \to J_3$ and there are no arrows $J_2 \to J_j$ for $j > 3$, the only possibility is that $f^3(z) = x_{k+2}$, and thus $J_3 = [f(z), f^3(z)] = I_{k+1}$.

Next, we claim that $f^4(z) < f^2(z)$. Indeed, if $f^4(z) > f^2(z)$ then also $f^4(z) > f^3(z) = x_{k+2}$, and hence $J_3 \to J_1$. This is not possible. Thus $f^4(z) < f^2(z)$. Since there is an arrow $J_3 \to J_4$ and there are no arrows $J_3 \to J_j$ for $j \geq 5$, the only possibility is that $f^4(z) = x_{k-2}$ and that $J_4 = [f^4(z), f^2(z)]$.

Continuing in a similar vein, we find that the points on the orbits are ordered as

$$f^{p-1}(z) < f^{p-3}(z) < \cdots < f^2(z) < z < f(z) < f^3(z) < \cdots < f^{p-2}(z).$$

The point z is necessarily the middle point on the orbit, as the above shows. Moreover $f(z) > z$. The graph $\mathsf{G}(\mathcal{O}_f(x))$ takes the form given in Figure 14.1.

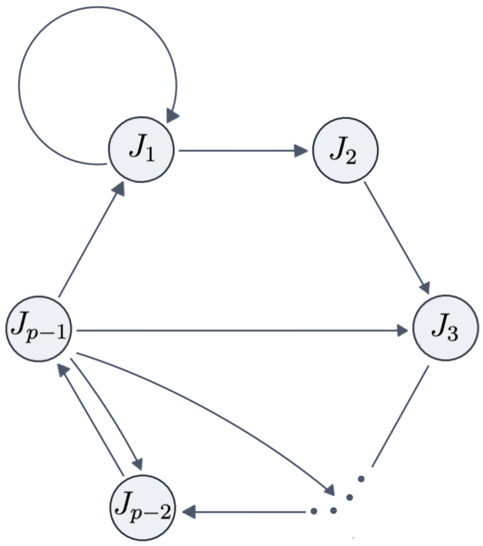

Figure 14.1: The fundamental cycle from Proposition 14.11.

Finally, if case (ii) holds then we end up with the reverse order

$$f^{p-2}(z) < \cdots < f^3(z) < f(z) < z < f^2(z) < \cdots f^{p-3}(z) < \cdots < f^{p-1}(z).$$

Again, z is necessarily the middle point on the orbit, and this time $f(z) < z$. This completes the proof. ∎

Before proceeding with the proof of the Sharkovsky Theorem, let us go back and prove Proposition 13.13. For convenience, we restate it here:

PROPOSITION 14.12. *Let $f\colon [0,1] \to [0,1]$ be mixing. Then f^2 is strictly turbulent.*

Proof. By Corollary 12.11 f has a periodic point x of odd minimal period at least 3. By Proposition 14.11 we may assume that there exists $z \in \mathcal{O}_f(x)$ such that, writing $z_i := f^i(z)$, one has either[3]

$$z_{p-1} < z_{p-3} < \cdots < z_2 < z_0 < z_1 < z_3 < \cdots < z_{p-2}$$

or the reverse order. We prove only the case where the order is as above; the argument for the reverse order is similar. We find four points a, b, c, d such that

$$z_{p-1} < a < b < z_{p-3} < c < d < z_1 \tag{14.3}$$

such that

$$f^2\big([a,b]\big) = f^2\big([c,d]\big) = [z_{p-1}, z_1]. \tag{14.4}$$

Combining (14.3) and (14.4) shows that $\big([a,b],[c,d]\big)$ form a strictly turbulent pair for f^2.

- *The point d:* We start with d. Since[4] $[z_0, z_1] \to_f [z_2, z_0]$, by Lemma 12.7 there exists $d \in (z_0, z_1)$ such that $f(d) = z_0$. Then $f^2(d) = f(z_0) = z_1$.
- *The point a:* Since $[z_{p-1}, z_{p-3}] \to_{f^2} [z_{p-1}, z_1]$, there exists a point $a \in (z_{p-1}, z_{p-3})$ such that $f^2(a) = z_1$.
- *The point b:* We have $[a, z_{p-3}] \to_{f^2} [z_{p-1}, d]$, and thus there exists $b \in (a, z_{p-3})$ such that $f^2(b) = z_{p-1}$.
- *The point c:* Finally, since $[z_{p-3}, d] \to_{f^2} [z_{p-1}, z_1]$ there exists $c \in (z_{p-3}, d)$ such that $f^2(c) = z_{p-1}$.

This completes the proof. ∎

REMARK 14.13. Here is a simpler argument[5] that proves that if $f\colon [0,1] \to [0,1]$ is mixing then f^2 is turbulent (this weaker conclusion is in fact sufficient for Theorem 12.1, cf. Corollary 13.16).

If f is mixing then f has a periodic point of odd order at least three by Corollary 12.11. Thus by Corollary 14.10, f has a periodic point of period 2. Thus f^2 has at least two fixed points. Finally, Problem C.3 tells us that f^2 is transitive, and hence Proposition 13.15 tells us that f^2 is turbulent.

With this out the way, we now move onto the proof of the Sharkovsky Theorem. We need one more lemma, whose proof is trivial and left as an exercise.

LEMMA 14.14. *Let $f\colon [0,1] \to [0,1]$ be a dynamical system.*

[3] As in the statement of Proposition 14.11, if $p = 3$ this should be read as $z_2 < z_0 < z_1$. The rest of the proof is formally identical.

[4] Note these arrows are not bold, cf. Remark 14.4.

[5] Thanks to A. Musso for this argument.

(i) *Suppose x is a periodic point of f with minimal period p. Then for any $k \geq 1$, x is also periodic for f^k. The minimal period of x for f^k is given by $\frac{p}{r}$, where $r := \gcd(p, k)$.*

(ii) *Suppose x is a periodic point of f^k with minimal period q. Then there exists a divisor r of k with $\gcd(r, q) = 1$ such that x is a periodic point of f with minimal period $\frac{qk}{r}$.*

The following proof is non-examinable.

(♣) *Proof of the Sharkovsky Theorem 14.2.* We prove the result in two steps.

1. In this first step, we will show that if $p \geq 3$ is an odd integer and f has a point of period p then for any $k \preceq p$, f also has a periodic point of minimal period k.

We may assume (by definition of the Sharkovsky ordering) that p is the smallest odd period. For such a p, the graph of the periodic orbit is given by Figure 14.1. We continue to use the notation established in the proof of Proposition 14.11. If k is even and $2 \leq k \leq p - 1$ then

$$J_{p-k} \to J_{p-k+1} \to \cdots \to J_{p-1} \to J_{p-k}$$

is a primitive cycle of length k, and thus by Lemma 14.9, f has a point of period k. If instead k is greater than p, we add $k - p$ times the cycle $J_1 \to J_1$ onto the end of the fundamental cycle to obtain a primitive cycle of length k. Lemma 14.9 then gives the desired periodic point again.

2. We now prove the general case. Suppose f has a periodic point x with minimal period $n = 2^r p$, where $p \geq 1$ is odd. Suppose $k \preceq n$. We will prove that f has a point of period k. There are four cases to consider:

(i) If $k = 1$ then we are done by Lemma 12.7.

(ii) Suppose $p = 1$ and $k = 2^s$ for some $0 < s < r$. Then by part (i) of Lemma 14.14, x is periodic for $g := f^{k/2}$, with minimal period $2^{r-s+1} > 1$. By Corollary 14.10, g has a periodic point y of minimal period 2. Then y is a periodic point of f with minimal period $2^s = k$ by part (ii) of Lemma 14.14.

(iii) Suppose $p > 1$ and $k = 2^r m$ for $m \geq 2$ an even number. By part (i) of Lemma 14.14, x is periodic for $g = f^{2^r}$ with minimal period p. Since p is odd and greater than 1, g has a periodic point y of period m by Step 1. Then y is periodic for f with minimal period $k = 2^r m$ by part (ii) of Lemma 14.14.

(iv) Suppose $p > 1$ and $k = 2^r m$ where $q > p$ is an odd number. By part (i) of Lemma 14.14 again, x is periodic for $g = f^{2^r}$ of minimal period p, and then as before Step 1 tells us that g has a periodic point y of minimal period q. Then by part (ii) of Lemma 14.14 there exists an integer $1 \leq s \leq r$ such that y is periodic for f with minimal period $2^s q$. If $s = r$ we are done. If not, set $m := 2^{r-s} q$, so that $k = 2^s m$ with m even. We have just shown that f has a periodic point of period $2^s m$, and thus by case (iii) applied to y, we find another periodic point z with minimal period $2^r m = k$. This completes the proof. ■

We conclude our discussion of dynamical systems on the interval by explaining how the Sharkovsky Theorem is in some sense, sharp.

Definition 14.15. Let $f\colon [0,1] \to [0,1]$ be a dynamical system. We define the **type** of f to be the subset

$$\wp(f) \coloneqq \{k \in \mathbb{N} \mid f \text{ has a periodic point of period } k\}.$$

The Sharkovsky Theorem therefore tells us that for any system f, the type of f is either of the form $\wp(f) = \{k \in \mathbb{N} \mid k \preceq p\}$ for some $p \in \mathbb{N}$, or $\wp(f) = \left\{2^k \mid k \geq 0\right\}$. In fact, all such possibilities are realised. This is the content of the next theorem, which sadly we do not have time to prove.

Theorem 14.16. *There exist dynamical systems on $[0,1]$ of all possible types. More precisely:*

(i) *Let $p \in \mathbb{N}$. There exists a dynamical system $f\colon [0,1] \to [0,1]$ such that*

$$\wp(f) \coloneqq \{k \in \mathbb{N} \mid k \preceq p\}.$$

(ii) *There exists a dynamical system $f\colon [0,1] \to [0,1]$ such that*

$$\wp(f) = \left\{2^k \mid k \geq 0\right\}.$$

Rotation Numbers

For the next few lectures we shift attention from the interval $[0,1]$ to the circle S^1. Moreover we will focus exclusively on reversible systems. This means that topological entropy is no longer a useful invariant (since all such systems have zero entropy, cf. Proposition 8.8). Nevertheless, some systems are more complicated than others.

Our model example is the circle rotation ρ_θ. As we have seen, for θ rational the dynamics are easy to understand, meanwhile for θ irrational the dynamics can be much wilder.

The main goal of today's lecture is to associate a **rotation number** to an arbitrary orientation-preserving (see Definition 15.2 below) reversible dynamical system $f\colon S^1 \to S^1$, denoted by $\operatorname{rot}(f)$. The rotation number is an element of S^1. The main properties of the rotation number are:

- The rotation number of a circle rotation is given by (surprise!) $\operatorname{rot}(\rho_\theta) = \theta$.
- If $\operatorname{rot}(f)$ is rational, the dynamics are simple: f has periodic points, and every orbit is either periodic or asymptotic to a periodic orbit. Thus f is not transitive.
- If $\operatorname{rot}(f)$ is irrational, the dynamics are more complicated: either all orbits are dense or all orbits are asymptotic to a Cantor set.

Perhaps most amazingly of all, the Poincaré Classification Theorem (which we will prove as Theorem 17.6 in Lecture 17) shows that for transitive orientation-preserving systems, the rotation number is a **complete dynamical invariant**, in the sense that two such systems are conjugate *if and only if* they have the same rotation number.

Throughout our discussion on rotation numbers, we will typically use z, w to indicates points in S^1 and x, y to indicate points in $\mathbb{R}$. Denote by $\pi\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z} = S^1$ the projection. Given $z \in S^1$, we say a point $x \in \mathbb{R}$ is a **lift** of z if $\pi(x) = z$. It is often convenient to identify an element of S^1 (which is, formally, an equivalence class) with its representative in $[0,1)$. With this convention,

$$\pi(x) = x - \lfloor x \rfloor.$$

PROPOSITION 15.1. *Let $f\colon S^1 \to S^1$ be a reversible dynamical system. There exists a reversible dynamical system $F\colon \mathbb{R} \to \mathbb{R}$ such that $\pi \circ F = f \circ \pi$:*

$$\begin{tikzcd} \mathbb{R} \arrow[r, "F"] \arrow[d, "\pi"'] & \mathbb{R} \arrow[d, "\pi"] \\ S^1 \arrow[r, "f"'] & S^1 \end{tikzcd}$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

The map F is strictly monotone and unique up to an additive integer constant.

We call F a **lift** of f. Note that f is a factor of F. The following proof uses some elementary algebraic topology, and is therefore non-examinable.

(♣) *Proof.* The map $f \circ \pi \colon \mathbb{R} \to S^1$ can be lifted to a continuous map

$$\begin{tikzcd}
& & \mathbb{R} \arrow[d, "\pi"] \\
\mathbb{R} \arrow[r, "\pi"'] \arrow[urr, dashed, "F"] & S^1 \arrow[r, "f"'] & S^1
\end{tikzcd} \tag{15.1}$$

This follows from the fact that $\pi \colon \mathbb{R} \to S^1$ is the **universal cover**[1] of S^1. The map F is not unique, but it is up to an integer constant. Indeed, if F' is another map such that $\pi \circ F' = f \circ \pi$ then the function $F - F' \colon \mathbb{R} \to \mathbb{R}$ is a continuous function that takes values in $\mathbb{Z} \subset \mathbb{R}$. Such a function is necessarily constant. It remains to show that F is reversible. Suppose G is a lift of f^{-1}, so that the following diagram commutes:

$$\begin{tikzcd}
& & \mathbb{R} \arrow[d, "\pi"] \\
\mathbb{R} \arrow[r, "\pi"'] \arrow[urr, dashed, "G"] & S^1 \arrow[r, "f^{-1}"'] & S^1
\end{tikzcd} \tag{15.2}$$

Then by concatenating the commuting diagrams (15.1) and (15.2) together we see that the following commutes:

$$\begin{tikzcd}
& & & \mathbb{R} \arrow[dd, "\pi"] \\
& & \mathbb{R} \arrow[d, "\pi"] \arrow[ur, "G"] & \\
\mathbb{R} \arrow[r, "\pi"'] \arrow[urr, "F"] & S^1 \arrow[r, "f"'] & S^1 \arrow[r, "g"'] & S^1
\end{tikzcd}$$

But since $f \circ g = \mathrm{id}$ the following diagram also commutes

$$\begin{tikzcd}
& & \mathbb{R} \arrow[d, "\pi"] \\
\mathbb{R} \arrow[r, "\pi"'] \arrow[urr, "G \circ F"] & S^1 \arrow[r, "\mathrm{id}"'] & S^1
\end{tikzcd}$$

[1] This is covered in any introductory text on algebraic topology. For instance, I did it in my lectures notes here.

Another map that makes this diagram commute is the identity map on $\mathbb{R}$:

$$\begin{tikzcd}
 & & \mathbb{R} \arrow[d, "\pi"] \\
\mathbb{R} \arrow[urr, "\mathrm{id}"] \arrow[r, "\pi"'] & S^1 \arrow[r, "\mathrm{id}"'] & S^1
\end{tikzcd}$$

Thus by the already established uniqueness of lifts we see that $G \circ F = \mathrm{id} + k$ for some constant $k \in \mathbb{Z}$. Therefore if we replace G by $G'(x) := G(x) - k$ then G is another lift of f^{-1} and $G' \circ F = \mathrm{id}$. Similarly $F \circ G' = \mathrm{id}$. Thus F is reversible, as claimed. ■

DEFINITION 15.2. We say that a reversible dynamical system $f\colon S^1 \to S^1$ is **orientation-preserving** if F is increasing for some (and hence any) lift F of f. Similarly we say that a reversible dynamical system $f\colon S^1 \to S^1$ is **orientation-reversing** if F is decreasing for some (and hence any) lift F of f.

EXAMPLE 15.3. The circle rotation ρ_θ is orientation-preserving. Indeed, a lift of ρ_θ is given by

$$R_\theta(x) := x + \theta, \tag{15.3}$$

(where we think of θ as a number in $[0, 1)$) which is increasing. On the other hand, the map

$$\tilde{\rho}_\theta(x) := -x + \theta \bmod 1$$

is orientation-reversing.

(♣) REMARK 15.4. The notion of orientability is one of those concepts that regularly confuses students, since there are so many different ways to define it. For instance, in Algebraic Topology one typically defines an orientation-preserving map $f\colon S^1 \to S^1$ as one for which the induced map $f_\star\colon H_1(S^1; \mathbb{Z}) \to H_1(S^1; \mathbb{Z})$ is multiplication by a positive constant. Meanwhile in Differential Geometry an orientation-preserving map is one that preserves an equivalence class of volume forms. Nevertheless, you will be reassured to know that *all these definitions are equivalent*. You are invited to try proving this.

From now on we will focus on orientation-preserving dynamical systems. On Problem Sheet H the orientation-reversing case is explored.

LEMMA 15.5. *Suppose $f\colon S^1 \to S^1$ is an orientation-preserving reversible dynamical system. Let $F\colon \mathbb{R} \to \mathbb{R}$ denote a lift of f. Then*

$$F(x + k) = F(x) + k, \qquad \forall\, x \in \mathbb{R},\ k \in \mathbb{Z}.$$

Proof. Immediate from the proof of Proposition 15.1. ■

PROPOSITION 15.6. *Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system. Let F denote a lift of f, and define*

$$\mathrm{rot}(F) := \lim_{k \to \infty} \frac{F^k(x) - x}{k}.$$

Then the limit $\mathrm{rot}(F)$ *exists and is a finite real number which is independent of the choice of* $x \in \mathbb{R}$. *Moreover if*

$$\mathrm{rot}(f) := \pi(\mathrm{rot}(F))$$

then $\mathrm{rot}(f)$ *is independent of the choice of lift* F.

One should think of $\mathrm{rot}(f)$ as measuring the average speed of f.

Proof. Let us first check that $\mathrm{rot}(F)$ is independent of $x \in \mathbb{R}$. Since $F(x+1) = F(x) + 1$, it is sufficient to prove independence of $x \in [0, 1)$. If $x, y \in [0, 1)$ then one has $|F(x) - F(y)| < 1$, and hence also $|F^k(x) - F^k(y)| < 1$ for all $k \in \mathbb{N}$. Thus for such x, y,

$$\begin{aligned} \left| \frac{1}{k}\big(F^k(x) - x\big) - \frac{1}{k}\big(F^k(y) - y\big) \right| &\le \frac{1}{k}\left(|F^k(x) - F^k(y)| + |x - y|\right) \\ &\le \frac{2}{k} \to 0. \end{aligned}$$

Now let us show that the limit exists. Fix $x \in \mathbb{R}$ and set $x_k := F^k(x)$. Set $y_k := x_k - x$, and note that

$$y_k \ge \min_{z \in [0,1]} \big(F(t) - t\big)$$

by Lemma 15.5. In particular, (y_k) is uniformly bounded below. Next we compute

$$\begin{aligned} y_{k+n} =& F^{k+n}(x) - x \\ =& F^n(x_k) - x_k + x_k - x \\ \overset{(\heartsuit)}{=}& \underbrace{F^n(x_k) - F^n(x + \lfloor y_k \rfloor)}_{\le 1} + \underbrace{F^n(x + \lfloor y_k \rfloor) - (x + \lfloor y_k \rfloor)}_{= y_n} \\ & + \underbrace{x_k - x}_{= y_k} + \underbrace{x - x_k + \lfloor y_k \rfloor}_{\le 0} \\ \le& 1 + y_n + y_k, \end{aligned}$$

where in $(\heartsuit)$ we used Lemma 15.5 to identify the second term with y_n. We would like to apply Fekete's Lemma 7.7. Unfortunately the sequence (y_k) is not quite subadditive, but if we define $a_k := y_k + 1$ then (a_k) is subadditive and bounded below. Therefore by Lemma 7.7 the sequence $\big(\frac{1}{k} a_k\big)$ converges to $\inf_k \frac{1}{k} a_k$, and hence also $\big(\frac{1}{k} y_k\big)$ converges to $\inf_k \frac{1}{k} y_k$. This proves that $\mathrm{rot}(F)$ is a well-defined finite number.

Finally, if G is another lift of g then $G = F + k$ for some $k \in \mathbb{Z}$. It is clear from the definition that $\mathrm{rot}(F + k) = \mathrm{rot}(F) + k$, and hence $\mathrm{rot}(f) := \pi(\mathrm{rot}(F))$ is independent of the choice of lift F. This completes the proof. ■

Note that this shows that there exists a unique lift F of f such that $\mathrm{rot}(F) = \mathrm{rot}(f)$ (i.e. there exists a unique lift F of f such that $\mathrm{rot}(F) \in [0, 1)$.)

Definition 15.7. Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system. We call the number $\mathrm{rot}(f) \in S^1$ the **rotation number** of f.

A trivial example is given by an actual rotation.

Example 15.8. The rotation number of a circle rotation is given by

$$\mathrm{rot}(\rho_\theta) = \theta.$$

Indeed, this is immediate if we take the lift R_θ of ρ_θ from (15.3):

$$\begin{aligned}\mathrm{rot}(R_\theta) &= \lim_{k\to\infty} \frac{R_\theta^k(x) - x}{k} \\ &= \lim_{k\to\infty} \frac{k\theta}{k} \\ &= \theta.\end{aligned}$$

The rotation number is a dynamical invariant of orientation-preserving reversible dynamical systems.

Proposition 15.9. *The rotation number is invariant under conjugacies that preserve orientation.*

Proof. Let $h\colon S^1 \to S^1$ denote an orientation-preserving homeomorphism. We will show that $\mathrm{rot}(h^{-1}fh) = \mathrm{rot}(f)$. Let $H\colon \mathbb{R} \to \mathbb{R}$ and $F\colon \mathbb{R} \to \mathbb{R}$ denote lifts of h and f respectively. Then since

$$\pi \circ H^{-1} = h^{-1} \circ h \circ \pi \circ H^{-1} = h^{-1} \circ \pi,$$

we see that H^{-1} is a lift of h^{-1}. Similarly since

$$\pi \circ H^{-1} \circ F \circ H = h^{-1} \circ \pi \circ F \circ H = h^{-1} \circ f \circ \pi \circ H = h^{-1} \circ f \circ h \circ \pi,$$

we see that $H^{-1}FH$ is a lift of $h^{-1}fh$. We may assume that H is chosen so that $H(0) \in [0,1)$. We want to estimate

$$\left| \left(H^{-1}FH\right)^k(x) - F^k(x) \right| = \left| H^{-1}F^kH(x) - F^k(x) \right|.$$

For $x \in [0,1)$, we have

$$-1 < H(x) - x \le H(x) < H(1) < 2,$$

and hence by periodicity we have

$$|H(x) - x| < 2, \qquad \forall\, x \in \mathbb{R}.$$

Similarly $|H^{-1}(x) - x| < 2$ for all $x \in \mathbb{R}$. Next, observe that if $|y - x| < 2$ then also $\left|F^k(y) - F^k(x)\right| < 3$. Indeed, if $|y - x| < 2$ then also $|\lfloor y \rfloor - \lfloor x \rfloor| \le 2$, and hence

$$\begin{aligned}-3 \le \lfloor y \rfloor - \lfloor x \rfloor - 1 &\overset{(\heartsuit)}{=} F^k\big(\lfloor y \rfloor\big) - F^k\big(\lfloor x \rfloor + 1\big) \\ &< F^k(y) - F^k(x) \\ &< F^k\big(\lfloor y \rfloor + 1\big) - F^k\big(\lfloor x \rfloor\big) \\ &= \lfloor y \rfloor + 1 - \lfloor x \rfloor \le 3,\end{aligned}$$

where (♡) once again used Lemma 15.5. Thus

$$\begin{aligned}\left|H^{-1}F^kH(x) - F^k(x)\right| &\leq \left|H^{-1}F^kH(x) - F^kH(x)\right| + \left|F^kH(x) - F^k(x)\right| \\ &\leq 2 + 3 = 5,\end{aligned}$$

and hence

$$\frac{\left|\left(H^{-1}FH\right)^k(x) - F^k(x)\right|}{k} \leq \frac{5}{k} \to 0,$$

which shows that $\operatorname{rot}(H^{-1}FH) = \operatorname{rot}(F)$ as required. ■

LECTURE 16

Rational Rotation Numbers

In this lecture we continue our investigation of rotation numbers, focusing on the (less interesting) case where the rotation number is rational. Our first observation is that rationality of the rotation number determines whether an orientation-preserving reversible dynamical system on S^1 has periodic points or not.

PROPOSITION 16.1. *Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system. Then $\mathsf{per}(f) \neq \emptyset$ if and only if $\operatorname{rot}(f)$ is a rational number.*

Proof. Let us first show that if f has a periodic point $z \in S^1$ of period q then $\operatorname{rot}(f) \in \mathbb{Q}$. Let $x \in \pi^{-1}(z)$, and let F denote a lift of f. Then there exists $p \in \mathbb{Z}$ such that

$$F^q(x) = x + p.$$

Then for $k \in \mathbb{N}$ one has

$$\begin{aligned} \frac{F^{kq}(x) - x}{kq} &= \frac{1}{kq} \sum_{i=0}^{k-1} \big(F^q(F^{iq}(x)) - F^{iq}(x) \big) \\ &= \frac{1}{kq} \sum_{i=0}^{k-1} p = \frac{kp}{kq} = \frac{p}{q}. \end{aligned}$$

Thus $\operatorname{rot}(f) = \frac{p}{q} \in \mathbb{Q}$. Conversely, suppose $\operatorname{rot}(f) = \frac{p}{q} \in \mathbb{Q}$. Observe that for any $n \in \mathbb{N}$, one has

$$\begin{aligned} \operatorname{rot}(F^n) &= \lim_{k \to \infty} \frac{1}{k} \big((F^n)^k(x) - x \big) \\ &= n \lim_{k \to \infty} \frac{1}{nk} \big(F^{nk}(x) - x \big) \\ &= n \operatorname{rot}(F). \end{aligned}$$

Therefore if $\operatorname{rot}(f) = \frac{p}{q}$, then $\operatorname{rot}(f^q) = 0$ (recall $\operatorname{rot}(f) = \pi(\operatorname{rot}(F))$.)

It thus suffices to show that if g is an orientation-preserving reversible dynamical system with $\operatorname{rot}(g) = 0$ then g has a fixed point. Suppose for contradiction that this is not the case. Let G be the unique lift of g with $G(0) \in [0, 1)$. Since g has no fixed points, $G(x) - x$ is never an integer, and hence $0 < G(x) - x < 1$ for all $x \in \mathbb{R}$. Since $G - \mathrm{id}$ is continuous on $[0, 1]$, it achieves its minimum and maximum. Thus there exists $\delta > 0$ such that $\delta \leq G(x) - x \leq 1 - \delta$ for all $x \in [0, 1]$, and then by periodicity, also for all $x \in \mathbb{R}$. Now set $x_k := G^k(0)$. Since

$$x_k = \sum_{i=0}^{k-1} \left(G(x_i) - x_i \right),$$

it follows that $k\delta \leq x_k \leq (1-\delta)k$ for all $k \in \mathbb{N}$, and hence

$$\delta \leq \frac{x_k}{k} \leq 1-\delta, \qquad \forall\, k \in \mathbb{N}.$$

This implies that $\mathrm{rot}(G) \in [\delta, 1-\delta]$, which is a contradiction. The proof is complete. ∎

For the rest of this lecture we consider the case where the rotation number is rational. Roughly speaking, the moral of this story is that rational rotation numbers are boring. The dynamics of f will turn out to be trivial: any orbit is asymptotic to a periodic orbit and any two periodic orbits have the same period.

More precisely, suppose $\mathrm{rot}(f) = \frac{p}{q}$ where p and q are relatively prime. We will show that the dynamics of f are completely determined by $\mathrm{rot}(f)$, the topology of the set $\mathsf{per}(f)$ of periodic points of f, and the "direction" of the dynamics of f^q on each of the connected components of $S^1 \setminus \mathsf{per}(f)$.

Proposition 16.2. *Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with rational rotation number $\mathrm{rot}(f) = \frac{p}{q}$ where p and q relatively prime. Then every point $x \in \mathsf{per}(f)$ has minimal period q:*

$$\mathsf{per}(f) = \mathsf{fix}(f^q).$$

In fact, if F denotes the unique lift of f with $\mathrm{rot}(F) = \frac{p}{q}$, then a point $x \in \mathbb{R}$ is a fixed point of $F^q - p$ if and only if $\pi(x)$ is a periodic point of f.

Proof. Let F denote the unique lift of f with $\mathrm{rot}(F) = \frac{p}{q}$. Let z denote a periodic point, and let x denote a lift of z. Suppose z has minimal period n. Then $F^n(x) = x + m$ for some integer m. Since

$$\frac{p}{q} = \mathrm{rot}(F) = \lim_{k\to\infty} \frac{F^{kn}(x) - x}{kn} = \lim_{k\to\infty} \frac{mk}{nk} = \frac{m}{n},$$

we see that $m = dp$ and $n = dq$ for some d. Thus $F^{dq}(x) = x + dp$. To complete the proof we claim that $d = 1$.

Indeed, suppose $d > 1$. Then $F^q(x) \neq x + p$. Suppose $F^q(x) > x + p$. Then by monotonicity

$$F^{2q}(x) - 2p = F^q(F^q(x) - p) - p > F^q(x) - p > x,$$

and hence by induction $F^{jq}(x) > x + jp$ for all j. Taking $j = d$ gives a contradiction. The same argument also gives a contradiction if $F^q(x) < x + p$. These contradictions imply that $d = 1$, and this completes the proof. ∎

Now let us explain how to "order" an q-tuple of points in S^1.

Definition 16.3. Given a q-tuple $(z_0, z_1, \dots, z_{q-1})$ of distinct points in S^1, choose a lift x_0 of z_0 in $\mathbb{R}$ and let $x_1, \dots, x_{q-1}$ denote the unique lifts of $z_1, \dots, z_{q-1}$ that belong to the interval $[x_0, x_0 + 1)$. Since the x_i are real numbers, they are ordered, say

$$x_0 < x_{\sigma(1)} < x_{\sigma(2)} < \cdots < x_{\sigma(q-1)},$$

where here σ is a permutation of $\{1, 2, \dots, q-1\}$. We then define the **ordering** of $(z_0, z_1, \dots, z_{q-1})$ to be $(z_0, z_{\sigma(1)}, \dots, z_{\sigma(q-1)})$.

Informally, we say that $z_{\sigma(1)}$ is the *next point to the right* of z_0, and similarly $z_{\sigma(i+1)}$ is the *next point to the right* of $z_{\sigma(i)}$. Similarly after having ordered our q-tuple, it makes sense to speak of the **intervals** $[z_{\sigma(i)}, z_{\sigma(i+1)}] \subset S^1$, which explicitly are given by

$$[z_{\sigma(i)}, z_{\sigma(i+1)}] := \pi\left([x_{\sigma(i)}, x_{\sigma(i+1)}]\right).$$

EXAMPLE 16.4. Let p, q be relatively prime integers, and let $r := \frac{p}{q}$. Consider the rotation ρ_r, and consider the q-tuple

$$\left(0, \rho_r(0), \rho_r^2(0), \dots, \rho_r^{q-1}(0)\right). \tag{16.1}$$

Let m denote the unique integer between 0 and q such that

$$mp = 1 \bmod q, \tag{16.2}$$

and let σ be the permutation

$$\sigma(k) = mk \bmod q. \tag{16.3}$$

Then the ordering of (16.1) is given by $\left(0, \rho_r^{\sigma(1)}(0), \dots, \rho_r^{\sigma(q-1)}(0)\right)$.

Our next result tells us that the ordering of a periodic orbit of *any* orientation-preserving reversible dynamical system f on S^1 is the same as the ordering of the orbit of 0 under the rotation $\rho_{\mathrm{rot}(f)}$ by the (necessarily rational) rotation number $\mathrm{rot}(f)$ of f.

PROPOSITION 16.5. *Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with rational rotation number $\mathrm{rot}(f) = \frac{p}{q}$ where p and q are relatively prime. Let $z \in S^1$ be a periodic point of f (and hence a fixed point of f^q by Proposition 16.1). Then the ordering of the q-tuple*

$$\left(z, f(z), f^2(z), \dots, f^{q-1}(z)\right)$$

is the same as the ordering of the q-tuple (16.1)*, namely the permutation σ specified by* (16.2) *and* (16.3).

Proof. Suppose that the next point to the right of z is given by $f^d(z)$. Then necessarily the next point to the right of $f^d(z)$ is given by $f^{2d}(z)$ (where $2d$ should be read mod q), since if instead it was $f^m(z)$ then we would have $f^m(z) \in [f^d(z), f^{2d}(z)]$, and hence $f^{m-d}(z)$ (where $m-d$ should be read mod q) would lie in the interval $[z, f^d(z)]$, contradicting the fact that $f^d(z)$ was the next point to the right of z. Thus (again reading the multiples of d mod q), the ordering is given by $\left(z, f^d(z), f^{2d}(z), \dots, f^{(q-1)d}(z)\right)$, and it remains to determine d.

Fix a lift x of z. Since f^d carries each interval $[f^{kd}(z), f^{(k+1)d}(z)]$ to its successor, and there are q of these intervals, there is a lift G of f^d such that $G^q(x) = x + 1$.

Now let F denote the lift of f such that $F^q(x) = x + p$ (i.e. so that $\mathrm{rot}(F) = \frac{p}{q}$.) Then F^d is also a lift of f^d, and hence there exists $j \in \mathbb{Z}$ such that $F^d = G + j$. Then since

$$x + dp = F^{qd}(x) = (G + j)^q(x) = G^q(x) + qj = x + 1 + qj,$$

we have $dp = 1 + qj$, and hence d is the unique number between 0 and q such that $dp = 1 \bmod q$. Thus d agrees with m from (16.2). This completes the proof. ■

The next lemma is not needed for the discussion that follows (and hence its proof is relegated to Problem Sheet H). However it helps to put Definitions 16.7 and 16.8 in context.

Lemma 16.6. *Let $f\colon X \to X$ be a dynamical system on a compact metric space, and suppose $x \in X$ has the property that there exists $y \in \mathsf{per}(f)$ such that $\mathcal{O}_f(y) \subseteq \omega_f(x)$. Then $\omega_f(x) = \mathcal{O}_f(y)$. Conversely if $\omega_f(x)$ is finite then there exists $y \in \mathsf{per}(f)$ such that $\omega_f(x) = \mathcal{O}_f(y)$.*

Definition 16.7. Let $f\colon X \to X$ denote a dynamical system and let $x \in X$. If $\omega_f(x) = \{y\}$ then we say that x is **positively asymptotic** to y. (In this case $y \in \mathsf{fix}(f)$ by Lemma 16.6). If f is reversible then we say that x is **negatively asymptotic** to y if $\alpha_f(x) = \{y\}$.

Definition 16.8. Let $f\colon X \to X$ denote a reversible dynamical system. Suppose that $x \in X$ has the property that is positively asymptotic to y and negatively asymptotic to z. If $y \neq z$ then we say that x is a **heteroclinic point**. If $y = z$ then we say that x is a **homoclinic point**.

Remark 16.9. Next semester we will see that for differentiable dynamical systems the existence of a special type of homoclinic point, called a *transverse homoclinic point*, implies the existence of an infinite mesh of such points (known as a "homoclinic tangle"). As the name suggests, this "tangle" has extremely complicated dynamics, and it forces the topological entropy of the system to be positive.

For a general dynamical system, most points will be neither heteroclinic nor homoclinic. However the next result shows for a reversible system f on S^1 with rational rotation number $\operatorname{rot}(f) = \frac{p}{q}$, *every* non-periodic point is heteroclinic or homoclinic under f^q.

Proposition 16.10. *Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with rational rotation number $\operatorname{rot}(f) = \frac{p}{q}$ with p and q relatively prime. Then there are two possible types of non-periodic orbits for f:*

(i) *Suppose f has exactly one periodic orbit. If $q > 1$ then every other point is heteroclinic under f^q to two points on this periodic orbit. Meanwhile if $q = 1$ then all other points are homoclinic to the fixed point.*

(ii) *If f has more than one periodic orbit, then each non-periodic point is heteroclinic under f^q to two points on different periodic orbits.*

The proof of Proposition 16.10 requires the following preliminary lemma, whose proof is similar to several of the results from Lectures 13 and 14.

Lemma 16.11. *Suppose that $f\colon [0,1] \to [0,1]$ is dynamical system which is non-decreasing. Then all $x \in [0,1]$ are positively asymptotic to a fixed point of f. If f is onto and strictly increasing (and hence reversible), all $x \in [0,1]$ are either fixed or positively and negatively asymptotic to adjacent fixed points of f.*

Proof. We may assume that $f \neq \mathrm{id}$, and hence that the open set $[0,1]\backslash\mathsf{fix}(f)$ is non-empty. So let $x \in [0,1]\backslash\mathsf{fix}(f)$ and let $(a,b) \subset [0,1]\backslash\mathsf{fix}(f)$ be the maximal open[1] interval containing x. Since f is non-decreasing we must have $f(a,b) \subseteq (a,b)$, and by the intermediate value theorem we either have $f(y) > y$ or $f(y) < y$ for all $y \in (a,b)$. Assume we are in the first case (the second case is similar). Then $(f^k(x))_{k\in\mathbb{N}} \subset (a,b)$ is a non-decreasing sequence and therefore $x_0 := \lim_{k\to\infty} f^k(x) \in (a,b]$ exists. Now

$$f(x_0) = f\Big(\lim_{k\to\infty} f^k(x)\Big) = \lim_{k\to\infty} f^{k+1}(x) = x_0,$$

so in fact $x_0 = b$ and b is a fixed point. This finishes the proof of the first part. For the second part we can apply the same argument both to f and f^{-1} to see that

$$\lim_{k\to\infty} f^k(x) = b \qquad \text{and} \qquad \lim_{k\to\infty} f^{-k}(x) = a.$$

This completes the proof. ■

Proof of Proposition 16.10. We can identify f^q with a homeomorphism of an interval by taking a lift x of a fixed point z of f^q and restricting a lift $F^q(\cdot) - p$ of f^q to $[x, x+1]$. Now the result follows from Lemma 16.11, apart from the last point of part (ii), which claimed that the two periodic orbits found are distinct.

Suppose this is not the case. This would mean there is an interval $[x_1, x_2] \subset \mathbb{R}$ such that x_1 and x_2 are adjacent fixed points of $F^q - p$ and such that x_1, x_2 project to the same periodic orbit. But if x_1 projects to z and x_2 projects to $f^k(z)$, then

$$\bigcup_{i=0}^{q-1} f^{ik}(\pi(x_1, x_2))$$

covers the complement of $\mathcal{O}_f(z)$ in S^1 and contains no periodic points. Thus f only has the one periodic orbit. ■

We conclude with the following enhancement of Proposition 16.10. The proof is left for you on Problem Sheet H.

PROPOSITION 16.12. *Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with rational rotation number $\mathrm{rot}(f) = \frac{p}{q}$ with p and q relatively prime. Suppose $z \in S^1$ is not a periodic point for f. Let $w_1, w_2 \in \mathsf{per}(f)$ denote the periodic points such that z is positively asymptotic to w_1 and negatively asymptotic to w_2 under f^q. Then for each $1 \leq i \leq q-1$, $f^i(z)$ is positively asymptotic to $f^i(w_1)$ and negatively asymptotic to $f^i(w_2)$ under f^q.*

[1] By "open" we mean open in $[0,1]$. If either $a = 0$ or $b = 1$ minor changes in the notation are needed, which we leave up to you.

LECTURE 17

The Poincaré Classification Theorem

In this lecture we turn to the case of an *irrational* rotation number. Our main result is the Poincaré Classification Theorem (proved as Theorem 17.6 below), which tells us if $f\colon S^1 \to S^1$ is an orientation-preserving reversible dynamical system with irrational rotation number $\theta := \mathrm{rot}(f)$, then the corresponding irrational rotation ρ_θ is a factor of f, and if f is transitive then f is conjugate to ρ_θ.

We begin with the following statement, which is the irrational analogue of Proposition 16.10. In the statement (and for the rest of this lecture) we will typically identify θ with its unique representative in $[0, 1)$.

Proposition 17.1. *Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with rotation number $\vartheta := \mathrm{rot}(f) \in S^1 \setminus \mathbb{Q}$, and let $F\colon \mathbb{R} \to \mathbb{R}$ denote a lift of f. Then for any $a, b, c, d \in \mathbb{Z}$ and $x \in \mathbb{R}$, one has*

$$a\theta + b < c\theta + d \qquad \Leftrightarrow \qquad F^a(x) + b < F^c(x) + d.$$

Remark 17.2. Thus if f is an orientation-preserving reversible dynamical system with irrational rotation number θ, for *any* point $z \in S^1$, the orbit of z is ordered in the same way as it would be under the irrational rotation ρ_θ. Compare this to Proposition 16.10, which proved the same thing for periodic points when the rotation number was rational.

Proof of Proposition 17.1. It suffices to prove the proposition for the unique lift F of f with $\mathrm{rot}(F) = \theta$, since any other lift of f differs from F by a constant. Moreover we may assume that $a \neq c$, otherwise the result is trivial.

First observe that for any $a, b, c, d \in \mathbb{Z}$, the expression $\lambda(x) := F^a(x) + b - F^c(x) - d$ never changes sign, and hence the inequality on the right-hand side is independent of x. Indeed, if $\lambda(x) = 0$ for some $x \in \mathbb{R}$ then $\pi(x) \in S^1$ is a periodic point of f, since $F^a(x) - F^c(x) \in \mathbb{Z}$. But this contradicts the fact that $\mathsf{per}(f) = \emptyset$ by Proposition 16.1.

Now assume that $F^a(0) + b < F^c(0) + d$. Setting $y = F^c(0)$ this is equivalent to saying

$$F^{a-c}(y) - y < d - b. \tag{17.1}$$

As before, if (17.1) holds for one $y \in \mathbb{R}$, then it holds for all $y \in \mathbb{R}$, and in particular it holds for $y = 0$, whence we obtain

$$F^{a-c}(0) < d - b.$$

Then applying (17.1) to $y = F^{a-c}(0)$ we obtain

$$F^{2(a-c)}(0) < d - b + F^{a-c}(0) < 2(d - b),$$

and arguing by induction we obtain

$$F^{k(a-c)}(0) < k(d-b), \qquad \forall\, k \geq 1,$$

which tells us that

$$\theta = \lim_{k\to\infty} \frac{F^{k(a-c)}(0)}{k(a-c)} \leq \lim_{k\to\infty} \frac{k(d-b)}{k(a-c)} = \frac{d-b}{a-c}$$

In fact, since θ is irrational, the inequality is strict, and hence

$$a\theta + b < c\theta + d.$$

This proves $\Leftarrow$. Conversely, the same argument shows that if $F^a(0)+b > F^c(0)+d$ then $a\theta + b > c\theta + d$. Since equality can never hold on either the left-hand side or the right-hand side (as θ is irrational and f has no periodic points), $\Rightarrow$ follows as well. This completes the proof. ∎

Before stating the next result, note that for any given two points $z \neq w$ in S^1 there are two connected components of $S^1 \setminus \{z, w\}$.

Proposition 17.3. *Let $f \colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with irrational rotation number. Let $p \neq q$ be two integers and let $z \in S^1$. Let I denote the closure of a connected component of $S^1 \setminus \{f^p(z), f^q(z)\}$ (note $f^p(z) \neq f^q(z)$ as there are no periodic points.) Then for any $w \in S^1$, one has both $\mathcal{O}_f(w) \cap I \neq \emptyset$ and $\mathcal{O}_f^-(w) \cap I \neq \emptyset$.*

Proof. We give the proof for the forward orbit only. Assume without loss of generality that $p > q$. It suffices to show that

$$S^1 = \bigcup_{k=0}^{\infty} f^{-k}(I).$$

For this, set $I_k := f^{-k(q-p)}(I)$, and observe that for each $k \geq 1$ the intervals I_k and I_{k+1} have a common endpoint. Suppose that $S^1 \neq \bigcup_k I_k$. Then since the intervals I_k abut at the endpoints, it follows that $f^{-k(q-p)}\big(f^p(z)\big)$ converges monotonically to a point $z_0 \in S^1$. But then z_0 is a fixed point of f^{q-p}:

$$\begin{aligned}
z_0 &= \lim_{k\to\infty} f^{-k(q-p)}\big(f^p(z)\big) \\
&= \lim_{k\to\infty} f^{(-k+1)(q-p)}\big(f^p(z)\big) \\
&= \lim_{k\to\infty} f^{q-p}\Big(f^{-k(q-p)}\big(f^p(z)\big)\Big) \\
&= f^{q-p}\left(\lim_{k\to\infty} f^{-k(q-p)}\big(f^p(z)\big)\right) \\
&= f^{q-p}(z_0).
\end{aligned}$$

This contradicts the fact that $\mathsf{per}(f) = \emptyset$. ∎

We now look at the ω-limit sets of a system with irrational rotation number.

Proposition 17.4. *Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with irrational rotation number. Then the set $\omega_f(z)$ is independent of $z \in S^1$. If f is transitive it is equal to all of S^1. If f is not transitive it is a nowhere dense set which has no isolated points and is totally disconnected.*

Remark 17.5. Since any compact totally disconnected metric space without isolated points is homeomorphic to a Cantor set (cf. Remark 4.17), Proposition 17.4 can be stated more concisely as: the ω-limit set is either all of S^1 or a Cantor set.

Proof of Proposition 17.4. First let us show that $\omega_f(z)$ is independent of z. Suppose $z_1 \neq z_2$ are two points in S^1 and suppose $w \in \omega_f(z_1)$. Thus there exists a sequence $k_n \to \infty$ such that $f^{k_n}(z_1) \to w$. Let I_n denote the shorter of the two connected components of $S^1 \setminus \{f^{k_n}(z_1), f^{k_{n+1}}(z_1)\}$. By the previous proposition there exists a sequence i_n such that $f^{i_n}(z_2)$ lies in the interval I_n. Since the length of the I_n's goes to zero, we must have $i_n \to \infty$ and $\lim_{n\to\infty} f^{i_n}(z_2) = w$. Thus $w \in \omega_f(z_2)$. This shows that $\omega_f(z_1) \subseteq \omega_f(z_2)$ and by symmetry they are equal.

Now let $A := \omega_f(z_1)$. Observe that A is the only minimal set for $f|_A$ by Corollary 3.3. Since the boundary ∂A is another closed invariant set, we must have either $\partial A = \emptyset$ or $\partial A = A$. In the former case we have $A = S^1$, since A is then open and closed. In this case f is necessarily transitive (cf. Corollary 2.11 and Corollary 3.4). In the latter case A is nowhere dense. Since A is closed (Proposition 3.2), it follows that A is totally disconnected[1].

To see that A has no isolated points, fix $z \in A$. Then since $A = \omega_f(z)$ this means there exists a sequence $k_n \to \infty$ such that $f^{k_n}(z) \to z$. Since $\mathrm{rot}(f)$ is irrational, there are no periodic orbits, and hence $f^{k_n}(z) \neq z$ for each n. Thus z is an accumulation point of A, since each point $f^{k_n}(z)$ belongs to A by invariance. ■

This result shows how different the irrational case is from the rational one. In the rational case, we saw last lecture that every orbit was either periodic or asymptotic to a periodic orbit. Meanwhile in the irrational case either all orbits are dense or all orbits as asymptotic to a Cantor set. We are now ready to state and prove the main result of today's lecture, which is due to Poincaré.

Theorem 17.6 (Poincaré Classification Theorem). *Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with irrational rotation number $\theta := \mathrm{rot}(f)$. Then the irrational rotation ρ_θ is a factor of f. Moreover if f is transitive then f is conjugate to ρ_θ.*

Proof. Let $F\colon \mathbb{R} \to \mathbb{R}$ denote the unique lift of f with $\mathrm{rot}(F) = \theta$. Fix $x \in \mathbb{R}$ and let $B \subset \mathbb{R}$ denote the complete lift of the orbit of $\pi(x)$:

$$B := \{F^p(x) + q \mid p, q \in \mathbb{Z}\}.$$

[1] S^1 has the property that any compact nowhere dense subset is totally disconnected. This is not true for all compact metric spaces X. For example, it fails for

$$X := \{(x, \sin \tfrac{1}{x}) \mid 0 < x \leq 1\} \cup \{(0, y) \mid -1 \leq y \leq 1\},$$

regarded as a subset of $\mathbb{R}^2$. *Exercise:* Formulate a general criterion that guarantees when a compact metric space has the property that every compact nowhere dense subset is totally disconnected.

Define a map $H\colon B \to \mathbb{R}$ by

$$H\big(F^p(x) + q\big) := p\theta + q.$$

It follows from Proposition 17.1 that the map H is monotone. Since θ is irrational, the image $H(B)$ is dense in $\mathbb{R}$ (cf. the solution to part (ii) of Problem A.2).

Let $R_\theta\colon \mathbb{R} \to \mathbb{R}$ denote the map $R_\theta(x) = x + \theta$, so that R_θ is a lift of the irrational rotation ρ_θ (cf. (15.3)). Then since

$$\begin{aligned} H \circ F\big(F^p(x) + q\big) &= H\big(F^{p+1}(x) + q\big) \\ &= (p+1)\theta + q \end{aligned}$$

and

$$R_\theta \circ H\big(F^p(x) + q\big) = R_\theta(p\theta + q) = (p+1)\theta + q,$$

we see that H is conjugacy between F and R_θ on B.

We now extend H to a map defined on all of $\mathbb{R}$ by setting

$$H(y) := \sup\{p\theta + q \mid F^p(x) + q < y\}.$$

Equivalently,

$$H(y) = \inf\{p\theta + q \mid F^p(x) + q > y\},$$

since otherwise $\mathbb{R} \setminus H(B)$ would contain an interval, contradicting the fact that $H(B)$ is dense in $\mathbb{R}$. We now prove that H is continuous. Firstly, if $y \in \overline{B}$ then

$$H(y) = \sup\{H(x) \mid x \in B,\, x < y\}$$

and also

$$H(y) = \inf\{H(x) \mid x \in B,\, x > y\}.$$

Thus H is continuous on $\overline{B}$. If I is an interval in $\mathbb{R}\setminus\overline{B}$ then H is constant on I and the constant agrees with the values of H at the endpoints. Thus H is continuous on all of $\mathbb{R}$. Moreover by construction H is surjective and non-decreasing. Since

$$\begin{aligned} H(y+1) &= \sup\{p\theta + q \mid F^p(x) + q < y + 1\} \\ &= \sup\{p\theta + q \mid F^p(x) + (q-1) < y\} \\ &= H(y) + 1, \end{aligned}$$

we see that H descends to define a map $h\colon S^1 \to S^1$. The computation above shows that $h \circ f = \rho_\theta \circ h$, and hence h is a semi-conjugacy. This proves that ρ_θ is a factor of f.

Finally, if f was topologically transitive then we could have chosen our original point x to have dense orbit (by Proposition 2.9), which would then have given $\overline{B} = \mathbb{R}$. Then the map h would have been bijective, and thus a true conjugacy. This completes the proof. ■

Remark 17.7. In the non-topologically transitive case, the theorem can be understood as follows. Let A denote the ω-limit set of some (and hence any) point in S^1, which is a Cantor set by Proposition 17.4. If we run the proof of Theorem 17.6 starting with a point $x \in \pi^{-1}(A)$, we end up with $\pi(\overline{B}) = A$. Thus if $A/\sim$ denotes the quotient space where the endpoints of complementary intervals are identified, then f induces a map on $A/\sim$ which is conjugate to the irrational rotation. Thus one should think of the non-topologically transitive case as follows: take an irrational rotation and "blow up" some orbits to intervals whose union makes up the complement of the set A.

LECTURE 18

The Denjoy Theorem

We conclude our discussion on rotation numbers by showing that if we insist on more regularity (i.e. differentiability), any orientation-preserving reversible dynamical system on S^1 with irrational rotation number is automatically transitive.

It won't be until Dynamical Systems II next semester that we discuss what it means for a continuous map between manifolds to be differentiable. Nevertheless, for dynamical systems on the circle, the definition is transparent.

DEFINITION 18.1. Let $f\colon S^1 \to S^1$ be a dynamical system. We say that f is of **class** C^1 if the derivative $f'\colon S^1 \to \mathbb{R}$ exists and is continuous. A reversible dynamical system is said to be a C^1**-diffeomorphism** if both f and f^{-1} are of class C^1.

(♣) REMARK 18.2. The derivative $f'\colon S^1 \to \mathbb{R}$ is defined as you expect it to be:

$$f'(z) := \lim_{t \to 0} \frac{f(z+t) - f(z)}{t}.$$

Note that this becomes a map $f'\colon S^1 \to \mathbb{R}$ (and not $S^1 \to S^1$). If instead we regard S^1 as a smooth manifold, then the derivative is a linear map $Df(z)\colon T_zS^1 \to T_{f(z)}S^1$. So how are the two definitions related? Identifying T_zS^1 and $T_{f(z)}S^1$ with $\mathbb{R}$, the map $Df(x)$ is simply multiplication by a constant. It is easy to check that this constant is $f'(z)$:

$$Df(z)[v] = f'(z)v$$

(♣) REMARK 18.3. In Differential Geometry by a "diffeomorphism" one typically means a bijective map for which both the map and its inverse are of class C^∞. In Dynamical Systems however, we try to get away with the minimal regularity possible. We will say more about this next semester, but for almost everything we do C^1 or C^2 regularity is sufficient.

DEFINITION 18.4. Given a continuous map $g\colon S^1 \to \mathbb{R}$, we define the **variation** of g to be the (possibly infinite) number

$$\operatorname{var}(g) := \sup \sum_{k=0}^{n} |g(x_k) - g(x_{k+1})|,$$

where the supremum is taken over all finite partitions

$$0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1, \qquad n \geq 0.$$

We say that g is of **bounded variation** if $\operatorname{var}(g)$ is finite.

Here is our promised result.

Theorem 18.5 (The Denjoy Theorem). *Let $f\colon S^1 \to S^1$ be an orientation-preserving C^1-diffeomorphism whose derivative $f'\colon S^1 \to \mathbb{R}$ has bounded variation. Assume that $\mathrm{rot}(f)$ is an irrational number. Then f is transitive.*

Remark 18.6. Note that if $g\colon S^1 \to \mathbb{R}$ is Lipschitz continuous then g is of bounded variation. Thus Theorem 18.5 also holds for diffeomorphisms of class $C^{1,1}$ (i.e. f is C^1, and f' is Lipschitz). However if $\alpha \in (0,1)$ then there exists a $C^{1,\alpha}$ orientation-preserving diffeomorphism with irrational rotation number which is *not* topologically transitive (i.e. f is C^1, and f' is α-Hölder continuous). Thus the regularity requirement in Theorem 18.5 is essentially sharp.

The proof of Theorem 18.5 will require two preliminary results. Recall that given an ordered pair z, w of distinct points in S^1, the open interval[1] (z, w) is by definition $\pi((x, y))$, where $x \in \mathbb{R}$ is any lift of z, and y is then unique lift of w that satisfies $x < y < x + 1$.

Proposition 18.7. *Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with irrational rotation number. There exist infinitely many $p \in \mathbb{N}$ with the property that for any $z \in S^1$, the $p+1$ intervals*

$$\big(z, f^{-p}(z)\big),\ \big(f(z), f^{1-p}(z)\big),\ \big(f^2(z), f^{2-p}(z)\big), \dots, \big(f^p(z), z\big)$$

are all pairwise disjoint.

Proof. Fix $z \in S^1$ and abbreviate $I(p, z) \coloneqq \big(z, f^{-p}(z)\big)$. Note that $f^i\big(I(p,z)\big) = \big(f^i(z), f^{i-p}(z)\big)$ since f is orientation preserving. Thus we wish to find conditions that guarantee that the intervals $f^i\big(I(p,z)\big)$ are all pairwise disjoint for all $0 \le i \le p$. This is the case if and only if the endpoints of $f^i\big(I(p,z)\big)$ do not belong to $f^j\big(I(p,z)\big)$ for all pairs (i,j) with $0 \le j < i \le p$. This in turn is equivalent to asking that

$$f^i(z) \notin I(p,z), \qquad \forall\, |i| \le p.$$

Since we want this to hold for all $z \in S^1$, we must show that there exist infinitely many $p \in \mathbb{N}$ such that

$$f^i(z) \notin I(p,z), \qquad \forall\, |i| \le p,\ \forall\, z \in S^1. \tag{18.1}$$

The key point now is that (18.1) only depends on the ordering of the orbit $\mathcal{O}_f(z)$. Let $\theta \coloneqq \mathrm{rot}(f)$. Then by Proposition 17.1, (18.1) is equivalent to requiring that

$$\rho_\theta^i(z) \notin \big(z, \rho_\theta^{-p}(z)\big), \qquad \forall\, |i| \le p,\ \forall\, z \in S^1. \tag{18.2}$$

Finally, by Lemma 1.10 the orbit $\mathcal{O}_{\rho_\theta}(z)$ is dense in S^1 for every $z \in S^1$. Thus there are infinitely many $p \in \mathbb{N}$ such that (18.2) holds. This completes the proof. ∎

Proposition 18.8. *Let $f\colon S^1 \to S^1$ be an orientation-preserving C^1-diffeomorphism whose derivative $f'\colon S^1 \to \mathbb{R}$ has bounded variation. Suppose $I = (z, w)$ is an interval in S^1 with the property that the intervals $I, f(I), \dots, f^{p-1}(I)$ are all pairwise disjoint. Set*[2] *$g \coloneqq \log f'$. Then g has bounded variation, and moreover*

$$\mathrm{var}(g) \ge \left| \log \frac{(f^p)'(z)}{(f^p)'(w)} \right|. \tag{18.3}$$

[1]For the remainder of this lecture, unless stated otherwise by "interval" in S^1 we mean an open non-empty interval (this is in contrast to Lecture 14!)

[2]Since f is orientation preserving and f is a diffeomorphism, $f' > 0$ and thus g is well defined.

Proof. We first show that $\operatorname{var}(g)$ is finite. Since $f'\colon S^1 \to \mathbb{R}$ is continuous and positive, by compactness $\inf f' > 0$. Then for any $u, v \in S^1$ we can estimate

$$|g(u) - g(v)| = \big|\log f'(u) - \log f'(v)\big| \le \frac{|f'(u) - f'(v)|}{\inf f'},$$

and hence

$$\operatorname{var}(g) \le \frac{\operatorname{var}(f')}{\inf f'} < \infty.$$

We now prove (18.3). Since the intervals $I, f(I), \dots, f^{p-1}(I)$ are all pairwise disjoint we have

$$\begin{aligned}
\operatorname{var}(g) &\ge \sum_{k=0}^{p-1} \Big| g\big(f^k(z)\big) - g\big(f^k(w)\big)\Big| \\
&\ge \left| \sum_{k=0}^{p-1} g\big(f^k(z)\big) - g\big(f^k(w)\big)\right| \\
&= \left| \log \prod_{k=0}^{p-1} f'\big(f^k(z)\big) - \log \prod_{k=0}^{p-1} f'\big(f^k(w)\big)\right| \\
&= \left| \log \frac{(f^p)'(z)}{(f^p)'(w)}\right|.
\end{aligned}$$

This completes the proof. ∎

We are now in a position to prove the Denjoy Theorem.

Proof of the Denjoy Theorem 18.5. Assume for contradiction that f is not transitive. Then by Proposition 17.4 the set $A := \omega_f(z_0)$ is nowhere dense set without isolated points (for some and hence any point $z_0 \in S^1$). Thus $S^1 \setminus A$ is a union of intervals. Since f is reversible, the image and preimage of any one of these intervals (i.e. connected components) is another such interval. Let I be one of these intervals. Then we claim that the intervals $f^p(I)$ for $p \in \mathbb{Z}$ are all pairwise disjoint. Indeed, if $f^p(I) \cap f^q(I) \neq \emptyset$ then $f^{p-q}(I) \cap I \neq \emptyset$, and by the previous remark this implies that $f^{p-q}(I) = I$. By continuity, $f^{p-q}(\overline{I}) = \overline{I}$. Then Lemma 12.7 implies that f^{p-q} has a fixed point, which contradicts Proposition 16.1.

Thus the $f^p(I)$ are indeed all disjoint. In particular, denoting by

$$\operatorname{length}\big(f^p(I)\big) := \int_I (f^p)'(z)\, dz,$$

we have

$$\sum_{p \in \mathbb{Z}} \operatorname{length}(f^p(I)) \le 1. \tag{18.4}$$

Next, by Proposition 18.7, there exists an infinite set $S \subset \mathbb{N}$ such that for each $p \in S$ and every $z \in S^1$, the intervals

$$\big(z, f^{-p}(z)\big),\ \big(f(z), f^{1-p}(z)\big),\ \big(f^2(z), f^{2-p}(z)\big), \dots, \big(f^p(z), z\big)$$

are all pairwise disjoint. Fix $p \in S$ and $z \in I$. By Proposition 18.8, applied with $w = f^{-p}(z)$, we have

$$\operatorname{var}(g) \geq \left| \log \Big((f^p)'(z)(f^{-p})'(z) \Big) \right|, \tag{18.5}$$

where we used the chain rule. Thus we can estimate

$$\begin{aligned} \operatorname{length}\big(f^p(I)\big) + \operatorname{length}\big(f^{-p}(I)\big) &= \int_I (f^p)'(z)\, dz + \int_I (f^{-p})'(z)\, dz \\ &= \int_I \big((f^p)'(z) + (f^{-p})'(z) \big)\, dz \\ &\overset{(\heartsuit)}{\geq} \int_I \sqrt{(f^p)'(z)(f^{-p})'(z)}\, dz \\ &\overset{(\diamondsuit)}{\geq} \int_I \sqrt{\exp\left(-\operatorname{var}(g)\right)}\, dz \\ &= \exp\left(-\frac{1}{2} \operatorname{var}(g) \right) \operatorname{length}(I), \end{aligned}$$

where ($\heartsuit$) used the arithmetic-geometric inequality $ab \leq \frac{a^2+b^2}{2}$ and ($\diamondsuit$) used (18.5). This implies that

$$\sum_{p \in S} \Big(\operatorname{length}\big(f^p(I)\big) + \operatorname{length}\big(f^{-p}(I)\big) \Big) = \infty,$$

which contradicts (18.4). This completes the proof. ■

This concludes the section of the course on topological dynamics. Starting next lecture, we will commence our study of **measure-theoretic dynamics**. Since "measure-theoretic dynamics" is rather unwieldy, this subject often goes by another name: **ergodic theory**. Before then, however, we present:

> REVISION OF MEASURE THEORY: The rest of today's notes consists of a summary of the results in measure theory that we will need for the rest of the course. No proofs will be given, but hopefully this material is at least vaguely familiar to most of you. None of this material is *directly* examinable.

DEFINITION 18.9. Let X be a set. A **sigma-algebra** on X is a collection $\mathscr{A}$ of subsets of X satisfying the following three conditions:

(i) $X \in \mathscr{A}$.

(ii) If $A \in \mathscr{A}$ then $X \setminus A \in \mathscr{A}$.

(iii) If $(A_k) \subset \mathscr{A}$ is a countable sequence then $\bigcup_k A_k \in \mathscr{A}$.

We call the pair $(X, \mathscr{A})$ a **measurable space**.

Definition 18.10. Let $(X, \mathscr{A})$ be a measurable space. A **measure** on $(X, \mathscr{A})$ is a function $\mu\colon \mathscr{A} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and if $(A_k) \subseteq \mathscr{A}$ are a sequence of pairwise disjoint elements of $\mathscr{A}$ then

$$\mu\left(\bigcup_k A_k\right) = \sum_k \mu(A_k). \tag{18.6}$$

We say that μ is **finite** if $\mu(X) < \infty$. We say that μ is a **probability measure** if $\mu(X) = 1$. In this case the triple $(X, \mathscr{A}, \mu)$ is called a **probability space**.

In this course we will always restrict to working with probability measures. This is no less general than working with finite measures, since if ν is any finite measure then one can define a probability measure μ from ν via

$$\mu(A) := \frac{\nu(A)}{\nu(X)}.$$

The fact that we work only with finite measure spaces should be thought as being analogous to the fact that when looking at topological dynamics we were mainly interested in compact metric spaces.

Definition 18.11. A set $A \in \mathscr{A}$ is said to be a **null set** for μ if $\mu(A) = 0$. It follows from (18.6) that a countable union of null sets is again a null set.

One can produce new probability spaces by restricting to subsets of positive measure.

Example 18.12. Let $(X, \mathscr{A}, \mu)$ be a probability space and let $A \in \mathscr{A}$ have positive measure. Define a new sigma-algebra $\mathscr{A}_A$ on A as

$$\mathscr{A}_A := \{B \cap A \mid B \in \mathscr{A}\},$$

and define a probability measure μ_A on $(A, \mathscr{A}_A)$ by setting

$$\mu_A(C) := \frac{1}{\mu(A)}\mu(C), \qquad \forall\, C \in \mathscr{A}_A.$$

We call the probability space $(A, \mathscr{A}_A, \mu_A)$ the **restriction** of $(X, \mathscr{A}, \mu)$ to A.

Here is a standard way of producing a sigma-algebra.

Definition 18.13. Let X be a set. A **semi-algebra** on X is a collection $\mathscr{S}$ of subsets of X satisfying the following three conditions.

(i) $X \in \mathscr{S}$.

(ii) If $A, B \in \mathscr{S}$ then $A \cap B \in \mathscr{S}$.

(iii) If $A \in \mathscr{S}$ then there exist finitely many pairwise disjoint $B_1, B_2, \dots, B_k \in \mathscr{S}$ such that $X \setminus A = \bigcup_{i=1}^k B_i$.

Example 18.14. If $X = [0, 1]$ then the collection of all subintervals $(a, b]$ and $[0, b]$ where $0 \le a < b \le 1$, together with the empty set, forms a semi-algebra.

Definition 18.15. Let X be a set and $\mathscr{S}$ a semi-algebra on X. A **probability pre-measure** on $\mathscr{S}$ is a function $\hat{\mu}\colon \mathscr{S} \to [0,1]$ such that:

(i) $\hat{\mu}(\emptyset) = 0$.

(ii) If $(A_k) \subset \mathscr{S}$ are is a countable sequence of pairwise disjoint subsets of X such that $\bigcup_k A_k \in \mathscr{S}$ then $\hat{\mu}\left(\bigcup_k A_k\right) = \sum_k \hat{\mu}(A_k)$.

(iii) If $B_1, B_2, \dots B_k$ are pairwise disjoint elements of $\mathscr{S}$ such that $X = \bigcup_{i=1}^k B_i$ then $\sum_{i=1}^k \hat{\mu}(B_i) = 1$.

Notation. We write $A \vartriangle B$ for the **symmetric difference** of A and B:

$$A \vartriangle B \coloneqq (A \cup B) \setminus (A \cap B).$$

Definition 18.16. Let $\mathscr{S}$ be a semi-algebra on a set X. The **sigma-algebra generated by** $\mathscr{S}$, written $\mathscr{A}(\mathscr{S})$, is the smallest[3] sigma-algebra on X that contains $\mathscr{S}$. Equivalently, $\mathscr{A}(\mathscr{S})$ is the intersection of all sigma-algebras that contain $\mathscr{S}$.

The next result[4] is arguably the cornerstone of the entire subject.

Theorem 18.17 (Main Theorem of Measure Theory). *Let X be a set, and suppose $\mathscr{S}$ a semi-algebra on X and $\hat{\mu}$ a probability pre-measure on $\mathscr{S}$. Then $\hat{\mu}$ uniquely extends to a probability measure μ on $\mathscr{A}(\mathscr{S})$, i.e.*

$$\mu(A) = \hat{\mu}(A), \qquad \forall\, A \in \mathscr{S} \subseteq \mathscr{A}(\mathscr{S}).$$

Moreover, for each $A \in \mathscr{A}(\mathscr{S})$ and each $\epsilon > 0$, there exists finitely many disjoint sets $B_1, B_2, \dots B_k \in \mathscr{S}$ such that if $B \coloneqq \bigcup_{i=1}^k B_i$ then

$$\mu(A \vartriangle B) < \epsilon.$$

Example 18.18. If $\mathscr{S}$ is the semi-algebra of subintervals $(a,b]$ and $[0,b]$ of $[0,1]$ from Example 18.14 and $\hat{\mu}$ is defined by $\hat{\mu}(a,b] = b-a$ and $\hat{\mu}[0,b] = b$ then Theorem 18.17 produces the **Lebesgue measure** λ on $[0,1]$.

Remark 18.19. It can be shown that the Lebesgue measure λ is the unique *translation-invariant* measure on $[0,1]$. This will be useful in Lecture 25.

More generally, we have:

Definition 18.20. Let X be a metric space. Let $\mathscr{S}$ denote the semi-algebra on X obtained by taking finite intersections and complements of the open sets of X. The **Borel sigma-algebra** $\mathscr{B} = \mathscr{B}(X)$ on X is the sigma-algebra generated by this semi-algebra.

Definition 18.21. Let μ be a probability measure on $(X, \mathscr{A})$. We say that a subset $A \in \mathscr{A}$ is an **atom** for μ if $\mu(A) > 0$ and if $B \subset A$ is any measurable set strictly contained in A then $\mu(B) = 0$. We say that μ is **purely atomic** if $\mu(A) > 0$ implies that A contains an atom. We say that μ is **atomless** if μ has no atoms.

[3] The fact that $\mathscr{A}(\mathscr{S})$ is well-defined is not entirely obvious. Luckily for us, this is not a course on measure theory, and so we will not discuss this.

[4] There are many different ways to formulate Theorem 18.17. We use semi-algebras purely for convenience.

Example 18.22. Let X be a metric space, endowed with its Borel sigma-algebra $\mathcal{B}$. Define a probability measure δ_x on $(X, \mathscr{B})$ be declaring that for $U \in \mathscr{B}$,

$$\delta_x(U) = \begin{cases} 1, & x \in U, \\ 0, & x \notin U, \end{cases}$$

We call δ_x the **Dirac measure at** x. This measure is purely atomic.

We now define a measurable function.

Definition 18.23. Let $(X, \mathscr{A})$ be a measurable space. A function $u \colon X \to \mathbb{R}$ is called **measurable** if $u^{-1}(B) \in \mathscr{A}$ for all $B \in \mathscr{B}(\mathbb{R})$. A complex-valued function is called measurable if both its real and imaginary parts are measurable.

Definition 18.24. We say that two measurable functions are equal μ**-almost everywhere** if $\{x \in X \mid f(x) \neq g(x)\}$ is a null set (note this set is automatically measurable). When μ is understood, we just say **almost everywhere**.

One has the following easy result.

Proposition 18.25. *Let X be a metric space equipped with the Borel sigma-algebra $\mathscr{B}$. Then a continuous function $u \colon X \to \mathbb{R}$ is measurable.*

Notation. Let $(X, \mathscr{A})$ be a measurable space. Given $A \in \mathscr{A}$, we denote by $\mathbb{1}_A$ the **characteristic function** of A, defined by

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in X \setminus A. \end{cases}$$

Note that $\mathbb{1}_A$ is measurable.

Let us recall how integration works.

Definition 18.26. A function $u \colon X \to \mathbb{R}$ is called simple if there exists finitely many measurable pairwise disjoint sets $A_1, \dots, A_k$ and real numbers $a_1, \dots, a_k$ such that $u = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$ almost everywhere. Simple functions are obviously measurable, and we define the μ**-integral** of a simple function $u = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$ to be

$$\int_X u \, d\mu = \sum_{i=1}^k a_i \mu(A_i).$$

When μ is understood we say simply **integral** instead of μ-integral. Note the value of $\int_X u \, d\mu$ is independent of the representation $\sum_{i=1}^k a_i \mathbb{1}_{A_i}$.

Now suppose $u \colon X \to \mathbb{R}$ is a measurable function which is non-negative almost everywhere. Then it is easy to see there exists a sequence (u_k) of simple functions such that $u_k \leq u$ and $u_k \to u$ almost everywhere. Indeed, set

$$u_k(x) := \begin{cases} \frac{i-1}{2^k}, & \text{if } \frac{i-1}{2^k} \leq u(x) < \frac{i}{2^k}, \qquad i = 1, \dots, k2^k, \\ k, & \text{if } u(x) \geq k. \end{cases}$$

Definition 18.27. Let $u\colon X \to \mathbb{R}$ is a measurable function which is non-negative almost everywhere. We define the **integral** of f as the limit

$$\int_X u\,d\mu = \lim_{k\to\infty} \int_X u_k\,d\mu \in [0,\infty],$$

where (u_k) is any sequence of simple functions u_k such that $u_k \le u$ and $u_k \to u$ almost everywhere. The value of $\int_X u\,d\mu$ is independent of the sequence (u_k), and we say f is integrable if $\int_X u\,d\mu < \infty$.

Next suppose $u\colon X \to \mathbb{R}$ is an arbitrary measurable function. Set

$$u^+(x) := \max\{u(x), 0\}, \qquad u^-(x) := \max\{-u(x), 0\}.$$

Note that $u^\pm \ge 0$ and $u = u^+ - u^-$.

Definition 18.28. We say that a real-valued function u is **integrable** if both u^+ and u^- are, and in this case we define the **integral** of u to be the well-defined number

$$\int_X u\,d\mu := \int_X u^+\,d\mu - \int_X u^-\,d\mu,$$

Thus u is integrable if and only if $|u|$ is integrable.

Definition 18.29. If $u\colon X \to \mathbb{C}$ is complex valued measurable function, then writing $u = v + iw$, we say that u is **integrable** if both v and w are, and we define the **integral** of u to be the complex number

$$\int_X u\,d\mu = \int_X v\,d\mu + i\int_X w\,d\mu.$$

Integration has the property that if u, v are two measurable functions such that $u = v$ almost everywhere then u is integrable if and only if v is, and if they are integrable then $\int_X u\,d\mu = \int_X v\,d\mu$.

Definition 18.30. Let $(X, \mathscr{A}, \mu)$ be a probability space and $1 \le p < \infty$. The space $L^p(X, \mathscr{A}, \mu; \mathbb{R})$ is the space of all equivalence classes of measurable functions $u\colon X \to \mathbb{R}$ with the property that $|u|^p$ is integrable, where the equivalence relation is given by being equal almost everywhere. The space $L^p(X, \mathscr{A}, \mu; \mathbb{C})$ is defined similarly.

Typically we will omit the non-important parts of the notation in $L^p(X, \mathscr{A}, \mu; \mathbb{R})$, and just write $L^p(\mu)$ or similar.

Proposition 18.31. *The space $L^p(\mu)$ is a Banach space, with norm*

$$\|u\|_p := \left(\int_X |u|^p\,d\mu\right)^{1/p}.$$

Moreover $L^2(\mu; \mathbb{C})$ is a complex Hilbert space with inner product

$$\langle\!\langle u, v\rangle\!\rangle := \int_X u\,\overline{v}\,d\mu$$

(here $\overline{v}$ denotes the complex conjugate of v.)

There are three basic results on integrating that we will need.

THEOREM 18.32 (Monotone Convergence Theorem). *Let $(X, \mathscr{A}, \mu)$ be a probability space. Suppose (u_k) is an increasing sequence of integrable functions (i.e. $u_k(x) \leq u_{k+1}(x)$ for almost every x). If there exists $C > 0$ such that $\int_X u_k \, d\mu < C$ for all k then the limit $\lim_k u_k$ exists almost everywhere and hence defines a measurable function u. Moreover u is integrable with*

$$\int_X u \, d\mu = \lim_{k \to \infty} \int_X u_k \, d\mu.$$

Next we have:

THEOREM 18.33 (Fatou's Lemma). *Let $(X, \mathscr{A}, \mu)$ be a probability space. Suppose (u_k) is a sequence of measurable functions which is bounded below by an integrable function. If $\liminf_k \int_X u_k \, d\mu < \infty$ then $u := \liminf_k u_k$ is integrable and*

$$\int_X u \, d\mu \leq \liminf_{k \to \infty} \int_X u_k \, d\mu.$$

Finally, we have:

THEOREM 18.34 (Dominated Convergence Theorem). *Let $(X, \mathscr{A}, \mu)$ be a probability space. Suppose v is an integrable function and (u_k) is a sequence of measurable functions with $|u_k| \leq v$ almost everywhere. Suppose $\lim_k u_k = u$ almost everywhere. Then u is integrable and*

$$\int_X u \, d\mu = \lim_{k \to \infty} \int_X u_k d\mu.$$

We now move onto a slightly more advanced topic: the *Radon-Nikodym Theorem.*

DEFINITION 18.35. Let $(X, \mathscr{A})$ be a measurable space. Suppose μ and ν are two probability measures on $(X, \mathscr{A})$. We say that μ is **absolutely continuous** with respect to ν, if for any set $A \in \mathscr{A}$,

$$\nu(A) = 0 \quad \Rightarrow \quad \mu(A) = 0.$$

In words: a null set for ν is also a null set for μ. We write $\mu \ll \nu$ to indicate that μ is absolutely continuous with respect to ν. We say that μ and ν are **equivalent** if $\mu \ll \nu$ and $\nu \ll \mu$.

Given $u \in L^1(\mu)$ and $A \in \mathscr{A}$, we use the notation

$$\int_A u d\mu := \int_X u \, \mathbb{1}_A \, d\mu.$$

THEOREM 18.36 (Radon-Nikodym Theorem). *Let μ and ν be two probability measures on the measurable space $(X, \mathscr{A})$. Then $\mu \ll \nu$ if and only if there exists a function $u_{\mu,\nu} \in L^1(\nu)$ such that $u_{\mu,\nu} \geq 0$ and $\int_X u_{\mu,\nu} d\nu = 1$, and such that*

$$\mu(A) = \int_A u_{\mu,\nu} \, d\nu, \qquad \forall \, A \in \mathscr{A}.$$

The function $u_{\mu,\nu}$ is unique ν-almost everywhere, in the sense that any other function with these properties is equal to $u_{\mu,\nu}$ ν-almost everywhere).

Theorem 18.36 suggests the following definition.

Definition 18.37. Let μ and ν be two probability measures on the measurable space $(X, \mathscr{A})$, and suppose $\mu \ll \nu$. We call the function $u_{\mu,\nu}$ appearing in Theorem 18.36 the **Radon-Nikodym derivative** of μ with respect to ν and use the suggestive notation

$$u_{\mu,\nu} = \frac{d\mu}{d\nu}.$$

See Remark 18.40 below for the reason for this notation.

Remark 18.38. The Radon-Nikodym Theorem 18.36 does not actually require μ or ν to be probability measures. If they are not probability measures however then the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$ satisfies $\int_X \frac{d\mu}{d\nu}\, d\nu = \mu(X)$.

The next example is the origin of the terminology "absolutely continuous".

Example 18.39. Let our measurable space be $[0, 1]$, equipped with its usual Borel sigma-algebra (Example 18.18), and μ be a given probability measure. If we denote by u the measurable function

$$u(x) \coloneqq \mu((0, x]),$$

then one can show that μ is absolutely continuous with respect to the Lebesgue measure λ if and only if the function u is absolutely continuous. Moreover in this case $\frac{d\mu}{d\lambda}$ is equal to the derivative u' of u λ-almost everywhere.

We now explain the motivation behind the notation $\frac{d\mu}{d\nu}$.

Remark 18.40. Let μ and ν be two probability measures on the measurable space $(X, \mathscr{A})$, and suppose $\mu \ll \nu$. Suppose $u \in L^1(\mu)$. Then $u\frac{d\mu}{d\nu} \in L^1(\nu)$ and

$$\int_X u\, d\mu = \int_X u \frac{d\mu}{d\nu}\, d\nu$$

(i.e. formally you can "cancel" the $d\nu$'s). Moreover if $\mu \ll \rho$ and $\nu \ll \rho$ then

$$\frac{d(\mu+\nu)}{d\rho} = \frac{d\mu}{d\rho} + \frac{d\mu}{d\rho}, \qquad \rho\text{-almost everywhere.}$$

Moreover if $\mu \ll \nu$ and $\nu \ll \rho$ then the "chain rule" holds:

$$\frac{d\mu}{d\rho} = \frac{d\mu}{d\nu}\frac{d\nu}{d\rho}, \qquad \rho\text{-almost everywhere.}$$

In particular if μ and ν are two equivalent measures then

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1} \qquad \nu\text{-almost everywhere}$$

The "opposite" notion of absolute continuity is the following.

Definition 18.41. Two probability measures μ and ν on $(X, \mathscr{A})$ are **mutually singular**, written $\mu \perp \nu$, if there exists some $A \in \mathscr{A}$ such that

$$\mu(A) = 0, \qquad \text{and} \qquad \nu(X \setminus A) = 0.$$

We then have the following result.

THEOREM 18.42 (Lebesgue Decomposition Theorem). *Let μ and ν be two probability measures on $(X, \mathscr{A})$. Then there exists $c \in [0,1]$ and two probability measures μ_1 and μ_2 on $(X, \mathscr{A})$ such that*

$$\mu = c\mu_1 + (1-c)\mu_2, \qquad \text{and} \qquad \mu_1 \ll \nu, \qquad \mu_2 \perp \nu.$$

The number c and the probability measures μ_1 and μ_2 are uniquely determined.

Here the notation $\mu = c\mu_1 + (1-c)\mu_2$ means that for all $A \in \mathscr{A}$, one has

$$\mu(A) = c\mu_1(A) + (1-c)\mu_2(A).$$

We conclude this lecture by defining the notion of a countable basis.

DEFINITION 18.43. A probability space $(X, \mathscr{A}, \mu)$ has a **countable basis** if there exists a sequence $(B_k) \subset \mathscr{A}$ such that for any $A \in \mathscr{A}$ and any $\epsilon > 0$ there exists B_k such that $\mu(A \bigtriangleup B_k) < \epsilon$.

EXAMPLE 18.44. Suppose X is a separable metric space and $\mathscr{B}$ denotes the Borel sigma-algebra. Then $(X, \mathscr{B}, \mu)$ has a countable basis for any probability measure μ.

Recall that a Hilbert space is **separable** if it has a countable dense subset. Our final result for today is:

PROPOSITION 18.45. *A probability space $(X, \mathscr{A}, \mu)$ has a countable basis if and only if the Hilbert space $L^2(\mu; \mathbb{C})$ is separable.*

Ergodicity

We begin by giving the "measure-theoretic" notion of a dynamical system. First, a remark on the terminology.

REMARK 19.1. It is customary in measure theory to refer to maps between probability spaces as *transformations* rather than functions. This is purely by convention.

DEFINITION 19.2. Suppose $(X, \mathscr{A}, \mu)$ is a probability space. A transformation $f\colon X \to X$ is called **measurable** if $f^{-1}(A) \in \mathscr{A}$ for all $A \in \mathscr{A}$.

To reduce the profligation of parentheses, we will usually write simply $f^{-1}A$ instead of the more formally correct $f^{-1}(A)$.

EXAMPLE 19.3. Let X be a metric space. Equip X with its Borel sigma-algebra $\mathscr{B}$. Then any continuous map $f\colon X \to X$ is measurable. This is proved[1] in a similar fashion to Proposition 18.25.

DEFINITION 19.4. Let $(X, \mathscr{A}, \mu)$ be a probability space. A **measure-preserving dynamical system** is a measurable transformation $f\colon X \to X$ such that

$$\mu(f^{-1}A) = \mu(A), \qquad \forall\, A \in \mathscr{A}.$$

If in addition f is bijective and f^{-1} is also measure-preserving then we call f a **reversible measure-preserving dynamical system**.

Most of the time we will omit the phrase *measure-preserving* and simply call f a **dynamical system**. This is analogous to the way that we always omitted the adjective "topological" from Definition 1.1. Since this is a slightly non-standard convention, we emphasise it once more:

CONVENTION: A **dynamical system** f on a probability space $(X, \mathscr{A}, \mu)$ is (by definition) a measure-preserving transformation $f\colon X \to X$.

Starting in Proposition 19.18 below and throughout Lectures 24 to 28, we will consider both topological dynamical systems and measure-preserving dynamical systems at the same time. Whenever ambiguity is possible, we will not drop the relevant adjective.

REMARK 19.5. Let $(X, \mathscr{A}, \mu)$ be a probability space. Suppose $\mathscr{S} \subseteq \mathscr{A}$ is a semi-algebra that generates $\mathscr{A}$ (i.e. such that $\mathscr{A}(\mathscr{S}) = \mathscr{A}$). If $f\colon X \to X$ is "measure-preserving on $\mathscr{S}$" in the sense that for each $A \in \mathscr{S}$ we have $f^{-1}A \in \mathscr{A}$ and moreover that

$$\mu(f^{-1}A) = \mu(A), \qquad \forall\, A \in \mathscr{S},$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]Apart from the fact that Proposition 18.25 was stated without proof. . .

then in fact f is measure-preserving on all of $\mathscr{A}$, and hence a dynamical system. This is proved using abstract measure-theoretic results (similar to those needed to prove Theorem 18.17), which we will not discuss.

Therefore it suffices to check the measure-preserving property on a semi-algebra that generates the sigma-algebra. This will occasionally be useful. For example, given a transformation $f\colon [0,1] \to [0,1]$, in order to show that f is measure-preserving with respect to the Lebesgue measure, it suffices to check this property on the intervals $(a,b]$ and $[0,b]$ for $0 \le a < b \le 1$, cf. Example 18.18.

Definition 19.6. Let $(X, \mathscr{A}, \mu)$ be a probability space and $f\colon X \to X$ a dynamical system. A measurable set A is called **invariant** if $f^{-1}A = A$.

Note we must use $f^{-1}A$ in the definition rather than $f(A)$, since $f(A)$ may not be an element of $\mathscr{A}$.

Remark 19.7. **Warning:** This definition is slightly at odds with the definition of an invariant set for a topological dynamical system (cf. Definition 1.14). This is unfortunate, but the terminology is too entrenched to try and change.

Remark 19.8. Suppose f is a dynamical system on $(X, \mathscr{A}, \mu)$, and A is an invariant set with positive measure. Then $f|_A$ defines a dynamical system on the restricted probability space $(A, \mathscr{A}_A, \mu_A)$ from Example 18.12.

The following result is our main way of producing invariant sets.

Proposition 19.9. *Let f be dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Given $A \in \mathscr{A}$, set*

$$E_n := \bigcup_{k=n}^{\infty} f^{-k}A, \qquad E := \bigcap_{n=0}^{\infty} E_n.$$

Then E is an invariant set with $\mu(E) = \mu(E_0)$ and $\mu(A \cap E) = \mu(A)$.

Proof. Observe that for $m \ge n$, $E_m = f^{n-m}(E_n)$, and hence as f is measure preserving we have $\mu(E_n) = \mu(E_0)$ for all $n \ge 0$. Since $E_n \subseteq E_{n-1}$ for each n, it follows that[2] $\mu(E) = \mu(E_0)$. Moreover by definition E is invariant:

$$\begin{aligned} f^{-1}(E) &= \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} f^{-(k+1)}(A) \\ &= \bigcap_{n=0}^{\infty} \bigcup_{k=n+1}^{\infty} f^{-k}A \\ &= \bigcap_{n=1}^{\infty} E_n \\ &= E, \end{aligned}$$

[2]This is a standard elementary piece of measure theory, but it is important you realise that this only holds because we are working on a finite measure space. Indeed, if $(X, \mathscr{A}, \mu)$ is any (not necessarily finite) measure space and $(E_n)_{n\ge 0}$ is a family of measurable subsets such that $E_n \subseteq E_{n-1}$ for $n \ge 1$ and $\mu(E_0) < \infty$ then

$$\mu\left(\bigcap_{n=0}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

Exercise: Prove this!

as $E_n \subseteq E_0$ for each $n \geq 0$. Finally $\mu(A \cap E) = \mu(A \cap E_0) = \mu(A)$ since $A \subseteq E_0$. This completes the proof. ■

To illustrate the power of bringing a measure into play, let us prove the following famous result, which is also[3] due to Poincaré.

THEOREM 19.10 (Poincaré Recurrence Theorem). *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Suppose $A \in \mathscr{A}$ has $\mu(A) > 0$. Then almost all points in A return to A under f infinitely many times. That is, there exists a subset $A^* \subseteq A$ with $\mu(A^*) = \mu(A)$ such that for each $x \in A^*$ there exists a sequence (k_n) of numbers such that $k_n \to \infty$ and such that $f^{k_n}(x) \in A^*$ for each $n \geq 1$.*

Proof. Let E be as in Proposition 19.9. Then E is the set of points in X that enter A infinitely many times under positive iterates of f. Set $A^* := A \cap E$. If $x \in A^*$ then there exists a sequence $0 < k_1 < k_2 < \dots$ with $f^{k_n}(x) \in A$ for each n. In fact, $f^{k_n}(x) \in A^*$, since $f^{k_m}(x) = f^{k_m - k_n}\big(f^{k_n}(x)\big) \in A$ for every $m > n$. Moreover $\mu(A^*) = \mu(A)$ by Proposition 19.9. This completes the proof. ■

Suppose f is a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and suppose one can find an invariant set A for f. Then as in the topological case, we can then reduce the study of the dynamics of f to the (hopefully simpler) systems $f|_A$ and $f|_{X \setminus A}$. But in a measure-theoretic setting, there is an additional subtlety, in that we ignore sets of measure zero. Thus we are only interested in invariant sets A with the property that both $\mu(A)$ and $\mu(X \setminus A)$ are positive. It therefore makes sense to single out those dynamical systems for which one cannot simplify things by restricting to an invariant subset. This gives rise to the measure-theoretic analogue[4] of transitivity, which is called *ergodicity*.

DEFINITION 19.11. Let $(X, \mathscr{A}, \mu)$ be a probability space. A dynamical system $f \colon X \to X$ is called **ergodic** if the only invariant measurable sets either have full measure or zero measure: if $A \in \mathscr{A}$ then

$$f^{-1}A = A \qquad \Rightarrow \qquad \mu(A) \in \{0, 1\}.$$

Here is a first result about ergodicity.

PROPOSITION 19.12. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then the following are equivalent:*

(i) *f is ergodic.*

(ii) *If $A \in \mathscr{A}$ satisfies $\mu\big(f^{-1}A \vartriangle A\big) = 0$ then $\mu(A) \in \{0, 1\}$.*

(iii) *If $A \in \mathscr{A}$ has $\mu(A) > 0$ then $\mu\left(\bigcup_{k=1}^{\infty} f^{-k}A\right) = 1$.*

[3] A great number of definitions and theorems in Dynamical Systems bear Poincaré's name. This is not due to a lack of imagination in textbook writers, but more a testament to Poincaré's brilliance. He single-handedly invented what we refer to as "Dynamical Systems" during his work on the Three-Body Problem in the late nineteenth century (more on this next semester). Amazingly enough, Poincaré was also responsible for inventing Algebraic Topology. (After all, why invent one field of mathematics when you can invent two?) An absolute legend.

[4] Actually at a first glance this may appear closer to the notion of minimality; see Remark 19.14 below.

(iv) *If $A, B \in \mathscr{A}$ and $\mu(A), \mu(B) > 0$ then there exists $k \geq 1$ such that $\mu\big(f^{-k}A \cap B\big) > 0$.*

The proof will use the following easy observation: for any measurable sets A, B, one has:

$$|\mu(A) - \mu(B)| \leq \mu(A \vartriangle B). \tag{19.1}$$

Indeed, this is immediate from the fact that $\mu(A) = \mu(A \setminus B) + \mu(A \cap B)$ and $\mu(B) = \mu(B \setminus A) + \mu(A \cap B)$.

Proof. We first prove that (i) implies (ii). For this let us first note that for any set $A \in \mathscr{A}$ and any $k \geq 1$, one has

$$f^{-k}A \vartriangle A \subseteq \bigcup_{i=0}^{k-1} \Big(f^{-(i+1)}(A) \vartriangle f^{-i}A \Big). \tag{19.2}$$

Indeed, suppose $x \in A$ but $f^k(x) \notin A$. If $f(x) \notin A$ then $x \in f^{-1}A \vartriangle A$ and we are done. Thus we may assume that $f(x) \in A$. Then if $f^2(x) \notin A$ then we are done, as then $x \in f^{-2}A \vartriangle f^{-1}A$. This process will eventually produce some $0 \leq i \leq k-1$ such that $f^i(x) \in A$ but $f^{i+1}(x) \notin A$, as otherwise we would end up assuming that $f^k(x) \in A$, contrary to our hypotheses. A similar argument works if we assume that $f^k(x) \in A$ but $x \notin A$, and thus (19.2) is proved.

Now assume that $A \in \mathscr{A}$ satisfies $\mu(f^{-1}A \vartriangle A) = 0$. We claim that

$$\mu(f^{-k}A \vartriangle A) = 0, \qquad \forall\, k \geq 1. \tag{19.3}$$

To see this, use (19.2) to obtain

$$\begin{aligned} f^{-k}A \vartriangle A &\subseteq \bigcup_{i=0}^{k-1} \Big(f^{-(i+1)}(A) \vartriangle f^{-i}A \Big) \\ &= \bigcup_{i=0}^{k-1} f^{-i}(f^{-1}A \vartriangle A), \end{aligned}$$

and then as f is measure-preserving we get

$$\mu(f^{-k}A \vartriangle A) \leq k\mu(f^{-1}A \vartriangle A) = 0.$$

Now let E and E_n be defined as in Proposition 19.9. We claim that in this case one actually has $\mu(A) = \mu(E)$ (rather than just $\mu(A \cap E) = \mu(A)$.) Indeed, by (19.3) we have

$$\mu(A \vartriangle E_n) \leq \sum_{k=n}^{\infty} \mu(A \vartriangle f^{-k}A) = 0,$$

and thus as $E \subseteq E_n$ and $\mu(E) = \mu(E_n)$ by Proposition 19.9 we obtain $\mu(A \vartriangle E) = 0$. Thus by (19.1), we have $\mu(A) = \mu(E)$. Since E is invariant, by ergodicity $\mu(E) \in \{0, 1\}$, and hence the same is true of $\mu(A)$. This proves (ii).

Now we prove that (ii) implies (iii). Suppose $A \in \mathscr{A}$ has $\mu(A) > 0$. Consider E_1 as defined in Proposition 19.9. Then since $f^{-1}E_1 \subseteq E_1$ and $\mu(f^{-1}E_1) = \mu(E_1)$, we have $\mu(f^{-1}E_1 \vartriangle E_1) = 0$. Thus by (ii) we have $\mu(E_1) \in \{0, 1\}$. Since $f^{-1}A \subseteq E_1$ we cannot have $\mu(E_1) = 0$, whence $\mu(E_1) = 1$. This proves (iii).

Now let us prove that (iii) implies (iv). Suppose $A, B \in \mathscr{A}$ both have positive measure. Then with E_1 as above, by (iii) one has $\mu(E_1) = 1$, and thus

$$0 < \mu(B) = \mu(B \cap E_1) = \mu\left(\bigcup_{k=1}^{\infty} (B \cap f^{-k} A)\right).$$

Hence there must exist $k \geq 1$ with $\mu(f^{-k} A \cap B) > 0$. This proves (iv).

Finally let us show that (iv) implies (i). Suppose $A \in \mathscr{A}$ is invariant. If $0 < \mu(A) < 1$ then

$$0 = \mu\big(A \cap (X \setminus A)\big) = \mu\big(f^{-k} A \cap (X \setminus A)\big)$$

for all $k \geq 1$, which contradicts (iv). This completes the proof. ∎

Remark 19.13. The strength of part (ii) of Proposition 19.12 is the following. Suppose $A \in \mathscr{A}$ satisfies $A \subseteq f^{-1} A$ (or $f^{-1} A \subseteq A$). Then $\mu(A \triangle f^{-1} A) = 0$, and hence if f is ergodic then A has measure 0 or 1.

Another important point to note is the following.

Remark 19.14. At a first glance, it may appear that the natural topological analogue of ergodicity is minimality, not topological transitivity. However in the measure-theoretic world, we are free to ignore things that happen on sets of measure zero. This corresponds to asking sets to be dense in the topological world—compare part (iii) of Proposition 19.13 with part (iii) or Proposition 2.4 and part (iii) of Proposition 2.14. Nevertheless, the correspondence between ergodicity and transitivity is not "perfect"; see Proposition 19.18 and Remark 19.19 below.

We will shortly give another equivalent set of characterisations of ergodicity, but let us first show how to a measure-preserving transformation f one can associate an isometry of the Hilbert space L^2.

Definition 19.15. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Given a (real or complex-valued) measurable function u, we denote by $f^*(u)$ the measurable function defined by

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

Note that f^* is a linear operator and

$$f^*(uv) = f^*(u) f^*(v).$$

If u is real-valued then so is $f^*(u)$, and if $u \geq 0$ then $f^*(u) \geq 0$. Slightly less obviously, we have:

Proposition 19.16. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then for any measurable function u, one has*

$$\int_X f^*(u)\, d\mu = \int_X u\, d\mu \tag{19.4}$$

(where one side doesn't exist or is infinite if and only if the other is). If $p \in [1, \infty)$ and $u \in L^p(\mu)$ then $\|f^*(u)\|_p = \|u\|_p$, and hence f^* is a linear isometry $f^* \colon L^p(\mu) \to L^p(\mu)$. In particular, if f is reversible then f^* is a unitary operator on the Hilbert space $L^2(\mu; \mathbb{C})$.

Proof. It suffices to prove (19.4) when f is real-valued by considering the real and imaginary parts separately. Similarly if suffices to prove (19.4) when u is non-negative. If u is a simple function, then the result is immediate as f is measure preserving and

$$f^*(\mathbb{1}_A) = \mathbb{1}_{f^{-1}A}.$$

If u is a non-negative measurable function and u_k a sequence of simple functions increasing to u then $f^*(u_k)$ is a sequence of simple functions increasing to $f^*(u)$. Then

$$\begin{aligned}\int_X f^*(u)\,d\mu &= \lim_k \int_X f^*(u_k)\,d\mu \\ &= \lim_k \int_X u_k\,d\mu \\ &= \int_X u\,d\mu,\end{aligned}$$

which proves (19.4) in this case. Next, if $u \in L^p(\mu)$ then apply (19.4) to $v := |u|^p$ to see that $f^*(u) \in L^p(\mu)$ and $\|f^*(u)\|_p = \|u\|_p$. If f is invertible then f^* is surjective as $f^*(f^{-1})^*(u) = (u)$. Moreover if $u, v \in L^2(\mu; \mathbb{C})$ then applying (19.4) to $w = u\overline{v}$ and using the fact that $f^*(w) = f^*(u) \cdot \overline{f^*(v)}$ shows that

$$\langle\!\langle f^*(u), f^*(v)\rangle\!\rangle = \langle\!\langle u, v\rangle\!\rangle.$$

Thus f^* is unitary. This completes the proof. ∎

As promised, here is another set of equivalent characterisations of ergodicity. On Problem Sheet J you will find another one.

Proposition 19.17. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. The following are equivalent:*

(i) *f is ergodic.*

(ii) *For any measurable u, if $f^*(u) = u$ almost everywhere then u is constant almost everywhere.*

(iii) *For any $u \in L^2(\mu)$, if $f^*(u) = u$ almost everywhere then u is constant almost everywhere.*

Proof. It is obvious that (ii) ⇒ (iii), so we need only show that (i) ⇒ (ii) and (iii) ⇒ (i).

We start with (i) ⇒ (ii). Assume that $f^*(u) = u$ almost everywhere. We may assume, by taking real and imaginary parts, that u is real-valued. Define for $k \in \mathbb{Z}$ and $n > 0$ a set

$$X_{k,n} := \left\{ x \in X \mid \frac{k}{2^n} \le u(x) < \frac{k+1}{2^n} \right\}.$$

Then

$$f^{-1}X_{k,n} \bigtriangleup X_{k,n} \subseteq \{x \in X \mid f^*(u)(x) \neq u(x)\}.$$

Thus by part (ii) of Proposition 19.12 and the assumption that f is ergodic, we obtain that $\mu(X_{k,n}) \in \{0,1\}$. Since for any $n > 0$ we have

$$X = \bigsqcup_{k=-\infty}^{\infty} X_{k,n},$$

as a disjoint union, it follows that for each $n > 0$ there exists a unique k_n such that $\mu(X_{k_n,n}) = 1$. Set

$$Y := \bigcap_{n=1}^{\infty} X_{k_n,n}.$$

Then $\mu(Y) = 1$. But by construction, u is constant on Y, and hence u is constant almost everywhere. This proves (ii).

We now prove (iii) $\Rightarrow$ (i). Suppose $E \subseteq X$ is invariant. Then $\mathbb{1}_E \in L^2(\mu)$ satisfies $f^*(\mathbb{1}_E) = \mathbb{1}_E$ everywhere. Thus by (iii) the function $\mathbb{1}_E$ is constant almost everywhere. Thus either $\mathbb{1}_E$ equals 0 almost everywhere, or $\mathbb{1}_E$ equals 1 almost everywhere. In either case we have

$$\mu(E) = \int_X \mathbb{1}_E \, d\mu \in \{0,1\}.$$

This completes the proof. ■

We conclude this lecture by making an explicit connection between ergodicity and transitivity.

PROPOSITION 19.18. *Let X be a metric space, and let $\mathscr{B}$ denote the Borel sigma-algebra. Suppose $f \colon X \to X$ is a topological dynamical system on X, and*[5] *assume there exists a probability measure μ on X for which f is an ergodic measure-preserving dynamical system with respect to μ. Assume in addition that $\mu(A) > 0$ for each non-empty open set A. Then f is transitive.*

Proof. Let U and V be any two non-empty open subsets of X. By part (iv) of Proposition 19.17 there exists $k > 0$ such that $\mu(f^{-k}U \cap V) > 0$. Thus $f^{-k}(U) \cap V \neq \emptyset$, and hence also $U \cap f^k(V) \neq \emptyset$. ■

REMARK 19.19. We will prove in Lecture 22 that when X is compact the red part of the hypotheses of Proposition 19.18 is actually **not** an assumption: if $f \colon X \to X$ a topological dynamical system on a compact metric space then there *always* exists a probability measure μ on the Borel sigma-algebra $\mathscr{B}$ such that f is an ergodic measure-preserving dynamical system with respect to μ. However in general there is no reason why this measure should be positive on open sets. For instance, suppose x is a fixed point of X. Then the Dirac measure δ_x from Example 18.22 is a probability measure for which f is both measure-preserving and ergodic.

[5]Why is this in red? See Remark 19.19.

LECTURE 20

The Birkhoff Ergodic Theorem

In this lecture we state and prove the famous *Birkhoff Ergodic Theorem*.

DEFINITION 20.1. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Given $u \in L^1(\mu)$, we define the **time average of u with respect to f** to be the function

$$\widehat{u}(x) \coloneqq \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)), \qquad \text{if the limit exists.}$$

When f is understood, we simply call $\widehat{u}$ the **time average** of u.

In fact, for any $u \in L^1(\mu)$, the limit exists almost everywhere, and $\widehat{u} \in L^1(\mu)$. This is a consequence of the following famous theorem.

THEOREM 20.2 (Birkhoff Ergodic Theorem). *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then for any $u \in L^1(\mu)$, the time average $\widehat{u}$ is a well defined integrable function which is f-invariant: $f^*(\widehat{u}) = \widehat{u}$ almost everywhere. Moreover*

$$\int_X u \, d\mu = \int_X \widehat{u} \, d\mu$$

An immediate corollary of Theorem 20.2 is:

COROLLARY 20.3. *Let f be an ergodic dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then for any $u \in L^1(\mu)$, the time average $\widehat{u}$ is constant almost everywhere, and equal to $\int_X u \, d\mu$.*

Proof. If f is ergodic then $f^*(\widehat{u}) = \widehat{u}$ almost everywhere implies that $\widehat{u}$ is almost everywhere constant by part (ii) of Proposition 19.17. ■

REMARK 20.4. One often calls the integral $\int_X u \, d\mu$ the **space average** of the function u. Thus in the ergodic case, Theorem 20.2 can be concisely stated as saying that:

$$\boxed{\text{time average} = \text{space average.}}$$

Many arguments in statistical mechanics implicitly assume that the time average is equal to the space average, and thus for these arguments to be mathematically valid, one needs to verify the dynamical system in question is ergodic.

Before embarking on the proof of Theorem 20.2, let us see two applications of Theorem 20.2. Further results of this form are on Problem Sheet J.

Recall that any number $x \in [0,1)$ has a binary expansion

$$x = \sum_{k=1}^{\infty} \frac{x_k}{2^k}$$

where $x_k \in \{0,1\}$ for each k. For all but countably many $x \in [0,1)$, the binary representation is unique.

Definition 20.5. We say that $x \in [0,1)$ is **normal to base 2** if the frequency of 1's in the binary expansion of x is $\frac{1}{2}$.

It is easy to write down examples of numbers that are *not* normal to base 2, for instance $x = \frac{1}{2}$ has a binary expansion with $x_1 = 1$ and $x_k = 0$ for all $k \geq 2$. Nevertheless, it is a somewhat remarkable fact that most numbers are normal:

Proposition 20.6. *Almost all numbers (with respect to the Lebesgue measure λ) in $[0,1)$ are normal to base 2.*

Proof. Let $e_2 \colon [0,1) \to [0,1)$ denote the doubling map, $e_2(x) = 2x \bmod 1$. By[1] Problem J.1, e_2 is ergodic with respect to Lebesgue measure λ. Let $X \subset [0,1)$ denote the set of points whose binary expansion is unique. Then $\lambda(X) = 1$, since the complement of X is countable. If $x \in X$ then writing

$$e_2(x) = e_2\left(\sum_{k=1}^{\infty} \frac{x_k}{2^k}\right) = \sum_{k=1}^{\infty} \frac{x_{k+1}}{2^k},$$

we observe that if $u = \mathbb{1}_{[1/2,1)}$ one has

$$u(e_2^i(x)) = \begin{cases} 1, & x_{i+1} = 1, \\ 0, & x_{i+1} = 0. \end{cases}$$

Thus for $x \in X$, the number of 1's in the first k digits of x is $\sum_{i=0}^{k-1} u(e_2^i(x))$. By Corollary 20.3, one has

$$\frac{1}{k}\sum_{i=0}^{k-1} u(e_2^i(x)) \to \int_X \mathbb{1}_{[1/2,1)}\, d\lambda = \frac{1}{2}, \qquad \text{almost everywhere.}$$

This says that the frequency of 1's in the binary expansion of almost every $x \in X$ is $\frac{1}{2}$, and thus completes the proof. ∎

Here is another application of Theorem 20.2, which will be useful in later lectures.

[1] Strictly speaking, Problem J.1 is formulated on the circle instead of $[0,1)$, but this makes no difference.

Proposition 20.7. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then f is ergodic if and only if for all $A, B \in \mathscr{A}$ one has*

$$\frac{1}{k}\sum_{i=0}^{k-1}\mu(f^{-i}A\cap B)\to\mu(A)\mu(B). \tag{20.1}$$

Proof. Suppose f is ergodic. Choose $u = \mathbb{1}_A$ and apply Corollary 20.3 to obtain

$$\frac{1}{k}\sum_{i=0}^{k-1}\mathbb{1}_A(f^i(x))\to\mu(A), \qquad \text{almost everywhere.}$$

Multiply both sides by $\mathbb{1}_B$ to obtain

$$\frac{1}{k}\sum_{i=0}^{k-1}\mathbb{1}_A(f^i(x))\mathbb{1}_B\to\mu(A)\mathbb{1}_B, \qquad \text{almost everywhere.}$$

Now apply the Dominated Convergence Theorem 18.34 to obtain

$$\frac{1}{k}\sum_{i=0}^{k-1}\mu(f^{-i}A\cap B)\to\mu(A)\mu(B)$$

as required.

For the converse, suppose E is an invariant set. Set $A = B = E$ and apply (20.1) to obtain $\frac{1}{k}\sum_{i=0}^{k-1}\mu(E)\to\mu(E)^2$. Thus $\mu(E) = \mu(E)^2$, which implies $\mu(E)\in\{0,1\}$. This completes the proof. ■

We will now get started on the proof of Theorem 20.2. The key step is the following rather strange series of constructions. Let $(X, \mathscr{A}, \mu)$ be a probability space. Suppose $P\colon L^1(\mu)\to L^1(\mu)$ is a *positive* linear operator with $\|P\|_{\mathrm{op}}\le 1$. Explicitly, this means that

$$u\ge 0 \qquad \Rightarrow \qquad Pu\ge 0$$

and

$$\|Pu\|_1\le\|u\|_1, \qquad \forall\, u\in L^1(\mu).$$

Let us now fix a function $u\in L^1(\mu;\mathbb{R})$. Using P, we define a sequence of functions (u_k) for $k = 0, 1, 2, \dots$. Namely, let us first set $u_0 := 0$ and then define inductively for $k\ge 1$

$$u_k := \sum_{i=0}^{k-1}P^i u.$$

We then define for $n = 0, 1, 2, \dots$ a function

$$v_n(x) := \max_{0\le k\le n}u_k(x).$$

Note that $v_n\ge 0$ since $u_0 = 0$. Finally we set

$$v \colon = \sup_{n\ge 0} v_n.$$

Note that clearly $v_n \in L^1(\mu)$, and hence the function v is measurable (but not necessarily in $L^1(\mu)$). Next, we define measurable sets $A_n \in \mathscr{A}$ by

$$A_n \coloneqq \{x \in X \mid v_n(x) > 0\},$$

so that $A_n \subseteq A_{n+1}$, and the union

$$A \coloneqq \bigcup_{n=0}^{\infty} A_n$$

is exactly the set on which v is positive. We then claim:

THEOREM 20.8 (Maximal Ergodic Theorem). *It holds that*

$$\int_A u \, d\mu \geq 0.$$

At the moment this result probably looks completely random to you (and utterly undeserving of a special name!), but fear not: all will be revealed soon.

Proof. Since $A_n \subseteq A_{n+1}$, by the Dominated Convergence Theorem 18.34 it suffices to show that

$$\int_{A_n} u \, d\mu \geq 0, \qquad \forall n \in \mathbb{N}.$$

Indeed, $u\mathbb{1}_{A_n} \to u\mathbb{1}_A$ and $|u\mathbb{1}_A| \leq |u|$, so

$$\int_{A_n} u \, d\mu = \int_X u\mathbb{1}_{A_n} \, d\mu \to \int_X u\mathbb{1}_A \, d\mu = \int_A u \, d\mu.$$

So let us fix $n \geq 1$. For $0 \leq k \leq n$ one has by definition that $v_n \geq u_k$, and hence by positivity $Pv_n \geq Pu_k$, which implies that

$$Pv_n + u \geq u_{k+1}, \qquad \forall\, 0 \leq k \leq n.$$

Thus if $x \in A_n$ then

$$Pv_n(x) + u(x) \geq \max_{1 \leq k \leq n} u_k(x) = \max_{0 \leq k \leq n} u_k(x) = v_n(x),$$

where the second equality used the fact that $v_n(x) > 0$ for $x \in A_n$, and hence the maximum cannot be achieved by $u_0 = 0$.

In other words, we have shown

$$u \geq v_n - Pv_n, \qquad \text{on} \qquad A_n,$$

and hence

$$\begin{aligned}
\int_{A_n} u \, d\mu &\geq \int_{A_n} v_n \, d\mu - \int_{A_n} Pv_n \, d\mu \\
&= \int_X v_n \, d\mu - \int_{A_n} Pv_n \, d\mu, && \text{since } v_n = 0 \text{ on } X \setminus A_n, \\
&\geq \int_X v_n \, d\mu - \int_X Pv_n \, d\mu, && \text{since } Pv_n \geq 0, \\
&\geq 0, && \text{since } \|P\| \leq 1.
\end{aligned}$$

This completes the proof. ■

We now use the Maximal Ergodic Theorem to prove another technical looking result.

PROPOSITION 20.9. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$ and let $w \in L^1(\mu; \mathbb{R})$. Given $a \in \mathbb{R}$, set*

$$Y_a := \left\{ x \in X \,\middle|\, \sup_{k \geq 1} \frac{1}{k} \sum_{i=0}^{k-1} w(f^i(x)) > a \right\}.$$

Then if E is any invariant subset for f one has

$$\int_{E \cap Y_a} w \, d\mu \geq a\mu(E \cap Y_a).$$

Proof. We first prove the result in the special case where $E = X$. Set $u := w - a$. Then Y_a is precisely the set A from the Maximal Ergodic Theorem 20.8, and thus $\int_{Y_a} u \, d\mu \geq 0$, which implies that $\int_{Y_a} w \, d\mu \geq a\mu(Y_a)$. For the general case we may assume that $\mu(E) > 0$, otherwise there is nothing to prove. We then consider the dynamical system $f|_E$ on the restricted probability space $(E, \mathscr{A}_E, \mu_E)$ (cf. Remark 19.8) and apply the case we have already proved. This completes the proof. ∎

We are now ready for the proof of Theorem 20.2. This proof is non-examinable—not because it is particularly hard, but because it is rather long.

(♣) *Proof.* We may as usual assume that u is real-valued, by taking real and imaginary parts. We prove the result in three steps.

1. Define

$$u_{\sup}(x) := \limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)),$$

and

$$u_{\inf}(x) := \liminf_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)).$$

In this step we show that

$$u_{\sup} = u_{\inf} \qquad \text{almost everywhere.}$$

First observe that both $u_{\sup}$ and $u_{\inf}$ are invariant under f. Indeed, if $u_k(x) := \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x))$ then

$$\frac{k+1}{k} u_{k+1}(x) - u_k(f(x)) = \frac{1}{k} u(x),$$

and thus taking lim sup and lim inf as $k \to \infty$ establishes our claim.

To show that $u_{\sup} = u_{\inf}$ almost everywhere, define for $a, b \in \mathbb{R}$ the set

$$X_{a,b} := \{x \in X \mid u_{\inf}(x) < b \text{ and } u_{\sup}(x) > a\}.$$

Since

$$\{x \in X \mid u_{\inf}(x) < u_{\sup}(x)\} \subseteq \bigcup_{a,b \in \mathbb{Q},\ b<a} X_{a,b},$$

it suffices to show that $\mu(X_{a,b}) = 0$ if $b < a$. Since u_{sup} and u_{inf} are invariant, one has $f^{-1}X_{a,b} = X_{a,b}$. If f was ergodic then it would immediately follow that $\mu(X_{a,b}) = 0$, but for the full statement we need to use Proposition 20.9.

For this, set as in Proposition 20.9

$$Y_a := \left\{ x \in X \,\middle|\, \sup_{k \geq 1} \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)) > a \right\}.$$

Then

$$X_{a,b} \cap Y_a = X_{a,b}.$$

By Proposition 20.9 we obtain

$$\begin{aligned} \int_{X_{a,b}} u \, d\mu &= \int_{X_{a,b} \cap Y_a} u \, d\mu \\ &\geq a\mu(X_{a,b} \cap Y_a) \\ &= a\mu(X_{a,b}). \end{aligned}$$

But now replacing u, a, b with $-u, -b, -a$ respectively, and using $(-u)_{\text{sup}} = -u_{\text{inf}}$ and $(-u)_{\text{inf}} = -u_{\text{sup}}$ we get also that

$$\int_{X_{a,b}} u \, d\mu \leq b\mu(X_{a,b}).$$

Thus $a\mu(X_{a,b}) \leq b\mu(X_{a,b})$. This is true for any a, b. If $b < a$ then it forces $\mu(X_{a,b}) = 0$.

2. It follows that $\widehat{u}$ exists almost everywhere and agrees with both u_{sup} and u_{inf}. Now we show that $\widehat{u} \in L^1(\mu)$. For this set

$$v_k(x) := \left| \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)) \right|.$$

Then $v_k \geq 0$ and by Proposition 19.16 one has $\int_X v_k \, d\mu \leq \int_X |u| \, d\mu$. Thus by Fatou's Lemma 18.33 we see that $\liminf_k v_k = |u_{\text{inf}}|$ is integrable, i.e. that $u_{\text{inf}} \in L^1(\mu)$.

3. In this final step we show that $\int_X \widehat{u} \, d\mu = \int_X u \, d\mu$. For this we play a similar game to the above. Set

$$Z_{n,k} := \left\{ x \in X \,\middle|\, \frac{k}{n} \leq u_{\text{sup}}(x) < \frac{k+1}{n} \right\}$$

for $k \in \mathbb{Z}$ and $n \in \mathbb{N}$. Given $\varepsilon > 0$, we have

$$Z_{n,k} \cap Y_{\frac{k}{n} - \varepsilon} = Z_{n,k},$$

where Y_a is as defined above. Thus by Proposition 20.9 once more we obtain

$$\int_{Z_{n,k}} u \, d\mu \geq \left(\frac{k}{n} - \varepsilon \right) \mu(Z_{n,k}).$$

Since ε was arbitrary, it follows that

$$\int_{Z_{n,k}} u \, d\mu \geq \frac{k}{n} \mu(Z_{n,k}).$$

Using the definition of $Z_{n,k}$ we then have

$$\begin{aligned}\int_{Z_{n,k}} u_{\text{sup}} \, d\mu &\leq \frac{k+1}{n} \mu(Z_{n,k}) \\ &\leq \frac{1}{n} \mu(Z_{n,k}) + \int_{Z_{n,k}} u \, d\mu.\end{aligned}$$

Summing over k gives

$$\int_X u_{\text{sup}} \, d\mu \leq \frac{1}{n} + \int_X u \, d\mu.$$

Since this holds for any n, we obtain

$$\int_X u_{\text{sup}} \, d\mu \leq \int_X u \, d\mu. \tag{20.2}$$

Now apply this to $-u$ to obtain

$$\int_X (-u)_{\text{sup}} \, d\mu \leq \int_X (-u) \, d\mu,$$

or equivalently that

$$\int_X u_{\text{inf}} \, d\mu \geq \int_X u \, d\mu. \tag{20.3}$$

Since $u_{\text{sup}} = u_{\text{inf}}$ almost everywhere by Step 1, combining (20.2) and (20.3) shows that

$$\int_X \widehat{u} \, d\mu = \int_X u \, d\mu.$$

This completes the proof. ∎

Mixing from a Measure-Theoretic Viewpoint

In this lecture we investigate the measure-theoretic analogues of the mixing and weakly mixing properties from Lecture 5. The starting point for this discussion is Proposition 20.7, which tells us that if f is an ergodic dynamical system on a probability space $(X, \mathscr{A}, \mu)$ then

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \mu(f^{-i}A \cap B) = \mu(A)\mu(B).$$

This motivates the following definitions.

DEFINITION 21.1. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. We say that f is:

(i) **weakly mixing** if for all $A, B \in \mathscr{A}$

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \left|\mu(f^{-i}A \cap B) - \mu(A)\mu(B)\right| = 0.$$

(ii) **mixing** if for all $A, B \in \mathscr{A}$,

$$\lim_{k\to\infty} \mu(f^{-k}A \cap B) = \mu(A)\mu(B).$$

REMARK 21.2. Whilst the mixing condition is very natural, you may be forgiven for thinking that the weakly mixing condition is somewhat contrived. Moreover at first glance it would not to have anything to do with the topological definition of weak mixing (cf. Definition 5.6). However fear not: by the end of the lecture the weak mixing condition will seem much more natural, and the correspondence between the topological and measure-theoretic definitions will be clear.

LEMMA 21.3. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then*

$$f \text{ is mixing} \quad \Rightarrow \quad f \text{ is weakly mixing} \quad \Rightarrow \quad f \text{ is ergodic}.$$

Proof. Observe that if $(a_k)_{k\geq 0}$ is any sequence of real numbers then $\lim_k a_k = 0$ implies that $\lim_k \frac{1}{k}\sum_{i=0}^{k-1} |a_i| = 0$, and this latter condition implies that $\lim_k \frac{1}{k}\sum_{i=0}^{k-1} a_i = 0$. The lemma follows with $a_k := \mu(f^{-k}A \cap B) - \mu(A)\mu(B)$. ■

Recall from Remark 19.5 that given a transformation $f\colon X \to X$ on a probability space $(X, \mathscr{A}, \mu)$, in order to check whether f is measure-preserving it suffices to test it on any semi-algebra generating $\mathscr{A}$. In fact, the same is true for ergodicity, weak mixing and mixing, as we now prove.

Proposition 21.4. *Suppose $(X, \mathscr{A}, \mu)$ is a probability space and suppose $\mathscr{S} \subseteq \mathscr{A}$ is a semi-algebra that generates $\mathscr{A}$. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then:*

(i) *f is ergodic if and only if for all $A, B \in \mathscr{S}$, one has*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \mu(f^{-k} A \cap B) = \mu(A)\mu(B).$$

(ii) *f is weakly mixing if and only if for all $A, B \in \mathscr{S}$, one has*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} |\mu(f^{-i} A \cap B) - \mu(A)\mu(B)| = 0.$$

(iii) *f is mixing if and only if for all $A, B \in \mathscr{S}$, one has*

$$\lim_{k\to\infty} \mu(f^{-k} A \cap B) = \mu(A)\mu(B).$$

Proof. Firstly, it is clear that if all three properties hold for elements of $\mathscr{S}$ then they also hold for finite disjoint unions of elements of $\mathscr{S}$. Now let $A, B \in \mathscr{A}$, and fix $\varepsilon > 0$. By Theorem 18.17 we can find $A_1, \dots, A_p$ and $B_1, \dots B_q$ that belong to $\mathscr{S}$ such that $A_i \cap A_j = \emptyset$ if $i \neq j$ and $B_i \cap B_j = \emptyset$ if $i \neq j$ and such that if $A_0 := \bigcup_{i=1}^p A_i$ and $B_0 := \bigcup_{i=1}^q B_i$ then $\mu(A \vartriangle A_0) < \varepsilon$ and $\mu(B \vartriangle B_0) < \varepsilon$. Note that for $i \geq 0$, one has

$$\big(f^{-i} A \cap B\big) \vartriangle \big(f^{-i} A_0 \cap B_0\big) \subseteq \big(f^{-i} A \vartriangle f^{-i} A_0\big) \cup \big(B \vartriangle B_0\big),$$

and hence we have

$$\mu\Big(\big(f^{-i} A \cap B\big) \vartriangle \big(f^{-i} A_0 \cap B_0\big)\Big) < 2\varepsilon,$$

and hence

$$\big|\mu(f^{-i} A \cap B) - \mu(f^{-i} A_0 \cap B_0)\big| < 2\varepsilon$$

by (19.1). Thus we have

$$\begin{aligned}
\big|\mu(f^{-i} A \cap B) - \mu(A)\mu(B)\big| &\leq \big|\mu(f^{-i} A \cap B) - \mu(f^{-i} A_0 \cap B_0)\big| \\
&\quad + \big|\mu(f^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)\big| \\
&\quad + \big|\mu(A)\mu(B) - \mu(A)\mu(B_0)\big| \\
&\quad + \big|\mu(A)\mu(B_0) - \mu(A_0)\mu(B_0)\big| \\
&< 4\varepsilon + \big|\mu(f^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)\big|,
\end{aligned}$$

where the last used used (19.1) again to estimate $|\mu(A) - \mu(A_0)| < \varepsilon$ and $|\mu(B) - \mu(B_0)| < \varepsilon$. From this, both (ii) and (iii) follow, since A_0 and B_0 are finite unions of elements in $\mathscr{S}$. The proof of (i) proceeds along similar lines: arguing as above we find

$$\left|\frac{1}{k} \sum_{i=0}^{k-1} \mu(f^{-i} A \cap B) - \mu(A)\mu(B)\right| < 4\varepsilon + \left|\frac{1}{k} \sum_{i=0}^{k-1} \mu(f^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)\right|,$$

and then the conclusion follows as before. This completes the proof. ∎

We now take a short detour and discuss some convergence properties of sequences of real numbers.

Definition 21.5. A subset $K \subseteq \{0,1,2,\dots\}$ has **density zero** if

$$\lim_{k\to\infty} \frac{\# K \cap \{0,1,2,\dots,k-1\}}{k} = 0.$$

Proposition 21.6. *Let $(a_k)_{k\geq 0}$ be a bounded sequence of real numbers. Then the following are equivalent:*

(i) $\lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} |a_i| = 0$,

(ii) *There exists a subset $K \subseteq \{0,1,2,\dots\}$ of density zero such that $\lim_{k\to\infty} a_k = 0$ provided $k \notin K$.*

(iii) $\lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} |a_i|^2 = 0$.

By a slight abuse of notation, we will write the conclusion of (ii) as

$$\lim_{k\notin K} a_k = 0.$$

This proof is non-examinable.

(♣) *Proof.* Given any subset $J \subseteq \{0,1,2,\dots\}$, let us denote by $N_J(k)$ the cardinality of $J \cap \{0,1,2,\dots,k-1\}$, so that J has density zero if and only if $\lim_{k\to\infty} \frac{1}{k} N_J(k) = 0$.

We first prove that (i) $\Rightarrow$ (ii). Given $n \geq 1$, let J_n denote the set of $k \in \{0,1,2,\dots\}$ such that $|a_k| \geq \frac{1}{n}$. Then $J_1 \subseteq J_2 \subseteq J_3 \subseteq \cdots$. Each set J_n has density zero, since

$$\frac{1}{k}\sum_{i=0}^{k-1} |a_i| \geq \frac{1}{nk} N_{J_n}(k).$$

This means there exists integers $0 = j_0 < j_1 < j_2 < \cdots$ such that

$$\frac{1}{k} N_{J_{n+1}}(k) < \frac{1}{n+1}, \qquad \forall\, k \geq j_n.$$

Now set

$$K := \bigcup_{n=0}^{\infty} \big(J_{n+1} \cap [j_n, j_{n+1}]\big).$$

We claim that K has density zero. Since $J_1 \subseteq J_2 \subseteq \cdots$, if $j_n \leq k < j_{n+1}$ we have

$$\begin{aligned} K \cap [0,k) &= \big(K \cap [0,j_n)\big) \cup \big(K \cap [j_n,k)\big) \\ &\subseteq \big(J_n \cap [0,j_n)\big) \cup \big(J_{n+1} \cap [0,k)\big), \end{aligned}$$

and therefore

$$\begin{aligned} \frac{1}{k} N_K(k) &\leq \frac{1}{k}\big(N_{J_n}(j_n) + N_{J_{n+1}}(k)\big) \\ &\leq \frac{1}{k}\big(N_{J_n}(k) + N_{J_{n+1}}(k)\big) \\ &\leq \frac{1}{n} + \frac{1}{n+1}. \end{aligned}$$

Thus K has density zero as required. Moreover if $k > j_n$ and $k \notin K$ then $k \notin J_{n+1}$ and hence $|a_k| < \frac{1}{n+1}$. Thus $\lim_{k \notin K} |a_k| = 0$. This proves (ii).

Now let us prove that (ii) $\Rightarrow$ (i). Let K be a set of density zero such that $\lim_{k \notin K} a_k = 0$. Since by assumption (a_k) is bounded, there exists $C > 0$ such that

$$|a_k| \le C, \qquad \forall\, k \ge 0. \tag{21.1}$$

Let $\varepsilon > 0$. Then there exists $n = n(\varepsilon)$ such that if

$$|a_k| < \varepsilon, \qquad \forall\, k \ge n \text{ such that } k \notin K, \tag{21.2}$$

and

$$\frac{1}{k} N_K(k) < \varepsilon, \qquad \forall\, k \ge n. \tag{21.3}$$

Now assume that

$$k > \frac{n}{\varepsilon}. \tag{21.4}$$

Then by splitting the sum $\frac{1}{k}\sum_{i=0}^{k-1} |a_i|$ into three pieces we see that

$$\begin{aligned}
\frac{1}{k}\sum_{i=0}^{k-1} |a_i| &= \frac{1}{k}\left(\sum_{i \in K \cap \{0,1,2,\dots,k-1\}} |a_i| + \sum_{i \in \{0,1,2,\dots,n-1\} \setminus K} |a_i| \quad + \sum_{i \in \{n,\dots,k-1\} \setminus K} |a_i| \right) \\
&< \frac{C}{k} N_K(k) + C\frac{n}{k} + \varepsilon \qquad\qquad \text{by (21.1) and (21.2)} \\
&< C\varepsilon + C\varepsilon + \varepsilon \qquad\qquad \text{by (21.3) and (21.4)} \\
&= (2C+1)\varepsilon.
\end{aligned}$$

Thus (i) follows. Finally, (iii) $\Rightarrow$ (i) is obvious, and to see that (i) $\Rightarrow$ (iii), we observe that if K is as in (ii) then

$$\lim_{k \notin K} |a_k| = 0 \qquad \Rightarrow \qquad \lim_{k \notin K} |a_k|^2 = 0.$$

Thus completes the proof. ■

An immediate corollary of Proposition 21.6 is the following alternative characterisations of the weak mixing property.

Corollary 21.7. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. The following are equivalent.*

(i) *f is weakly mixing.*

(ii) *For every pair $A, B \in \mathscr{A}$ there exists a set $K(A,B) \subseteq \{0,1,2\dots\}$ of density zero such that*

$$\lim_{k \notin K(A,B)} \mu(f^{-k}A \cap B) = \mu(A)\mu(B).$$

(iii) *For every pair $A, B \in \mathscr{A}$ one has*

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \left| \mu(f^{-i}A \cap B) - \mu(A)\mu(B) \right|^2 = 0.$$

Proof. Apply Proposition 21.6 with $a_k := \mu(f^{-k}A \cap B) - \mu(A)\mu(B)$. ∎

REMARK 21.8. In Problem Sheet K you will show that if $(X, \mathscr{A}, \mu)$ has a countable basis (Definition 18.43) then part (ii) of Corollary 21.7 can be improved to show that the set $K(A, B)$ can be chosen independently of A and B.

In the topological world, weak mixing was defined in terms of the product dynamical system $f \times f$. A similar thing is true in the measure-theoretic world, as we now explain.

DEFINITION 21.9. Suppose $(X, \mathscr{A})$ is a measurable space. Let $\mathscr{S}$ denote the semi-algebra on $X \times X$ given by sets of the form $A \times B$ for $A, B \in \mathscr{A}$. The sigma-algebra generated (cf. Definition 18.16) by this semi-algebra is denoted by $\mathscr{A} \times \mathscr{A}$ and is called the **product sigma-algebra**. Next, if μ is a probability measure on $\mathscr{A}$ then

$$\tilde{\mu}(A \times B) := \mu(A)\mu(B)$$

is a probability pre-measure on $\mathscr{S}$. Thus Theorem 18.17 tells us that there is a unique probability measure on $\mathscr{A} \times \mathscr{A}$ that agrees with $\tilde{\mu}$ on $\mathscr{S} \subseteq \mathscr{A} \times \mathscr{A}$. We denote this probability measure by $\mu \times \mu$ and call it the **product measure**.

We adopt the convention that if $(X, \mathscr{A}, \mu)$ is a probability space then $X \times X$ should always be considered with the product sigma-algebra $\mathscr{A} \times \mathscr{A}$ and the product measure $\mu \times \mu$, even if this is not explicitly stated.

We conclude this lecture by proving the measure-theoretic analogue of Furstenberg's Theorem 6.5.

THEOREM 21.10 (Measure-theoretic version of Furstenberg's Theorem). *Let f be a measure-preserving transformation on a probability space $(X, \mathscr{A}, \mu)$. Then the following are equivalent:*

(i) *f is weakly mixing.*

(ii) *$f \times f$ is ergodic.*

(iii) *$f \times f$ is weakly mixing.*

REMARK 21.11. The equivalence of (i) and (ii) explains why this definition of weak mixing is analogous to the topological one (Definition 5.6). The equivalence of (i) and (iii) is the analogue of Furstenberg's Theorem 6.5.

Proof. Let us first prove that (i) implies (iii). Let $A, B, C, D \in \mathscr{A}$. Since f is weakly mixing, by part (ii) of Corollary 21.7 there exist subsets $K_1 = K(A, B)$ and $K_2 = K(C, D)$ of density zero such that

$$\lim_{k \notin K_1} \mu(f^{-k}A \cap B) = \mu(A)\mu(B),$$

and

$$\lim_{k \notin K_2} \mu(f^{-k}C \cap D) = \mu(C)\mu(D).$$

Let $K := K_1 \cup K_2$. Then K is also a set of density zero, and we have

$$\begin{aligned}\lim_{k \notin K} (\mu \times \mu)\Big((f \times f)^{-k}(A \times C) \cap (B \times D)\Big) &= \lim_{k \notin K} \mu(f^{-k}A \cap B)\mu(f^{-k}C \cap D) \\ &= \mu(A)\mu(B)\mu(C)\mu(D) \\ &= (\mu \times \mu)(A \times C)(\mu \times \mu)(B \times D).\end{aligned}$$

Since sets of the form $A \times C$ form a semi-algebra that generates the sigma-algebra on $X \times X$, it follows from Proposition 21.4 and Corollary 21.7 that $f \times f$ is weakly mixing. This proves (iii). The fact that (iii) implies (ii) is obvious. Let us now prove that (ii) implies (i). Let $A, B \in \mathscr{A}$. We have

$$\begin{aligned}\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \mu\left(f^{-i}A \cap B\right) &= \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} (\mu \times \mu)\Big((f \times f)^{-i}(A \times X) \cap (B \times X)\Big) \\ &\overset{(\heartsuit)}{=} (\mu \times \mu)(A \times X)(\mu \times \mu)(B \times X), \\ &= \mu(A)\mu(B)\end{aligned}$$

where $(\heartsuit)$ used that $f \times f$ is ergodic. Similarly we have

$$\begin{aligned}\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \mu(f^{-i}A \cap B)^2 &= \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} (\mu \times \mu)\Big((f \times f)^{-i}(A \times A) \cap (B \times B)\Big) \\ &\overset{(\heartsuit)}{=} (\mu \times \mu)(A \times A)(\mu \times \mu)(B \times B) \\ &= \mu(A)^2\mu(B)^2\end{aligned}$$

where $(\heartsuit)$ again used that $f \times f$ is ergodic. Putting these two together gives

$$\begin{aligned}&\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \Big(\mu(f^{-i}A \cap B) - \mu(A)\mu(B)\Big)^2 \\ &\quad = \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \Big(\mu(f^{-i}A \cap B)^2 - 2\mu(f^{-i}A \cap B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2\Big) \\ &\quad = 2\mu(A)^2\mu(B)^2 - 2\mu(A)^2\mu(B)^2 \\ &\quad = 0.\end{aligned}$$

Thus by Corollary 21.7 we see that f is weakly mixing. This proves (i), and thus completes the proof. ∎

Spectral properties of Dynamical Systems

In this lecture we discuss a "functional analytic" interpretation of mixing and weakly mixing for measure-preserving dynamical systems. Unfortunately we are only able to scratch the surface, since most of the results in this direction require considerably more functional analysis that this course assumes[1] as a prerequisite.

We begin with the following statement.

PROPOSITION 22.1. *Let f be a measure-preserving transformation of a probability space $(X, \mathscr{A}, \mu)$. Then the following are equivalent:*

(i) *f is mixing.*

(ii) *For all $u, v \in L^2(\mu; \mathbb{C})$, one has*

$$\lim_{k\to\infty} \langle\!\langle (f^*)^k(u), v \rangle\!\rangle = \int_X u \, d\mu \int_X \overline{v} \, d\mu.$$

(iii) *For all $u \in L^2(\mu; \mathbb{C})$, one has*

$$\lim_{k\to\infty} \langle\!\langle (f^*)^k(u), u \rangle\!\rangle = \int_X u \, d\mu \int_X \overline{u} \, d\mu.$$

Proof. To see that (ii) implies (i), given $A, B \in \mathscr{A}$ take $u = \mathbb{1}_A$ and $v = \mathbb{1}_B$. To see that (i) implies (iii), observe that (i) implies that for any $A, B \in \mathscr{A}$, one has

$$\lim_{k\to\infty} \langle\!\langle (f^*)^k(\mathbb{1}_A), \mathbb{1}_B \rangle\!\rangle = \mu(A)\mu(B).$$

Fixing B, we see that for any simple function v, one has

$$\lim_{k\to\infty} \langle\!\langle (f^*)^k(v), \mathbb{1}_B \rangle\!\rangle = \int_X v \, d\mu \cdot \mu(B).$$

Then fixing v we see that

$$\lim_{k\to\infty} \langle\!\langle (f^*)^k(v), v \rangle\!\rangle = \int_X v \, d\mu \int_X \overline{v} \, d\mu,$$

i.e. that (iii) is true for any simple function v. Now given $u \in L^2(\mu; \mathbb{C})$ and $\varepsilon > 0$, choose a (possibly complex valued) simple function v such that $\|u - v\|_2 < \varepsilon$. Then choose $n = n(\varepsilon) > 0$ such that for $k \geq n$ one has

$$\left| \langle\!\langle (f^*)^k(v), v \rangle\!\rangle - \int_X v \, d\mu \int_X \overline{v} \, d\mu \right| < \varepsilon.$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]Which is, essentially, zero.

To ease the notation set

$$a := \int_X u \, d\mu, \qquad \text{and} \qquad b := \int_X v \, d\mu.$$

Then if $k \geq n$ one has

$$\begin{aligned}
\left| \langle\!\langle (f^*)^k(u), u \rangle\!\rangle - a\bar{a} \right| \leq& \left| \langle\!\langle (f^*)^k(u), u \rangle\!\rangle - \langle\!\langle (f^*)^k(v), u \rangle\!\rangle \right| \\
&+ \left| \langle\!\langle (f^*)^k(v), u \rangle\!\rangle - \langle\!\langle (f^*)^k(v), v \rangle\!\rangle \right| + \left| \langle\!\langle (f^*)^k(v), v \rangle\!\rangle - b\bar{b} \right| \\
&+ |b\bar{b} - a\bar{b}| + |a\bar{b} - a\bar{a}| \\
\leq& \left| \langle\!\langle (f^*)^k(u-v), u \rangle\!\rangle \right| + \left| \langle\!\langle (f^*)^k(v), u - v \rangle\!\rangle \right| \\
&+ \varepsilon + |\bar{b}| \left| \int_X (u - v) \, d\mu \right| + |a| \left| \int_X \overline{(v - u)} \, d\mu \right| \\
\leq& \|u - v\|_2 \|u\|_2 + \|u - v\|_2 \|v\|_2 + \varepsilon \\
&+ \|v\|_2 \|u - v\|_2 + \|u\|_2 \|v - u\|_2 \\
\leq& \varepsilon \|u\|_2 + \varepsilon(\|u\|_2 + \varepsilon) + \varepsilon + (\|u\|_2 + \varepsilon)\varepsilon + \varepsilon \|u\|_2 \\
=& 4\varepsilon \|u\|_2 + 2\varepsilon^2 + \varepsilon.
\end{aligned}$$

This proves (iii). Finally to see that (iii) implies (ii), we argue as follows. Fix $u \in L^2(\mu; \mathbb{C})$ and let

$$\mathcal{L}_u := \left\{ v \in L^2(\mu; \mathbb{C}) \;\middle|\; \lim_{k \to \infty} \langle\!\langle (f^*)^k(u), v \rangle\!\rangle = \int_X u \, d\mu \int_X \bar{v} \, d\mu \right\}.$$

Then $\mathcal{L}_u$ is a closed f^*-invariant subspace of $L^2(\mu; \mathbb{C})$ which contains both u (by (iii)) and the constant functions. We must show that actually $\mathcal{L}_u = L^2(\mu; \mathbb{C})$.

To prove this, let $\mathcal{S}_u \subset L^2(\mu; \mathbb{C})$ denote the smallest closed f^*-invariant subspace containing u and the constant functions. Then $\mathcal{S}_u \subseteq \mathcal{L}_u$ by definition. Now consider the orthogonal complement

$$\mathcal{S}_u^\perp = \left\{ v \in L^2(\mu; \mathbb{C}) \mid \langle\!\langle v, w \rangle\!\rangle = 0, \ \forall \, w \in \mathcal{S}_u \right\}.$$

Suppose $v \in \mathcal{S}_u^\perp$. Since the constant function $\mathbb{1}_X$ belongs to $\mathcal{S}_u$ we have

$$0 = \langle\!\langle v, \mathbb{1}_X \rangle\!\rangle = \int_X v \, d\mu. \tag{22.1}$$

Similarly since $(f^k)^*(u) \in \mathcal{S}_u$ for any $k \geq 0$ (as $u \in \mathcal{S}_u$ and $\mathcal{S}_u$ is f^*-invariant) we also have

$$\langle\!\langle v, (f^*)^k(u) \rangle\!\rangle = 0, \qquad \forall \, k \geq 0. \tag{22.2}$$

Combining (22.1) and (22.2) shows that if $v \in \mathcal{S}_u^\perp$ then $v \in \mathcal{L}_u$. Thus $\mathcal{S}_u^\perp \subseteq \mathcal{L}_u$. But now we are done, since

$$L^2(\mu; \mathbb{C}) = \mathcal{S}_u + \mathcal{S}_u^\perp \subseteq \mathcal{L}_u.$$

This establishes (ii), and so completes the proof. ■

The next two results can be proved in exactly the same way as Proposition 22.1. The proof is left to the interested reader.

Proposition 22.2. *Let f be a measure-preserving transformation of a probability space $(X, \mathscr{A}, \mu)$. Then the following are equivalent:*

(i) f is ergodic.

(ii) For all $u, v \in L^2(\mu; \mathbb{C})$, one has

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \langle\!\langle (f^*)^i(u), v \rangle\!\rangle = \int_X u \, d\mu \int_X \overline{v} \, d\mu.$$

(iii) For all $u \in L^2(\mu; \mathbb{C})$, one has

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \langle\!\langle (f^*)^i(u), u \rangle\!\rangle = \int_X u \, d\mu \int_X \overline{u} \, d\mu.$$

Proposition 22.3. *Let f be a measure-preserving transformation of a probability space $(X, \mathscr{A}, \mu)$. Then the following are equivalent:*

(i) *f is weakly mixing.*

(ii) *For all $u, v \in L^2(\mu; \mathbb{C})$, one has*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \left| \langle\!\langle (f^*)^i(u), v \rangle\!\rangle - \int_X u \, d\mu \int_X \overline{v} \, d\mu \right| = 0.$$

(iii) *For all $u \in L^2(\mu; \mathbb{C})$, one has*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \left| \langle\!\langle (f^*)^i(u), u \rangle\!\rangle - \int_X u \, d\mu \int_X \overline{u} \, d\mu \right| = 0.$$

Let us now look at *eigenvalues* of a measure-preserving dynamical system.

Definition 22.4. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. We say a complex number ζ is an **eigenvalue** of f if it is an eigenvalue of the isometry $f^* \colon L^2(\mu; \mathbb{C}) \to L^2(\mu; \mathbb{C})$. Thus ζ is an eigenvalue of f is and only if there exists a non-zero $u \in L^2(\mu; \mathbb{C})$ such that

$$u(f(x)) = \zeta u(x), \qquad \text{for almost every } x \in X.$$

Such a function u is called an **eigenfunction** corresponding to the eigenvalue ζ.

Here are some elementary properties.

Lemma 22.5. *Let f be a dynamical system on $(X, \mathcal{A}, \mu)$. Then:*

(i) *$\zeta = 1$ is always an eigenvalue of f.*

(ii) *Every eigenvalue ζ of f satisfies $|\zeta| = 1$.*

(iii) *If f is ergodic then any eigenfunction with eigenvalue 1 is constant almost everywhere.*

Proof. The proof of (i) is immediate, since any non-zero constant function is an eigenfunction with eigenvalue 1. To prove (ii) observe that since f^* is an isometry by Proposition 19.16, if u is an eigenfunction with eigenvalue ζ then

$$\|u\|_2^2 = \|f^*(u)\|_2^2 = \langle\!\langle \zeta u, \zeta u \rangle\!\rangle = |\zeta|^2 \|u\|_2^2.$$

Thus $|\zeta| = 1$. Finally (iii) follows immediately from part (ii) of Proposition 19.17. ■

PROPOSITION 22.6. *Let f be a measure-preserving transformation of a probability space $(X, \mathscr{A}, \mu)$. If f is weakly mixing then the only eigenvalue of f is $\zeta = 1$.*

Proof. Suppose for contradiction that there exists an eigenvalue $\zeta \neq 1$. Let u denote an eigenfunction corresponding to ζ. Then

$$\int_X u \, d\mu = \int_X f^*(u) \, d\mu = \zeta \int_X u \, d\mu$$

by Proposition 19.16, and hence $\int_X u \, d\mu = 0$. Since f is weakly mixing, by Proposition 22.3 one thus has

$$\frac{1}{k} \sum_{i=0}^{k-1} |\langle\!\langle (f^*)^i(u), u \rangle\!\rangle| \to 0,$$

and hence

$$\frac{1}{k} \sum_{i=0}^{k-1} |\langle\!\langle \zeta^i u, u \rangle\!\rangle| \to 0.$$

Since $|\zeta| = 1$, this implies that $\langle\!\langle u, u \rangle\!\rangle = 0$, and hence $u = 0$ almost everywhere. This contradicts u being an eigenfunction, and thus completes the proof. ■

If we assume that f is invertible, then the converse to this result holds:

THEOREM 22.7. *Let f be an invertible measure-preserving transformation of a probability space $(X, \mathscr{A}, \mu)$. Then f is weakly mixing if and only if f is ergodic and the only eigenvalue of f is $\zeta = 1$.*

The proof of Theorem 22.7 uses tools from functional analysis, and therefore:

The rest of this lecture is **non-examinable**.

More precisely, we require the following version[2] of the Spectral Theorem.

[2]The Spectral Theorem is one of those results that you have probably seen proved in numerous guises, starting in Linear Algebra. The version we use here is not remotely the most general one, but it is formulated in a somewhat more advanced fashion than you may be used to.

Theorem 22.8 (Spectral Theorem for Unitary Operators). *Let $(H, \langle\cdot,\cdot\rangle)$ be a complex Hilbert space and $L\colon H \to H$ a bounded unitary operator. Then for each $u \in H$ there exists a unique finite Borel measure ν_u on S^1 such that*

$$\langle L^k u, u\rangle = \int_{S^1} z^k \, d\nu_u, \qquad \text{for all } k \in \mathbb{Z}.$$

If V denotes the closure of the linear span of all eigenvectors of L and $u \in V^\perp$ then the measure ν_u is atomless.

We now prove Theorem 22.7.

(♣) *Proof.* We need only prove that if f is ergodic and the only eigenvalue of f is 1 then f is weakly mixing. Take $u \in L^2(\mu; \mathbb{C})$. We must show that

$$\frac{1}{k}\sum_{i=0}^{k-1}\left| \langle\!\langle (f^*)^i(u), u\rangle\!\rangle - \int_X u\, d\mu \int_X \overline{u}\, d\mu \right| \to 0. \tag{22.3}$$

If u is constant almost everywhere then (22.3) is immediate. Thus without loss of generality we may assume that $\int_X u\, d\mu = 0$. Since f is reversible and measure-preserving, by Proposition 19.16 the operator $f^*\colon L^2(\mu;\mathbb{C}) \to L^2(\mu;\mathbb{C})$ is unitary, and hence the Spectral Theorem is applicable.

Denote by ν_u the finite Borel measure on S^1 given to us via Theorem 22.8. It thus suffices to show that

$$\frac{1}{k}\sum_{i=0}^{k-1}\left|\int_{S^1} z^i \, d\nu_u\right|^2 \to 0.$$

For this we compute[3]:

$$\begin{aligned}
\frac{1}{k}\sum_{i=0}^{k-1}\left|\int_{S^1} z^i d\nu_u\right|^2 &= \frac{1}{k}\sum_{i=0}^{k-1}\left(\int_{S^1} z^i d\nu_u \cdot \int_{S^1} z^{-i} d\nu_u\right) \\
&\overset{(\heartsuit)}{=} \frac{1}{k}\sum_{i=0}^{k-1}\int_{S^1}\int_{S^1} (z\overline{w})^i d(\nu_u \times \nu_u) \\
&= \int_{S^1}\int_{S^1} \frac{1}{k}\left(\sum_{i=0}^{k-1}(z\overline{w})^i\right) d(\nu_u \times \nu_u),
\end{aligned}$$

where (♡) used Fubini's Theorem. Let

$$\varphi_k\colon S^1 \times S^1 \to \mathbb{R}, \qquad \varphi(z,w) := \frac{1}{k}\sum_{i=0}^{k-1} z\overline{w}^i,$$

so that the computation above shows that

$$\frac{1}{k}\sum_{i=0}^{k-1}\left|\int_{S^1} z^i d\nu_u\right|^2 = \lim_{k\to\infty}\int_{S^1\times S^1} \varphi_k \, d(\nu_u \times \nu_u).$$

[3]In this computation z, and later, w, are elements of S^1.

To complete the proof we show that

$$\lim_{k\to\infty} \varphi_k(z,w) \to 0, \qquad \text{for } \nu_u \times \nu_u \text{ almost every } (z,w), \tag{22.4}$$

from which (22.3) follows by the Dominated Convergence Theorem 18.34. To show (22.4), note first that if $z \neq w$ then

$$\varphi_k(z,w) = \frac{1-(z\overline{w})^k}{k(1-(z\overline{w})} \to 0$$

as $k \to \infty$.

This argument doesn't work for $z = w$ however, and this is where we finally need to use the assumptions of the theorem (so far we have only used the fact that f is reversible and measure-preserving). Since f is ergodic and 1 is the only eigenvalue of f, the only eigenfunctions of f are the constants by part (iii) of Lemma 22.5. Thus the assumption $\int_X u\, d\mu = 0$ implies that u is orthogonal to all eigenfunctions by (cf. (22.1)). This means that the measure ν_u is atomless by the last statement of the Spectral Theorem . Therefore the measure of the diagonal $\{(z,z) \mid z \in S^1\}$ (as a subset of $S^1 \times S^1$) under $\nu_u \times \nu_u$ is zero. Thus (22.4) follows, and hence so does (22.3). The proof is complete. ■

LECTURE 23

Measures on Metric Spaces

In the last few lectures we started with a probability space $(X, \mathscr{A}, \mu)$, and then looked at transformations $f \colon X \to X$ which preserve μ. Over the next two lectures we flip this on its head. Rather than starting with the measure μ and then restricting attention to transformations f that preserve μ, now we will start with f and look for measures for which f is measure-preserving.

Such a paradigm shift is possible purely within the measure-theoretic world, but it is maximally profitable if we begin in a topological setting. Indeed, suppose we are given a topological dynamical system f on a compact metric space (X, d). Let $\mathscr{B}$ denote the Borel sigma-algebra. Can we find a probability measure μ on $(X, \mathscr{B})$ for which f becomes[1] measure-preserving? If we are successful in our quest to find such a μ, the dynamical system f will then simultaneously be a topological dynamical system and a measure-preserving dynamical system. The benefits of this approach should be clear: we can then bring all the results from both topological and measure-theoretic dynamics to bear when studying the dynamics of f.

In order to have effective methods to "find" measures for which our given topological dynamical system f is measure-preserving, we need to understand what properties the space of *all* probability measures on the Borel sigma-algebra of X has. For example, does it carry a topology? If so, is it compact? In fact, the answer to both of these questions is yes: the space of all probability measures on the Borel sigma-algebra of X is itself a compact metric space. We will prove this today.

> Throughout our discussion of measures on metric spaces, we will always assume that the underlying metric spaces are **compact**.

DEFINITION 23.1. Let X be a compact metric space, and let $\mathscr{B}$ denote the Borel sigma-algebra on X. We denote by $\mathcal{M}(X)$ the space of all probability measures μ on $(X, \mathscr{B})$.

We call elements of $\mathcal{M}(X)$ simply *Borel probability measures* on X (this saves having to constantly explicitly label $\mathscr{B}$). We will prove shortly that $\mathcal{M}(X)$ is itself a compact metric space. One can think of $\mathcal{M}(X)$ as an "enlargement" of the space X. Indeed, there is an obvious inclusion $X \hookrightarrow \mathcal{M}(X)$ given by sending a point to its corresponding Dirac measure (Example 18.22):

$$\imath \colon X \to \mathcal{M}(X), \qquad x \mapsto \delta_x. \tag{23.1}$$

On Problem Sheet L you will prove that—when $\mathcal{M}(X)$ is given its metric space structure discussed in this lecture—the map (23.1) is an embedding (i.e. a homeomorphism onto its image). In general the map $\imath$ is not surjective; the space $\mathcal{M}(X)$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1] Note that since f is continuous (by definition of a topological dynamical system), f is automatically measurable with respect to any measure on $(X, \mathscr{B})$ by Example 19.3.

can be much larger than X. Nevertheless, in some ways the space $\mathcal{M}(X)$ is easier to handle than X itself. For example:

LEMMA 23.2. *The space $\mathcal{M}(X)$ is convex: given $\mu, \nu \in \mathcal{M}(X)$ and $c \in [0,1]$, the measure $c\mu + (1-c)\nu$ belongs to $\mathcal{M}(X)$.*

Here $c\mu + (1-c)\nu$ is defined as you would guess:

$$(c\mu + (1-c)\nu)(A) := c\mu(A) + (1-c)\nu(A), \qquad \forall\, A \in \mathscr{B}.$$

The proof of Lemma 23.2 is immediate. In contrast, note that for an arbitrary metric space X, it does *not* make sense to form the "sum" $x + y$ of two points, let alone form convex combinations.

REMARK 23.3. The development of statistical mechanics from Hamiltonian mechanics essentially began with the idea that one could replace points by measures. That is, instead of considering the "states" that a given system can be in to be given by elements of some set X, one instead considers the states to be probability measures on X. (Compare this to discussion at the beginning of Lecture 1). This new viewpoint is more general, since as we have just observed, there are more measures than points.

The proof that $\mathcal{M}(X)$ is itself a compact metric space requires some measure-theoretic preliminaries.

PROPOSITION 23.4. *Let $\mu \in \mathcal{M}(X)$. Then for any set $A \in \mathscr{B}$ and any $\varepsilon > 0$ there exists an open set U and a closed set C such that $C \subseteq A \subseteq U$ and such that $\mu(U \setminus C) < \varepsilon$.*

Proof. Let us denote by $\mathscr{A}$ the collection of subsets $A \in \mathscr{B}$ for which the stated property holds. We claim that $\mathscr{A}$ is itself a sigma-algebra. Clearly $X \in \mathscr{A}$. Now suppose $A \in \mathscr{A}$. Let us show $X \setminus A \in \mathscr{A}$. Fix $\varepsilon > 0$ and choose U open and C closed such that $C \subseteq A \subseteq U$ with $\mu(U \setminus C) < \varepsilon$. Then

$$X \setminus U \subseteq X \setminus A \subseteq X \setminus C$$

and

$$(X \setminus C) \setminus (X \setminus U) = U \setminus C,$$

so that $\mu((X \setminus C) \setminus (X \setminus U)) < \varepsilon$. Since $X \setminus U$ is closed and $X \setminus C$ is open, this shows that $X \setminus A \in \mathscr{A}$.

Next we show that $\mathscr{A}$ is closed under countable unions. Suppose $(A_k)_{\{k \in \mathbb{N}\}} \subset \mathscr{A}$. Set $A := \bigcup_k A_k$, and fix $\varepsilon > 0$. For each k, we may choose

$$C_k \subseteq A_k \subseteq U_k,$$

with

$$\mu(U_k \setminus C_k) < \frac{\varepsilon}{3^k}.$$

Let $U := \bigcup_k U_k$ and $C := \bigcup_k C_k$. Then U is open but C may not be closed. Nevertheless, we may choose $n \geq 1$ such that $C' := \bigcup_{k=1}^n C_k$ is closed and such that $\mu(C \setminus C') < \varepsilon/2$. Then $C' \subseteq A \subseteq U$, and

$$\begin{aligned}\mu(U \setminus C') &\leq \mu(U \setminus C) + \mu(C \setminus C') \\ &\leq \sum_{k=1}^{\infty} \mu(U_k \setminus C_k) + \frac{\varepsilon}{2} \\ &\leq \sum_{k=1}^{\infty} \frac{\varepsilon}{3^k} + \frac{\varepsilon}{2} \\ &\leq \varepsilon.\end{aligned}$$

To complete the proof is suffices to show that $\mathscr{A}$ contains all the closed sets, since $\mathscr{B}$ is the smallest sigma-algebra with this property (by definition) and $\mathscr{A} \subseteq \mathscr{B}$ (again by definition), whence $\mathscr{A} = \mathscr{B}$. To see this, let A be a closed set and fix $\varepsilon > 0$. In this case we may take $C = A$, so it suffices to find an open set U containing A such that $\mu(U \setminus A) < \varepsilon$. This is not so hard: set

$$U_k := \{x \in X \mid d(x, A) < \tfrac{1}{k}\}.$$

Then each U_k is open, with $U_{k+1} \subseteq U_k$. Since $\bigcap_{k=1}^{\infty} U_k = A$, it follows that for k large enough, we have $\mu(U_k \setminus A) < \varepsilon$ (cf. the footnote from the proof of Proposition 19.9). This completes the proof. ∎

Corollary 23.5. *For any $\mu \in \mathcal{M}(X)$ and any $A \in \mathscr{B}$, we have $\mu(A) = \sup_C \mu(C)$, where the supremum is taken over all closed sets $C \subseteq A$, and similarly $\mu(A) = \inf_U \mu(U)$, where the infimum is taken over all open sets U with $A \subseteq U$.*

Definition 23.6. We abbreviate $\mathcal{C}(X) := C^0(X, \mathbb{C})$ for the space of all continuous functions $u\colon X \to \mathbb{C}$. This is a complex vector space under pointwise addition and scalar multiplication by constants. In fact, $\mathcal{C}(X)$ is a complex algebra, since we can also multiple elements together pointwise. We endow $\mathcal{C}(X)$ with the supremum norm

$$\|u\|_\infty := \sup_{x \in X} |u(x)|.$$

This makes $(\mathcal{C}(X), \|\cdot\|_\infty)$ into a Banach algebra. It follows from the Stone-Weierstrass Theorem that $\mathcal{C}(X)$ is separable.

Proposition 23.7. *Let $\mu, \nu \in \mathcal{M}(X)$. If*

$$\int_X u d\mu = \int_X u d\nu, \qquad \forall\, u \in \mathcal{C}(X)$$

then $\mu = \nu$.

Proof. It suffices to show that $\mu(C) = \nu(C)$ for any closed set C by Corollary 23.5. Let C be a closed set and let $\varepsilon > 0$. We may assume $C \neq X$, as otherwise there is nothing to prove. Then there is an open set U containing C with $\mu(U \setminus C) < \varepsilon$ by Proposition 23.4. Now define $u\colon X \to \mathbb{R}$ by

$$u(x) := \frac{d(x, X \setminus U)}{d(x, X \setminus U) + d(x, C)}. \tag{23.2}$$

This is well defined and continuous as the denominator in the second case can never vanish. Since u is 0 on $X \setminus U$ and 1 on C and satisfies $0 \leq u(x) \leq 1$ for all $x \in X$, one has

$$\begin{aligned} \nu(C) &\leq \int_X u \, d\nu \\ &= \int_X u d\mu \\ &\leq \mu(U) \\ &< \mu(C) + \varepsilon. \end{aligned}$$

Thus $\nu(C) < \mu(C) + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we see that $\nu(C) \leq \mu(C)$. But by symmetry one also has $\mu(C) \leq \nu(C)$. This completes the proof. ■

Denote by $\mathcal{C}(X)^*$ the dual space of $\mathcal{C}(X)$. That is, $\mathcal{C}(X)^*$ is the space of continuous linear functionals on $\mathcal{C}(X)$. This is another Banach space under the norm

$$\|T\|_* := \sup \left\{ |Tu| \mid \|u\|_\infty \leq 1 \right\}.$$

Definition 23.8. A continuous linear functional $T \in \mathcal{C}(X)^*$ is said to be **positive** if $Tu \geq 0$ whenever u is real-valued and non-negative. A positive continuous linear functional T is said to be **normalised** if $T(\mathbb{1}_X) = 1$. We denote by $\mathcal{T}(X) \subseteq \mathcal{C}(X)^*$ the space of positive normalised linear functionals.

The next statement is another foundational result in measure theory.

Theorem 23.9 (Riesz Representation Theorem). *Let X denote a compact metric space. Then for each $T \in \mathcal{T}(X)$ there exists a unique $\mu \in \mathcal{M}(X)$ such that*

$$Tu = \int_X u \, d\mu, \qquad \forall \, u \in \mathcal{C}(X). \tag{23.3}$$

We will merely give a brief sketch of the proof of Theorem 23.9, since filling in all the details would take all lecture.

(♣) *Proof (sketch).* Here is how to define μ: given an open set U, set

$$\mu(U) := \sup \left\{ Tu \,\middle|\, u \colon X \to [0,1] \text{ continuous, with } u|_{X \setminus U} = 0 \right\}.$$

Then for an arbitrary Borel set A, define

$$\mu(A) = \inf\{\mu(U) \mid A \subseteq U, \ U \text{ open}\}$$

(note we have no choice here if μ is to satisfy Corollary 23.5). Uniqueness of μ follows from Proposition 23.7. ■

Remark 23.10. Rather confusingly, there are two different results in functional analysis and measure theory that are usually referred to as the "Riesz Representation Theorem". These two results are not the same—their only relation is that they were both proved by the Hungarian mathematician F. Riesz. One of these results is Theorem 23.9 (which is not stated in anything like maximal generality). The "other" Riesz Representation Theorem is rather more elementary. It states that a complex Hilbert space H is isometrically (anti)-isomorphic to its dual space H^*.

Corollary 23.11. *The space $\mathcal{T}(X)$ is a closed convex subset of the unit ball*[2] *in* $\mathcal{C}(X)^*$.

Proof. It is clear that $\mathcal{T}(X)$ is closed and convex. To prove that $\|T\|_* = 1$, we use Theorem 23.9. If $\mu \in \mathcal{M}(X)$ satisfies (23.3) then for any $u \in \mathcal{C}(X)$ we have

$$|Tu| = \left| \int_X u \, d\mu \right| \leq \|u\|_\infty \cdot \mu(X) = \|u\|_\infty.$$

This shows that $\|T\|_* \leq 1$. Finally, since $T(\mathbb{1}_X) = 1$ we also have $\|T\|_* \geq 1$, and thus $\|T\|_* = 1$. This completes the proof. ■

Definition 23.12. Let X be a compact metric space. The **weak star topology** on $\mathcal{C}(X)^*$ is defined by declaring that a sequence T_k converges to T if and only if

$$T_k u \to Tu, \qquad \forall\, u \in \mathcal{C}(X).$$

It is customary to use the notation $T_k \rightharpoonup T$ to indicate that T_k converges to T in the weak star topology.

As the name suggests, the weak star topology is coarser than the norm topology. This means that if $\|T_k - T\|_* \to 0$ then also $T_k \rightharpoonup T$, but the converse need not be true.

We now transfer this topology to $\mathcal{M}(X)$. By the Riesz Representation Theorem there is a bijective map

$$\mathcal{T} \colon \mathcal{M}(X) \to \mathcal{T}(X), \qquad \mathcal{T}(\mu) := T_\mu$$

where

$$T_\mu(u) := \int_X u \, d\mu.$$

Definition 23.13. Consider $\mathcal{T}(X)$ equipped with the weak star topology. We define a topology on $\mathcal{M}(X)$ by declaring that $\mathcal{T}$ is a homeomorphism. This means that a sequence $(\mu_k) \subset \mathcal{M}(X)$ converges to an element $\mu \in \mathcal{M}(X)$ if and only if the corresponding sequence $\mathcal{T}(\mu_k)$ converges to $\mathcal{T}(\mu)$ in $\mathcal{T}(X)$. More explicitly, μ_k converges to μ if and only if

$$\int_X u \, d\mu_k \to \int_X u \, d\mu, \qquad \forall\, u \in \mathcal{C}(X).$$

This is called the **weak star topology** on $\mathcal{M}(X)$. We again use the notation $\mu_k \rightharpoonup \mu$ to indicate that μ_k converges to μ in the weak star topology.

This is the smallest topology making the maps $\mu \mapsto \int_X u \, d\mu$ continuous for each fixed $u \in \mathcal{C}(X)$. A basis is given by the collection of sets of the form

$$\mathcal{N}_\mu(v_1, \dots, v_p; \varepsilon) := \left\{ \nu \in \mathcal{M}(X) \;\middle|\; \left| \int_X v_i \, d\mu - \int_X v_i \, d\nu \right| < \varepsilon, \ \forall\, 1 \leq i \leq p \right\},$$

[2] In functional analysis the “unit ball” is (by definition) the set of vectors of norm 1. This is a somewhat unfortunate discrepancy with topological terminology (the unit “ball” is not a ball, but rather a sphere!)

where $\mu \in \mathcal{M}(X)$, $p \geq 1$, $v_i \in \mathcal{C}(X)$ and $\varepsilon > 0$.

We will now prove that this topology is metrisable. For this first choose a countable dense subset $(w_k) \subset \mathcal{C}(X)$ (recall from Definition 23.6 that $\mathcal{C}(X)$ is separable as X is compact).

Proposition 23.14. *The following is a metric on $\mathcal{M}(X)$ inducing the weak star topology:*

$$d_{\mathcal{M}}(\mu, \nu) := \sum_{k=1}^{\infty} \frac{\left|\int_X w_k \, d\mu - \int_X w_k \, d\nu\right|}{2^k \|w_k\|_\infty}.$$

This proof is non-examinable.

(♣) *Proof.* We prove the result in three steps.

1. In this first step we prove that $d_{\mathcal{M}}$ is a metric. The only non-trivial part is to show that $d_{\mathcal{M}}(\mu, \nu) = 0$ implies $\mu = \nu$. By Proposition 23.7 it suffices to show that

$$d_{\mathcal{M}}(\mu, \nu) = 0 \qquad \Rightarrow \qquad \int_X u \, d\mu = \int_X u \, d\nu, \qquad \forall\, u \in \mathcal{C}(X).$$

Since (w_k) is a dense sequence, we can pick a subsequence (w_{k_n}) such that $w_{k_n} \to u$ uniformly. Then by the Dominated Convergence Theorem 18.34 we have

$$\begin{aligned}\left|\int_X u \, d\mu - \int_X u \, d\nu\right| &= \lim_{n\to\infty} \left|\int_X w_{k_n} d\mu - \int_X w_{k_n} \, d\nu\right| \\ &\leq \lim_{n\to\infty} 2^{k_n} \|w_{k_n}\|_\infty \, d_{\mathcal{M}}(\mu, \nu) \\ &= 0.\end{aligned}$$

2. In this second step, we show that for any $u \in \mathcal{C}(X)$, the functional

$$L_u \colon \mathcal{M}(X) \to \mathbb{R}, \qquad L_u(\mu) := \int_X u \, d\mu = T_\mu(u) \tag{23.4}$$

is continuous with respect to $d_{\mathcal{M}}$. For this $u \in \mathcal{C}(X)$ and $\varepsilon > 0$. As before we find a subsequence $(w_{k_n}) \subseteq (w_k)$ such that w_{k_n} converges to u uniformly. Thus there exists $m = m(\varepsilon)$ such that for all $n \geq m$ one has

$$\|u - w_{k_n}\|_\infty < \varepsilon. \tag{23.5}$$

Fix an $n \geq m$ and set

$$C := 2^{k_n} \|w_{k_n}\|_\infty.$$

Then we have

$$\begin{aligned}\left|L_u(\mu) - L_u(\nu)\right| &\leq \left|L_u(\mu) - L_{w_{k_n}}(\mu)\right| + \left|L_{w_{k_n}}(\mu) - L_{w_{k_n}}(\nu)\right| + \left|L_u(\nu) - L_{w_{k_n}}(\nu)\right| \\ &\leq \|u - w_{k_n}\|_\infty \, \mu(X) + 2^k \|w_k\|_\infty \, d_{\mathcal{M}}(\mu, \nu) + \|u - w_{k_n}\|_\infty \, \nu(X) \\ &\leq C d_{\mathcal{M}}(\mu, \nu) + 2\varepsilon.\end{aligned}$$

3. In the final step we show that $d_{\mathcal{M}}$ induces the weak star topology on $\mathcal{M}(X)$. Given a basis set $\mathcal{N}_\mu(v_1, \ldots, v_p; \varepsilon)$, we observe that if I_i denotes the interval

$$I_i := \left(\int_X v_i \, d\mu - \varepsilon, \int_X v_i \, d\mu + \varepsilon \right)$$

then

$$\mathcal{N}_\mu(v_1, \ldots, v_p; \varepsilon) = \bigcap_{i=1}^{k} L_{v_i}^{-1}(I_i).$$

The right-hand side is open in the topology generated by d, since the L_{v_i} are d-continuous by Step 2. Finally we show that for any $\mu \in \mathcal{M}(X)$ and $\delta > 0$ there exist a basis set $\mathcal{N}_\mu(v_1, \ldots, v_p; \varepsilon)$ such that

$$\mathcal{N}_\mu(v_1, \ldots, v_p; \varepsilon) \subseteq B_d(\mu, \delta).$$

For this first choose p such that

$$\sum_{k=p+1}^{\infty} \frac{2}{2^k} < \frac{\delta}{2},$$

and then set

$$\varepsilon := \frac{\delta}{2} \left(\sum_{k=1}^{p} \frac{1}{2^k \|w_k\|_\infty} \right)^{-1}.$$

Then taking $v_i := w_i$ for $1 \le i \le p$, we have

$$\mathcal{N}_\mu(w_1, \ldots, w_p; \varepsilon) \subseteq B_{d_{\mathcal{M}}}(\mu, \delta).$$

This completes the proof. ∎

Right at the very end of the course (see Step 3 of the proof of Theorem 28.2) we will need the following fact. Recall that for any subset $A \subseteq X$, the (topological) boundary is $\partial A := \overline{A} \setminus A^\circ$.

Proposition 23.15. *If $\mu_k \rightharpoonup \mu$ then for any set $A \in \mathscr{B}$ with $\mu(\partial A) = 0$ one has $\mu_k(A) \to \mu(A)$.*

Proof. We will first show that if C is a closed subset of X then $\limsup_k \mu_k(C) \le \mu(C)$. Let $U_n := B(C, 1/n)$ denote the open ball about C of radius $1/n$. Then $\mu(U_n) \to \mu(C)$. As in (23.2), choose $u_n \in \mathcal{C}(X)$ such that $0 \le u_n \le 1$ and such that $u_n = 1$ on C and $u_n = 0$ on $X \setminus U_n$. Then

$$\begin{aligned} \limsup_{k \to \infty} \mu_k(C) &\le \limsup_{k \to \infty} \int_X u_n \, d\mu_k \\ &= \int_X u_n \, d\mu \\ &\le \mu(U_n) \end{aligned}$$

and hence $\limsup_k \mu_k(C) \le \mu(C)$. Similarly if U is an open set then

$$\limsup_k \mu_k(X \setminus U) \le \mu(X \setminus U)$$

and hence

$$\liminf_k \mu_k(U) \geq \mu(U).$$

Now if $A \in \mathscr{B}$ has $\mu(\partial A) = 0$ then $\mu(A^\circ) = \mu(\overline{A}) = \mu(A)$ and thus

$$\begin{aligned}
\limsup_{k\to\infty} \mu_k(A) &\leq \limsup_{k\to\infty} \mu_k(\overline{A}) \\
&\leq \mu(\overline{A}) \\
&= \mu(A^\circ) \\
&\leq \liminf_{k\to\infty} \mu_k(A^\circ) \\
&\leq \liminf_{k\to\infty} \mu_k(A),
\end{aligned}$$

which implies that $\lim \mu_k(A) = \mu(A)$. This completes the proof. ∎

REMARK 23.16. In fact, Proposition 23.15 is an "if and only if" statement. We will not need the converse direction in this course, but it is good to know. So you will no doubt be pleased to learn that this is on Problem Sheet K.

We are now ready to prove that $\mathcal{M}(X)$ is a compact metric space.

THEOREM 23.17. *The space $\mathcal{M}(X)$ is compact in the weak star topology.*

Theorem 23.17 follows immediately from the Banach-Alaoglu Theorem, which states that the closed unit ball of the dual space of a topological vector space is compact in the weak star topology. However we will give a direct proof.

Proof. Suppose (μ_k) is a sequence in $\mathcal{M}(X)$. We will show it has a convergent subsequence in three steps. To simplify the notation let us write $T_k = \mathcal{T}(\mu_k)$, so that

$$T_k u = \int_X u d\mu_k.$$

1. Let (w_i) denote a dense subset of $\mathcal{C}(X)$. In this step we will show that, up to passing to a subsequence, the sequence $(T_k w_i)_{k\in\mathbb{N}}$ of complex numbers is convergent for any i.

Firstly, the sequence of complex numbers $(T_k w_1)_{k\in\mathbb{N}}$ is bounded by $\|w_1\|_\infty$ and hence has a convergent subsequence, call it $\big(T_{k_n^1} w_1\big)_{n\in\mathbb{N}}$. Then the sequence $\big(T_{k_n^1} w_2\big)_{n\in\mathbb{N}}$ also has a convergent subsequence, call it $\big(T_{k_n^2} w_2\big)_{n\in\mathbb{N}}$. In this way we obtain sequences $(k_n^i)_{n\in\mathbb{N}}$ such that $(k_n^{i+1})_{n\in\mathbb{N}}$ is a subsequence of $(k_n^i)_{n\in\mathbb{N}}$ and such that the sequence $\big(T_{k_n^i} w_i\big)_{n\in\mathbb{N}}$ is convergent. Now we use the usual "diagonal" argument trick: set $k'_n := k_n^n$. Then $\big(T_{k'_n} w_i\big)_{n\in\mathbb{N}}$ converges for every i.

To keep the notation under control, we now relabel this subsequence simply by k again. With this understood, we have shown that $(T_k w_i)_{k\in\mathbb{N}}$ converges for every i. This completes Step 1.

2. In this step we show that for *any* $u \in \mathcal{C}(X)$ the sequence $(T_k u)_{k\in\mathbb{N}}$ is convergent. Indeed, fix $u \in \mathcal{C}(X)$ and $\varepsilon > 0$. Choose i such that $\|u - w_i\|_\infty < \varepsilon/4$. Since $\big(T_k w_i\big)_{k\in\mathbb{N}}$ converges it is Cauchy; thus there exists $p = p(\varepsilon)$ such that

$$\big|T_m w_i - T_n w_i\big| < \frac{\varepsilon}{2}, \qquad \forall\, m, n \geq p.$$

Then we estimate that for $m, n \geq p$ one has

$$\begin{aligned}
\left|T_m u - T_n u\right| &\leq \left|T_m u - T_m w_i\right| + \left|T_m w_i - T_n w_i\right| + \left|T_n w_i - T_n u\right| \\
&\leq \|T_n\|_* \|u - w_i\|_\infty + \frac{\varepsilon}{2} + \|T_m\|_* \|u - w_i\|_\infty \\
&\overset{(\heartsuit)}{=} \|u - w_i\|_\infty + \frac{\varepsilon}{2} + \|u - w_i\|_\infty \\
&\leq \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} \\
&= \varepsilon,
\end{aligned}$$

where $(\heartsuit)$ used the fact that $\|T_n\|_* = \|T_m\|_* = 1$ by Proposition 23.11. This shows that $\left(T_k u\right)_{k \in \mathbb{N}}$ is a Cauchy sequence, and hence it converges.

3. We are ready to complete the proof. For a given $u \in \mathcal{C}(X)$ define

$$Tu \coloneqq \lim_n T_n u.$$

Clearly $T \colon \mathcal{C}(X) \to \mathbb{C}$ is linear. Moreover $|Tu| \leq \|u\|_\infty$ since the same is true of each T_k. Similarly $Tu \geq 0$ if $u \geq 0$ and $T(\mathbb{1}_X) = 1$. It thus follows from the Riesz Representation Theorem 23.9 that $Tu = \int_X u \, d\mu$ for some $\mu \in \mathcal{M}(X)$. By definition of the weak star topology, $\mu_k \rightharpoonup \mu$. This completes the proof. ■

For the rest of the course, the space $\mathcal{M}(X)$ should always be understood to be carry the weak star topology, even if this is not explicitly stated.

Finding Invariant Measures

We begin the programme outlined at the beginning of the previous lecture and look for measures which a given topological dynamical system preserves.

DEFINITION 24.1. Let $f\colon X \to X$ denote a topological dynamical system on a compact metric space. Let $\mathcal{M}(f) \subseteq \mathcal{M}(X)$ denote those Borel probability measures μ for which f is measure-preserving. We call an element $\mu \in \mathcal{M}(f)$ an **invariant measure** for f.

Thus if $\mu \in \mathcal{M}(f)$ then f is a measure-preserving dynamical system on the probability space $(X, \mathscr{B}, \mu)$. Our first task at hand is to show that $\mathcal{M}(f)$ is non-empty. To do this we will show that f induces a map $f_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$ with the property that μ is an invariant measure for f if and only if μ is a fixed point of f_*. The existence of such a fixed point will then be an application of a classical fixed point theorem.

DEFINITION 24.2. Let $f\colon X \to X$ denote a dynamical system. We denote by

$$f_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$$

the map given by

$$\mu \mapsto f_*\mu, \qquad \text{where} \qquad f_*\mu(A) := \mu(f^{-1}A), \qquad \forall\, A \in \mathscr{B}.$$

This is well defined (i.e. $f_*\mu$ is a measure) thanks to Example 19.3. It is immediate from the definition that μ is invariant if and only if $f_*\mu = \mu$. Moreover if $\imath\colon X \to \mathcal{M}(X)$ denotes the embedding from (23.1) (the fact that $\imath$ is an embedding is Problem L.1), it is clear that the following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\imath"'] & X \arrow[d, "\imath"] \\ \mathcal{M}(X) \arrow[r, "f_*"'] & \mathcal{M}(X) \end{tikzcd}$$

Thus following result is a generalisation of Proposition 19.16, and can be proved in exactly the same way. The pleasing visual appearance of (24.1) is the reason why we chose the f^* and f_* notation.

PROPOSITION 24.3. *Let $f\colon X \to X$ denote a topological dynamical system on a compact metric space. Then for any $\mu \in \mathcal{M}(X)$ and any measurable u, one has*

$$\int_X f^*(u)\, d\mu = \int_X u\, d(f_*\mu), \tag{24.1}$$

where one side doesn't exist or is infinite if and only if the other is.

Proof. As usual, it suffices to deal with real-valued u. If u is the indicator function $\mathbb{1}_A$ of a measurable set A then both sides are equal to $\mu(f^{-1}A)$. Then by linearity, the result holds for simple functions. Then as in Proposition 19.16, if u is a non-negative measurable function and (v_k) a sequence of simple functions increasing to u then $(f^*(v_k))$ is a sequence of simple functions increasing to $f^*(u)$. Then

$$\begin{aligned}\int_X f^*(u)\,d\mu &= \lim_k \int_X f^*(v_k)\,d\mu \\ &= \lim_k \int_X v_k\,d(f_*\mu) \\ &= \int_X u\,d(f_*\mu).\end{aligned}$$

Finally the general case is proved by considering the positive and negative parts of u. This completes the proof. ∎

Corollary 24.4. *Let $f\colon X \to X$ denote a topological dynamical system. Then $\mu \in \mathcal{M}(f)$ if and only if*

$$\int_X u\,d\mu = \int_X u\,d(f_*\mu), \qquad \forall\, u \in \mathcal{C}(X).$$

The only content of this statement is the fact that it suffices to check that the equality holds for all *continuous* u rather than all measurable u in order to conclude $\mu \in \mathcal{M}(f)$.

Proof. If $\mu \in \mathcal{M}(f)$ then the claim holds by Proposition 19.16 and Proposition 24.3, and the converse follows from Proposition 23.7. ∎

Proposition 24.5. *Let $f\colon X \to X$ denote a dynamical system. Then $f_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$ is a continuous affine map.*

Proof. If $\mu_k \rightharpoonup \mu$ and $u \in \mathcal{C}(X)$ then

$$\begin{aligned}\int_X u\,d(f_*\mu_k) &= \int_X f^*(u)\,d\mu_k \\ &\to \int_X f^*(u)\,d\mu \\ &= \int_X u\,d(f_*\mu)\end{aligned}$$

by Proposition 24.3 and the definition of the weak star topology. Since u was arbitrary it thus follows that $f_*\mu_k \rightharpoonup f_*\mu$, again by definition of the weak star topology.

To show f_* is affine, simply note that if $\mu, \nu \in \mathcal{M}(X)$ and $c \in [0,1]$ then for any $A \in \mathscr{B}$, one has

$$\begin{aligned}f_*(c\mu + (1-c)\nu)(A) &= c\mu(f^{-1}A) + (1-c)\nu(f^{-1}A) \\ &= cf_*\mu(A) + (1-c)f_*\nu(A).\end{aligned}$$

Since A was arbitrary one has $f_*(c\mu + (1-c)\nu) = cf_*\mu + (1-c)f_*\nu$. This completes the proof. ∎

We are now ready to prove that $\mathcal{M}(f)$ is non-empty. The proof that follows is needlessly wordy—this is because we are trying to avoid quoting any theorems from functional analysis.

THEOREM 24.6 (Markov-Kakutani Fixed Point Theorem). *Let K be a compact convex subset of a Banach space E. Let $g\colon K \to K$ be a continuous affine map. Then g has a fixed point in K.*

Proof. We show that the following recipe always produces a fixed point of g:

- Let (x_k) denote any sequence in K. Define a new sequence

$$y_k := \frac{1}{k}\sum_{i=0}^{k-1} g^i(x_k).$$

  Note that $y_k \in K$ since K is convex.

- Since K is compact, we may assume that $y_k \to y$ for some point $y \in K$.

- Then we claim that y is a fixed point of g.

To show this, let E^* denote the dual of E, equipped with its dual norm $\|\cdot\|_*$, and let $S \subseteq E^*$ denote any subset of E^* that *separates* the points of K in the sense that if $z, w \in K$ then

$$s(z) = s(w), \qquad \forall\, s \in S \qquad \Rightarrow \qquad z = w.$$

Such a set S always exists—for example, one could take $S = E^*$, by the Hahn-Banach Theorem. In the case of interest for us, however, we will explicitly construct such an S, and so the Hahn-Banach Theorem is not needed.

Now let $s \in S$ be arbitrary. We will show that

$$s(g(y)) = s(y),$$

whence it follows that $g(y) = y$ as claimed. Set

$$c := \sup_{z\in K} \|z\|.$$

Note $c < \infty$ as K is compact. Then we compute

$$\big|s(g(y_k)) - s(y_k)\big| = \frac{1}{k}|s(g^k(x_k)) - s(x_k)| \le \frac{2c\|s\|_*}{k},$$

since the sum telescopes. Letting $k \to \infty$ shows that $s(g(y)) = s(y)$ as required. ∎

COROLLARY 24.7. *Let $f\colon X \to X$ be a topological dynamical system on a compact metric space X. Then the space $\mathcal{M}(f)$ of invariant measures for f is non-empty.*

Proof. Take $E = \mathcal{C}(X)^*$, and identify $\mathcal{M}(X)$ with the compact convex set $\mathcal{T}(X) \subseteq \mathcal{C}(X)^*$ via the Riesz Representation Theorem. We take[1]

$$S := \{L_u \mid u \in \mathcal{C}(X)\},$$

[1]Note in this case S is not equal to E^*, since $\mathcal{C}(X)$ is not reflexive as a Banach space (unless X is a finite set).

where L_u is defined as in (23.4):

$$L_u \colon \mathcal{M}(X) \to \mathbb{C}, \qquad L_u(\mu) := \int_X u \, d\mu.$$

Proposition 23.7 tells us that S separates points of $\mathcal{M}(X)$. Thus Theorem 24.6 tells us that f_* has a fixed point. In fact, if $\mu \in \mathcal{M}(X)$ is any Borel probability measure then the recipe from the proof of Theorem 24.6 tells us that any limit point of the sequence $\mu_k := \frac{1}{k}\sum_{i=0}^{k-1} f_*^i \mu$ will be an invariant measure for f. ∎

Remark 24.8. We remind you that merely proving that $\mathcal{M}(f)$ is non-empty may be of little practical use. Recall from Remark 19.19 that if $x \in \mathsf{fix}(f)$ then the Dirac measure δ_x belongs to $\mathcal{M}(f)$. Studying the dynamical system f on the probability space $(X, \mathscr{B}, \delta_x)$ is pointless—the only information we could gleam from its measure-theoretic properties is that x is a fixed point of f, which we already know.

(♣) Remark 24.9. The Markov-Kakutani Theorem is valid in more general situations than was stated in Theorem 24.6. Namely, instead of asking for E to be a Banach space, it is sufficient to ask that E is a locally convex topological vector space.

Moreover the assumption that g is affine in Theorem 24.6 can actually be dropped entirely, although this requires quoting a more advanced fixed point theorem, called the Schauder Fixed Point Theorem. This states that any (not necessarily affine) continuous self-map on a compact convex subset of a Hausdorff topological vector space has a fixed point.

Now that we know $\mathcal{M}(f)$ is non-empty, we can also prove it is compact.

Lemma 24.10. *Let f be a topological dynamical system on a compact metric space. Then the space $\mathcal{M}(f)$ is a non-empty closed compact convex subset of $\mathcal{M}(X)$.*

Proof. We already know that $\mathcal{M}(f)$ is non-empty, and convexity is obvious. Suppose $(\mu_k) \subset \mathcal{M}(f)$ with $\mu_k \rightharpoonup \mu \in \mathcal{M}(X)$. Then for any $u \in C(X)$, one has

$$\begin{aligned}
\int_X u \, d(f_*(\mu)) &= \int_X f^*(u) \, d\mu \\
&= \lim_{k\to\infty} \int_X f^*(u) \, d\mu_k \\
&= \lim_{k\to\infty} \int_X u \, d(f_*\mu_k) \\
&= \lim_{k\to\infty} \int_X u \, d\mu_k \\
&= \int_X u \, d\mu,
\end{aligned}$$

from which it follows from Corollary 24.4 that μ belongs to $\mathcal{M}(f)$. Thus $\mathcal{M}(f)$ is closed, and hence compact. This completes the proof. ∎

Definition 24.11. Let us write $\mathcal{E}(f) \subseteq \mathcal{M}(f)$ for the set of **ergodic measures** for f. Here by a slight abuse of language we say that a measure μ is an ergodic measure for f if f is an ergodic measure-preserving dynamical system on $(X, \mathscr{B}, \mu)$.

It turns out it is easy to characterise the elements of $\mathcal{E}(f)$.

Definition 24.12. Let K be a non-empty convex subset of a Banach space E. A point $x \in K$ is said to be an **extremal point** of K if writing $x = cy + (1-c)z$ for $0 < c < 1$ and $y, z \in K$ implies that $y = z$.

Remark 24.13. In finite dimensions, it is geometrically obvious that such extremal points exist. In infinite dimensions this is still true, but it is much harder to prove. The Krein-Milman Theorem from functional analysis tells us that if K is any non-empty compact convex subset of a Banach space[2] E then K is the convex hull of its extremal points, and thus in particular such extremal points exist. As with the other statements in this lecture, however, we will refrain from unnecessarily quoting difficult theorems from functional analysis, and give a direct proof for this fact that works for the case in hand.

Theorem 24.14. *Let f be a topological dynamical system on a compact metric space. Then*

(i) *The extremal points of $\mathcal{M}(f)$ are precisely the ergodic measures, i.e. the elements of $\mathcal{E}(f)$.*

(ii) *Such extremal points exist: $\mathcal{E}(f) \neq \emptyset$.*

(iii) *Given $\mu, \nu \in \mathcal{E}(f)$, either $\mu = \nu$ or $\mu \perp \nu$ (i.e. μ and ν are mutually singular).*

Proof. We begin by proving (i). Suppose $\mu \in \mathcal{M}(f)$ is not ergodic. This means there exists $A \in \mathscr{B}$ with $0 < \mu(A) < 1$ such that $f^{-1}A = A$. Given such an A, we can define two new measures μ_1 and μ_2 by

$$\mu_1(B) := \frac{\mu(B \cap A)}{\mu(A)}, \qquad \mu_2(B) := \frac{\mu(B \cap (X \setminus A))}{\mu(X \setminus A)}, \qquad \forall\, B \in \mathscr{A}.$$

By construction both μ_1 and μ_2 belong to $\mathcal{M}(f)$, and we have

$$\mu = \mu(A)\mu_1 + (1 - \mu(A))\mu_2.$$

Since obviously $\mu_1 \neq \mu_2$, this proves μ is not an extremal point of $\mathcal{M}(f)$.

Now suppose $\mu \in \mathcal{E}(f)$, and suppose we are given $0 < c < 1$ and $\mu_1, \mu_2 \in \mathcal{M}(f)$ with

$$\mu = c\mu_1 + (1-c)\mu_2.$$

We must show that $\mu = \mu_1 = \mu_2$. Clearly $\mu_1 \ll \mu$ (i.e. μ_1 is absolutely continuous with respect to μ), and hence the Radon-Nikodym derivative $\frac{d\mu_1}{d\mu}$ exists (cf. Definition 18.37). Recall $\frac{d\mu_1}{d\mu} \in L^1(\mu)$ is a non-negative function with $\int_X \frac{d\mu_1}{d\mu}\, d\mu = 1$ such that

$$\mu_1(A) = \int_A \frac{d\mu_1}{d\mu} d\mu, \qquad \forall\, A \in \mathscr{B}.$$

[2] Actually it is enough to ask that E is a locally convex topological vector space.

Since μ_1 is invariant, for any $A \in \mathscr{B}$ we have

$$\begin{aligned}\int_A \frac{d\mu_1}{d\mu}\, d\mu &= \mu_1(A) \\ &= \mu_1(f^{-1}A) \\ &= \int_{f^{-1}A} \frac{d\mu_1}{d\mu}\, d\mu \\ &= \int_A f^*\left(\frac{d\mu_1}{d\mu}\right) d\mu.\end{aligned}$$

Since A was arbitrary, this implies that $f^*\left(\frac{d\mu_1}{d\mu}\right) = \frac{d\mu_1}{d\mu}$. Similarly the Radon-Nikodym derivative $\frac{d\mu_2}{d\mu}$ exists and has integral 1. Suppose for contradiction that $\mu_1 \neq \mu_2$. Then $\frac{d\mu_1}{d\mu} \neq \frac{d\mu_2}{d\mu}$. If $\frac{d\mu_1}{d\mu}$ was constant μ-almost everywhere then the identity

$$c\frac{d\mu_1}{d\mu} + (1-c)\frac{d\mu_2}{d\mu} = \mathbb{1}_X,$$

shows that $\frac{d\mu_2}{d\mu}$ is also constant μ-almost everywhere. Since they both have total integral 1, this constant must be 1, and this contradicts $\frac{d\mu_1}{d\mu}$. This shows that $\frac{d\mu_1}{d\mu}$ is an f^*-invariant function which is not constant μ-almost everywhere. By Proposition 19.17 this implies that μ is not ergodic. This completes the proof of (i).

We next prove (iii), since this is a similar argument to the proof of (i). Given $\mu, \nu \in \mathcal{E}(f)$ we use the Lebesgue Decomposition Theorem 18.42 to uniquely write

$$\mu = c\mu_1 + (1-c)\mu_2, \qquad \text{where} \qquad \mu_1 \ll \nu, \quad \mu_2 \perp \nu.$$

But then also

$$\mu = f_*(\mu) = cf_*(\mu_1) + (1-c)f_*(\mu_2),$$

and moreover $f_*(\mu_1) \ll f_*(\nu)$ and $f_*(\mu_2) \perp f_*(\nu)$. The uniqueness part of the Lebesgue Decomposition Theorem thus implies that $f_*(\mu_1) = \mu_1$ and $f_*(\mu_2) = \mu_2$. Thus $\mu_1, \mu_2 \in \mathcal{M}(f)$ with $\mu = c\mu_1 + (1-c)\mu_2$. By part (i) we know that μ is necessarily an extremal point of $\mathcal{M}(f)$, which implies that either $c = 0$ or $c = 1$ (since clearly $\mu_1 \neq \mu_2$). If $c = 0$ we are done. If $c = 1$ then $\mu \ll \nu$, and the same argument as in (i) shows that $\frac{d\mu}{d\nu} = 1$ almost everywhere, whence $\mu = \nu$. This proves (iii).

Finally we prove (ii). Let $(w_k)_{k\in\mathbb{N}}$ denote a countable dense subset of $C^0(X;\mathbb{R}) \subset \mathcal{C}(X)$. Consider the map

$$\mathcal{M}(f) \to \mathbb{R}, \qquad \mu \mapsto \int_X w_1\, d\mu.$$

Since this map is continuous (by definition of the weak-star topology) and $\mathcal{M}(f)$ is compact, there exists at least one measure ν such that

$$\int_X w_1\, d\nu = \sup_{\mu\in\mathcal{M}(f)} \int_X w_1\, d\mu.$$

Thus the set

$$\mathcal{M}_1 := \left\{ \nu \in \mathcal{M}(f) \;\middle|\; \int_X w_1\, d\nu = \sup_{\mu\in\mathcal{M}(f)} \int_X w_1\, d\mu \right\}$$

is a non-empty set, which is also closed and compact. We now inductively define new non-empty compact subsets $\mathcal{M}_k$ for $k \geq 2$ by setting

$$\mathcal{M}_k := \left\{ \nu \in \mathcal{M}_{k-1} \,\middle|\, \int_X w_k \, d\nu = \sup_{\mu \in \mathcal{M}_{k-1}} \int_X w_k \, d\mu \right\}.$$

This defines a nested sequence $\mathcal{M}_1 \supseteq \mathcal{M}_2 \supseteq \mathcal{M}_3 \dots$ of subsets of $\mathcal{M}(f)$. By compactness (Proposition 2.18) the space

$$\mathcal{M}_\infty := \bigcap_{k=1}^{\infty} \mathcal{M}_k$$

is non-empty. Now suppose $\mu_\infty \in \mathcal{M}_\infty$. We claim that μ_∞ is an extremal point of $\mathcal{M}(f)$. Indeed, suppose $\mu_\infty = c\nu_1 + (1-c)\nu_2$ for $\nu_1, \nu_2 \in \mathcal{M}(f)$ and $0 < c < 1$. Since (w_k) is dense in $C^0(X;\mathbb{R})$, as in the proof of Proposition 23.14 or Theorem 23.17, it suffices to show that that

$$\int_X w_k \, d\nu_1 = \int_X w_k \, d\nu_2, \qquad \text{for all } k \geq 1. \tag{24.2}$$

For this first consider w_1. By assumption we have

$$\int_X w_1 \, d\mu_\infty = c \int_X w_1 \, d\nu_1 + (1-c) \int_X w_1 \, d\nu_2.$$

Thus in particular we have

$$\int_X w_1 \, d\mu_\infty \leq \max \left\{ \int_X w_1 \, d\nu_1, \int_X w_1 \, d\nu_2 \right\}$$

But since $\mu_\infty \in \mathcal{M}_\infty \subseteq \mathcal{M}_1$, we have

$$\int_X w_1 d\, \mu_\infty = \sup_{\mu \in \mathcal{M}(f)} \int_X w_1 \, d\mu.$$

This implies that

$$\int_X w_1 \, d\nu_1 = \int_X w_1 \, d\nu_2,$$

and moreover that ν_1, ν_2 both belong to $\mathcal{M}_1 \subseteq \mathcal{M}(f)$. Now we repeat this with w_2: we have

$$\int_X w_2 \, d\mu_\infty = c \int_X w_2 \, d\nu_1 + (1-c) \int_X w_2 \, d\nu_2,$$

and thus in particular we have

$$\int_X w_2 \, d\mu_\infty \leq \max \left\{ \int_X w_2 \, d\nu_1, \int_X w_2 \, d\nu_2 \right\}.$$

But since $\mu_\infty \in \mathcal{M}_\infty \subseteq \mathcal{M}_2$, we have

$$\int_X w_2 \, d\mu_\infty = \sup_{\mu \in \mathcal{M}_1} \int_X w_2 \, d\mu.$$

This implies that

$$\int_X w_2 \, d\nu_1 = \int_X w_2 \, d\nu_2,$$

and moreover that ν_1, ν_2 both belong to $\mathcal{M}_2 \subseteq \mathcal{M}_1$. Arguing inductively, the same argument shows for each $k \geq 2$, $\int_X w_k \, d\nu_1 = \int_X w_k \, d\nu_2$ and $\nu_1, \nu_2 \in \mathcal{M}_k$. This establishes (24.2), and thus completes the proof. ■

Recall that we can think of $\mathcal{M}(X)$ as an enlargement of X via the topological embedding $\imath \colon X \to \mathcal{M}(X)$ from (23.1). Moreover as we have already remarked,

$$\imath(\mathsf{fix}(f)) \subset \mathcal{M}(f).$$

But what about $\mathsf{per}(f)$? If $x \in \mathsf{per}(f)$ is a periodic point that is not a fixed point then $\delta_x = \imath(x)$ will not belong to $\mathcal{M}(f)$. However it is easy to see how to rectify this.

Definition 24.15. Let $x \in X$ and $p \geq 1$. Define a measure

$$\wp_{x,p} := \frac{1}{p} \sum_{k=0}^{p-1} \delta_{f^k(x)}.$$

If $f^p(x) = x$ then we call $\wp_{x,p}$ the **periodic orbit measure** associated to the periodic point x and the period p.

Lemma 24.16. *Let f be a topological dynamical system on a compact metric space. Let $p \in \mathbb{N}$ and $x \in X$. Then*

$$f^p(x) = x \qquad \Leftrightarrow \qquad \wp_{x,p} \in \mathcal{M}(f).$$

Proof. It is obvious that if $f^p(x) = x$ then $\wp_{x,p}$ is invariant. For the converse we apply Corollary 24.4 with $\mu = \wp_{x,p}$. This tells us that if $\wp_{x,p} \in \mathcal{M}(f)$ then

$$\frac{1}{p} \sum_{k=0}^{p-1} u(f^{k+1}(x)) = \frac{1}{p} \sum_{k=0}^{p-1} u(f^k(x)), \qquad \forall \, u \in \mathcal{C}(X),$$

which is equivalent to saying that

$$u(f^p(x)) = u(x), \qquad \forall \, u \in \mathcal{C}(X).$$

If $f^p(x) \neq x$ then there exists[3] a continuous function u such that $u(f^p(x)) \neq u(x)$. This completes the proof. ■

On Problem Sheet L you will upgrade Lemma 24.16 in two ways. Firstly, you will prove:

Lemma 24.17. *Let f be a topological dynamical system on a compact metric space. Let $p \in \mathbb{N}$ and $x \in X$. Suppose $f^p(x) = x$. Then the invariant measure $\wp_{x,p}$ is ergodic.*

[3]See (23.2). This result is known as Urysohn's Lemma.

A purely atomic measure $\mu \in \mathcal{M}(X)$ is necessarily of the form

$$\mu = \sum_{k=1}^{\infty} c_k \, \delta_{x_k}$$

where $x_k \in X$ and $c_k > 0$ are real numbers satisfying $\sum_{k=1}^{\infty} c_k = 1$. The next result is again on Problem Sheet L.

Proposition 24.18. *Let f be a topological dynamical system on a compact metric space.*

(i) *Suppose $\mu \in \mathcal{M}(f)$ is a purely atomic measure. Then μ is a (possibly countably infinite) convex combination of periodic orbit measures.*

(ii) *Suppose $\mu \in \mathcal{E}(f)$ is a purely atomic measure. Then μ is a periodic orbit measure.*

LECTURE 25

Unique Ergodicity

We have seen that a topological dynamical system on a compact metric space always admits at least one invariant measure. In this lecture we investigate the case where there is *exactly* one invariant measure.

DEFINITION 25.1. A topological dynamical system $f\colon X \to X$ on a compact metric space is called **uniquely ergodic** if there is precisely one invariant measure.

The reason for the name is explained by the following simple lemma.

LEMMA 25.2. *Let $f\colon X \to X$ be a topological dynamical system on a compact metric space. Then f is uniquely ergodic if and only if $\mathcal{E}(f)$ consists of exactly one point.*

Proof. If $\mathcal{M}(f)$ consists of exactly one point then this point is necessarily an extremal point, and thus by Theorem 24.14 it is ergodic. Conversely if $\#\mathcal{M}(f) \geq 2$ then by convexity it has at least two extremal points[1], and thus also $\#\mathcal{E}(f) \geq 2$. ∎

Irrational circle rotations are uniquely ergodic.

LEMMA 25.3. *A rotation $\rho_\theta\colon S^1 \to S^1$ is uniquely ergodic if and only if θ is irrational. When θ is irrational, the unique ergodic measure is the Lebesgue measure.*

Proof. If θ is rational, say $\theta = \frac{p}{q}$ then every point is periodic, and by Lemma 24.17 these give rise to many ergodic measures for ρ_θ. Now suppose θ is irrational. Then by Problem J.2 the Lebesgue measure λ is ergodic. It remains therefore to show that this is the unique ergodic measure.

Suppose $\mu \in \mathcal{M}(\rho_\theta)$ is an arbitrary invariant measure. Let ρ_β denote another circle rotation. We will show that μ also belongs to $\mathcal{M}(\rho_\beta)$. To see this, first fix a point $z_0 \in S^1$. By Lemma 1.10 there exists a sequence $k_n \to \infty$ such that

$$\lim_{n\to\infty} \rho_\theta^{k_n}(z_0) = \rho_\beta(z_0). \tag{25.1}$$

Since both sides are rotations, (25.1) actually holds for all $z \in S^1$. Now fix $u \in \mathcal{C}(S^1)$. Then by the Dominated Convergence Theorem 18.34 and Proposition 24.3 we have

$$\begin{aligned} \int_{S^1} \rho_\beta^*(u)\, d\mu &= \lim_{n\to\infty} \int_{S^1} \left(\rho_\theta^{k_n}\right)^*(u)\, d\mu \\ &= \lim_{n\to\infty} \int_{S^1} u\, d\big((\rho_\theta^{k_n})_*\mu\big) \\ &= \int_{S^1} u\, d\mu, \end{aligned}$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]This follows from the Krein-Milman Theorem, which tells us that $\mathcal{M}(f)$ is the convex hull of its extremal points. Alternatively one can argue directly as in the proof of part (ii) of Theorem 24.14.

since $\mu \in \mathcal{M}(\rho_\theta)$. But now since Corollary 24.4 was an "if and only if" statement and u was arbitrary, it follows that $\mu \in \mathcal{M}(\rho_\beta)$.

Thus we have shown that if θ is irrational and $\mu \in \mathcal{M}(\rho_\theta)$ then μ is an invariant measure for every circle rotation. Put differently, we have shown that μ is a *translation invariant* measure on S^1. But now we are done, for the Lebesgue measure is the *unique* translation invariant measure on S^1 (cf. Remark 18.19). This completes the proof. ■

REMARK 25.4. On Problem Sheet M you will improve Lemma 25.3 to show that any reversible topological dynamical system on S^1 with irrational rotation number is uniquely ergodic.

The goal for the rest of this lecture is to improve the Birkhoff Ergodic Theorem 20.2 and Corollary 20.3. We first prove a minor improvement that is valid for any topological dynamical system f on a compact metric space and any $\mu \in \mathcal{E}(f)$. We then present a substantial improvement that is valid under the additional assumption of unique ergodicity.

This will require several preliminary results. Firstly, we now modify the statement of Propositions 22.1, 22.2 and 22.3, and Problem K.3 to take advantage of the fact that we are not on a topological space.

NOTATION. In this lecture we will *not* use the $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ notation, and will instead write the integral out in full. This is for two reasons:

- We prefer to reserve the notation $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ for the inner product that makes $L^2(\mu; \mathbb{C})$ into a complex Hilbert space. In this lecture we will usually not be working with $L^2(\mu; \mathbb{C})$.
- The complex conjugate in the second factor is unnecessary for our purposes today.

With this in mind, recall for instance that in Proposition 22.2 we proved that f is ergodic if and only if for all $u, v \in L^2(\mu)$, one has

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u)\, v\, d\mu = \int_X u\, d\mu \int_X v\, d\mu.$$

Here is convenient to change where u and v lie:

PROPOSITION 25.5. *Let $f \colon X \to X$ denote a topological dynamical system on a compact metric space, and let $\mu \in \mathcal{M}(f)$. Then:*

(i) *f is ergodic if and only if for all $u \in C(X)$ and $v \in L^1(\mu)$, one has*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u)\, v\, d\mu = \int_X u\, d\mu \int_X v\, d\mu. \tag{25.2}$$

(ii) *f is weakly mixing with respect*[2] *to μ if and only if there exists a set $K \subset \{0,1,2,\dots\}$ of density zero such that for all $u \in \mathcal{C}(X)$ and $v \in L^1(\mu)$ one has*

$$\lim_{k \notin K} \int_X (f^*)^k(u)\, v\, d\mu = \int_X u\, d\mu \int_X v\, d\mu.$$

(iii) *f is mixing with respect to μ if and only if for all $u \in \mathcal{C}(X)$ and $v \in L^1(\mu)$ one has*

$$\lim_{k \to \infty} \int_X (f^*)^k(u)\, v\, d\mu = \int_X u\, d\mu \int_X v\, d\mu.$$

For all three cases, the idea is to insist on additional regularity for u (continuous instead of merely L^2) and at the same time require less regularity from v (L^1 instead of L^2).

Proof. Let us give a complete proof for (i). Suppose the stated convergence condition holds, and let $\tilde{u}, \tilde{v} \in L^2(\mu)$. Then in particular $\tilde{v} \in L^1(\mu)$, and hence for all $u \in \mathcal{C}(X)$, one has

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u)\, \tilde{v}\, d\mu = \int_X u\, d\mu \int_X \tilde{v}\, d\mu. \tag{25.3}$$

Now we use the fact that any L^2 function can be approximated by continuous functions. Fix $\varepsilon > 0$ and choose $u_0 \in \mathcal{C}(X)$ such that

$$\|u_0 - \tilde{u}\|_2 < \varepsilon. \tag{25.4}$$

By (25.3) there exists $n \in \mathbb{N}$ such that for all $k \geq n$, one has

$$\left| \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u_0)\, \tilde{v}\, d\mu - \int_X u_0\, d\mu \int_X \tilde{v}\, d\mu \right| < \varepsilon.$$

Then for $k \geq n$ we estimate:

$$\begin{aligned}
&\left| \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(\tilde{u})\, \tilde{v}\, d\mu - \int_X \tilde{u}\, d\mu \int_X \tilde{v}\, d\mu \right| \\
&\qquad\qquad \leq \int_X \frac{1}{k} \sum_{i=0}^{k-1} \left|(f^*)^i(\tilde{u} - u_0)\right| \left|\tilde{v}\right| d\mu \\
&\qquad\qquad\quad + \left| \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u_0)\, \tilde{v}\, d\mu - \int_X u_0\, d\mu \int_X \tilde{v}\, d\mu \right| \\
&\qquad\qquad\quad + \left| \int_X u_0\, d\mu \int_X \tilde{v}\, d\mu - \int_X \tilde{u}\, d\mu \int_X \tilde{v}\, d\mu \right| \\
&\qquad\qquad \leq 2\varepsilon \|\tilde{v}\|_2 + \varepsilon,
\end{aligned}$$

[2] The phrase "weakly mixing with respect to μ" should naturally be interpreted as weakly mixing in the measure-theoretic sense (and similarly for mixing). If we wish to refer to the topological definitions we will explicitly call them "topologically weakly mixing" and "topologically mixing".

where the last line used (25.4) and Proposition 19.16. Since ε was arbitrary it follows that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(\tilde{u})\, \tilde{v}\, d\mu = \int_X \tilde{u}\, d\mu \int_X \tilde{v}\, d\mu.$$

Then since $\tilde{u}$ and $\tilde{v}$ were arbitrary, Proposition 22.2 implies that f is ergodic.

Conversely, suppose that f is ergodic. Let $u \in \mathcal{C}(X)$. Then automatically $u \in L^2(\mu)$, and hence by Proposition 22.2 for any $\tilde{v} \in L^2(\mu)$ one has

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u)\, \tilde{v}\, d\mu = \int_X u\, d\mu \int_X \tilde{v}\, d\mu.$$

If $v \in L^1(\mu)$ then we can choose $(\tilde{v}_n) \in L^2(\mu)$ such that $\tilde{v}_n \to v$ in $L^1(\mu)$. Then arguing as above we see that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \int_X (f^*)^i(u)\, v\, d\mu = \int_X u\, d\mu \int_X v\, d\mu.$$

This completes the proof of (i). The proofs of (ii) and (iii) are similar, and use Problem K.3 and Proposition 22.1 instead of Proposition 22.2. Note that in (ii) the hypotheses of Problem K.3 are applicable as every compact metric space is separable (cf. Example 18.44). ■

We can now prove the following new characterisation:

Proposition 25.6. *Let $f\colon X \to X$ be a topological dynamical system on a compact metric space X, and let $\mu \in \mathcal{M}(f)$. Then:*

(i) *f is ergodic with respect to μ if and only if for all $\nu \in \mathcal{M}(X)$ such that $\nu \ll \mu$ one has*

$$\frac{1}{k} \sum_{i=0}^{k-1} f_*^i \nu \rightharpoonup \mu.$$

(ii) *f is weakly mixing with respect to μ if and only if there exists a set $K \subset \{0, 1, 2, \dots\}$ of density zero such that for all $\nu \in \mathcal{M}(X)$ such that $\nu \ll \mu$ one has*

$$f_*^k \nu \xrightharpoonup[k \notin K]{} \mu.$$

(iii) *f is mixing with respect to μ if and only if for all $\nu \in \mathcal{M}(X)$ such that $\nu \ll \mu$ one has*

$$f_*^k \nu \rightharpoonup \mu.$$

Proof. Again, we will prove (i) only. Suppose $\mu \in \mathcal{E}(f)$ and suppose $\nu \in \mathcal{M}(X)$

satisfies $\nu \ll \mu$. If $u \in \mathcal{C}(X)$ then we compute

$$\begin{aligned}
\int_X u \, d\left(\frac{1}{k}\sum_{i=0}^{k-1} f_*^i \nu\right) &\overset{(\heartsuit)}{=} \frac{1}{k}\sum_{i=0}^{k-1}\int_X (f^*)^i(u)\, d\nu \\
&\overset{(\diamondsuit)}{=} \frac{1}{k}\sum_{i=0}^{k-1}\int_X (f^*)^i(u)\, \frac{d\nu}{d\mu}\, d\mu \\
&\overset{(25.2)}{\to} \int_X u\, d\mu \underbrace{\int_X \frac{d\nu}{d\mu}\, d\mu}_{=1} \\
&= \int_X u\, d\mu,
\end{aligned}$$

where $(\heartsuit)$ used Proposition 24.3 and $(\diamondsuit)$ used the Radon-Nikodym Theorem 18.36. Note that it is crucial here that we first established Proposition 25.5, since the Radon-Nikodym derivative $\frac{d\nu}{d\mu}$ typically only belongs to $L^1(\mu)$ and not $L^2(\mu)$. Since u was arbitrary, it follows from the definition of the weak-star topology that $\frac{1}{k}\sum_{i=0}^{k-1} f_*^i \nu \rightharpoonup \mu$.

Conversely, suppose the stated condition holds. Let $u \in \mathcal{C}(X)$ and $v \in L^1(\mu)$. We will show that (25.2) holds. First assume that $v \geq 0$. Abbreviate

$$a := \int_X v \, d\mu. \tag{25.5}$$

If $a = 0$ then $v = 0$ almost everywhere, and (25.2) readily follows. If $a > 0$, define a measure $\nu \in \mathcal{M}(X)$ by setting

$$\nu(A) := \frac{1}{a}\int_A v\, d\mu.$$

Then $\nu \ll \mu$ and

$$\frac{d\nu}{d\mu} = \frac{1}{a} v. \tag{25.6}$$

By assumption

$$\frac{1}{k}\sum_{i=0}^{k-1} f_*^i \nu \rightharpoonup \mu. \tag{25.7}$$

We now argue as above, only backwards:

$$\begin{aligned}
\frac{1}{k}\sum_{i=0}^{k-1}\int_X (f^*)^i(u)\, v\, d\mu &\overset{(25.6)}{=} \frac{a}{k}\sum_{i=0}^{k-1}\int_X (f^*)^i(u)\frac{d\nu}{d\mu} d\mu \\
&\overset{(\diamondsuit)}{=} \frac{a}{k}\sum_{i=0}^{k-1}\int_X (f^*)^i(u) d\nu \\
&\overset{(\heartsuit)}{=} a\int_X u\, d\left(\frac{1}{k}\sum_{i=0}^{k-1} f_*^i\nu\right) \\
&\overset{(25.7)}{\to} a\int_X u\, d\mu \\
&\overset{(25.5)}{=} \int_X v\, d\mu \int_X u\, d\mu.
\end{aligned}$$

This proves (25.2) in the case where $v \geq 0$. If v is real-valued, then we can apply the above reasoning to the positive and negative parts, and finally if v is complex valued, we can apply the above to the real and imaginary parts. It thus follows from part (i) of Proposition 25.5 that $\mu \in \mathcal{E}(f)$.

Finally, the proofs of (ii) and (iii) are similar, and use parts (ii) and (iii) of Proposition 25.5 instead. The details are left to the interested reader. ■

We now improve prove a minor enhancement of the Birkhoff Ergodic Theorem (Corollary 20.3) to the setting of topological dynamical systems on compact metric spaces. Let $\mu \in \mathcal{E}(f)$ and suppose $u \in \mathcal{C}(X)$. Recall we denote by

$$\widehat{u}(x) := \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x))$$

the time average of u, which exists (and is constant) μ-almost everywhere.

Proposition 25.7. *Let $f\colon X \to X$ be a topological dynamical system on a compact metric space. If $\mu \in \mathcal{E}(f)$ then there exists a measurable set Y with $\mu(Y) = 1$ such that*

$$\widehat{u}(x) = \int_X u \, d\mu, \qquad \text{for all } x \in Y \text{ and for all } u \in \mathcal{C}(X).$$

The reason the Proposition 25.7 is stronger than Corollary 20.3 is that the set Y does *not* depend on u.

Proof. Choose a countable dense subset (w_k) of $\mathcal{C}(X)$. By Corollary 20.3, for each k there exists a measurable set $Y_k \subset X$ with $\mu(Y_k) = 1$ such that

$$\widehat{w}_k(x) = \int_X w_k \, d\mu, \qquad \text{for all } x \in Y_k.$$

Now put

$$Y = \bigcap_{k=1}^{\infty} Y_k.$$

Then $\mu(Y) = 1$ and

$$\widehat{w}_k(x) = \int_X w_k \, d\mu, \qquad \text{for all } x \in Y \text{ and for each } k \in \mathbb{N}.$$

The result now follows by approximating an arbitrary $u \in \mathcal{C}(X)$ by members of the (w_k). This completes the proof. ■

This allows us to prove the following result.

Proposition 25.8. *Let $f\colon X \to X$ be a topological dynamical system on a compact metric space and $\mu \in \mathcal{M}(f)$. Then $\mu \in \mathcal{E}(f)$ if and only if*

$$\frac{1}{k} \sum_{i=0}^{k-1} \delta_{f^i(x)} \rightharpoonup \mu, \qquad \text{for } \mu\text{-almost every } x \in X.$$

Proof. If $\mu \in \mathcal{E}(f)$ then Proposition 25.7 tells us that $\frac{1}{k}\sum_{i=0}^{k-1}\delta_{f^i(x)} \rightharpoonup \mu$ for all $x \in Y$. Conversely, suppose there exists a measurable set $Y \subset X$ with $\mu(Y) = 1$ such that $\frac{1}{k}\sum_{i=0}^{k-1}\delta_{f^i(x)} \rightharpoonup \mu$ for all $x \in Y$. Then for all $x \in Y$ and for all $u \in \mathcal{C}(X)$, one has

$$\widehat{u}(x) = \int_X u\, d\mu.$$

Thus if $x \in Y$, $u \in \mathcal{C}(X)$ and $v \in L^1(\mu)$ then

$$\frac{1}{k}\sum_{i=0}^{k-1} u(f^i(x))v(x) \to v(x)\int_X u\, d\mu.$$

Thus by the Dominated Convergence Theorem 18.34, the conclusion of part (i) of Proposition 25.5 is satisfied, and hence μ is ergodic. This completes the proof. ∎

As promised, we now further improve the Birkhoff Ergodic Theorem in the uniquely ergodic case.

THEOREM 25.9. *Let f be a topological dynamical system on a compact metric space. Then the following are equivalent:*

(i) *f is uniquely ergodic.*

(ii) *For every $u \in \mathcal{C}(X)$, $\frac{1}{k}\sum_{i=0}^{k-1} u(f^i(x))$ converges uniformly to a constant.*

(iii) *For every $u \in \mathcal{C}(X)$, $\frac{1}{k}\sum_{i=0}^{k-1} u(f^i(x))$ converges pointwise to a constant.*

(iv) *There exists $\mu \in \mathcal{M}(f)$ such that for all $u \in \mathcal{C}(X)$ and for all $x \in X$ one has*

$$\frac{1}{k}\sum_{i=0}^{k-1} u(f^i(x)) \to \int_X u\, d\mu.$$

Note that important part in (iv) is that the conclusion holds for *all* $x \in X$ (and not just almost all). In other words, time average = space average *everywhere*.

Proof. Clearly (ii) implies (iii). Let us first prove that (iii) implies (iv). For this we define a map $T\colon \mathcal{C}(X) \to \mathbb{C}$ by

$$Tu = \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1} u(f^i(x)).$$

Then T is a bounded (and hence continuous) linear operator which satisfies $T\mathbb{1}_X = 1$ and $u \geq 0$ implies $Tu \geq 0$. Thus by the Riesz Representation Theorem 23.9, there exist $\mu \in \mathcal{M}(X)$ such that

$$Tu = \int_X u\, d\mu.$$

Since clearly $T(f^*(u)) = Tu$, the measure μ belongs to $\mathcal{M}(f)$ by Proposition 24.3 and Corollary 24.4. This proves (iv).

Now let us prove that (iv) implies (i). Suppose that $\nu \in \mathcal{M}(f)$. Let $u \in \mathcal{C}(X)$ and set $a := \int_X u \, d\mu$. By assumption $\frac{1}{k} \sum_{i=0}^{k-1} u(f^i(x)) \to a$ for all $x \in X$, and thus in particular it also converges for ν-almost all $x \in X$. Therefore by Birkhoff Ergodic Theorem 20.2 applied to the invariant measure ν, we have

$$\int_X u \, d\nu = a = \int_X u \, d\mu.$$

Since u was arbitrary we then have $\mu = \nu$ by Proposition 23.7. Thus f is uniquely ergodic.

Finally let us prove that (i) implies (ii). First note that for a given $v \in \mathcal{C}(X)$, *if* $\frac{1}{k} \sum_{i=0}^{k-1} v(f^i(x))$ converges uniformly to a constant, then this constant must be $\int_X v \, d\mu$. Thus if (ii) does not hold then there exists $v \in \mathcal{C}(X)$ and $\epsilon > 0$ such that for all $n \geq 1$ there exists $k \geq n$ and $x_k \in X$ such that

$$\left| \frac{1}{k} \sum_{i=0}^{k-1} v(f^i(x_k)) - \int_X v \, d\mu \right| > \epsilon.$$

Now set

$$\mu_k := \frac{1}{k} \sum_{i=0}^{k-1} \delta_{f^i(x_k)} = \frac{1}{k} \sum_{i=0}^{k-1} f_*^i \delta_{x_k}.$$

Then for each k one has

$$\left| \int_X v \, d\mu_k - \int_X v \, d\mu \right| > \epsilon.$$

Now observe from the recipe from the proof[3] of the Markov-Kakutani Fixed Point Theorem 24.6 that the sequence (μ_k) has a convergent subsequence, which moreover converges to an element μ_∞ of $\mathcal{M}(f)$. But then

$$\left| \int_X v \, d\mu_\infty - \int_X v \, d\mu \right| > \epsilon.$$

Thus $\mu_\infty \neq \mu$, which contradicts f being uniquely ergodic. This completes the proof. ■

[3]This is where it is important that the recipe from the proof of Theorem 24.6 allowed us to start with an arbitrary *sequence* of points. In this case the sequence is δ_{x_k}.

LECTURE 26

Partitions and the Rokhlin Metric

In this lecture we introduce a "measure-theoretic" version of entropy. In this lecture we go back to working on an arbitrary probability space $(X, \mathscr{A}, \mu)$. This entire discussion should be contrasted with the similar notions introduced in Lecture 10 for open covers.

DEFINITION 26.1. Let $\mathscr{P} = \mathscr{P}(X, \mathscr{A}, \mu)$ denote the space of all (equivalence classes of) finite measurable **partitions** (which we will henceforth simply call "partition") of $(X, \mathscr{A}, \mu)$. Thus elements of $\mathscr{P}$ are (equivalence classes) of finite tuples

$$\xi = \{C_1, \dots, C_p\}$$

where each set $C_k \in \mathscr{A}$ is a measurable set, and

$$\mu(C_i \cap C_j) = 0 \qquad \text{if } i \neq j, \qquad \text{and} \qquad \mu\left(X \setminus \bigcup_{k=1}^{p} C_k\right) = 0.$$

The equivalence relation in question is the following: two partitions ξ and η are equivalent if for all $C \in \xi$ with $\mu(C) > 0$ there exists $D \in \eta$ such that $\mu(C \vartriangle D) = 0$, and conversely. In other words, ξ and η are equivalent if there exists a set A of measure zero such that the restrictions of ξ and η to $X \setminus A$ coincide.

Like all good mathematicians however, for the remainder of this section we will suppress the equivalence classes and just think of elements of $\mathscr{P}$ as genuine partitions (a bit like we do with the L^p spaces of integrable functions). If f is a dynamical system[1] on $(X, \mathscr{A}, \mu)$ then given $\xi \in \mathscr{P}$ there is a well-defined pull-back partition given by

$$f^{-1}\xi = \{f^{-1}C \mid C \in \xi\}.$$

DEFINITION 26.2. Given two partitions ξ and η, we say that η is a **refinement** of ξ, or equivalently, that ξ is **subordinate** to η, and write $\xi \preceq \eta$ if for all $D \in \eta$ there exists $C \in \xi$ such that $D \subseteq C$.

LEMMA 26.3. *The operation $\preceq$ defines a partial order on $\mathscr{P}$.*

Proof. If $\xi \preceq \eta$ and $\eta \preceq \xi$ then $\xi = \eta$ (recall we are really working with equivalence classes). The other properties are clear. ∎

DEFINITION 26.4. Given two partitions ξ and η, we define their **join**, written $\xi \vee \eta$, to be the partition

$$\xi \vee \eta := \{C \cap D \mid C \in \xi,\ D \in \eta\}.$$

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]This lecture we are exclusively working in the measure-theoretic world, so there is no danger of ambiguitiy in just writing "dynamical system" to mean "measure-preserving dynamical system".

Note that $\xi \preceq \xi \vee \eta$ for any η. The next lemma is immediate from the definitions (compare to Lemma 10.12).

Lemma 26.5. *Let f be a dynamical system on $(X, \mathscr{A}, \mu)$, and suppose $\xi, \eta \in \mathscr{P}$. Then*

$$f^{-1}(\xi \vee \eta) = f^{-1}(\xi) \vee f^{-1}(\eta), \tag{26.1}$$

and

$$\xi \preceq \eta \qquad \Rightarrow \qquad f^{-1}(\xi) \preceq f^{-1}(\eta). \tag{26.2}$$

Definition 26.6. We say that two partitions ξ and η are **independent** if $\mu(C \cap D) = \mu(C) \cdot \mu(D)$ for all $C \in \xi$ and $D \in \eta$.

By convention we will say that $0 \log 0 = 0$ in the following. We now define the *entropy* of a partition.

Definition 26.7. Let $\xi = \{C_1, \dots, C_p\}$ denote a partition. The **entropy** of ξ, written $\mathsf{H}(\xi)$ is the non-negative real number

$$\mathsf{H}(\xi) := -\sum_{i=1}^{p} \mu(C_i) \log \mu(C_i).$$

The next lemma is immediate from the definition.

Lemma 26.8. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Then*

$$\mathsf{H}(f^{-1}\xi) = \mathsf{H}(\xi), \qquad \forall \xi \in \mathscr{P}.$$

Definition 26.9. Given two partitions $\xi = \{C_1, \dots, C_p\}$, and $\eta = \{D_1, \dots, D_q\}$, we define the **conditional entropy** of ξ with respect to η, written $\mathsf{H}(\xi|\eta)$, by

$$\mathsf{H}(\xi|\eta) := -\sum_{j=1}^{q} \mu(D_j) \sum_{i=1}^{p} \frac{\mu(C_i \cap D_j)}{\mu(D_j)} \log \frac{\mu(C_i \cap D_j)}{\mu(D_j)}.$$

Here we assumed that $\mu(D_j) > 0$ for each j, which is harmless since we are working with equivalence classes.

Motivating the definition of entropy: We will now attempt to motivate these definitions. Suppose we are performing an "experiment" on our system $(X, \mathscr{A})$. The possible outcomes of this experiment are determined by a partition $\xi = \{C_1, \dots, C_p\}$, and the probability of C_i happening is given by $\mu(C_i)$. Let us pretend for a while that we haven't yet come up with Definitions 26.7 and 26.9 and attempt to "guess" how to define the entropy.

Indeed, the entropy $\mathsf{H}(\xi)$ is intended to measure the amount of "uncertainty" we have when we made the experiment. It is reasonable to assume that $\mathsf{H}(\xi)$ should depend only on the numbers $\{\mu(C_1), \dots, \mu(C_p)\}$. So let us write

$$\mathsf{H}(\xi) \stackrel{?}{=} \Phi\big(\mu(C_1), \dots, \mu(C_p)\big),$$

where Φ is a function to be discovered.

Suppose moreover that we have two experiments $\xi = \{C_1, \dots, C_p\}$ and $\eta = \{D_1, \dots, D_q\}$. The conditional entropy $\mathsf{H}(\xi|\eta)$ is intended to measure the uncertainty about the outcome of ξ under the assumption that we *already* know what happened when we did η.

If we know that D_j happened then the probability of C_i happening is given by $\frac{\mu(C_i \cap D_j)}{\mu(D_j)}$, and thus the uncertainty about the outcome of ξ given that D_j happened is given by

$$\Phi\left(\frac{\mu(C_1 \cap D_j)}{\mu(D_j)}, \frac{\mu(C_2 \cap D_j)}{\mu(D_j)}, \dots, \frac{\mu(C_p \cap D_j)}{\mu(D_j)}\right).$$

This means that uncertainty about ξ, given that we know the outcome of η, should be given by the number

$$\mathsf{H}(\xi|\eta) \overset{\text{must be}}{=} \sum_{j=1}^{q} \mu(D_j)\Phi\left(\frac{\mu(C_1 \cap D_j)}{\mu(D_j)}, \frac{\mu(C_2 \cap D_j)}{\mu(D_j)}, \dots, \frac{\mu(C_p \cap D_j)}{\mu(D_j)}\right). \tag{26.3}$$

It turns out that if we make a couple more "reasonable" assumptions then we essentially have no choice about the definition of Φ.

Theorem 26.10 (Khinchin's Theorem). *Let $\Delta_p \subset \mathbb{R}^p$ denote the simplex*

$$\Delta_p := \left\{ (x_1, \dots, x_p) \,\middle|\, x_i \geq 0,\ \sum_{i=1}^{p} x_i = 1 \right\} \subset \mathbb{R}^p.$$

Let $\Delta := \bigcup_{p=1}^{\infty} \Delta_p$, and suppose $\Phi \colon \Delta \to \mathbb{R}$ is a function such that

(i) $\Phi(x_1, \dots x_p) \geq 0$, *and* $\Phi(x_1, \dots, x_p) = 0$ *if and only if some* $x_i = 1$.

(ii) $\Phi(x_1, \dots, x_p, 0) = \Phi(x_1, \dots, x_p)$.

(iii) *For each $p \geq 1$, $\Phi|_{\Delta_p}$ is continuous and symmetric.*

(iv) *For each $p \geq 1$, $\Phi|_{\Delta_p}$ has its largest value at $(1/p, \dots, 1/p)$.*

(v) *If H is defined from Φ via (26.3) then $\mathsf{H}(\xi \vee \eta) = \mathsf{H}(\xi) + \mathsf{H}(\eta|\xi)$.*

Then there exists $c > 0$ such that $\Phi(x_1, \dots, x_p) = -c\sum_{i=1}^{p} x_i \log x_i$.

These are indeed "reasonable" assumptions:

- Property (i) says there should be no uncertainty when there is only one possible outcome.
- Property (ii) tells us that if we can experiment with p possible outcomes then by adding an impossible (i.e. probability 0) outcome we can also regard it as an experiment with $p+1$ outcomes.
- Property (iii) says that the uncertainty of an experiment shouldn't depend on the ordering of the outcomes we chose (and continuity is a basic mathematical requirement!)
- Property (iv) says that the uncertainty is maximised when all the outcomes are equally probable.

- Finally property (v) says that the uncertainty from performing two experiments ξ and η is the same as the uncertainty from performing ξ plus the uncertainty from performing η already knowing that ξ had been performed.

The proof of Khinchin's Theorem 26.10 is "easy" in the sense that it involves no advanced concepts. But it is a little fiddly, and would take us the entire lecture to prove. Since Khinchin's Theorem is only used to motivate the definition of entropy, and is not needed in any of the material that follows, we will skip the proof. We refer the interested reader pages 9–13 of Khinchin's book Mathematical Foundations of Information Theory.

Let us now go back to the general theory. The following result contains some useful properties of entropy. Many of these properties are immediate from Khinchin's Theorem, yet since we did not prove the theorem we will prove them directly[2]

Proposition 26.11. *Let $(X, \mathscr{A}, \mu)$ denote a probability space. Fix three partitions*

$$\xi = \{C_1, \dots, C_p\}, \qquad \eta = \{D_1, \dots, D_q\}, \qquad \zeta = \{E_1, \dots, E_r\}.$$

Then:

(i) *It holds that*

$$0 \le -\log\left(\max_{1 \le i \le p} \mu(C_i)\right) \le \mathsf{H}(\xi) \le \log p.$$

The first equality is strict if ξ is not the trivial partition, and the second equality is strict unless $\mu(C_i) = \frac{1}{p}$ for each i.

(ii) $0 \le \mathsf{H}(\xi|\eta) \le \mathsf{H}(\xi)$.

(iii) $\mathsf{H}(\xi|\eta) = 0$ *if and only if* $\xi \preceq \eta$.

(iv) $\mathsf{H}(\xi|\eta) = \mathsf{H}(\xi)$ *if and only if ξ and η are independent.*

(v) *If $\eta \preceq \zeta$ then* $\mathsf{H}(\xi|\zeta) \le \mathsf{H}(\xi|\eta)$.

(vi) $\mathsf{H}(\xi \vee \eta|\zeta) = \mathsf{H}(\xi|\zeta) + \mathsf{H}(\eta|\xi \vee \zeta)$. *In particular, taking ζ to be the trivial partition $\zeta = \{X\}$, we have* $\mathsf{H}(\xi \vee \eta) = \mathsf{H}(\xi) + \mathsf{H}(\eta|\xi)$.

(vii) $\mathsf{H}(\xi \vee \eta|\zeta) \le \mathsf{H}(\xi|\zeta) + \mathsf{H}(\eta|\zeta)$, *and hence taking ζ to be the trivial partition we have* $\mathsf{H}(\xi \vee \eta) \le \mathsf{H}(\xi) + \mathsf{H}(\eta)$.

(viii) $\mathsf{H}(\xi|\zeta) \le \mathsf{H}(\xi|\eta) + \mathsf{H}(\eta|\zeta)$.

We will prove (i), (ii), (iii), (iv) and (vii). The remaining three parts—(v), (vi) and (viii)—are on Problem Sheet M.

[2]In fact Proposition 26.11 is used to prove Khinchin's Theorem, so it would be a circular argument to use Khinchin's Theorem to prove Proposition 26.11!

Proof. Clearly $\mathsf{H}(\xi) \geq 0$. Moreover if ξ contains at least two elements of positive measure then $\mathsf{H}(\xi) > 0$. Thus $\mathsf{H}(\xi) = 0$ if and only if ξ is the trivial partition (recall we are really working with equivalence classes). It is obvious that

$$\mathsf{H}(\xi) \geq -\log\left(\max_{1\leq k\leq p} \mu(C_k)\right).$$

To prove that $\mathsf{H}(\xi) \leq \log p$, consider the function $\psi\colon [0,\infty) \to \mathbb{R}$ given by

$$\psi(x) := \begin{cases} x\log x, & x > 0 \\ 0, & x = 0. \end{cases} \tag{26.4}$$

See Figure 26.1. Then $\psi''(x) = 1/x > 0$ and hence ψ is strictly convex. Thus

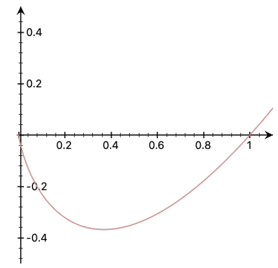

Figure 26.1: The function $\psi(x) = x\log x$.

$$\psi\left(\sum_i a_i x_i\right) \leq \sum_i a_i \psi(x_i)$$

for non-negative a_i such that $\sum_i a_i = 1$. Now taking $a_i = 1/p$ and $x_i = \mu(C_i)$ for $1 \leq i \leq p$ we obtain

$$\begin{aligned} -\frac{1}{p}\log p &= \psi\left(\frac{1}{p}\right) \\ &= \psi\left(\frac{1}{p}\sum_{i=1}^{p}\mu(C_i)\right) \\ &\leq \sum_{i=1}^{p}\frac{1}{p}\psi(\mu(C_i)) \\ &= -\frac{1}{p}\mathsf{H}(\xi). \end{aligned}$$

Finally since ψ is strictly convex we obtain equality only when the $x_i = \mu(C_i)$ are all equal to $\frac{1}{p}$. This finishes the proof of (i).

Now we prove (ii). We may assume that $\mu(C_i),\ \mu(D_j) > 0$ for each i, j, since we are working with equivalence classes. Again, using the fact that ψ is strictly convex, we compute:

$$\begin{aligned}
0 \le \mathsf{H}(\xi|\eta) &= -\sum_{j=1}^{q}\mu(D_j)\sum_{i=1}^{p}\psi\left(\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right)\\
&= -\sum_{i,j}\mu(D_j)\psi\left(\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right)\\
&\le -\sum_{i=1}^{p}\psi\left(\sum_{j=1}^{q}\mu(D_j)\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right)\\
&= -\sum_{i=1}^{p}\psi(\mu(C_i))\\
&= \mathsf{H}(\xi).
\end{aligned}$$

To prove (iii) observe that $\psi(x) < 0$ for $0 < x < 1$, and hence if $\mathsf{H}(\xi|\eta) = 0$ then for every j we have $\psi\left(\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right) = 0$ for each i. Thus $\xi \preceq \eta$. The other direction of (iii) is clear.

To prove (iv), if $\mathsf{H}(\xi|\eta) = \mathsf{H}(\xi)$ then we must have equality in the previous displayed equation for each term of the summation over i, which means that

$$\begin{aligned}
\psi(\mu(C_i)) &= \psi\left(\sum_{j=1}^{q}\mu(D_j)\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right)\\
&= \sum_{j=1}^{q}\mu(D_j)\psi\left(\frac{\mu(C_i\cap D_j)}{\mu(D_j)}\right).
\end{aligned}$$

By strict convexity of ψ this implies that for all i, j one has $\frac{\mu(C_i\cap D_j)}{\mu(D_j)} = \mu(C_i)$, which implies that ξ and η are independent.

Parts (v) and (vi) are on Problem Sheet M. Part (vii) is an immediate consequence of (vi) and the inequality

$$\mathsf{H}(\eta|\xi\vee\zeta) \le \mathsf{H}(\eta|\zeta)$$

which follows from (iii) since $\zeta \preceq \xi\vee\zeta$. Finally (viii) is also on Problem Sheet M. This completes the proof. ∎

Following the recent general theme of the course, we now turn $\mathscr{P}(X)$ into a metric space.

Definition 26.12. We define

$$d_{\mathrm{R}}\colon \mathscr{P}\times\mathscr{P}\to\mathbb{R}^+$$

by

$$d_{\mathrm{R}}(\xi,\eta) := \mathsf{H}(\xi|\eta) + \mathsf{H}(\eta|\xi).$$

The "R" stands for the Russian mathematician Rokhlin.

Corollary 26.13. *The function d_{R} defines a metric on $\mathscr{P}$ called the Rokhlin metric.*

Proof. That $d_{\mathrm{R}}(\xi, \eta) \geq 0$ with equality if and only if $\xi = \eta$ follows from parts (ii) and (iii) of Proposition 26.11. It is immediate that d_{R} is symmetric. To see the triangle inequality, we use part (viii) of Proposition 26.11:

$$\begin{aligned} d_{\mathrm{R}}(\xi, \zeta) &= \mathsf{H}(\xi|\zeta) + \mathsf{H}(\zeta|\xi) \\ &\leq \mathsf{H}(\xi|\eta) + \mathsf{H}(\eta|\zeta) + \mathsf{H}(\zeta|\eta) + \mathsf{H}(\eta|\xi) \\ &= d_{\mathrm{R}}(\xi, \eta) + d_{\mathrm{R}}(\eta, \zeta). \end{aligned}$$

This completes the proof. ∎

REMARK 26.14. Let us emphasise that the space $\mathscr{P} = \mathscr{P}(X, \mathscr{A}, \mu)$ of partitions is always a metric space for *any* probability space $(X, \mathscr{A}, \mu)$. Thus we are *not* assuming that X is itself a metric space.

Although the metric d_{R} is conceptually nice, it is a bit inconvenient to use in computations. Here is another approach to defining a metric on the space of partitions.

DEFINITION 26.15. Given $p \in \mathbb{N}$, we denote by $\mathscr{P}_p \subset \mathscr{P}$ those partitions with exactly p elements.

There is an easy way to define a metric on $\mathscr{P}_p$.

DEFINITION 26.16. We define

$$\tilde{d}_p \colon \mathscr{P}_p \times \mathscr{P}_p \to \mathbb{R}^+$$

as follows. Given $\xi = \{C_1, \dots, C_p\}$ and $\eta = \{D_1, \dots, D_p\}$ in $\mathscr{P}_p$, we set

$$\tilde{d}_p(\xi, \eta) \coloneqq \min_{\sigma \in \mathfrak{S}(p)} \sum_{i=1}^{p} \mu(C_i \vartriangle D_{\sigma(i)}),$$

where the sum is over the symmetric group of permutations of $\{1, 2, \dots, p\}$.

LEMMA 26.17. *For any $p \in \mathbb{N}$, $\tilde{d}_p$ is a metric on $\mathscr{P}_p$.*

The proof of Lemma 26.17 is left for you on Problem Sheet M. Here we will prove the following more tricky result.

PROPOSITION 26.18. *For any $p \in \mathbb{N}$, the inclusion*

$$(\mathscr{P}_p, \tilde{d}_p) \hookrightarrow (\mathscr{P}, d_{\mathrm{R}})$$

is uniformly continuous.

Proof. Fix $\varepsilon > 0$. Choose $0 < \delta < \frac{1}{4}$ small enough such that

$$- p(p-1)\delta \log \delta - (1-\delta)\log(1-\delta) < \frac{\varepsilon}{2}. \tag{26.5}$$

Suppose $\xi = \{C_1, \dots, C_p\}$ and $\eta = \{D_1, \dots, D_p\}$ in $\mathscr{P}_p$ satisfy

$$\tilde{d}_p(\xi, \eta) < \delta.$$

We claim that $d_{\mathrm{R}}(\xi,\eta) < \varepsilon$. First, without loss of generality by relabelling η we may assume that

$$\sum_{i=1}^{p} \mu(C_i \vartriangle D_i) < \delta.$$

Now consider the partition ζ with $p(p-1)+1$ elements given by the sets $C_i \cap D_j$ for $i \neq j$ and $\bigcup_{i=1}^{p}(C_i \cap D_i)$. Since for $i \neq j$ one has

$$C_i \cap D_j \subset \bigcup_{i=1}^{p}(C_i \vartriangle D_i)$$

it follows that

$$\mu(C_i \cap D_j) < \delta, \quad \text{for} \quad i \neq j, \quad \text{and} \quad \mu\left(\bigcup_{i=1}^{p}(C_i \cap D_i)\right) > 1-\delta.$$

In other words, the partition ζ consists of $p(p-1)$ "small" sets and one "large" set. Thus from the definition of $\mathsf{H}(\zeta)$ we obtain

$$\mathsf{H}(\zeta) < -p(p-1)\delta \log \delta - (1-\delta)\log(1-\delta),$$

and hence $\mathsf{H}(\zeta) < \frac{\varepsilon}{2}$ by (26.5). Next observe that by construction one has $\xi \vee \eta = \zeta \vee \eta$. Therefore applying parts (vi) and (vii) of Proposition 26.11 we obtain

$$\begin{aligned}
\mathsf{H}(\eta) + \mathsf{H}(\xi|\eta) &= \mathsf{H}(\xi \vee \eta) \\
&= \mathsf{H}(\eta \vee \zeta) \\
&\leq \mathsf{H}(\eta) + \mathsf{H}(\zeta) \\
&< \mathsf{H}(\eta) + \frac{\varepsilon}{2}.
\end{aligned}$$

This implies that

$$\mathsf{H}(\xi|\eta) < \frac{\varepsilon}{2}.$$

By symmetry (since $\xi \vee \eta = \xi \vee \zeta$) we also have

$$\mathsf{H}(\eta|\xi) < \frac{\varepsilon}{2},$$

and hence

$$d_{\mathrm{R}}(\xi,\eta) = \mathsf{H}(\xi|\eta) + \mathsf{H}(\eta|\xi) < \varepsilon.$$

This completes the proof. ■

In fact, the converse to this result is true.

Proposition 26.19. *Given $\xi,\eta \in \mathscr{P}_p$ and $\varepsilon > 0$ there exists $\delta > 0$ such that $d_{\mathrm{R}}(\xi,\eta) < \delta$ implies that $\tilde{d}_p(\xi,\eta) < \varepsilon$.*

The proof of Proposition 26.19 is on Problem Sheet M. It therefore follows that the two metrics $\tilde{d}_p$ and d_{R} are strongly equivalent on $\mathscr{P}_p$.

Measure-Theoretic Entropy

In this lecture we stick in the measure-theoretic setting, and define the *measure-theoretic entropy* of a measure-preserving dynamical system on a probability space. The construction is analogous to the definition of topological entropy via open covers given in Lecture 10, where now partitions play the role that open covers did previously. The similarity in the definitions is no accident[1]: next lecture we will prove the *Variational Principle*, which says that the topological entropy of a topological dynamical system f on a compact metric space is the supremum over all $\mu \in \mathcal{M}(f)$ of all the measure-theoretic entropies.

We start with the partition version of Definition 10.10.

DEFINITION 27.1. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Let ξ be a partition of $(X, \mathscr{A}, \mu)$. Given $k \geq 1$, we define

$$\xi_f^k := \xi \vee f^{-1}\xi \vee \cdots \vee f^{-(k-1)}\xi.$$

Explicitly, if $\xi = \{C_1, \dots, C_p\}$ then an element of ξ_f^k is a set of the form

$$\bigcap_{i=1}^{k-1} f^{-i} C_{j_i}, \qquad j_i \in \{1, \dots, p\}.$$

For any $k, n \geq 0$ we have

$$\xi_f^{k+n} = \xi^k \vee f^{-k}\xi^n. \tag{27.1}$$

We then have the following version of Proposition 10.16.

PROPOSITION 27.2. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and let $\xi \in \mathscr{P}$. Then the limit $\lim_{k\to\infty} \frac{1}{k}\mathsf{H}\big(\xi_f^k\big)$ exists.*

Proof. This is another application of Fekete's Lemma 7.7. Define $\alpha \colon \mathbb{N} \to [0, \infty)$ by

$$\alpha(k) := \mathsf{H}\big(\xi_f^k\big).$$

We will show that α is subadditive. Indeed,

$$\begin{aligned} \alpha(k+n) &= \mathsf{H}\big(\xi_f^{k+n}\big) \\ &\leq \mathsf{H}\big(\xi_f^k\big) + \mathsf{H}\big(f^{-k}\xi_f^n\big) \\ &= \alpha(k) + \alpha(n) \end{aligned}$$

where the first inequality used (27.1) and part (vii) of Proposition 26.11 and the second inequality used Lemma 26.8. Now the result follows from Fekete's Lemma 7.7. ∎

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1] In fact the similarity is so extreme that large swathes of this lecture is essentially copy-pasted from Lecture 10—modulo replacing $\mathfrak{U}$ with ξ and updating the references!

This allows us to define the measure-theoretic entropy relative to a partition, just as in Definition 10.17.

Definition 27.3. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and let $\xi \in \mathscr{P}$. We define the **measure-theoretic entropy of f relative to ξ** by

$$\mathsf{h}_\mu(f, \xi) := \lim_{k\to\infty} \frac{\mathsf{H}\big(\xi_f^k\big)}{k} \in [0, \infty).$$

To define the measure-theoretic entropy of f we now take the supremum over all partitions ξ, just as in Definition 10.18.

Definition 27.4. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. We define the **measure-theoretic entropy of f** to be the quantity

$$\mathsf{h}_\mu(f) := \sup_{\xi\in\mathscr{P}} \mathsf{h}_\mu(f, \xi).$$

In general the measure-theoretic entropy could be infinite. When there is more than one measure in play (which will be the case next lecture), we call $\mathsf{h}_\mu(f)$ the **measure-theoretic entropy of f with respect to μ.**

Remark 27.5. Some textbooks refer to measure-theoretic entropy as *metric entropy*. Whilst this has the advantage of being shorter and less cumbersome, we will avoid this terminology as it clashes with the "metric" in metric spaces.

We first show that the entropy $\mathsf{h}_\mu(f, \xi)$ can be computed via a different limit. After proving this statement we will attempt to motivate the definition of measure-theoretic entropy.

Proposition 27.6. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and let $\xi \in \mathscr{P}$. Then*

$$\mathsf{h}_\mu(f, \xi) = \lim_{k\to\infty} \mathsf{H}\big(\xi \big| f^{-1}\xi_f^k\big).$$

Proof. Set

$$a_k := \mathsf{H}\big(\xi_f^k\big) \qquad \text{and} \qquad b_k := \mathsf{H}\big(\xi\big|f^{-1}\xi_f^k\big).$$

Then

$$\mathsf{h}_\mu(f, \xi) = \lim_{k\to\infty} \frac{a_k}{k},$$

and we are required to prove that

$$\mathsf{h}_\mu(f, \xi) \stackrel{?}{=} \lim_{k\to\infty} b_k.$$

First note that

$$\xi_f^{k+1} = \xi \vee f^{-1}\xi_f^k$$

by (27.1). Since $\mathsf{H}(f^{-1}\zeta) = \mathsf{H}(\zeta)$ and $\mathsf{H}(\zeta \vee \eta) = \mathsf{H}(\zeta) + \mathsf{H}(\eta|\zeta)$ by Lemma 26.8 and part (vi) of Proposition 26.11 respectively, we have

$$\begin{aligned} a_{k+1} &= \mathsf{H}\big(\xi_f^{k+1}\big) \\ &= \mathsf{H}\big(f^{-1}\xi_f^k \vee \xi\big) \\ &= \mathsf{H}(\xi_f^k) + \mathsf{H}\big(\xi\big|f^{-1}\xi_f^k\big) \\ &= a_k + b_k. \end{aligned}$$

Thus

$$a_{k+1} - a_k = b_k$$

and so summing over k yields

$$a_k = \mathsf{H}(\xi) + \sum_{i=1}^{k-1} b_i.$$

Therefore we have

$$\mathsf{h}_\mu(f, \xi) = \lim_{k\to\infty} \frac{a_k}{k} = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} b_i.$$

We are nearly done. By part (v) of Proposition 26.11 we have $\mathsf{H}(\xi|\zeta) \le \mathsf{H}(\xi|\eta)$ when $\eta \preceq \zeta$. This implies that $b_k \le b_{k-1}$, and hence (b_k) is a bounded sequence of real numbers and so $\lim_k b_k$ exists. Now recall from real analysis[2] that if the Cesaro limit of a convergent sequence exists, then the Cesaro limit is necessarily equal to the normal limit. Thus

$$\lim_{k\to\infty} b_k = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} b_i = \mathsf{h}_\mu(f, \xi).$$

This completes the proof. ■

REMARK 27.7. Recall that the entropy $\mathsf{H}(\xi)$ can be thought as measuring the uncertainty of an "experiment" ξ. If f is a dynamical system which governs the behaviour of the system under time, then the partition ξ_f^k represents the combined experiment of performing ξ on k consecutive "days" (or whatever time unit an application of f represents). Thus the entropy $\mathsf{H}\big(\xi_f^k\big)$ can be thought of as the uncertainty inherent in performing ξ on k consecutive days.

Next, recall that $\mathsf{H}(\xi|\eta)$ represents the uncertainty in the experiment ξ given that we already know what happened when we performed η. Therefore Proposition 27.6 tells us that the measure-theoretic entropy $\mathsf{h}_\mu(f, \xi)$ can be thought of as an average uncertainty of performing the experiment ξ on a given day, given that we already know what happened on all the previous days.

Taking this one step further, this means that $\mathsf{h}_\mu(f)$ can be thought of measuring the maximum (over all possible experiments) of the average uncertainty of performing a given experiment every day, forever. In other words, we look for the "least accurate" experiment we can find in our system and then test it every single day and see on average how many mistakes we make in our predictions.

The next result summarises the basic properties of the measure-theoretic entropy.

PROPOSITION 27.8. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and let $\xi, \eta \in \mathscr{P}$. Then:*

[2] Here is the precise statement: Suppose $(c_k)_{k\ge 0}$ is any sequence of real numbers. Let $s_k := \frac{1}{k}\sum_{i=0}^{k-1} c_i$. Suppose that $c_k \to c$ and $s_k \to s$ with both c, s finite. Then $c = s$. The limit s is called the *Cesaro limit* of the sequence (c_k). If you have forgotten how to prove this, consider it an instructive exercise!

(i) *It holds that*

$$\limsup_{k\to\infty} -\frac{1}{k}\log\left(\max_{C\in\xi_f^k}\mu(C)\right) \leq \mathsf{h}_\mu(f,\xi) \leq \mathsf{H}(\xi). \tag{27.2}$$

(ii) $\mathsf{h}_\mu(f,\xi\vee\eta) \leq \mathsf{h}_\mu(f,\xi) + \mathsf{h}_\mu(f,\eta)$.

(iii) $\mathsf{h}_\mu(f,f^{-1}\xi) = \mathsf{h}_\mu(f,\xi)$ *and if* f *is reversible then* $\mathsf{h}_\mu(f,f\xi) = \mathsf{h}_\mu(f,\xi)$.

(iv) $\mathsf{h}_\mu(f,\xi) = \mathsf{h}_\mu\big(f,\xi_f^p\big)$ *for any* $p\in\mathbb{N}$.

(v) $\mathsf{h}_\mu(f,\xi) \leq \mathsf{h}_\mu(f,\eta) + \mathsf{H}(\xi|\eta)$. *Thus if* $\xi\preceq\eta$ *then* $\mathsf{h}_\mu(f,\xi)\leq\mathsf{h}_\mu(f,\eta)$.

Almost everything here follows readily from the various parts of Proposition 26.11, but let us check the details anyway.

Proof. The first inequality in (27.2) follows from part (i) of Proposition 26.11. The second inequality in (27.2) follows from Proposition 27.6 together with part (ii) of Proposition 26.11.

Next, since

$$(\xi\vee\eta)_f^k = \xi_f^k\vee\eta_f^k,$$

part (ii) follows from part (vii) of Proposition 26.11. Similarly, since

$$(f^{-1}\xi)_f^k = f^{-1}\xi_f^k$$

the first statement in part (iii) follows from Lemma 26.8, and the invertible case works in the same way.

Part (iv) is analogous to the end of the proof of Theorem 10.25. Using (27.1) we compute

$$\begin{aligned}
\mathsf{h}_\mu\big(f,\xi_f^p\big) &= \lim_{k\to\infty}\frac{1}{k}\mathsf{H}\left(\bigvee_{i=0}^{k-1} f^{-i}\xi_f^p\right)\\
&= \lim_{k\to\infty}\frac{1}{k}\mathsf{H}\big(\xi_f^{k+p}\big)\\
&= \lim_{n\to\infty}\frac{1}{n}\mathsf{H}\big(\xi_f^n\big)\\
&= \mathsf{h}_\mu(f,\xi).
\end{aligned}$$

This leaves us with part (v), which is considerably trickier. Using $\mathsf{H}(\xi\vee\eta|\zeta) = \mathsf{H}(\xi|\zeta)+\mathsf{H}(\eta|\xi\vee\zeta)$ and the fact that $\mathsf{H}(\xi|\eta)\leq\mathsf{H}(\xi|\zeta)$ if $\zeta\preceq\eta$ (these are parts (vi) and (v) of Proposition 26.11 respectively) repeatedly we have

$$\begin{aligned}
\mathsf{H}\big(\xi_f^k\big|\eta_f^k\big) &= \mathsf{H}\big(\xi\vee f^{-1}\xi_f^{k-1}\big|\eta_f^k\big)\\
&= \mathsf{H}\big(\xi\big|\eta_f^k\big) + \mathsf{H}\big(f^{-1}\xi_f^{k-1}\big|\xi\vee\eta_f^k\big)\\
&\leq \mathsf{H}(\xi|\eta) + \mathsf{H}\big(f^{-1}\xi_f^{k-1}\big|\eta_f^k\big), \qquad (27.3)
\end{aligned}$$

where we used that $\eta \preceq \eta_f^k$ and $\eta_f^k \preceq \xi \vee \eta_f^k$. Now the same argument applied to the second term on the right-hand side of (27.3) gives

$$\begin{aligned} \mathsf{H}\big(f^{-1}\xi_f^{k-1}\big|\eta_f^k\big) &\leq \mathsf{H}\big(f^{-1}\xi\big|f^{-1}\eta\big) + \mathsf{H}\big(f^{-2}\xi_f^{k-2}\big|\eta_f^k\big) \\ &= \mathsf{H}(\xi|\eta) + \mathsf{H}\big(f^{-2}\xi_f^{k-2}|\eta_f^k\big). \end{aligned} \tag{27.4}$$

Combining (27.3) and (27.4)

$$\mathsf{H}(\xi_f^k|\eta_f^k) \leq 2\mathsf{H}(\xi|\eta) + \mathsf{H}(f^{-2}\xi_f^{k-2}|\eta_f^k).$$

Continuing inductively, we eventually get

$$\begin{aligned} \mathsf{H}\big(\xi_f^k\big|\eta_f^k\big) &\leq (k-1)\mathsf{H}(\xi|\eta) + \mathsf{H}\big(f^{-(k-1)}\xi\big|\eta_f^k\big) \\ &\leq k\mathsf{H}(\xi|\eta). \end{aligned}$$

Thus we have

$$\begin{aligned} \mathsf{H}\big(\xi_f^k\big) &\leq \mathsf{H}\big(\xi_f^k \vee \eta_f^k\big) \\ &= \mathsf{H}\big(\eta_f^k\big) + \mathsf{H}\big(\xi_f^k\big|\eta_f^k\big) \\ &\leq \mathsf{H}\big(\eta_f^k\big) + k\mathsf{H}(\xi|\eta). \end{aligned}$$

Dividing by k and letting $k \to \infty$ gives

$$\mathsf{h}_\mu(f,\xi) \leq \mathsf{h}_\mu(f,\eta) + \mathsf{H}(\xi|\eta),$$

which proves part (v). This completes the proof. ∎

Corollary 27.9. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. The function $\mathsf{h}_\mu(f,\cdot)\colon \mathscr{P} \to [0,\infty)$ is Lipschitz continuous function with respect to the Rokhlin metric:*

$$\big|\mathsf{h}_\mu(f,\xi) - \mathsf{h}_\mu(f,\eta)\big| \leq d_{\mathrm{R}}(\xi,\eta). \tag{27.5}$$

We will refer to (27.5) as the **Rokhlin inequality** in what follows.

Proof. This is immediate from from part (v) of Proposition 27.8. ∎

Now let us isolate a special class of partitions that can be used to compute the entropy.

Definition 27.10. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. A family Ξ of partitions is a **generating family** for f if the space of all partitions subordinate to a partition of the form ξ_f^k for some $\xi \in \Xi$ and $k \geq 1$ is dense in the space $\mathscr{P}$ (equipped with the Rokhlin metric).

Explicitly, a family Ξ is generating if for any $\eta \in \mathscr{P}$ and any $\varepsilon > 0$ there exists a partition $\zeta \in \mathscr{P}$ and $\xi \in \Xi$ such that

$$d_{\mathrm{R}}(\eta,\zeta) < \varepsilon \qquad \text{and} \qquad \zeta \preceq \xi_f^k \text{ for all } k \text{ sufficiently large.} \tag{27.6}$$

Equivalently one could ask that $\zeta \preceq \xi_f^k$ for *some* k, since then for any $n \geq k$ one has $\zeta \preceq \xi_f^k \preceq \xi_f^n$.

Proposition 27.11. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Suppose Ξ is a generating family of partitions for f. Then*

$$\mathsf{h}_\mu(f) = \sup_{\xi \in \Xi} \mathsf{h}_\mu(f, \xi).$$

Proof. Let $\eta \in \mathscr{P}$ and $\varepsilon > 0$. There exists a partition $\zeta \in \mathscr{P}$ and $\xi \in \Xi$ such that $\zeta \preceq \xi_f^k$ for all k sufficiently large, and such that $d_{\mathrm{R}}(\eta, \zeta) < \varepsilon$. Then

$$\mathsf{h}_\mu(f, \eta) \leq \mathsf{h}_\mu(f, \zeta) + \varepsilon$$

by the Rokhlin inequality (27.5), and thus

$$\mathsf{h}_\mu(f, \zeta) \leq \mathsf{h}_\mu(f, \xi_f^k) = \mathsf{h}_\mu(f, \xi)$$

by parts (iv) and (v) of Proposition 27.8. Since η and ε were arbitrary, the result follows. ∎

Now let us make the analogy of Definition 10.1.

Definition 27.12. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. A partition ξ is called a **generator** for f if $\Xi = \{\xi\}$ is a generating family.

The following corollary (which is a special case of Proposition 27.11) has its own special name. It is the measure-theoretic analogue of Theorem 10.25.

Corollary 27.13 (Kolmogorov-Sinai). *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and suppose ξ is a generator for f. Then $\mathsf{h}_\mu(f) = \mathsf{h}_\mu(f, \xi)$.*

If we begin on a metric space then we can produce generating families in the same way as Proposition 10.22.

Proposition 27.14. *Let (X, d) be a compact metric space and suppose $\mu \in \mathcal{M}(X)$. Suppose (ξ_n) is a sequence of partitions of X such that*[3]

$$\operatorname{diam} \xi_n := \max\{\operatorname{diam} C \mid C \in \xi_n\} \to 0, \qquad \text{as} \qquad n \to \infty.$$

Set $\Xi = \{\xi_n\}$. Then if $f \colon X \to X$ is any topological dynamical system which is measure-preserving with respect to μ then Ξ is a generating family of partitions for f.

Proof. Fix $\varepsilon > 0$. Suppose $\eta = \{C_1, \ldots, C_p\}$ is any partition of $(X, \mathscr{B}, \mu)$. By Proposition 26.18 it suffices to find a partition $\zeta = \{D_1, \ldots, D_p\}$ such that $\zeta \preceq \xi_n$ for all n sufficiently large and such that $\tilde{d}_p(\eta, \xi) < \varepsilon$.

Choose compact sets $K_i \subseteq C_i$ such that

$$\mu(C_i \setminus K_i) < \frac{\varepsilon}{p(p+1)}$$

(such sets exist by Proposition 23.4). Set

$$\delta := \inf_{i \neq j} d(K_i, K_j) > 0.$$

[3] Here we define the diameter of a disconnected space to be infinite. Thus the hypotheses imply that the elements of each ξ_n are all connected for large n.

Choose $m \in \mathbb{N}$ large enough such that

$$\operatorname{diam} \xi_m \leq \frac{\delta}{2}.$$

Now for $1 \leq i \leq p-1$ let D_i denote the union of elements of ξ_m which intersect K_i, and let D_p denote the union of the remaining elements of ξ_m. By the choice of δ, each element $C \in \xi_m$ can intersect at most one K_i. Thus ζ is a well defined partition which is obviously subordinate to ξ_m. For any $1 \leq i \leq p$, one has

$$\begin{aligned}
\mu(C_i \vartriangle D_i) &= \mu(C_i \setminus D_i) + \mu(D_i \setminus C_i) \\
&\leq \mu(C_i \setminus K_i) + \mu\left(X \setminus \bigcup_{j=1}^{p} K_j\right) \\
&< \frac{\varepsilon}{p(p+1)} + \frac{\varepsilon}{p+1} \\
&= \frac{\varepsilon}{p}.
\end{aligned}$$

Summing over k tells us that

$$\sum_{i=1}^{p} \mu(C_i \vartriangle D_i) < \varepsilon.$$

Thus we also have $\tilde{d}_p(\eta, \zeta) < \varepsilon$ as required. ∎

LECTURE 28

The Variational Principle

In this last lecture, we compare measure-theoretic entropy to topological entropy, culminating in the statement and proof of the famous *Variational Principle*.

As we have seen, measure-theoretic entropy gives a quantitative measure of the inherent uncertainty, or complexity, of a dynamical system, as seen via an invariant measure. Topological entropy was actually discovered later than measure-theoretic entropy, and its discovery was motivated by an attempt to extract the same information using topological dynamics only.

Nevertheless, the absence of a "natural measure" for the size of a set in topological dynamics makes the latter more crude. For example, in Proposition 8.5 we proved that the topological entropy of the union of two invariant sets is the *maximum* of their individual entropies. Meanwhile in the measure-theoretic setting, we have the following more refined result: the entropy is the weighted sum of the individual entropies.

NOTATION. Since there are now multiple measures in play, we will write $\mathsf{H}_\mu(\xi)$ for the entropy of a partition ξ with respect to μ instead of just $\mathsf{H}(\xi)$.

PROPOSITION 28.1. *Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Suppose A is an invariant set for f with $0 < \mu(A) < 1$. Then*

$$\mathsf{h}_\mu(f) = \mu(A)\mathsf{h}_{\mu_A}\big(f|_A\big) + (1 - \mu(A))\mathsf{h}_{\mu_{X\setminus A}}\big(f|_{X\setminus A}\big).$$

Here $f|_A$ is the measure-preserving dynamical system on the restricted probability space $\big(A, \mathscr{A}_A, \mu_A\big)$ (Example 18.12), and similarly for $f|_{X\setminus A}$.

Proof. Let $\zeta = \{A, X \setminus A\}$, and suppose ξ is a partition such that $\zeta \preceq \xi$. Then we have (with the obvious notation) that

$$\mathsf{H}_\mu(\xi) = \mu(A)\mathsf{H}_{\mu_A}\big(\xi|_A\big)+\mu(X\setminus A)\mathsf{H}_{\mu_{X\setminus A}}\big(\xi|_{X\setminus A}\big)+\mu(A)\log\mu(A)+\mu(X\setminus A)\log\mu(X\setminus A)$$

Since A is f-invariant, if $\zeta \preceq \xi$ then also $\zeta \preceq \xi_f^k$. Thus

$$\begin{aligned}
\mathsf{h}_\mu(f,\xi) &= \lim_{k\to\infty}\frac{1}{k}\mathsf{H}_\mu\big(\xi_f^k\big) \\
&= \mu(A)\lim_{k\to\infty}\frac{1}{k}\mathsf{H}_{\mu_A}\Big((\xi|_A)^k_{f|_A}\Big) + \mu(X\setminus A)\lim_{k\to\infty}\frac{1}{k}\mathsf{H}_{\mu_{X\setminus A}}\Big((\xi|_{X\setminus A})^k_{f|_{X\setminus A}}\Big) \\
&\quad + \lim_{k\to\infty}\frac{1}{k}\Big(\mu(A)\log\mu(A) + \mu(X\setminus A)\log\mu(X\setminus A)\Big) \\
&= \mu(A)\mathsf{h}_{\mu_A}\big(f|_A, \xi|_A\big) + (1-\mu(A))\mathsf{h}_{\mu_{X\setminus A}}\big(f|_{X\setminus A}, \xi|_{X\setminus A}\big) + 0.
\end{aligned}$$

Taking the supremum over such ξ we obtain

$$\sup_{\zeta\preceq\xi}\mathsf{h}_\mu(f,\xi) = \mu(A)\mathsf{h}_{\mu_A}\big(f|_A\big) + (1-\mu(A))\mathsf{h}_{\mu_{X\setminus A}}\big(f|_{X\setminus A}\big).$$

Finally we observe that for any partition η there is a partition ξ such that $\zeta \preceq \xi$ and $\eta \preceq \xi$ (for example, $\xi := \zeta \vee \eta$ works). Thus

$$\mathsf{h}_\mu(f) = \sup_{\zeta \preceq \xi} \mathsf{h}_\mu(f, \xi)$$

This completes the proof. ■

Comparing Proposition 8.5 and Proposition 28.1 suggests that topological entropy measures the *maximum dynamical complexity*, whereas measure-theoretic entropy measures the *average dynamical complexity*. This leads us to hypothesise:

- Topological entropy to be an upper bound for the measure-theoretic entropy;
- The measure-theoretic entropy should be maximised by a measure that assigns most weight to regions of maximal complexity.

This is indeed the case. Here is the statement of the Variational Principle.

THEOREM 28.2 (The Variational Principle). *Let X be a compact metric space and let $f\colon X \to X$ denote a dynamical system. Then*

$$\mathsf{h}_{\mathrm{top}}(f) = \sup_{\mu \in \mathcal{M}(f)} \mathsf{h}_\mu(f).$$

The proof of Theorem 28.2 will take some time, and we will need a number of auxiliary results.

IMPROVED DEFINITION: Since we are now restricting our attention once more to compact metric spaces, we can adopt a slightly simpler definition of a partition. Namely, let us henceforth declare that **partition** of $(X, \mathscr{B})$ is any finite collection $\xi = \{C_1, \dots, C_p\}$ of *disjoint* measurable sets whose union is all of X. We denote the set of all such partitions as $\mathscr{P}_{\mathrm{dis}}(X, \mathscr{B})$.

Thus an element $\xi \in \mathscr{P}_{\mathrm{dis}}(X, \mathscr{B})$ is an actual honest partition of X (in the normal sense of the word). And best of all—there are *no* equivalence relations involved!

The following facts are clear:

- If $\mu \in \mathcal{M}(X)$ is any Borel probability measure and $\xi \in \mathscr{P}_{\mathrm{dis}}(X, \mathscr{B})$ then ξ will also be a partition of $(X, \mathscr{B}, \mu)$ in the sense of Definition 26.1.
- Conversely if ξ is any partition of $(X, \mathscr{B}, \mu)$ in the sense of Definition 26.1 then there is another partition η that belongs to the same equivalence class as ξ in $\mathscr{P}(X, \mathscr{B}, \mu)$ and such that $\eta \in \mathscr{P}_{\mathrm{dis}}(X, \mathscr{B})$.

Thus all the results from Lectures 26 and 27 continue to hold with our new definition of a partition. Whilst the new definition is slightly more restrictive, it will be more efficient in this lecture. The reason for this is because we will be working with multiple measures at once, and it would quickly become tiresome to have to constantly check that all our purported partitions were indeed partitions (in the sense of Definition 26.1) for each individual measure involved. From now on, if we simply say "partition" it should be understood in the new sense of the word.

PROPOSITION 28.3. *Let X be a compact metric space, and let $\mu, \nu \in \mathcal{M}(X)$ be two probability measures on $(X, \mathscr{B})$.*

(i) *Let ξ be a partition of X. Then for any $c \in [0,1]$ one has*

$$c\,\mathsf{H}_\mu(\xi) + (1-c)\,\mathsf{H}_\nu(\xi) \le \mathsf{H}_{c\mu+(1-c)\nu}(\xi).$$

(ii) *Suppose f be a topological dynamical system on X such that $\mu, \nu \in \mathcal{M}(f)$. Then for any $c \in [0,1]$ one has*

$$c\,\mathsf{h}_\mu(f) + (1-c)\,\mathsf{h}_\nu(f) = \mathsf{h}_{c\mu+(1-c)\nu}(f).$$

Thus the entropy map

$$\mathsf{h}_\bullet(f)\colon \mathcal{M}(f) \to [0,\infty], \qquad \mu \mapsto \mathsf{h}_\mu(f)$$

is affine.

Proof. We begin with (i). Recall the convex function

$$\psi(x) := \begin{cases} x\log x, & x \ge 0 \\ 0, & x = 0, \end{cases}$$

from (26.4) and Figure 26.1. Let $A \in \mathscr{B}$ and abbreviate

$$a := \mu(A), \qquad b := \nu(A).$$

Then since ψ is convex we have

$$\begin{aligned}
0 &\ge \psi\Big(ca + (1-c)b\Big) - c\psi(a) - (1-c)\psi(b) \\
&= \Big(ca + (1-c)b\Big)\log\Big(ca + (1-c)b\Big) \\
&\quad - ca\log a - (1-c)b\log b \\
&= ca\underbrace{\Big(\log\big(ca + (1-c)b\big) - \log ca\Big)}_{\ge 0} \\
&\quad + (1-c)b\underbrace{\Big(\log\big(ca + (1-c)b\big) - \log\big((1-c)b\big)\Big)}_{\ge 0} \\
&\quad + ca\Big(\log ca - \log a\Big) \\
&\quad + (1-c)b\Big(\log\big((1-c)b\big) - \log b\Big) \\
&\ge ca\log c + (1-c)b\log(1-c),
\end{aligned}$$

where we used the fact that log is an increasing function. It follows that if ξ is any partition then

$$\begin{aligned}
0 &\le \mathsf{H}_{c\mu+(1-c)\nu}(\xi) - c\,\mathsf{H}_\mu(\xi) - (1-c)\,\mathsf{H}_\nu(\xi) \\
&\le -\Big(c\log c + (1-c)\log(1-c)\Big) \\
&\le \log 2.
\end{aligned}$$

This proves (i). Now if η is any partition, setting $\xi = \eta_f^k$ in the above gives

$$0 \le \mathsf{H}_{c\mu+(1-c)\nu}(\eta_f^k) - c\,\mathsf{H}_\mu(\eta_f^k) - (1-c)\,\mathsf{H}_\nu(\eta_f^k) \le \log 2.$$

Dividing by k and letting $k \to \infty$ tells us that

$$\mathsf{h}_{c\mu+(1-c)\nu}(f,\eta) = c\,\mathsf{h}_\mu(f,\eta) + (1-c)\,\mathsf{h}_\nu(f,\eta). \tag{28.1}$$

The right-hand side of (28.1) is bounded above by the respective measure-theoretic entropies:

$$\mathsf{h}_{c\mu+(1-c)\nu}(f,\eta) \le c\,\mathsf{h}_\mu(f) + (1-c)\,\mathsf{h}_\nu(f)$$

Then since η was arbitrary we obtain

$$\mathsf{h}_{c\mu+(1-c)\nu}(f) \le c\,\mathsf{h}_\mu(f) + (1-c)\,\mathsf{h}_\nu(f). \tag{28.2}$$

It remains to prove the other direction. Fix $\varepsilon > 0$, and choose a partition η and ζ such that

$$\mathsf{h}_\mu(f,\eta) > \begin{cases} \mathsf{h}_\mu(f) - \varepsilon, & \text{if } \mathsf{h}_\mu(f) < \infty, \\ \frac{1}{\varepsilon}, & \text{if } \mathsf{h}_\mu(f) = \infty. \end{cases}$$

and

$$\mathsf{h}_\nu(f,\zeta) > \begin{cases} \mathsf{h}_\nu(f) - \varepsilon, & \text{if } \mathsf{h}_\nu(f) < \infty, \\ \frac{1}{\varepsilon}, & \text{if } \mathsf{h}_\nu(f) = \infty. \end{cases}$$

Set $\xi := \eta \vee \zeta$. Then from (28.1) we have

$$\mathsf{h}_{c\mu+(1-c)\nu}(f,\xi) > \begin{cases} c\,\mathsf{h}_\mu(f) + (1-c)\,\mathsf{h}_\nu(f) - \varepsilon, & \text{if } \mathsf{h}_\mu(f) < \infty \text{ and } \mathsf{h}_\nu(f) < \infty, \\ \frac{1}{\varepsilon}, & \text{if either } \mathsf{h}_\mu(f) = \infty \text{ or } \mathsf{h}_\nu(f) = \infty. \end{cases}$$

Since ε was arbitrary we conclude that

$$\mathsf{h}_{c\mu+(1-c)\nu}(f) \ge c\,\mathsf{h}_\mu(f) + (1-c)\,\mathsf{h}_\nu(f).$$

This completes the proof. ■

REMARK 28.4. The entropy map $\mu \mapsto \mathsf{h}_\mu(f)$ is *not* necessarily continuous. That is, if $\mu_k, \mu \in \mathcal{M}(f)$ with $\mu_k \rightharpoonup \mu$ then in general $\mathsf{h}_{\mu_k}(f)$ may not converge to $\mathsf{h}_\mu(f)$.

However with a bit of work[1] one can prove that if f is either an expansive topological dynamical system or a reversible weakly expansive dynamical system (cf. Definitions 9.1 and 9.8) then in this case the entropy map is at least *upper semi-continuous*, in the sense that if $\mu_k \rightharpoonup \mu$ then

$$\limsup_{k\to\infty} \mathsf{h}_{\mu_k}(f) \le \mathsf{h}_\mu(f).$$

[1] In an ideal world I would prove this next week in "Lecture 29". Sadly though next week is Christmas, and thus I rather suspect none of you would turn up if I tried to lecture... So I will just content myself at stating this as an interesting remark.

Remark 28.5. It will be useful in the proof of the Variational Principle to observe that part (i) of Proposition 28.3 works for arbitrary finite convex combinations: Suppose $\mu_1, \dots, \mu_k$ belong to $\mathcal{M}(X)$ and $c_1, \dots, c_k$ are non-negative real numbers such that $\sum_{i=1}^k c_i = 1$. Set $\mu := \sum_{i=1}^k c_i \mu_i \in \mathcal{M}(X)$. Then for any partition ξ one has

$$\sum_{i=1}^{k} c_i \, \mathsf{H}_{\mu_i}(\xi) \le \mathsf{H}_{\mu}(\xi). \tag{28.3}$$

The proof of (28.3) is by induction on k, where part (i) of Proposition 28.3 does both the base case $k = 2$ and the inductive step.

We need two further preliminary results before we are ready to embark on the proof of the Variational Principle.

Lemma 28.6. *Let X be a compact metric space and let $\mu \in \mathcal{M}(X)$. For any $\delta > 0$ there exists a partition $\xi = \{C_1, \dots, C_p\} \in \mathscr{P}_{\mathrm{dis}}(X, \mathscr{B})$ such that*

$$\operatorname{diam} C_i < \delta \qquad \text{and} \qquad \mu(\partial C_i) = 0, \qquad \forall\, 1 \le i \le p.$$

Proof. First note that for any $x \in X$ and any $\delta > 0$ there exists $0 < \delta' < \delta$ such that $\mu\big(\partial B(x, \delta')\big) = 0$. Indeed, if this was not true we would have an uncountable collection of disjoint sets of positive measure, which contradicts μ being a probability measure. Thus we can find an open cover $\mathcal{U} = \{B_1, \dots, B_p\}$ of open balls of radius less than $\delta/2$ such that $\mu(\partial B_i) = 0$ for each i. Now let $C_1 = \overline{B_1}$ and for $i > 1$ let

$$C_i = \overline{B}_i \setminus \bigcup_{j=1}^{i-1} \overline{B}_j.$$

Then $\xi = \{C_1, \dots C_p\}$ is a partition of X of sets of diameter less than δ, and since

$$\partial C_i \subseteq \bigcup_{j=1}^{i} \partial B_i,$$

we have $\mu(\partial C_i) = 0$ for each i. This completes the proof. ■

Lemma 28.7. *Let $f \colon X \to X$ be a topological dynamical system on a compact metric space, and let $\mu \in \mathcal{M}(f)$. Suppose $A_0, \dots, A_q$ are any sets with $\mu(\partial A_i) = 0$ for each $0 \le i \le q$. Then*

$$\mu\left(\partial\left(\bigcap_{i=0}^{q} f^{-i}(A_i)\right)\right) = 0.$$

Proof. Since

$$\partial\left(\bigcap_{i=0}^{q} f^{-i}(A_i)\right) \subseteq \bigcup_{i=0}^{q} f^{-i}(\partial A_i),$$

the conclusion is immediate from the hypotheses. ■

We are now ready for the best[2] proof in the course! This proof is non-examinable.

[2]Where of course, best := longest.

(♣) *Proof of the Variational Principle 28.2.* We prove the result in four steps. The first two steps are devoted to showing that

$$\mathsf{h}_\mu(f) \le \mathsf{h}_{\mathrm{top}}(f), \qquad \forall\, \mu \in \mathcal{M}(f),$$

which thus proves one half of the Variational Principle. The last two steps deal with the other (harder) direction.

1. Fix $\mu \in \mathcal{M}(f)$, and let $\xi = \{C_1, \dots, C_p\}$ be a partition of $(X, \mathscr{B})$. In this first step we will construct an open cover $\mathfrak{U} = \{U_1, \dots, U_p\}$ of X such that

$$\mathsf{h}_\mu(f, \xi) \le \mathsf{h}^*(f, \mathfrak{U}) + \log 2 + 1. \tag{28.4}$$

Fix

$$0 < \varepsilon < \frac{1}{p \log p}, \tag{28.5}$$

and choose compact sets $K_i \subset \mathrm{int}(C_i)$ for $k = i, \dots, p$ such that $\mu(C_i \setminus K_i) < \varepsilon$ (such sets exist by Proposition 23.4). Let

$$K_0 := X \setminus \bigcup_{i=1}^{p} K_i,$$

so that $\mu(K_0) < p\varepsilon$, and define a new partition with $p+1$ elements:

$$\eta := \{K_0, K_1, \dots, K_p\}.$$

Let us compute the conditional entropy $\mathsf{H}_\mu(\xi|\eta)$:

$$\begin{aligned}
\mathsf{H}_\mu(\xi|\eta) &= -\sum_{i=0}^{p} \sum_{j=1}^{p} \mu(K_i) \frac{\mu(C_j \cap K_i)}{\mu(K_i)} \log\left(\frac{\mu(C_j \cap K_i)}{\mu(K_i)}\right) \\
&\overset{(\heartsuit)}{=} -\mu(K_0) \sum_{j=1}^{p} \frac{\mu(C_j \cap K_0)}{\mu(K_0)} \log\left(\frac{\mu(C_j \cap K_0)}{\mu(K_0)}\right) \\
&\le \mu(K_0) \log p \\
&< \varepsilon p \log p \\
&< 1.
\end{aligned} \tag{28.6}$$

Here $(\heartsuit)$ used the fact that for $i \ne 0$, $\frac{\mu(C_j \cap K_i)}{\mu(K_i)} \in \{0, 1\}$, and the last line used (28.5).

Observe that for each $1 \le i \le p$, the set

$$U_i := K_i \cup K_0$$

is an open subset of X, since it is equal to $X \setminus \bigcup_{j \ne i} K_j$, and thus $\mathcal{U} := \{U_1, \dots, U_p\}$ is an open cover of X.

Every element of the join cover $\mathcal{U}_f^k$ is of the form

$$U_{i_0} \cap f^{-1}(U_{i_1}) \cap \dots \cap f^{-(k-1)}(U_{i_{k-1}}),$$

for some tuple $(i_0, \dots i_{k-1})$ of integers in $\{1, \dots, p\}$. Such a set can be written as a pairwise disjoint union of 2^q elements of the join partition η_f^k (some of which may be empty sets). Thus every element of $\mathcal{U}_f^k$ contains at most 2^q elements of η_f^k. Taking the union over a sub cover of $\mathcal{U}_f^k$ with cardinality $\min \mathcal{U}_f^k$, we get that

$$\# \eta_f^k \leq 2^k \min \mathcal{U}_f^k,$$

and hence by part (i) of Proposition 26.11 we have

$$\begin{aligned} \mathsf{H}_\mu(\eta_f^k) &\leq \log \# \eta_f^k \\ &\leq \log(2^k \cdot \min \mathcal{U}_T^k) \\ &= \mathsf{H}(\mathcal{U}_f^k) + k \log 2, \end{aligned}$$

Thus we obtain

$$\mathsf{h}_\mu(f, \eta) \leq \mathsf{h}^*(f, \mathfrak{U}) + \log 2.$$

This is not quite what we want, since on the left-hand side we used the partition η and not our original partition ξ. However by part (v) of Proposition 27.8 and (28.6) we can estimate

$$\begin{aligned} \mathsf{h}_\mu(f, \xi) &\leq h_\mu(f, \eta) + \mathsf{H}_\mu(\xi | \eta) \\ &\leq \ \mathsf{h}^*(f, \mathfrak{U}) + \log 2 + 1. \end{aligned}$$

This proves (28.1), and concludes this first step.

2. Since ξ was arbitrary, it follows from (28.1) that

$$\begin{aligned} \mathsf{h}_\mu(f) &\leq \mathsf{h}^*(f, \mathfrak{U}) + \log 2 + 1 \\ &\leq \mathsf{h}_{\text{top}}(f) + \log 2 + 1, \end{aligned}$$

where we used Corollary 10.23. We are almost done, apart from the annoying $1 + \log 2$ term. But this can be got rid off with the following trick (which we have used before in the proof of Step *3* of Theorem 11.7). Namely, by repeat the same argument with f^q (for $q \in \mathbb{N}$ arbitrary) instead of f we obtain also that

$$\mathsf{h}_\mu(f^q) \leq \mathsf{h}_{\text{top}}(f^q) + \log 2 + 1. \tag{28.7}$$

Now by Problem D.2 and Problem N.1 respectively, we have

$$\mathsf{h}_\mu(f^q) = q \mathsf{h}_\mu(f) \qquad \text{and} \qquad \mathsf{h}_{\text{top}}(f^q) = q \mathsf{h}_{\text{top}}(f^q)$$

and thus dividing (28.7) by q and letting q tend to infinity gives

$$\mathsf{h}_\mu(f) \leq \mathsf{h}_{\text{top}}(f)$$

as desired.

3. Fix $\varepsilon > 0$. In this step we will find (modulo a technical detail that is postponed until Step 4) a measure $\mu \in \mathcal{M}(f)$ with

$$\mathsf{h}_\mu(f) \geq \mathsf{h}_\varepsilon^{\text{sep}}(f)^+ := \limsup_{k \to \infty} \frac{1}{k} \log \operatorname{sep}(f, k, \varepsilon) \tag{28.8}$$

Since $\varepsilon > 0$ is arbitrary, this shows that

$$\mathsf{h}_{\mathrm{top}}(f) \leq \sup_{\mu \in \mathcal{M}(f)} \mathsf{h}_\mu(f).$$

Let $A_k \subset X$ be a maximal (k, ε)-separated set, and let ν_k denote the atomic measure concentrated uniformly on the points of A_k:

$$\nu_k := \frac{1}{\mathrm{sep}(f, k, \varepsilon)} \sum_{x \in A_k} \delta_x.$$

Now let

$$\mu_k := \frac{1}{k} \sum_{i=0}^{k-1} f_*^i \nu_k.$$

By compactness of $\mathcal{M}(X)$ and the definition of the limit supremum, we can choose a sequence $k_j \to \infty$ such that

$$\lim_{j \to \infty} \frac{1}{k_j} \mathrm{sep}(f, k, \varepsilon) = \mathsf{h}_\varepsilon^{\mathrm{sep}}(f)^+$$

and such that $\mu_{k_j} \rightharpoonup \mu$ for some $\mu \in \mathcal{M}(f)$ (here we are using the recipe for constructing fixed points from the proof of the Markov-Kakutani Fixed Point Theorem 24.6.

Now by Lemma 28.6, let $\xi = \{C_1, \ldots, C_p\}$ denote a partition of X of sets of diameter smaller than ε such that $\mu(\partial C_i) = 0$ for each i. We claim that

$$\mathsf{H}_{\nu_k}\left(\xi_f^k\right) = \log \mathrm{sep}(f, k, \varepsilon). \tag{28.9}$$

Indeed, no member of ξ_f^k can contains more than one member of A_k, and so $\mathrm{sep}(f, k, \varepsilon)$ members of ξ_f^k each have ν_k-measure $\frac{1}{\mathrm{sep}(f,k,\varepsilon)}$, and all the others have measure zero. Thus (28.9) follows directly from the definition of H.

Now we would like to replace H_{ν_k} with H_{μ_k}. Fix $1 < q < k$. We claim that

$$q \log \mathrm{sep}(f, k, \varepsilon) \leq \sum_{i=0}^{k-1} \mathsf{H}_{f_i^* \nu_k}(\xi_f^q) + 2q^2 \log p. \tag{28.10}$$

Let us assume (28.10) for the time being. Using (28.3) and dividing through by k, we obtain

$$\frac{q}{k} \mathrm{sep}(f, k, \varepsilon) \leq \mathsf{H}_{\mu_k}(\xi_f^q) + \frac{2q^2}{k} \log p. \tag{28.11}$$

Now by Lemma 28.7, for each member $A \in \xi_f^q$, we have $\mu(A) = 0$. Thus by Proposition 23.15. we have $\mu_{k_i}(A) \to \mu(A)$ for each $A \in \xi_f^q$. Thus

$$\lim_{j \to \infty} \mathsf{H}_{\mu_{k_j}}(\xi_f^q) = \mathsf{H}_\mu(\xi_f^q).$$

Now by replacing k by k_j in (28.11) and letting $j \to \infty$ we obtain

$$q \mathsf{h}_\varepsilon^{\mathrm{sep}}(f)^+ \leq \mathsf{H}_\mu(\xi_f^q).$$

Finally, dividing by q and letting $q \to \infty$ we obtain (28.8).

4. It remains to prove (28.10). This is rather tedious, unfortunately. Fix $0 \le l \le q-1$, and set

$$c(l) := \left\lfloor \frac{k-l}{q} \right\rfloor .$$

We can partition the set $\{0, 1, 2, \dots, k-1\}$ into two sets: $A_l \cup B_l$, where A_l denotes those numbers that can be written as $l + aq + b$ for some $0 \le a \le c(l) - 1$ and some $0 \le b \le q-1$, and B_l denotes those numbers which cannot be written in this form. Thus B_l certainly contains $\{0, 1, \dots, q-1\}$, but it is not too large: we leave it as an easy exercise to check that

$$\# B_l \le 2q. \tag{28.12}$$

What is the point of this? Let us apply this decomposition of $\{0, 1, \dots, k-1\}$ to write

$$\xi_f^k = \left(\bigvee_{a=0}^{c(l)-1} f^{-(aq+l)} \xi_f^q \right) \vee \bigvee_{s \in B_l} f^{-s} \xi.$$

Thus by part (vii) of Proposition 26.11, we have

$$\begin{aligned} \mathsf{H}_{\nu_k}(\xi_f^k) &\le \sum_{a=0}^{c(l)-1} \mathsf{H}_{\nu_k}(f^{-(aq+l)} \xi_f^q) + \sum_{s \in B_l} \mathsf{H}_{\nu_k}(f^{-s} \xi) \\ &\le \sum_{a=0}^{c(l)-1} \mathsf{H}_{f_*^{aq+l} \nu_k}(\xi_f^q) + 2q \log p, \end{aligned}$$

where the last inequality used (28.12) and that $f^{-s}\xi$ has p elements (cf. part (i) of Proposition 26.11). Now sum this last equation over l from 0 to $q-1$ to obtain

$$q \mathsf{H}_{\nu_k}(\xi_f^k) \le \sum_{i=0}^{k-1} \mathsf{H}_{f_*^i \nu_k}(\xi_f^q) + 2q^2 \log p.$$

Since $\log \operatorname{sep}(f, k, \varepsilon) = \mathsf{H}_{\nu_k}(\xi_f^k)$ by (28.9), from this (28.10) follows. This completes this final step, and hence also the proof of the Variational Principle. ■

. . . and this also completes the course. Thank you all for attending, and enjoy your winter vacation!

LECTURE 29

Hyperbolic Linear Dynamical Systems

Welcome to Dynamical Systems II!

We began Dynamical Systems I by studying topological dynamical systems, i.e. continuous maps on metric spaces. In Dynamical Systems II we will upgrade these to **differentiable dynamical systems**.

Why study differential dynamics? As you no doubt had drilled into you during your first analysis course, the dual concepts of being able to measure the rate at which things change—*differentiation*—and being able to measure the total amount accumulated over time—*integration*—are arguably the most important mathematical ideas we humans have happened upon.

It should therefore not surprise you to learn that there is a rich and exciting subfield of dynamical systems concerned with differentiable systems. In this course will focus on a particularly interesting class of such systems, namely, *hyperbolic* differentiable dynamical systems. Hyperbolicity was first identified by Poincaré (130 years ago) during his work on the famous *Three Body Problem*. The precise definition is rather complicated (we will study a very special case later in this lecture), but very roughly speaking, hyperbolicity can be characterised by the presence of both expanding and contracting directions for the derivative (think saddle points).

Hyperbolicity is the main mechanism through which a differential dynamical system exhibits chaotic behaviour. It also has a remarkable "persistence" property: namely, if a given system is hyperbolic then so are all sufficiently "nearby" systems. We will see many instances of this in the course, starting with Proposition 30.14 next lecture.

Before we can begin with the study of hyperbolic dynamics, however, we first need to address differentiable dynamics in general. Unfortunately the process of moving from topological dynamics to differentiable dynamics is not entirely straightforward. The first thing to realise is that the definition of differentiability doesn't even *make sense* on an arbitrary metric space! Indeed, if $f\colon X \to X$ is a continuous map on a metric space then you might naively try to define the derivative of f using the same formula you learnt in high school:

$$f'(x) \overset{\text{nonsense}}{:=} \lim_{y \to 0} \frac{f(x+y) - f(x)}{y}. \tag{29.1}$$

However this is nonsense. Why?

- On a metric space one cannot simply "add" points together, so neither "$x+y$" nor "$f(x+y) - f(x)$" are defined.

- In a metric space there is no "zero element", and thus the expression "$y \to 0$" has no meaning.

The solution to these woes is to restrict our attention to special class of metric spaces where these expressions can be meaningfully interpreted. Such spaces are known as **(smooth) manifolds**, and they are commonly studied in Differential Geometry. If you have not taken a class on Differential Geometry before then **fear not**—we will cover all the necessary prerequisites from scratch in this course.

Before diving into manifolds, however, we will spend the first seven lectures focusing on an even more restrictive class of metric spaces for which you all already know how differentiation works: **vector spaces**.

Let us begin.

> CONVENTION: Throughout this course, all vector spaces are assumed to be real and finite-dimensional, unless stated otherwise.

DEFINITION 29.1. A **normed vector space** consists of a vector space E and a choice of norm $\| \cdot \|$ on E.

Since E is finite-dimensional, E is automatically a Banach space, i.e. the associated metric $d(v, w) := \|v - w\|$ is complete (Definition 6.7).

REMARK 29.2. In fact, everything in this section continues to make sense if we drop the assumption that E is finite-dimensional, and assume that E is a **Banach space** (although some of the proofs would need tweaking a bit). However since this course does not assume functional analysis as a prerequisite, we will (almost always) work with finite-dimensional spaces only. Nevertheless, the interested reader is invited to recast all the definition and theorems in a Banach space setting.

Recall that two norms $\| \cdot \|_1$ and $\| \cdot \|_2$ on E are called **equivalent** if there exists a constant $c \geq 0$ such that for all $v \in E$, one has

$$\frac{1}{c}\|v\|_1 \leq \|v\|_2 \leq c\|v\|_1.$$

In fact, since E is finite-dimensional[1], *all* norms are equivalent.

DEFINITION 29.3. Let $(E, \| \cdot \|)$ denote a finite-dimensional normed vector space and suppose $L\colon E \to E$ is a **linear** map. We will call L a **linear dynamical system**. Note $0 \in E$ is a automatically a fixed point of L.

NOTATION. Given a norm on E. we will use the notation $\| \cdot \|^{\mathrm{op}}$ to denote the operator norm of L with respect to the norm $\| \cdot \|$. Thus

$$\|L\|^{\mathrm{op}} := \sup \{\|Lv\| \mid \|v\| = 1\}.$$

[1] In the infinite-dimensional case one needs to make sure one only works with norms equivalent to our original one.

Definition 29.4. Let L be a reversible linear dynamical system on E. We say that L is a **hyperbolic linear dynamical system** if E splits into a direct sum

$$E = E^s \oplus E^u$$

which is L-invariant in the sense that

$$LE^s \subseteq E^s, \qquad LE^u \subseteq E^u,$$

and such that there exist constants $C \geq 1$ and $0 < \mu < 1$ such that

$$\|L^k v\| \leq C\mu^k \|v\|, \qquad \forall\, v \in E^s,\ \forall\, k \geq 0,$$

and such that

$$\|L^{-k} v\| \leq C\mu^k \|v\|, \qquad \forall\, v \in E^u,\ \forall\, k \geq 0,$$

or in terms of operator norms:

$$\left\|(L|_{E^s})^k\right\|^{\mathrm{op}} \leq C\mu^k, \qquad \left\|(L^{-1}|_{E^u})^k\right\|^{\mathrm{op}} \leq C\mu^k.$$

One calls E^s the **stable space** for L and one calls E^u the **unstable space**.

Remark 29.5. The connection between this definition and the definition of hyperbolicity we saw in Definition 8.12 will become clear soon; see Proposition 29.7. (Spoiler: The two definitions are the same.)

Since all norms are equivalent on E, the definition of hyperbolicity is independent of the choice of norm $\|\cdot\|$ used. Note that it could be the case that $E^s = \{0\}$ or $E^u = \{0\}$.

Remark 29.6. Since E^s is L-invariant, the inequalities in the definition hold not only for positive iterates but actually for all iterates:

$$\|L^k(L^n v)\| \leq C\mu^k \|L^n v\|, \qquad \forall\, v \in E^s,\ \forall\, n \in \mathbb{Z},\ \forall\, k \geq 0.$$

In particular, taking $n = -k$ gives

$$\|L^{-k} v\| \geq \frac{1}{C\mu^k} \|v\|, \qquad \forall\, v \in E^s,\ \forall\, k \geq 0.$$

A similar statement holds for E^u. Moreover as $0 < \mu < 1$, we see that $\|L^{-k} v\| \to \infty$ as $k \to \infty$ for $v \in E^s$, and $\|L^k v\| \to \infty$ as $k \to \infty$ for $v \in E^u$.

On Problem Sheet O you will prove the following equivalent definition of hyperbolicity.

Proposition 29.7. *A reversible linear dynamical system $L\colon E \to E$ is hyperbolic if and only if every eigenvalue λ of L has absolute value different to* 1.

As a corollary, we obtain:

Corollary 29.8. *The space of hyperbolic linear dynamical systems is an open subset in the space of reversible linear dynamical systems.*

We will prove shortly that the hyperbolic splitting is unique. First let us introduce some more notation. Suppose $E = E^s \oplus E^u$ is a direct sum (not necessarily a hyperbolic splitting). Denote $\pi_s \colon E \to E^s$ the projection onto E^s and $\pi_u \colon E \to E^u$ the projection onto E^u. We set

$$v_s = \pi_s v, \qquad v_u = \pi_u v.$$

If $\phi \colon E \to E$ is any map, we denote by $\phi_s := \pi_s \circ \phi \colon E \to E^s$ and $\phi_u := \pi_u \circ \phi \colon E \to E^u$. If $L \colon E \to E$ preserves the splitting $E = E^s \oplus E^u$, it makes sense to define

$$L_{ss} := L_s|_{E^s} \colon E^s \to E^s,$$

and similarly

$$L_{uu} := L_u|_{E^u} \colon E^u \to E^u.$$

Then for any $v \in E$,

$$L_s v = L_s v_s = L_{ss} v_s, \qquad L_u v = L_u v_u = L_{uu} v_u.$$

Definition 29.9. Suppose $L \colon E \to E$ is a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$. Given $\varepsilon > 0$, we define the **ε-cones** about E^s and E^u by:

$$\operatorname{cone}_\varepsilon(E^s) := \{ v \in E \mid \|v_u\| \le \varepsilon \|v_s\| \},$$

and

$$\operatorname{cone}_\varepsilon(E^u) := \{ v \in E \mid \|v_s\| \le \varepsilon \|v_u\| \},$$

Proposition 29.10. *Suppose $L \colon E \to E$ is a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$. Then E^s can be alternatively characterised as:*

$$\begin{aligned} E^s &= \left\{ v \in E \mid L^k v \to 0 \text{ as } k \to \infty \right\} \\ &= \left\{ v \in E \mid \exists\, r > 0,\ \|L^k v\| \le r,\ \forall\, k \ge 0 \right\} \\ &= \left\{ v \in E \mid \exists\, \varepsilon > 0,\ L^k v \in \operatorname{cone}_\varepsilon(E^s),\ \forall\, k \ge 0 \right\}. \end{aligned}$$

Similarly

$$\begin{aligned} E^u &= \left\{ v \in E \mid L^{-k} v \to 0 \text{ as } k \to \infty \right\} \\ &= \left\{ v \in E \mid \exists\, r > 0,\ \|L^{-k} v\| \le r \text{ for all } k \ge 0 \right\} \\ &= \left\{ v \in E \mid \exists\, \varepsilon > 0,\ L^{-k} v \in \operatorname{cone}_\varepsilon(E^u) \text{ for all } k \ge 0 \right\}. \end{aligned}$$

In particular, the hyperbolic splitting is unique: if $E = F^s \oplus F^u$ is another hyperbolic splitting for L then $E^s = F^s$ and $E^u = F^u$.

Proof. We give the proof for E^s only. Of the three sets on the right-hand side, it is clear that E^s is contained in the first subset, and that the first subset is contained in the second. Let us prove that the second subset is contained in the third. Indeed, suppose

$$u \notin \left\{ v \in E \mid \exists\, \varepsilon > 0,\ L^k v \in \operatorname{cone}_\varepsilon(E^s) \text{ for all } k \ge 0 \right\}$$

This means there exists $k_0 \geq 0$ such that $w := L^{k_0}u$ satisfies $w \notin \text{cone}_1(E^s)$. In particular, $w_u \neq 0$. Then we have $\|L^k w_u\| \to \infty$ and $\|L^k w_s\| \to 0$ as $k \to \infty$ (cf. Remark 29.6) . This means that

$$\|L^k w\| \geq \|L^k w_u\| - \|L^k w_s\| \to \infty,$$

Thus the sequence $\left(\|L^k w\|\right)_k$ is unbounded, and therefore w does not belong to the second set on the right-hand side.

To complete the proof, we show that the third set on the right-hand side is contained in E^s. Indeed, if $v \notin E^s$ then $v_u \neq 0$, and hence $\|L^k v_u\| \to \infty$ and $\|L^k v_s\| \to 0$ as $k \to \infty$ by Remark 29.6 again, which implies that for any $\varepsilon > 0$ one has $L^k v \notin \text{cone}_\varepsilon(E^s)$ for k large enough.

Since both the first and second sets on the right-hand side are manifestly independent of the splitting $E^s \oplus E^u$, it follows that E^s is unique. This completes the proof. ■

Up to changing the norm, one can assume that $C = 1$.

Proposition 29.11. *Let $L \colon E \to E$ be a hyperbolic linear dynamical system with hyperbolic splitting $E = E^s \oplus E^u$. There exists a norm $\|\cdot\|_{\mathrm{a}}$ on E and a constant $0 < \tau < 1$ such that*

$$\begin{aligned} \|Lv\|_{\mathrm{a}} &\leq \tau \|v\|_{\mathrm{a}}, \qquad \forall\, v \in E^s, \\ \|L^{-1}v\|_{\mathrm{a}} &\leq \tau \|v\|_{\mathrm{a}}, \qquad \forall\, v \in E^u, \end{aligned}$$

One calls such a norm $\|\cdot\|_{\mathrm{a}}$ an **adapted** *norm for L.*

Proof. Denote the original norm by $\|\cdot\|$, and let $C \geq 1$ and $0 < \mu < 1$ denote the original constants. Choose n large enough so that $C\mu^n < 1$. We define $\|\cdot\|_{\mathrm{a}}$ by

$$\|v\|_{\mathrm{a}} := \sum_{k=0}^{n-1} \|L^k v\|.$$

Then $\|\cdot\|_{\mathrm{a}}$ is obviously a norm on E. Setting $\alpha := \sum_{k=0}^{n-1} C\mu^k$, one has

$$\|v\|_{\mathrm{a}} \leq \alpha \|v\|, \qquad \forall\, v \in E^s,$$

and similarly

$$\|v\|_{\mathrm{a}} \leq \alpha \|L^{n-1}v\|, \qquad \forall\, v \in E^u.$$

Now suppose $v \in E^s$. Then

$$\begin{aligned} \|Lv\|_{\mathrm{a}} &= \|v\|_{\mathrm{a}} - \|v\| + \|L^n v\| \\ &\leq \|v\|_{\mathrm{a}} - (1 - C\mu^n)\, \|v\| \\ &\leq \left(1 - \frac{1}{\alpha}(1 - C\mu^n)\right) \|v\|_{\mathrm{a}}. \end{aligned}$$

Similarly if $v \in E^u$ one has

$$\begin{aligned} \|L^{-1}v\|_{\mathrm{a}} &= \|v\|_{\mathrm{a}} + \|L^{-1}v\| - \|L^{n-1}v\| \\ &\leq \|v\|_{\mathrm{a}} - (1 - C\mu^n)\, \|L^{n-1}v\| \\ &\leq \left(1 - \frac{1}{\alpha}(1 - C\mu^n)\right) \|v\|_{\mathrm{a}}. \end{aligned}$$

Set

$$\tau := \left(1 - \frac{1}{\alpha}\left(1 - C\mu^n\right)\right).$$

Since $\alpha \geq 1$ one has $0 < \tau < 1$. This completes the proof. ■

DEFINITION 29.12. Suppose $L\colon E \to E$ is a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$. Let $\|\cdot\|_{\mathrm{a}}$ be an adapted norm for L with associated operator norm $\|\cdot\|_{\mathrm{a}}^{\mathrm{op}}$. We define the **skewness** of L with respect to this norm by

$$\tau(L) := \max\left\{\|L_s\|_{\mathrm{a}}^{\mathrm{op}}, \|(L^{-1})_u\|_{\mathrm{a}}^{\mathrm{op}}\right\} < 1.$$

In fact, it is convenient to tweak $\|\cdot\|_{\mathrm{a}}$ even further. For this we need another more definition.

DEFINITION 29.13. Suppose $(E, \|\cdot\|)$ is a finite-dimensional normed vector space and $E = F \oplus G$ is a direct sum. We say that the norm $\|\cdot\|$ is of **box type** with respect to the splitting $F \oplus G$ if

$$\|v\| = \max\left\{\|v_F\|, \|v_G\|\right\},$$

where v_F and v_G are the components of v in this splitting. It is easy to make a box-type norm: if $\|\cdot\|$ is any norm then the function $\|\cdot\|_{\mathrm{b}}$ defined by

$$\|v\|_{\mathrm{b}} := \max\left\{\|v_F\|, \|v_G\|\right\}, \tag{29.2}$$

is another norm which is of box-type. One calls $\|\cdot\|_{\mathrm{b}}$ the **box-adjusted norm** of $\|\cdot\|$.

LEMMA 29.14. *Let $L\colon E \to E$ be a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$. There exists a norm $\|\cdot\|_{\mathrm{ab}}$ on E which is adapted to L and of box-type with respect to the hyperbolic splitting.*

Proof. Let $\|\cdot\|_{\mathrm{a}}$ denote an adapted norm for L, and then let $\|\cdot\|_{\mathrm{ab}}$ denote the box-adjusted norm defined by (29.2). Then $\|\cdot\|_{\mathrm{ab}}$ is another adapted norm with the same skewness as $\|\cdot\|_{\mathrm{ab}}$. ■

In the future, we will typically assume from the outset when discussing hyperbolic linear dynamical systems L that the norm $\|\cdot\|$ on E is both adapted to L and of box-type with respect to the hyperbolic splitting. In other words, we will omit the subscript "ab", and assume that the construction from Lemma 29.14 has already been performed.

LECTURE 30

Hyperbolic Fixed Points

In this lecture we continue to work on a finite-dimensional normed vector space $(E, \|\cdot\|)$, but we move outside the realm of linear dynamical systems. Again, all of the following continues to hold with only minor modifications if E is an arbitrary Banach space. We begin with some preliminaries on differentiable maps.

Suppose $\big(E, \|\cdot\|_E\big)$ and $\big(F, \|\cdot\|_F\big)$ are two normed vector spaces. Let $\mathcal{L}(E,F)$ denote the set of linear maps $L\colon E \to F$. Then $\mathcal{L}(E,F)$ is a vector space of dimension

$$\dim \mathcal{L}(E,F) = \dim E \cdot \dim F.$$

We make $\mathcal{L}(E,F)$ into a normed vector space via the operator norm:

$$\|L\|^{\mathrm{op}}_{E,F} := \sup\big\{\|Lv\|_F \mid \|v\|_E = 1\big\}.$$

When the norms on E and F are clear from the context, we will simply write $\|L\|^{\mathrm{op}}$.

> CONVENTION: In this lecture and beyond, the symbol Ω is reserved for an *open* subset of E, even if this is not explicitly said.

DEFINITION 30.1. Suppose $f\colon \Omega \subseteq E \to F$ is a continuous map. We say that f is **differentiable** at $u \in \Omega$ if there exists a **linear** map $L \in \mathcal{L}(E,F)$ such that

$$\lim_{\|v\|_E \to 0} \frac{\|f(u+v) - f(u) - Lv\|_F}{\|v\|_E} = 0.$$

Denoting L by $Df(u)$, if f is differentiable at every point $u \in \Omega$ then the **differential** of f is the map $\Omega \to \mathcal{L}(E,F)$ given by $u \mapsto Df(u)$. We say that f is **continuously differentiable** or **of class** C^1 if the map $Df\colon \Omega \to \mathcal{L}(E,F)$ is continuous with respect to the norms $\|\cdot\|_E$ and $\|\cdot\|^{\mathrm{op}}_{E,F}$ respectively[1].

EXAMPLE 30.2. If $L \in \mathcal{L}(E,F)$ then L is differentiable with $DL(v) = L$ for all $v \in E$.

DEFINITION 30.3. We say that a C^1 map f is a **diffeomorphism onto its image** if $f(\Omega) := \Omega'$ is an open subset of F, and there exists another C^1 map $g\colon \Omega' \to E$ with image $g(\Omega') = \Omega$, and such that $f \circ g = g \circ f = \mathrm{id}$.

Although we won't need them in this section, let us just quickly recall how one defines the higher derivatives[2].

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]Actually since all norms are equivalent on finite-dimensional spaces, it doesn't matter which norm we choose when checking whether Df is continuous.

[2]We will need this when we talk about manifolds later.

Definition 30.4. Given $p \geq 1$ we define $\mathcal{L}^p(E, F)$ to be the space of all **multilinear** maps

$$M : \underbrace{E \times \cdots \times E}_{p \text{ times}} \to F.$$

This is once again a normed finite-dimensional vector space, where this time the norm is given by:

$$\|M\|^{\mathrm{op},p}_{E,F} := \sup \big\{ \|M(v_1, \dots, v_k)\|_F \mid \|v_i\|_E = 1 \text{ for all } i = 1, \dots, p \big\}.$$

There is a natural linear isomorphism from $\mathcal{L}(E, \mathcal{L}(E, F))$ to $\mathcal{L}^2(E, F)$. Namely, given $L \in \mathcal{L}(E, \mathcal{L}(E, F))$ we define $M \in \mathcal{L}^2(E, F)$ via

$$M(u, v) := L(u)(v), \tag{30.1}$$

The same thing works for all higher p too: thus

$$\mathcal{L}^3(E, F) \cong \mathcal{L}(E, \mathcal{L}(E, \mathcal{L}(E, F))).$$

Definition 30.5. Suppose $f : \Omega \subseteq E \to F$ is a continuously differentiable map. Then we can ask whether the map $Df : \Omega \to \mathcal{L}(E, F)$ is itself differentiable. If it is, then its derivative, $D^2 f := D(Df)$ is a map

$$D^2 f : \Omega \to \mathcal{L}(E, \mathcal{L}(E, F)).$$

Using the identification (30.1), we usually regard $D^2 f$ as a map

$$D^2 f : \Omega \to \mathcal{L}^2(E, F).$$

If $D^2 f$ exists and is continuous, we say that f is of **class** C^2.

The other higher-order derivatives are defined inductively: we say f is of class C^p for $p \geq 2$ if f is of class C^{p-1}, and the map $D^{p-1} f : \Omega \to \mathcal{L}^{p-1}(E, F)$ is differentiable, and its derivative $D^p f := D(D^{p-1} f)$ is itself continuous. Finally we say f is **smooth**, or of class C^∞, if f is of class C^p for every $p \in \mathbb{N}$.

Remark 30.6. In differential geometry, the word "diffeomorphism" is usually reserved for a smooth map with a smooth inverse, contrary to Definition 30.3. In dynamical systems however it is important to keep track of the regularity, and typically C^1 is sufficient (the Denjoy Theorem 18.5 is an example of this).

Hopefully you are all familiar with the following foundational theorem.

Theorem 30.7 (The Inverse Function Theorem). *Suppose $\Omega \subseteq E$ is an open set and $f : \Omega \to F$ is a continuously differentiable map. Suppose $u \in \Omega$ has the property that $Df(u)$ is invertible*[3]*. Then f is locally a diffeomorphism onto its image near u. That is, there exists an open set*[4] *$\Omega_0 \Subset \Omega$ such that $u \in \Omega_0$ and $f|_{\Omega_0} : \Omega_0 \to E$ is a diffeomorphism onto its image.*

[3]Note that $Df(u)$ can only be invertible when $\dim E = \dim F$.

[4]The notation $\Omega_0 \Subset \Omega$ means that $\Omega_0 \subset \Omega$ and $\overline{\Omega}_0$ is compact.

Just as with topological dynamics and measure-theoretic dynamics, we will primarily be interested in the case $E = F$. Thus let us make the following definition.

DEFINITION 30.8. Let $(E, \|\cdot\|)$ be a normed vector space. A **local differentiable dynamical system** is a pair (f, Ω), where $\Omega \subseteq E$ is an open set, and $f\colon \Omega \to E$ is a diffeomorphism onto its image.

Just as we did previously with topological and measure-theoretic dynamics, when there is no danger of confusion, we will omit the adjectives and simply refer to $f\colon \Omega \to E$ as a **dynamical system**.

REMARK 30.9. **Why the word "local"?** Definition 30.3 is slightly at odds with our previous definitions of "dynamical systems", since we do not require f to be defined on the entire space E, merely on some open set Ω. The reason for this is twofold:

- Firstly, the Inverse Function Theorem 30.7 tells us that if $f\colon \Omega \to E$ is a continuously differentiable map, then up to shrinking Ω, we can always assume that f is a diffeomorphism onto its image. Thus one could think of a local differentiable dynamical system as being obtained in the following manner: start with an arbitrary continuously differentiable map f, defined on some open set of E. Find u in the domain of f so that $Df(u)$ is invertible. Then shrink the domain of f appropriately.

- The second reason will only make sense if you already know basic differential geometry (if you don't, don't worry—we will cover this material from scratch). Later on in the course, we will define a *differentiable dynamical system* to be diffeomorphism $f\colon M \to M$, where M is a smooth manifold. The local representation of such a map (i.e. when viewed in charts on M) is then a local differentiable dynamical system in the sense of Definition 30.3.

If $f\colon \Omega_1 \to E$ and $g\colon \Omega_2 \to E$ are two dynamical systems with $\Omega_1 \cap \Omega_2 \neq \emptyset$, then given any $u \in \Omega_1 \cap \Omega_2$ we can find $\Omega \Subset \Omega_1 \cap \Omega_2$ such that $u \in \Omega$ and both $f|_\Omega, g|_\Omega\colon \Omega \to E$ are dynamical systems. Thus when comparing two dynamical systems, we can without loss of generality assume that they are defined on the same open set.

DEFINITION 30.10. Suppose $f, g\colon \Omega \to E$ are two dynamical systems (defined on the same open set Ω). We define the **C^1-distance** between them to be

$$d_1(f,g) := \sup_{u\in\Omega} \max\left\{\|f(u) - g(u)\|, \|Df(u) - Dg(u)\|^{\mathrm{op}}\right\}. \tag{30.2}$$

Note the first term on the right-hand side of (30.2) is the norm on F, whereas the second norm is the operator norm of the linear operator $Df(u) - Dg(u)$.

If $f\colon \Omega \to E$ is a dynamical system, we denote by $\mathcal{B}_1(f, \Omega, \varepsilon)$ the set of dynamical systems g (with the same domain Ω) which satisfy $d_1(f,g) < \varepsilon$. Thus $\mathcal{B}_1(f, \Omega, \varepsilon)$ is the "open ball" about[5] f of radius ε in the space of local differentiable dynamical

[5]This is a slight abuse of language (hence the quotation marks), as this is not a true open call in the space of all local differentiable dynamical systems on E, since we are requiring the domain g of f to be same as Ω.

systems on E. Similarly $\overline{\mathcal{B}}_1(f, \Omega, \varepsilon)$ denotes the closed ball.

We need to quote one more result before we can get started on the meat of this lecture. As with the Inverse Function Theorem 30.7 this result should be familiar to all of you (possibly formulated slightly differently).

THEOREM 30.11 (The Mean Value Theorem). *Let $\Omega \subseteq E$ be a convex open set, and suppose $f\colon \Omega \to E$ is a C^1 map such that $\|Df(u)\|^{\mathrm{op}} \le C$ for all $u \in \Omega$. Then for any $v, w \in \Omega$,*

$$\|f(v) - f(w)\| \le C\|v - w\|.$$

Recall from Definition 11.11 that a map $\phi\colon \Omega \to E$ is called **Lipschitz** if there exists $\lambda \ge 0$ such that for all $v, w \in E$, one has

$$\|\phi(v) - \phi(w)\| \le \lambda\|v - w\|.$$

The minimal such λ is called the **Lipschitz constant** of ϕ and is denoted by $\operatorname{lip}(\phi)$. Any linear map is Lipschitz. More generally, the Mean Value Theorem 30.11 tells us that any continuously differentiable map defined on a convex open set with compact closure is automatically Lipschitz.

Let us finally prove something:

PROPOSITION 30.12. *Suppose $f\colon \Omega \to E$ is a dynamical system. Fix $u \in \Omega$. Then for any $\varepsilon > 0$ there exists $r > 0$ such that if $g\colon \Omega \to E$ is a dynamical system with $d_1(f, g) \le \frac{\varepsilon}{2}$ then*

$$\operatorname{lip}(g - Df(u)) \le \varepsilon \qquad \text{on } \overline{B}(u, r).$$

Here $\overline{B}(u, r)$ denotes the closed ball in E about u of radius r.

Proof. Recall that if L is a linear map than $DL(v) = L$ for all v (cf. Example 30.2). Thus if $g\colon \Omega \to E$ is a dynamical system then for any $u, v \in \Omega$ one has

$$D(g - Df(u))(v) = Dg(v) - Df(u).$$

Fix $u \in \Omega$. Since Df is continuous at u, there exists $r > 0$ such that

$$\|Df(u) - Df(v)\|^{\mathrm{op}} \le \frac{\varepsilon}{2}, \qquad \forall\, v \in \overline{B}(u, r).$$

Now suppose $d_1(f, g) \le \frac{\varepsilon}{2}$. Then for $v \in \overline{B}(u, r)$ we have:

$$\begin{aligned}\|D(g - Df(u))(v))\|^{\mathrm{op}} &= \|Dg(v) - Df(u)\|^{\mathrm{op}} \\ &\le \|Dg(v) - Df(v)\|^{\mathrm{op}} + \|Df(u) - Df(v)\|^{\mathrm{op}} \\ &\le d_1(f, g) + \|Df(u) - Df(v)\|^{\mathrm{op}} \\ &\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

The claim now follows from Mean Value Theorem 30.11. ∎

Let us now give the key definition of this lecture.

Definition 30.13. Suppose $f\colon \Omega \to E$ is a dynamical system, and $u \in \Omega$ is a fixed point of f. We say that u is a **hyperbolic fixed point** of f if the map $Df(u)\colon E \to E$ is a hyperbolic linear dynamical system.

In Corollary 29.8 we saw that a perturbation of a hyperbolic linear dynamical system is again another hyperbolic linear dynamical system. In other words, hyperbolicity is *persistent*. This is in fact the central theme of the course, and the reason why hyperbolic dynamics are interesting: they are stable under perturbation. The next result is a first step in this direction.

Proposition 30.14. *Let $f\colon \Omega \to E$ be a dynamical system. Suppose $u \in \Omega$ is a hyperbolic fixed point of f. Then there exists $\varepsilon_0, r_0 > 0$ such that any $g \in \overline{\mathcal{B}}_1(f, \Omega, \varepsilon_0)$ has at most one fixed point in $B(u, r_0)$. Moreover, if $0 < r < r_0$ then there exists $0 < \varepsilon(r) < \varepsilon_0$ such that any $g \in \overline{\mathcal{B}}_1(f, \Omega, \varepsilon(r))$ has at least one (and hence exactly one) fixed point in $B(u, r)$.*

Remark 30.15. Denoting the fixed point of g by u_g, Proposition 30.14 says that $u_g \to u$ as $g \to f$ in the C^1 distance. Thus u_g varies continuously in g. In fact, the fixed point u_g is a *hyperbolic* fixed point of g, and moreover the hyperbolic splitting also varies continuously with g, but we will not prove this until the next lecture.

The key step in the proof of Proposition 30.14—and numerous other results in this course—is the *Banach Fixed Point Theorem*. Let us recall the statement.

Definition 30.16. Let $f\colon X \to X$ be a continuous map on a metric space (X, d). We say that f is a **strict contraction** if there exists $0 \le \alpha < 1$ such that $d(f(x), f(y)) \le \alpha\, d(x, y)$ for all $x, y \in X$.

Clearly any strict contraction has at most one fixed point. However not all strict contractions have any fixed points. The next famous theorem gives a criterion for when they do.

Theorem 30.17 (The Banach Fixed Point Theorem). *Let (X, d) be a complete non-empty metric space. Then any strict contraction $f\colon X \to X$ has a fixed point (which is necessarily unique).*

(♣) *Proof.* Fix $x_0 \in X$ and set $x_k\colon f^k(x_0)$. Since f is a strict contraction, the sequence (x_k) is Cauchy. Since X is complete, there exists $y \in X$ such that (up to a subsequence) $x_k \to y$. This y is our desired fixed point. ■

Notation. We denote by $E(r) := \overline{B}(0, r)$ the closed ball of radius r about $0 \in E$.

Proposition 30.18. *Suppose $L\colon E \to E$ is a hyperbolic linear dynamical system with hyperbolic splitting $E = E^s \oplus E^u$. Assume the norm $\|\cdot\|$ on E is both adapted with respect to L and of box-type with respect to the splitting. Let $\tau = \tau(L)$ denote the skewness of L. Suppose $\phi\colon E(r) \to E$ is a Lipschitz continuous map satisfying*

$$\operatorname{lip}(\phi) < 1 - \tau. \tag{30.3}$$

Then $L + \phi$ has at most one fixed point in $E(r)$. If in addition one has

$$\|\phi(0)\| \le \big(1 - \tau - \operatorname{lip}(\phi)\big) r, \tag{30.4}$$

then $L + \phi$ has at least one fixed point in $E(r)$. Denoting this (necessarily unique) fixed point by u_ϕ, one has

$$\|u_\phi\| < \frac{\|\phi(0)\|}{1 - \tau - \operatorname{lip}(\phi)}. \tag{30.5}$$

Proof. We wish to solve the equation

$$(L + \phi)(v) = v, \tag{30.6}$$

for $v \in E(r)$ which is equivalent to

$$L_s v + \phi_s v = v_s, \qquad L_u v + \phi_u v = v_u,$$

or equivalently

$$L_{ss} v_s + \phi_s v = v_s, \qquad L_{uu}^{-1} v_u - L_{uu}^{-1} \phi_u v = v_u.$$

Define $X \colon E(r) \to E$ by setting

$$X(v) := \big(L_{ss} v_s + \phi_s v, L_{uu}^{-1} v_u - L_{uu}^{-1} \phi_u v\big).$$

Then v solves (30.6) if and only if v is a fixed point of X. To show that $L + \phi$ has at most one fixed point, it suffices to show that X is a strict contraction. For this we argue as follows, using the fact that the norm is of box type:

$$\begin{aligned}
\|X(v) - X(w)\| &\le \max\big\{\tau\|v_s - w_s\| + \operatorname{lip}(\phi)\|v - w\|, \tau\|v_u - w_u\| + \tau \operatorname{lip}(\phi)\|v - w\|\big\} \\
&\le (\tau + \operatorname{lip}(\phi))\|v - w\| \\
&< \|v - w\|,
\end{aligned}$$

where the last line used (30.3). Now assume in addition that (30.4) holds. To show that $L + \phi$ does indeed have a unique fixed point, it suffices to show that $X(E(r)) \subseteq E(r)$, since then the Banach Fixed Point Theorem 30.17 furnishes the desired fixed point. For this we note that

$$\|X(0)\| = \big\|\big(\phi_s(0), -L_{uu}^{-1}\phi_u(0)\big)\big\| \le \|\phi(0)\|.$$

Now fix $v \in E(r)$ and argue:

$$\begin{aligned}
\|X(v)\| &\le \|X(0)\| + \|X(v) - X(0)\| \\
&\le \|\phi(0)\| + (\tau + \operatorname{lip}(\phi))\|v\| \\
&\le r.
\end{aligned}$$

This proves the existence of a unique fixed point v_ϕ of X. Moreover the calculation above tells us that

$$\|v_\phi\| \le \|\phi(0)\| + (\tau + \operatorname{lip}(\phi))\|v_\phi\|,$$

and hence

$$\|v_\phi\| \le \frac{1}{1 - \tau - \operatorname{lip}(\phi)} \|\phi(0)\|.$$

This completes the proof. ∎

We now complete the proof of Proposition 30.14.

Proof of Proposition 30.14. Without loss of generality we may assume that $u = 0$. Set $L := Df(0)$, so that L is a hyperbolic linear dynamical system on E. If the statement holds for one norm on E then it holds for any norm, and hence without loss of generality we may assume that $\|\cdot\|$ is a norm which is adapted to L and of box-type with respect to the hyperbolic splitting of E. Let $0 < \tau < 1$ denote the skewness of L and choose $\tau < \mu < 1$. By Proposition 30.12 there exists ε_0 and r_0 such that for any $g \in \overline{\mathcal{B}}_1(f, \Omega, \varepsilon_0)$, the map $\phi_g \colon = g - L \colon E(r_0) \to E$ satisfies

$$\operatorname{lip}(\phi_g) \leq \mu - \tau.$$

Then by the first statement of Proposition 30.18, $g = L + \phi_g$ has at most one fixed point in $E(r_0)$. Now let $0 < r < r_0$, and set

$$\varepsilon(r) := \min\{\varepsilon_0, (1 - \mu)r\}.$$

Then if $g \in \overline{\mathcal{B}}_1(f, \Omega, \varepsilon(r))$ one has

$$\|\phi_g(0)\| = \|g(0)\| = \|g(0) - f(0)\| \leq d_1(f, g) \leq (1 - \mu)r.$$

Thus by the second part of Proposition 30.18, such a g has a unique fixed point $u_g \in E(r_0)$. In fact, $u_g \in E(r)$, since by (30.5),

$$\|u_g\| \leq \frac{(1 - \mu)r}{1 - \tau - \operatorname{lip}(\phi_g)} \leq r.$$

This completes the proof. ∎

Persistence of Hyperbolic Fixed Points

Suppose $f\colon \Omega \subseteq E \to E$ is a dynamical system, and $u \in \Omega$ is a hyperbolic fixed point. We proved last lecture that if g is another dynamical system sufficiently close to f, then g has a unique fixed point u_g which is close to u. In this lecture we prove that u_g is actually a hyperbolic fixed point of g. Here is the precise statement.

THEOREM 31.1 (The Local Persistence Theorem). *Let $f\colon \Omega \to E$ be a dynamical system. Suppose $u \in \Omega$ is a hyperbolic fixed point of f. There exists $\varepsilon_1, r_1 > 0$ such that every $g \in \overline{\mathcal{B}}_1(f, \Omega, \varepsilon_1)$ has a unique fixed point $u_g \in \overline{B}(u, r_1)$, which is a hyperbolic fixed point for g. Moreover, u_g, as well as the subspaces $E^s(Dg(u_g))$ and $E^u(Dg(u_g))$ all vary continuously with g.*

Despite the fact that we have already proved most of Theorem 31.1 last lecture in Proposition 30.14, the remaining assertion will take us the entire lecture. Just as Proposition 30.14 depended on the linear statement of Proposition 30.12, the proof of Theorem 31.1 comes down to the following statement about hyperbolic linear dynamical systems.

PROPOSITION 31.2. *Let $L\colon E \to E$ be a hyperbolic linear dynamical system. There exists $\delta > 0$ such that if $M\colon E \to E$ is another linear map such that $\|M - L\| < \delta$ then M is also a hyperbolic linear dynamical system. Moreover the stable and unstable spaces $E^s(M)$ and $E^u(M)$ vary continuously with M.*

The only content of Proposition 31.2 is the claim that the hyperbolic splitting varies continuously in M. We emphasise that this is a *much* deeper statement than the trivial assertion of Corollary 29.8.

Proof of Theorem 31.1, assuming Proposition 31.2. Use Proposition 31.2 and argue as in the proof of Proposition 30.14. ■

The proof of Proposition 31.2 will require several preliminary results and definitions.

DEFINITION 31.3. Suppose $h\colon E \to F$ is a homeomorphism between two finite-dimensional normed vector spaces. We say that h is **bi-Lipschitz** if both h and h^{-1} are Lipschitz.

Asking both h and h^{-1} to be Lipschitz is a fairly strong requirement (see Problem O.6 for an example of this).

DEFINITION 31.4. $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ are two normed vector spaces. Given $L \in \mathcal{L}(E, F)$, the **conorm** of L is defined by

$$\operatorname{co}(L) := \inf\left\{ \|Lv\|_F \mid v \in E,\ \|v\|_E = 1 \right\},$$

where $\|\cdot\|_E$ and $\|\cdot\|_F$ are the norms on E and F respectively.

Warning: The terminology is slightly abusive, as the conorm is *not* a norm! The following statements are left as an exercise:

LEMMA 31.5. . *Let $L \in \mathcal{L}(E, F)$. Then:*

(i) *The conorm is bounded by the operator norm:*

$$0 \le \operatorname{co}(L) \le \|L\|^{\mathrm{op}}.$$

(ii) *If L is invertible then $\operatorname{co}(L) > 0$. Indeed:*

$$\operatorname{co}(L) = \frac{1}{\|L^{-1}\|^{\mathrm{op}}}. \tag{31.1}$$

(iii) *Conversely if $\dim E = \dim F$ then if $\operatorname{co}(L) > 0$ then L is invertible and (31.1) holds.*

(iv) *If $L \in \mathcal{L}(E, F)$ and $M \in \mathcal{L}(F, G)$, then*

$$\operatorname{co}(L)\operatorname{co}(M) \le \operatorname{co}(ML) \le \|ML\|^{\mathrm{op}} \le \|M\|^{\mathrm{op}}\|L\|^{\mathrm{op}}.$$

(♣) REMARK 31.6. Another way to think about the relation between the operator norm and conorm is the following. Suppose that $\dim E = \dim F = n$. Then if $L \in \mathcal{L}(E, F)$ then the operator $\sqrt{L^*L}$ is a positive semi-definite operator. Enumerate the eigenvalues of $\sqrt{L^*L}$ as

$$\sigma_1(L) \ge \sigma_2(L) \ge \cdots \ge \sigma_n(L),$$

repeated according to multiplicity. These numbers are usually called the **singular values of** L. Then

$$\|L\|^{\mathrm{op}} = \sigma_1(L), \qquad \operatorname{co}(L) = \sigma_n(L).$$

This can be understood pictorially as follows: the image $LE(1)$ of the closed unit ball $E(1) \subseteq E$ is an ellipse in F. The lengths of the two semi-axis of this ellipse are precisely the operator norm and the conorm respectively. See Figure 31.1.

We now present a version of the Inverse Function Theorem for Lipschitz maps.

THEOREM 31.7 (The Lipschitz Inverse Function Theorem). *Let $L\colon E \to F$ be a reversible linear dynamical system, and let $\phi\colon E \to F$ be Lipschitz. If*

$$\operatorname{lip}(\phi) < \operatorname{co}(L), \tag{31.2}$$

then $L + \phi$ is bi-Lipschitz and

$$\operatorname{lip}\big((L+\phi)^{-1}\big) \le \frac{1}{\operatorname{co}(L) - \operatorname{lip}(\phi)}. \tag{31.3}$$

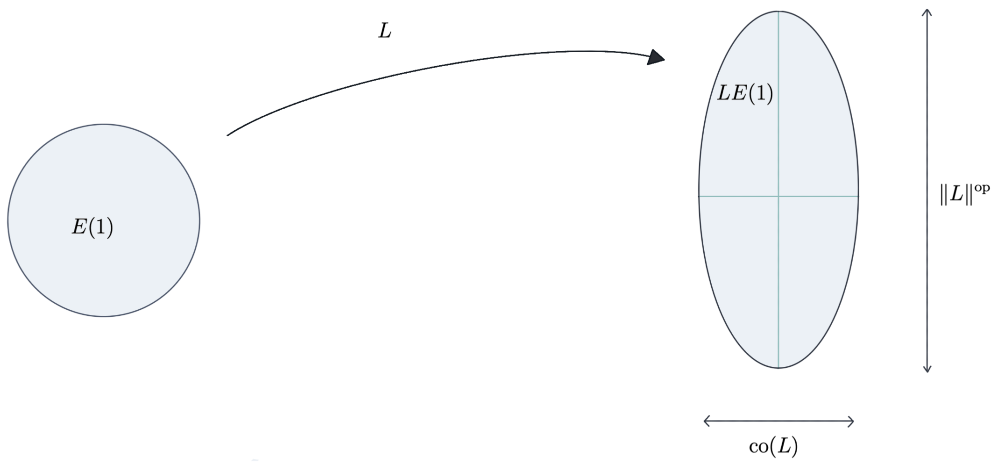

Figure 31.1: The operator norm and the conorm.

Proof. We first show that $L + \phi$ is bijective. That is, for any $w \in F$, we want to prove that there exists a unique $v \in E$ such that $(L+\phi)(v) = w$. This is equivalent to proving that the map

$$X_w \colon E \to E, \qquad X_w(v) := L^{-1}w - L^{-1}\phi(v)$$

has a unique fixed point. As in the proof of Proposition 30.18 it suffices by the Banach Fixed Point Theorem 30.17 to show that X_w is a strict contraction. For this we argue as follows:

$$\begin{aligned} \|X_w(u) - X_w(v)\|_E &\leq \|L^{-1}\phi(u) - L^{-1}\phi(v)\|_E \\ &\leq \|L^{-1}\|^{\mathrm{op}}_{F,E} \cdot \mathrm{lip}(\phi)\|u - v\|_E \\ &< \|u - v\|_E \end{aligned}$$

by the last line used (31.2).

It is clear that $L+\phi$ is Lipschitz. It remains to prove that $(L+\phi)^{-1}$ is Lipschitz, with Lipschitz constant satisfying (31.3). Given $u, v \in E$. we have

$$\begin{aligned} \|(L+\phi)(u) - (L+\phi)(v)\|_F &\geq \|L(u-v)\|_F - \|\phi(u) - \phi(v)\|_F \\ &\geq \big(\mathrm{co}(L) - \mathrm{lip}(\phi)\big)\|u - v\|_E. \end{aligned}$$

Thus writing $u = (L+\phi)^{-1}(w)$ and $v = (L+\phi)^{-1}(z)$, this gives

$$\|w - z\|_F \geq \big(\mathrm{co}(L) - \mathrm{lip}(\phi)\big)\big\|(L+\phi)^{-1}(w) - (L+\phi)^{-1}(z)\big\|,$$

which proves (31.3). This completes the proof. ∎

We will also need the following extension of the Banach Fixed Point Theorem 30.17.

Theorem 31.8 (Parametric Banach Fixed Point Theorem). *Suppose (X, d_X) and (Y, d_Y) are two metric spaces. Assume that Y is complete. Endow $X \times Y$ with the box metric*

$$d((x, y), (x', y')) := \max\big\{d_X(x, x'), d_Y(y, y')\big\}.$$

Suppose $\Phi \colon X \times Y \to Y$ is a continuous map with the property that there exists $0 < \alpha < 1$ such that

$$d_Y\big(\Phi(x, y_1), \Phi(x, y_2)\big) \le \alpha\, d_Y(y_1, y_2), \qquad \forall\,(x, y_1, y_2) \in X \times Y \times Y.$$

Then for each $x \in X$, the map $\Phi(x, \cdot) \colon Y \to Y$ has a unique fixed point. If we denote this fixed point by $\phi(x)$, then the map $\phi \colon X \to Y$ is continuous. Moreover if Φ is Lipschitz then so is ϕ.

Proof. Each map $\Phi(x, \cdot)$ is a strict contraction, and hence by the Banach Fixed Point Theorem 30.17 has a unique fixed point. Thus ϕ is well defined. Now fix $x_1, x_2 \in X$. We compute

$$\begin{aligned} d_Y\big(\phi(x_1), \phi(x_2)\big) &= d_Y\big(\Phi\big(x_1, \phi(x_1)\big), \Phi\big(x_2, \phi(x_2)\big)\big) \\ &\le d_Y\big(\Phi\big(x_1, \phi(x_1)\big)\big), \Phi\big(x_1, \phi(x_2)\big)\big) + d_Y\big(\Phi\big(x_1, \phi(x_2)\big), \Phi\big(x_2, \phi(x_2)\big)\big) \\ &\le \alpha\, d_Y\big(\phi(x_2), \phi(x_1)\big) + d_Y\big(\Phi\big(x_1, \phi(x_2)\big), \Phi\big(x_2, \phi(x_2)\big)\big), \end{aligned}$$

where we used the fact that $\phi(x_1)$ is a fixed point of $\Phi(x_1, \cdot)$ and analogously for $\phi(x_2)$. Since $0 < \alpha < 1$, we have

$$d_Y\left(\phi(x_1), \phi(x_2)\right) \le \frac{1}{1-\alpha}\, d_Y\left(\Phi\big(x_1, \phi(x_2)\big), \Phi\big(x_2, \phi(x_2)\big)\right).$$

Since $\Phi(\cdot, \phi(x_2))$ is continuous by assumption, the result follows. ∎

Notation. Given $r > 0$ we denote by

$$\mathcal{L}_r(E, F) := \{L \in \mathcal{L}(E, F) \mid \|L\|^{\mathrm{op}} \le r\}$$

the closed ball in the operator norm of radius r about the zero map (thus $\mathcal{L}_r(E, F) = \mathcal{L}(E, F)(r)$ in earlier notation.)

We now prove the main technical step needed to establish Proposition 31.2.

Proposition 31.9. *Suppose $E = F \oplus G$ is a normed vector space, endowed with a norm $\|\cdot\|_E$ which is of box-type with respect to the splitting. Let $L \colon E \to E$ be a reversible linear dynamical system, and write L in matrix form as*

$$L = \begin{pmatrix} A & B \\ C & D \end{pmatrix} : F \oplus G \to F \oplus G.$$

Suppose there exist two constants $\lambda, \varepsilon > 0$ such that

$$\lambda + \varepsilon < 1 \tag{31.4}$$

and

$$\max\left\{\|A^{-1}\|^{\mathrm{op}}_{F,F}, \|D\|^{\mathrm{op}}_{G,G}\right\} < \lambda, \tag{31.5}$$

$$\max\left\{\|B\|^{\mathrm{op}}_{G,F}, \|C\|^{\mathrm{op}}_{F,G}\right\} < \varepsilon, \tag{31.6}$$

Then there is a unique linear map $K_L \colon F \to G$ with $\|K_L\|^{\mathrm{op}}_{F,G} \le 1$ such that the linear subspace

$$\mathrm{gr}(K_L) := \{(v, K_L v) \mid v \in F\} \subset E$$

is L-invariant. Moreover for all $u \in \operatorname{gr}(K_L)$ one has

$$\|Lu\| \geq \left(\frac{1}{\lambda} - \varepsilon\right) \|u\|. \tag{31.7}$$

Finally, K_L—and hence also $\operatorname{gr}(K_L)$—depends continuously on L.

Proof. As before, we will reduce the problem to finding a fixed point for a map $X\colon \mathcal{L}_1(F,G) \to \mathcal{L}_1(F,G)$. It suffices to show there is a unique map $K \in \mathcal{L}_1(F,G)$ such that $L(\operatorname{gr}(K)) \subseteq \operatorname{gr}(K)$. Indeed, since L is invertible, such an inclusion is necessarily an equality, so

$$L(\operatorname{gr}(K)) = \operatorname{gr}(K).$$

Thus suppose $K \in \mathcal{L}_1(F,G)$ has the property that $L(\operatorname{gr}(K)) \subseteq \operatorname{gr}(K)$. Take $v \in F$. Then since

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} v \\ Kv \end{pmatrix} = \begin{pmatrix} Av + BKv \\ Cv + DKv \end{pmatrix},$$

we must have

$$K(Av + BKv) = Cv + DKv.$$

This holds for all $v \in F$ and hence

$$K(A + BK) = C + DK.$$

By (31.5) and (31.6) and the fact that $\|K_L\|^{\mathrm{op}}_{F,G} \leq 1$, we have

$$\operatorname{co}(A) \geq \frac{1}{\lambda}, \qquad \|BK\|^{\mathrm{op}}_{F,F} \leq \varepsilon,$$

and hence by the Lipschitz Inverse Function Theorem 31.7, the map $A + BK$ is invertible. Thus

$$K = (C + DK)(A + BK)^{-1},$$

and we are led to consider the map

$$X\colon \mathcal{L}_1(F,G) \to \mathcal{L}(F,G)$$

given by

$$X(K) = (C + DK)(A + BK)^{-1}. \tag{31.8}$$

We need to check that X is a contraction, and that X maps $\mathcal{L}_1(F,G)$ into itself. To see the latter point note that using (31.5) and (31.6), the Lipschitz Inverse Function Theorem tells us that

$$\|(A + BK)^{-1}\|^{\mathrm{op}}_{F,F} \leq \frac{1}{\frac{1}{\lambda} - \varepsilon},$$

and hence

$$\begin{aligned}\|X(K)\|^{\mathrm{op}}_{F,G} &\leq \|C + DK\|^{\mathrm{op}}_{F,G} \cdot \|(A + BM)^{-1}\|^{\mathrm{op}}_{F,F} \\ &\leq \frac{\lambda + \varepsilon}{\frac{1}{\lambda} - \varepsilon} < 1.\end{aligned}$$

Since for any $K \in \mathcal{L}_1(F,G)$ one has

$$X(K)(A+BK) = C + DK,$$

we have by rearranging that for any $K_1, K_2 \in \mathcal{L}_1(F,G)$,

$$X(K_1) - X(K_2) = (D - X(K_2)B)(K_1 - K_2)(A + BK_1)^{-1}.$$

Since $\|X(K)\|^{\text{op}}_{F,G} \leq 1$ we therefore have

$$\|X(K_1) - X(K_2)\|^{\text{op}}_{F,G} \leq \frac{\lambda + \varepsilon}{\frac{1}{\lambda} - \varepsilon} \|K_1 - K_2\|^{\text{op}}_{F,G},$$

which proves that X is a contraction.

Since the norm $\|\cdot\|_E$ on E is of box-type with respect to the splitting $F \oplus G$ and K has norm at most 1, the norm of a vector in $\text{gr}(K)$ is given by the first component. Since $L(v, Kv)$ belongs to $\text{gr}(K)$ we thus have

$$\|L(v, Kv)\|_E = \|Av + BKv\|_F \geq \left(\frac{1}{\lambda} - \varepsilon\right) \|v\|_F,$$

which proves (31.7).

Finally, we want to show that the map $L \mapsto K = K_L$ is continuous. Denoting by $\mathcal{S}$ the set of linear maps $L\colon E \to E$ that satisfy the hypotheses of Proposition 31.9, consider the map

$$\Phi\colon \mathcal{S} \times \mathcal{L}_1(F,G) \to \mathcal{L}_1(F,G), \qquad \Phi(L,K) = X_L(K),$$

where X_L is the map defined above for a given specific L. The map Φ is continuous with respect to both L and K, and the computation above shows that the hypotheses of the Parametric Banach Fixed Point Theorem 31.8 are satisfied, with

$$\frac{\lambda + \varepsilon}{\frac{1}{\lambda} - \varepsilon}.$$

Thus Theorem 31.8 tells us that the map $L \mapsto K_L$ is continuous, and hence the same is true of the map $L \mapsto \text{gr}(K_L) \subseteq E$. This completes the proof. ■

We conclude this lecture by proving Proposition 31.2.

Proof of Proposition 31.2. We may assume that the norm $\|\cdot\|$ on E is adapted to L and of box-type with respect to the hyperbolic splitting. Let $\tau = \tau(L)$ denote the skewness of L, so that

$$L = \begin{pmatrix} L_{ss} & 0 \\ 0 & L_{uu} \end{pmatrix}$$

with

$$\|L_{uu}^{-1}\|^{\text{op}} \leq \tau, \quad \|L_{ss}\|^{\text{op}} \leq \tau.$$

Fix $\tau < \lambda < 1$ and $0 < \varepsilon < 1 - \lambda\}$. Choose $\delta > 0$ small enough so that if $\|M - L\|^{\text{op}} \leq \delta$ then both M and M^{-1} are invertible and:

(i) M satisfies the hypotheses of Proposition 31.9 with respect to the decomposition $E = E^u \oplus E^s$ (note the order!)

(ii) M^{-1} satisfies[1] the hypotheses of Proposition 31.9 with respect to the decomposition $E = E^s \oplus E^u$ (again note the order!)

Thus by Proposition 31.9 there are two maps $K_M \in \mathcal{L}_1(E^u, E^s)$ and $K'_M \in \mathcal{L}_1(E^s, E^u)$ such that $\mathrm{gr}(K_M)$ and $\mathrm{gr}(K'_M)$ are both M-invariant, and such that $M|_{\mathrm{gr}(K_M)}$ is expanding and $M|_{\mathrm{gr}(K'_M)}$ is contracting. Therefore $\mathrm{gr}(K_M) \cap \mathrm{gr}(K'_M) = \{0\}$. Since the spaces have complementary dimensions, we see that E is the direct sum of $\mathrm{gr}(K_M)$ and $\mathrm{gr}(K'_M)$. Thus M is a hyperbolic linear dynamical system with hyperbolic splitting

$$E^s(M) \coloneqq \mathrm{gr}(K'_M), \qquad E^u(M) \coloneqq \mathrm{gr}(K_M).$$

Finally, the continuity statement follows from Proposition 31.9. This completes the proof. ∎

[1]Here we are using the fact that if $\|L - M\|^{\mathrm{op}}$ is small then $\|L^{-1} - M^{-1}\|^{\mathrm{op}}$ is also small, since

$$L^{-1} - M^{-1} = L^{-1}(M - L)M^{-1}.$$

LECTURE 32

The Hartman-Grobman Theorem

Suppose $f\colon \Omega \subseteq E \to E$ is a dynamical system, and that $u \in \Omega$ is a fixed point. As you learnt in Calculus, the differential $Df(u)$ can be thought of as the "best linear approximation" to f at u. That is, near u the function $g(v) \coloneqq f(v) - Df(u)(v-u)$ is[1] $o(v-u)$. From a dynamical point of view though, the fact that $Df(u)$ is a good linear approximation to f near u means practically nothing: in general there is no relation whatsoever between the dynamics of the non-linear system f and the linear system $Df(u)$.

In the presence of hyperbolicity however, it's another story entirely. The aim of today's lecture is to prove a celebrated theorem of Hartman and Grobman, which roughly speaking says that a dynamical system f is conjugate to its differential in a neighbourhood of a hyperbolic fixed point. To make this precise let us introduce the following notion.

Definition 32.1. Suppose $f\colon X \to X$ and $g\colon X \to X$ are dynamical systems on the same space X, and that $x \in X$ is a common fixed point of f and g. We say that f and g are **locally conjugate** at x if there exists a neighbourhood U of x and a continuous map $h\colon U \cup g(U) \to X$ which is a homeomorphism onto its image such that

$$h \circ g|_U = f \circ h|_U.$$

Note that the local conjugacy h is not a true conjugacy since neither $f|_U$ or $g|_{h(U)}$ is a dynamical system (i.e. we are not assuming that U is g-invariant or that $h(U)$ is f-invariant). Here is the statement of the Hartman-Grobman Theorem.

Theorem 32.2 (The Hartman-Grobman Theorem). *Let $f\colon \Omega \to E$ be a dynamical system. Assume that $u \in \Omega$ is a hyperbolic fixed point of f. Then f is locally conjugate to $Df(u)$ at u.*

The assumption $u = 0$ is made only to simplify the statement (since 0 is always a fixed point of $Df(0)$).

Remark 32.3. Theorem 32.2 tells us that the dynamics of a (perhaps arbitrarily complicated) map f near a hyperbolic fixed point are entirely determined by a *single* linear map. This is a remarkably strong result—it effectively reduces the study of the local hyperbolic dynamics of non-linear systems to the study of linear dynamical systems. Do not however let this fool you into thinking that hyperbolic dynamics are in any way "simple"! In fact, quite the opposite is true, as we shall see later on in the course when we explore the relationship between hyperbolicity and chaos and positive entropy.

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1] Using "little-oh" notation.

Notation. Let[2] $C^0(E)$ denote the space of continuous functions $\phi\colon E \to E$. Given $\phi \in C^0(E)$, we set

$$\|\phi\|_0 := \sup_{v\in E} \|\phi(v)\|$$

(the "0" stands for the fact that this is the C^0 norm). We denote by $C^0_b(E)$ the bounded functions:

$$C^0_b(E) := \left\{\phi \in C^0(E) \mid \|\phi\|_0 < \infty\right\}.$$

The space $\left(C^0_b(E), \|\cdot\|_0\right)$ is a Banach space.

The main step in the proof of Theorem 32.2 is the following result.

Proposition 32.4. *Let $L\colon E \to E$ be a hyperbolic linear dynamical system, and suppose $\|\cdot\|$ is a norm which is adapted to L and of box-type with respect to the hyperbolic splitting $E = E^s \oplus E^u$. Let $\tau = \tau(L)$ denote the skewness of L with respect to $\|\cdot\|$, and suppose $\phi, \psi \in C^0_b(E)$ are Lipschitz with*

$$\max\left\{\mathrm{lip}(\phi), \mathrm{lip}(\psi)\right\} < \min\left\{1-\tau, \mathrm{co}(L)\right\}. \tag{32.1}$$

Then there exists a unique $\eta \in C^0_b(E)$ such that $\mathrm{id}+\eta$ is a homeomorphism which serves as a topological conjugacy from $L+\psi$ to $L+\phi$:

$$\begin{tikzcd} E \arrow[r, "L+\phi"] \arrow[d, "\mathrm{id}+\eta"'] & E \arrow[d, "\mathrm{id}+\eta"] \\ E \arrow[r, "L+\psi"'] & E \end{tikzcd} \tag{32.2}$$

Remark 32.5. The assumption (32.1), together with the Lipschitz Inverse Function Theorem 31.7, tells us that both $L+\phi$ and $L+\psi$ are bi-Lipschitz homeomorphisms.

Proof of Proposition 32.4. Again, the proof strategy is to first reformulate the problem to finding a fixed point. For this observe that (32.2) is equivalent to

$$L + \phi + \eta(L+\phi) = L + L\eta + \psi(\mathrm{id}+\eta),$$

or

$$\phi + \eta(L+\phi) = L\eta + \psi(\mathrm{id}+\eta).$$

Writing this in terms of the splitting $E = E^s \oplus E^u$ we get

$$\phi_s + \eta_s(L+\phi) = L_{ss}\eta_s + \psi_s(\mathrm{id}+\eta),$$

and

$$\phi_u + \eta_u(L+\phi) = L_{uu}\eta_u + \psi_u(\mathrm{id}+\eta),$$

that is,

[2] We adopt slightly different notational conventions here than in Dynamical Systems I, as it will be important to keep track of the regularity.

$$\eta_s = \left(L_{ss}\eta_s + \psi_s(\mathrm{id}+\eta) - \phi_s\right)(L+\phi)^{-1},$$
$$\eta_u = L_{uu}^{-1}\left(\phi_u + \eta_u(L+\phi) - \psi_u(\mathrm{id}+\eta)\right).$$

Thus we are led to the map

$$X\colon C_b^0(E) \to C_b^0(E), \qquad X = (X_s, X_u), \tag{32.3}$$

where

$$X_s(\eta) := \big(L_{ss}\eta_s + \psi_s(\mathrm{id}+\eta) - \phi_s\big)(L+\phi)^{-1}$$
$$X_u(\eta) := L_{uu}^{-1}\big(\phi_u + \eta_u(L+\phi) - \psi_u(\mathrm{id}+\eta)\big)$$

Note that X is well defined, since as ϕ, ψ and η all belong to $C_b^0(E)$ so does[3] $X(\eta)$. We now verify that X is a strict contraction, which as usual then furnishes us a unique fixed point via the Banach Fixed Point Theorem 30.17. For this suppose $\eta, \zeta \in C_b^0(E)$. Write $X = (X_s, X_u)$. Since $\|\cdot\|$ is of box-type it suffices to estimate $\|X_s(\eta) - X_s(\zeta)\|_0$ and $\|X_u(\eta) - X_u(\zeta)\|_0$. We compute:

$$\begin{aligned}
\big\|X_s(\eta) - X_s(\zeta)\big\|_0 &= \big\|\big(L_{ss}(\eta_s - \zeta_s) + \psi_s(\mathrm{id}+\eta) - \psi_s(\mathrm{id}+\zeta)\big)(L+\phi)^{-1}\big\|_0 \\
&= \sup_{v\in E}\left\|\Big(L_{ss}(\eta_s - \zeta_s) + \psi_s(\mathrm{id}+\eta) - \psi_s(\mathrm{id}+\zeta)\Big)(L+\phi)^{-1}(v)\right\| \\
&= \sup_{w\in E}\left\|\Big(L_{ss}(\eta_s - \zeta_s) + \psi_s(\mathrm{id}+\eta) - \psi_s(\mathrm{id}+\zeta)\Big)(w)\right\| \\
&\le \sup_{w\in E}\Big(\tau\cdot\|\eta_s(w) - \zeta_s(w)\| + \mathrm{lip}(\psi)\cdot\|\eta(w) - \zeta(w)\|\Big) \\
&\le \big(\tau + \mathrm{lip}(\psi)\big)\|\eta - \zeta\|_0.
\end{aligned}$$

A similar computation shows that

$$\|X_u(\eta) - X_u(\zeta)\|_0 \le \big(\tau + \tau\,\mathrm{lip}(\psi)\big)\|\eta - \zeta\|_0. \tag{32.4}$$

Thus X is indeed a contraction, and hence has a unique fixed point η, and hence we have solved (32.2). It remains to show that $\mathrm{id}+\eta$ is a homeomorphism[4]. Interchanging ϕ and ψ, we find a unique ζ such that $\mathrm{id}+\zeta$ is a conjugacy from $L+\psi$ to $L+\phi$:

$$\begin{tikzcd}
E \arrow[r, "L+\psi"] \arrow[d, "\mathrm{id}+\zeta"'] & E \arrow[d, "\mathrm{id}+\zeta"] \\
E \arrow[r, "L+\phi"'] & E
\end{tikzcd} \tag{32.5}$$

[3]This can be seen explicitly by taking $\zeta = 0$ in (32.4).

[4]You have probably seen this argument in various other guises in your other courses. Abstractly, this comes down to a category-theoretic statement that *morphisms defined by universal properties are always isomorphisms*. Ignore this footnote.

If we splice the diagrams (32.2) and (32.5) together we obtain

$$\begin{tikzcd}
E \arrow[r, "L+\phi"] \arrow[d, "\mathrm{id}+\eta"'] & E \arrow[d, "\mathrm{id}+\eta"] \\
E \arrow[r, "L+\psi"] \arrow[d, "\mathrm{id}+\zeta"'] & E \arrow[d, "\mathrm{id}+\zeta"] \\
E \arrow[r, "L+\phi"] & E
\end{tikzcd}$$

we see that $(\mathrm{id}+\zeta)(\mathrm{id}+\eta)$ is a conjugacy from $L+\eta$ to itself:

$$\begin{tikzcd}
E \arrow[r, "L+\phi"] \arrow[d, "(\mathrm{id}+\zeta)(\mathrm{id}+\eta)"'] & E \arrow[d, "(\mathrm{id}+\zeta)(\mathrm{id}+\eta)"] \\
E \arrow[r, "L+\phi"'] & E
\end{tikzcd}$$

However another conjugacy from $L+\phi$ to itself is given by the identity id!

$$\begin{tikzcd}
E \arrow[r, "L+\phi"] \arrow[d, "\mathrm{id}"'] & E \arrow[d, "\mathrm{id}"] \\
E \arrow[r, "L+\phi"'] & E
\end{tikzcd}$$

Thus by uniqueness of such conjugacies, we must have

$$(\mathrm{id}+\zeta)(\mathrm{id}+\eta) = \mathrm{id}\,.$$

Similarly

$$(\mathrm{id}+\eta)(\mathrm{id}+\zeta) = \mathrm{id}\,.$$

This proves that $\mathrm{id}+\eta$ is a homeomorphism, and thus completes the proof. ■

(♣) Remark 32.6. Here is an alternative slicker proof of Proposition 32.4. As mentioned in Remark 29.2 there was no real need (other than to slightly simplify things) to restrict to finite-dimensional normed vector spaces in all our definitions. In particular, Proposition 30.18 is valid for hyperbolic operators on Banach spaces.

The operator $X\colon C^0_b(E) \to C^0_b(E)$ defined in the proof of Proposition 32.4 is actually itself a Lipschitz perturbation of a hyperbolic operator on the Banach space $C^0_b(E)$, and thus Proposition 32.4 is an immediate corollary of Proposition 30.18. Let us briefly explain the details.

Let $L\colon E \to E$ be a reversible linear dynamical system. Using L, we define another reversible linear dynamical system on T_L on the Banach space $C^0_b(E)$:

$$T_L\colon C^0_b(E) \to C^0_b(E), \qquad T_L(\eta) := L \circ \eta \circ L^{-1}.$$

With a bit of work, one can show that if L is hyperbolic with splitting $E = E^s \oplus E^u$ then so is T_L. The corresponding hyperbolic splitting of $C_b^0(E)$ is given by

$$C_b^0(E) = \mathcal{E}^s \oplus \mathcal{E}^u,$$

where

$$\begin{aligned} \mathcal{E}^s &= \left\{\phi \in C_b^0(E) \mid \phi(E) \subseteq E^s\right\}, \\ \mathcal{E}^u &= \left\{\phi \in C_b^0(E) \mid \phi(E) \subseteq E^u\right\}. \end{aligned}$$

Now consider a perturbed operator

$$T_{L,\phi}(\eta) := L \circ \eta \circ (L + \phi)^{-1}.$$

Provided $\operatorname{lip}(\phi)$ is small, $T_{L,\phi}$ is a Lipschitz perturbation of T_L, and hence $T_{L,\phi}$ is also hyperbolic (this follows from either Corollary 29.8 or Proposition 30.18).

Finally, the operator X from (32.3) can be written in the form

$$X = T_{L,\phi} + \xi,$$

where $\xi \in C_b^0(C_b^0(E))$ (i.e. a bounded function on the space of bounded functions). Moreover $\operatorname{lip}(\xi)$ can be estimated in terms of $\operatorname{lip}(\phi)$, and hence if $\operatorname{lip}(\phi)$ is small enough then X is a Lipschitz perturbation of the hyperbolic operator $T_{L,\phi}$. Using this, the desired fixed point could be found by applying Proposition 30.18 to $T_{L,\phi} + \xi$ on $C_b^0(E)$. See also Problem R.4, which studies the manifold version of this statement.

We now prove Theorem 32.2.

Proof of Theorem 32.2. Without loss of generality we may assume that $u = 0 \in \Omega$. If the theorem holds for one norm on E then it holds for all norms on E, so without loss of generality we may assume that the norm $\|\cdot\|$ is adapated to $Df(0)$ and of box-type with respect to the hyperbolic splitting $E = E^s \oplus E^u$. Let τ be the skewness of $Df(0)$ with respect to $\|\cdot\|$. The idea of the proof is very simple: since in a neighbourhood of 0, f is a Lipschitz small perturbation of $Df(0)$ (by the Mean Value Theorem 30.11), we hope to apply Proposition 32.4 with $L = Df(0)$, $\phi = f - Df(0)$ and $\psi = 0$.

There is a bug with this argument however: the Lipschitz constant of $f - Df(0)$ is only small near 0, and not on the entire[5] space E. Unfortunately, in the proof of Proposition 32.4 it was *essential* that we worked on the entire space E. Indeed, if we tried to prove the same result for maps defined only on a ball, then to apply the Banach Fixed Point Theorem 30.17, we would need to show that the operator X from (32.3) mapped the ball to itself—and this in general is false. So we need another idea.

Luckily, cutoff functions come to the rescue, and we proceed as follows: Choose a continuously differentiable function $\beta\colon E \to [0,1]$ such that

$$\beta(v) = \begin{cases} 1, & \|v\| \le \frac{1}{3}, \\ 0, & \|v\| \ge \frac{2}{3}, \end{cases} \qquad \text{with} \qquad |D\beta(v)| \le 3, \qquad \forall\, v \in E.$$

[5] Even worse: f is only defined on Ω, and not on all of E!

Now consider the map $\phi := f - Df(0)\colon \Omega \to E$. Then $\phi(0) = 0$ and $D\phi(0) = 0$. Choose $r_0 > 0$ small enough such that $E(r_0) \subset \Omega$. For $0 < r < r_0$ consider

$$\phi_r\colon E \to E, \qquad \phi_r(v) := \begin{cases} \beta\left(\frac{v}{r}\right) \cdot \phi(v), & \|v\| \le r_0, \\ 0, & \|v\| \ge r_0. \end{cases}$$

Then ϕ_r is a bounded continuous function on all of E, which agrees with ϕ on $E\left(\frac{r}{3}\right)$ and which vanishes outside of $E\left(\frac{2r}{3}\right)$.

We claim that $\operatorname{lip}(\phi_r) \to 0$ as $r \to 0$. By the Mean Value Theorem 30.11 it suffices to show that $\sup_{v\in E} \|D\phi_r(v)\|^{\mathrm{op}} \to 0$ as $r \to 0$. For this note that

$$\|D\phi_r(v)\|^{\mathrm{op}} \le \left|D\beta\left(\tfrac{v}{r}\right)\right| \cdot \frac{1}{r}\|\phi(v)\| + \left|\beta\left(\tfrac{v}{r}\right)\right| \|D\phi(v)\|^{\mathrm{op}}. \tag{32.6}$$

Since $\phi(0) = D\phi(0) = 0$, for any $\varepsilon > 0$ there exists $0 < r(\varepsilon) < r_0$ such that

$$\|v\| \le r(\varepsilon) \qquad \Rightarrow \qquad \|\phi(v)\| \le \frac{\varepsilon}{6}\|v\| \quad \text{and} \quad \|D\phi(v)\|^{\mathrm{op}} \le \frac{\varepsilon}{2}. \tag{32.7}$$

Since ϕ_r vanishes outside of $E\left(\frac{2r}{3}\right)$ we have

$$\sup_{v\in E} \|D\phi_r(v)\|^{\mathrm{op}} = \sup_{v\in E(r)} \|D\phi_r(v)\|^{\mathrm{op}}.$$

If $r < r(\varepsilon)$ then from (32.6) and (32.7) we obtain that for $v \in E(r)$,

$$\begin{aligned} \|D\phi_r(v)\|^{\mathrm{op}} &\le 3 \cdot \frac{1}{r} \cdot \frac{\varepsilon}{6} \cdot r + 1 \cdot \frac{\varepsilon}{2} \\ &\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This proves the claim.

In particular, if we set

$$\varepsilon = \min\left\{1 - \tau, \operatorname{co}(Df(0))\right\},$$

then for $r < r(\varepsilon)$ the function ϕ_r satisfies the hypotheses of Proposition 32.4 with $L = Df(0)$, $\phi = \phi_r$ as above and $\psi = 0$. Thus there exists a homeomorphism $h = \mathrm{id} + \eta\colon E \to E$ such that

$$h \circ (Df(0) + \phi_r) = Df(0) \circ h.$$

Setting $U := E\left(\frac{r}{3}\right)$, we have

$$h \circ f|_U = Df(0) \circ h|_U.$$

This completes the proof. ■

LECTURE 33

Stable and Unstable Manifolds

In this lecture we introduce the *stable* and *unstable* manifolds of a dynamical system.

DEFINITION 33.1. Suppose $f\colon \Omega \subseteq E \to E$ is a dynamical system, and suppose $u \in U$ is a hyperbolic fixed point of f. We define the **local stable manifold** of u with radius r to be

$$W^s_{\mathrm{loc},r}(u,f) := \left\{ v \in \Omega \,\middle|\, \|f^k(v) - u\| \le r \text{ for all } k \ge 0, \text{ and } \lim_{k\to\infty} f^k(v) = u \right\}.$$

Thus $W^s_{\mathrm{loc},r}(u,f)$ is the set of vectors $v \in \Omega$ such that $f^k(v) \in \overline{B}(u,r)$ for all $k \ge 0$, and which are eventually asymptotic to u. Similarly we define the **local unstable manifold** of u with radius r to be

$$W^u_{\mathrm{loc},r}(u,f) := \left\{ v \in \Omega \,\middle|\, \|f^{-k}(v) - u\| \le r \text{ for all } k \ge 0, \text{ and } \lim_{k\to\infty} f^{-k}(v) = u \right\}.$$

The definition also makes sense for $r = \infty$, but this has a different name.

DEFINITION 33.2. Suppose $f\colon \Omega \to E$ is dynamical system and $u \in \Omega$ is a hyperbolic fixed point of f. The **global stable manifold** of u is the set

$$W^s(u,f) := \left\{ v \in \Omega \,\middle|\, \lim_{k\to\infty} f^k(v) = u \right\},$$

and the **global unstable manifold** of u is the set

$$W^u(u,f) := \left\{ v \in \Omega \,\middle|\, \lim_{k\to\infty} f^{-k}(v) = u \right\}.$$

The local stable manifold depends on the choice of norm on E, whereas the global one does not. It is clear from the definition that

$$W^s(0,f) = \bigcup_{r>0} W^s_{\mathrm{loc},r}(0,f), \qquad W^u(0,f) = \bigcup_{r>0} W^u_{\mathrm{loc},r}(0,f). \tag{33.1}$$

Both the local and global stable manifolds are f-invariant sets:

$$f\big(W^s_{\mathrm{loc},r}(u,f)\big) \subseteq W^u_{\mathrm{loc},r}(u,f), \qquad f\big(W^s(u,f)\big) \subseteq W^s(u,f),$$

and the local and global unstable manifolds are f^{-1}-invariant sets:

$$f^{-1}\big(W^u_{\mathrm{loc},r}(u,f)\big) \subseteq W^u_{\mathrm{loc},r}(u,f), \qquad f^{-1}\big(W^u(u,f)\big) \subseteq W^u(u,f).$$

At the moment you should think of W^s and W^u as being sets. The name "stable manifold" is rather suggestive, and indeed, we will prove these are genuine manifolds next lecture[1]. In this lecture, however, we will simply regard them as subsets of E. In the linear case, the stable manifold coincides with the stable space:

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1] Amusingly enough, we will prove they are manifolds before defining the word "manifold"...

EXAMPLE 33.3. Let $L\colon E \to E$ be a hyperbolic linear dynamical system with hyperbolic splitting $E^s \oplus E^u$. Then

$$W^s(0, L) = E^s \qquad \text{and} \qquad W^u(0, L) = E^u.$$

Moreover if the norm $\|\cdot\|$ on E is adapted and of box-type with respect to the hyperbolic splitting and $E^s(r) := E(r) \cap E^s$ and $E^u(r) = E(r) \cap E^u$, then for any $0 < r < \infty$ one has

$$W^s_{\mathrm{loc},r}(0, L) = E^s(r) \qquad \text{and} \qquad W^u_{\mathrm{loc},r}(0, L) = E^u(r).$$

REMARK 33.4. In fact, these definition even make sense without the assumption that u is a fixed point. If $f\colon X \to X$ is a reversible dynamical system on a metric space, then for *any* point $x \in X$, we define the **stable manifold**

$$W^s(x, f) := \left\{ y \in X \;\Big|\; \lim_{k\to\infty} d\big(f^k(x), f^k(y)\big) = 0 \right\},$$

and the **unstable manifold**

$$W^u(x, f) := \left\{ y \in X \;\Big|\; \lim_{k\to\infty} d\big(f^{-k}(x), f^{-k}(y)\big) = 0 \right\}.$$

In this level of generality however, the name is a misnomer, since the sets $W^u(x, f)$ and $W^s(x, f)$ are typically *not* manifolds!

Let us suppose now that $u = 0 \in \Omega$ and f is of the special form

$$f = L + \phi, \qquad \text{where} \qquad \phi(0) = 0, \tag{33.2}$$

where L is a hyperbolic linear dynamical system and ϕ is a Lipschitz map satisfying with sufficiently small Lipschitz constant.

REMARK 33.5. Assumption (33.2) is not really a restriction, provided we are only interested in the local dynamics. Indeed, Proposition 30.12 tells us that any dynamical system f is of (33.2) in a small neighbourhood of a hyperbolic fixed point u, with $L = Df(u)$ and $\phi = f - Df(u)$.

We now present several alternative characterisations of the (local) stable manifolds, in a similar vein[2] to Proposition 29.10. The results are stated only for stable manifolds, but it is easy to reformulate them for the unstable manifolds.

PROPOSITION 33.6. *Let $L\colon E \to E$ be a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$ of skewness $0 < \tau < 1$ with respect to a norm $\|\cdot\|$ which is adapted to L and of box-type with respect to the splitting. Let $\phi\colon E(r) \to E$ be Lipschitz continuous map satisfying*

$$\operatorname{lip}(\phi) < 1 - \tau, \qquad \phi(0) = 0.$$

Set

$$f := L + \phi.$$

[2]Indeed, in the case $\phi = 0$, the next result reduces to Proposition 29.10, by Example 33.3.

Then the local stable manifold $W^s_{\mathrm{loc},r}(0,f)$ can be alternatively characterised as:

$$\begin{aligned} W^s_{\mathrm{loc},r}(0,f) &= \left\{ v \in E(r) \mid \|f^k(v)\| \le r \text{ for all } k \ge 0 \right\} \\ &= \left\{ v \in E(r) \mid f^k(v) \in E(r) \cap \mathrm{cone}_1(E^s) \text{ for all } k \ge 0 \right\} \\ &= \left\{ v \in E(r) \mid \|f^k(v)\| \le (\tau + \mathrm{lip}(\phi))^k \, \|v\| \text{ for all } k \ge 0 \right\}. \end{aligned}$$

Proof. First note that for any $v, w \in E(r)$, we have

$$\begin{aligned} \|f_s(v) - f_s(w)\| &= \left\| L_{ss}(v_s - w_s) + \phi_s(v) - \phi_s(w) \right\| \\ &\le (\tau + \mathrm{lip}(\phi)) \, \|v - w\|. \end{aligned} \tag{33.3}$$

Next, we claim that for any $v, w \in E(r)$, one has

$$v - w \notin \mathrm{cone}_1(E^s) \qquad \Rightarrow \qquad f(v) - f(w) \notin \mathrm{cone}_1(E^s). \tag{33.4}$$

Indeed,

$$\begin{aligned} \|f_u(v) - f_u(w)\| &= \left\| L_{uu}(v_u - w_u) + \phi_u(v) - \phi_u(w) \right\| \\ &\ge \frac{1}{\tau} \|v_u - w_u\| - \mathrm{lip}(\phi) \|v - w\|. \end{aligned}$$

Since the norm $\|\cdot\|$ is of box-type, if $v - w \notin \mathrm{cone}_1(E^s)$ then $\|v - w\| = \|v_u - w_u\|$, and hence

$$\|f_u(v) - f_u(w)\| \ge \left(\frac{1}{\tau} - \mathrm{lip}(\phi) \right) \|v - w\|. \tag{33.5}$$

Combining this with (33.3), we obtain

$$\|f_u(v) - f_u(w)\| \ge \frac{\left(\frac{1}{\tau} - \mathrm{lip}(\phi)\right)}{(\tau + \mathrm{lip}(\phi))} \|f_s(v) - f_s(w)\|.$$

Since $\tau + \mathrm{lip}(\phi) < 1$, we have $\frac{1}{\tau} - \mathrm{lip}(\phi) > 1$, and hence

$$\|f_u(v) - f_u(w)\| > \|f_s(v) - f_s(w)\|,$$

that is, $f(v) - f(w) \notin \mathrm{cone}_1(E^s)$. This proves (33.4).

We now prove the proposition. It is clear that $W^s_{\mathrm{loc},r}(0,f)$ is a subset of the first set on the right-hand side. Let us prove that the first set on the right-hand side is contained in the second. Suppose there exists u in the first set on the right-hand side, but such that u does not belong to the second set on the right-hand side. Thus $f^k(u) \in E(r)$ for all $k \ge 0$, but there exists an $n \ge 0$ such that $v := f^n(u) \notin \mathrm{cone}_1(E^s)$. By (33.5) applied with $w = 0$, we have $f(v) \notin \mathrm{cone}_1(E^s)$, and thus from (33.5) we have

$$\|f(v)\| \ge \left(\frac{1}{\tau} - \mathrm{lip}(\phi) \right) \|v\|.$$

Arguing inductively, we see that $f^k(v) \notin \mathrm{cone}_1(E^s)$ for all $k \ge 0$, and moreover

$$\|f^k(v)\| \ge \left(\frac{1}{\tau} - \mathrm{lip}(\phi) \right)^k \|v\|.$$

Since $v \notin \mathrm{cone}_1(E^s)$ we have in particular that $v \neq 0$. Thus $\big(\|f^k(u)\|\big)_{k \geq 0}$ is an unbounded sequence of real numbers, which contradicts the fact u belongs to the first set on the right-hand side.

Next, let us prove that the second set on the right-hand side is contained in the third. Assume v has the property that for any $k \geq 0$, $f^k(v) \in E(r) \cap \mathrm{cone}_1(E^s)$. Then by (33.3),

$$\|f(v)\| = \|f_s(v)\| \leq (\tau + \mathrm{lip}(\phi))\, \|v\|,$$

and thus inductively for any $k \geq 1$,

$$\|f^k(v)\| \leq (\tau + \mathrm{lip}(\phi))^k\, \|v\|.$$

Finally, it is clear that the third set on the right-hand side is contained in $W^s_{\mathrm{loc},r}(0, f)$. This completes the proof. ∎

Here is the global version.

Corollary 33.7. *Let $L\colon E \to E$ be a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$ of skewness $0 < \tau < 1$ with respect to a norm $\|\cdot\|$ which is adapted to L and of box-type with respect to the splitting. Let $\phi\colon E \to E$ be Lipschitz continuous map satisfying*

$$\mathrm{lip}(\phi) < 1 - \tau, \qquad \phi(0) = 0.$$

Set

$$f := L + \phi.$$

Then the global stable manifold $W^s(0, f)$ can be alternatively characterised as:

$$\begin{aligned} W^s(0,f) &= \big\{ v \in E \mid \exists\, r \geq 0 \text{ such that } \|f^k(v)\| \leq r \text{ for all } k \geq 0 \big\} \\ &= \big\{ v \in E \mid f^k(v) \in \mathrm{cone}_1(E^s) \text{ for all } k \geq 0 \big\} \\ &= \Big\{ v \in E \mid \|f^k(v)\| \leq (\tau + \mathrm{lip}(\phi))^k\, \|v\| \text{ for all } k \geq 0 \Big\}. \end{aligned}$$

Proof. Immediate from (33.1) and Proposition 33.6. ∎

Now let us apply Proposition 33.6 to the local stable manifold of an arbitrary dynamical system with a hyperbolic fixed point.

Proposition 33.8. *Let $f\colon \Omega \to E$ be a dynamical system and suppose $u \in \Omega$ is a hyperbolic fixed point of f. Then for r sufficiently small there exists $C \geq 1$ and $0 < \mu < 1$ such that the local stable manifold of radius r can be characterised as*

$$\begin{aligned} W^s_{\mathrm{loc},r}(u,f) &= \big\{ v \in \Omega \mid \|f^k(v) - u\| \leq r \text{ for all } k \geq 0 \big\} \\ &= \big\{ v \in \Omega \mid \|f^k(v) - u\| \leq r,\ \|f^k(v) - u\| \leq C\mu^k \|v - u\| \text{ for all } k \geq 0 \big\}. \end{aligned}$$

Similarly for r sufficiently small there exists $C \geq 1$ and $0 < \mu < 1$ such that the local unstable manifold of radius r can be characterised as

$$\begin{aligned} W^u_{\mathrm{loc},r}(u,f) &= \big\{ v \in \Omega \mid \|f^{-k}(v) - u\| \leq r \text{ for all } k \geq 0 \big\} \\ &= \big\{ v \in \Omega \mid \|f^{-k}(v) - u\| \leq r,\ \|f^{-k}(v) - u\| \leq C\mu^k \|v - u\| \text{ for all } k \geq 0 \big\}. \end{aligned}$$

Proof. We will prove the result for the stable manifold $W^s_{\mathrm{loc},r}(u,f)$ only. If the claim holds for one norm then it holds for all norms (albeit with different constants C and μ), and hence without loss of generality we may assume the norm $\|\cdot\|$ on E is adapted to $Df(u)$ and of box-type with respect to the hyperbolic splitting of E. The only non-obvious part is that the first set on the right-hand side is contained in the second set on the right-hand side for appropriate C, μ and all small r.

Without loss of generality we may assume $u = 0$. Let $0 < \tau < 1$ be the skewness of $Df(0)$ with respect to $\|\cdot\|$. Take $C = 1$ and let $\tau < \mu < 1$. By Proposition 30.12 for r sufficiently small $\phi := f - Df(0) \colon E(r) \to E$ satisfies $\operatorname{lip}(\phi) \le \mu - \tau$. Then if v satisfies $\|f^k(v)\| \le r$ for all $k \ge 0$ then

$$\|f^k(v)\| = \left\|(Df(0) + \phi)^k(v)\right\| \le (\tau + \operatorname{lip}(\phi))^k \,\|v\| \le \mu^k \|v\|,$$

by Proposition 33.6. This completes the proof. ∎

A corollary of this result is the following statement, which strengthens Theorem 31.1.

Corollary 33.9. *Let $f \colon \Omega \to E$ be a dynamical system and suppose $u \in \Omega$ is a hyperbolic fixed point of f. There exists $r > 0$ such that if $v \in \Omega$ satisfies*

$$\|f^k(v) - u\| \le r, \qquad \forall\, k \in \mathbb{Z},$$

then $v = u$. That is, for r sufficiently small, one has

$$W^s_{\mathrm{loc},r}(u,f) \cap W^u_{\mathrm{loc},r}(u,f) = \{u\}.$$

The proof of Corollary 33.9 is left for you on Problem Sheet P.

The Stable Manifold Theorem

In this lecture we will prove that the global stable manifold of a dynamical system $f = L + \phi$, where L is a hyperbolic linear dynamical system and ϕ is a Lipschitz continuous map whose Lipschitz constant is sufficiently small, has a differentiable structure. For variety this time round we will work with the unstable manifolds. One can think of the next result as the nonlinear version of Proposition 31.2 (cf. Example 33.3 and Remark 34.2 below).

THEOREM 34.1. *Suppose $L\colon E \to E$ is a hyperbolic linear dynamical system with splitting $E = E^s \oplus E^u$, and let $\|\cdot\|$ be a norm which is adapted to L and of box-type with respect to the splitting. There exists $\delta > 0$ such that if $\phi\colon E \to E$ is a Lipschitz map with*

$$\operatorname{lip}(\phi) < \delta, \qquad \phi(0) = 0,$$

then if we set

$$f := L + \phi,$$

then there is a Lipschitz continuous map $\xi\colon E^u \to E^s$ satisfying

$$\operatorname{lip}(\xi) \leq 1, \qquad \xi(0) = 0,$$

and such that

$$W^u(0, f) = \operatorname{gr}(\xi).$$

Finally, if ϕ is C^1 then so is ξ.

We will not prove the red part of Theorem 34.1. Most of the interesting ideas are contained in the Lipschitz statement, and verifying that ξ is C^1 if ϕ is as well is tedious and rather technical. Nevertheless, the consequences of the red statement are very important. Indeed, it is precisely this statement that tells us that $W^u(0, f)$ is a manifold—thus justifying the name "unstable manifold"—since the graph of a C^1 map is itself a C^1-submanifold. In fact, with a bit more work one can show that if ϕ is C^p for $p \geq 1$ then so is ξ, and hence if ϕ is smooth then $W^u(0, f)$ is a smooth embedded submanifold. If you are not happy about what this means, fear not: we will discuss manifolds in the next lecture.

Proof of Theorem 34.1. As usual, we will first reformulate the statement into a fixed point problem. Let $0 < \tau < 1$ denote the skewness of L with respect to $\|\cdot\|$. Set

$$\delta := \min\left\{\frac{1-\tau}{2}, \operatorname{co}(L)\right\},$$

and suppose $\phi\colon E \to E$ is a Lipschitz map with $\phi(0) = 0$ and $\operatorname{lip}(\phi) < \delta$. We will look for a Lipschitz continuous map $\xi\colon E^u \to E^s$ satisfying $\xi(0) = 0$ and $\operatorname{lip}(\xi) \leq 1$ such that

$$f(\operatorname{gr}(\xi)) \subseteq \operatorname{gr}(\xi). \tag{34.1}$$

Then at the end of the proof we will show that (34.1) actually implies that $\mathrm{gr}(\xi) = W^u(0, f)$. We argue in three steps.

1. In this step, we show that (34.1) can be reformulated as a fixed point problem. The condition (34.1) is equivalent to saying that for every $v \in E^u$,

$$\xi\big(f_u(v, \xi(v))\big) = f_s(v, \xi(v)).$$

Since $L_u(\xi(v)) = 0$ and $L_s v = 0$, this reduces to

$$\xi\big(L_{uu}v + \phi_u(v, \xi(v))\big) = L_{ss}\xi(v) + \phi_s(v, \xi(v)),$$

or, denoting by id_u the identity map on E^u,

$$\xi(L_{uu} + \phi_u(\mathrm{id}_u, \xi)) = L_{ss}\xi + \phi_s(\mathrm{id}_u, \xi).$$

Since we assume $\mathrm{lip}(\xi) \leq 1$, we have

$$\mathrm{lip}(\phi_u(\mathrm{id}_u, \xi)) \leq 2\,\mathrm{lip}(\phi) < 2\delta \leq 1 - \tau,$$

and since $\mathrm{co}(L_{uu}) \geq \frac{1}{\tau}$, by the Lipschitz Inverse Function Theorem 31.7 the map $L_{uu} + \phi_u(\mathrm{id}_u, \xi)$ is invertible. Thus

$$\xi = \big(L_{ss}\xi + \phi_s(\mathrm{id}_u, \xi)\big)\big(L_{uu} + \phi_u(\mathrm{id}_u, \xi)\big)^{-1}.$$

This tells us we should consider the map X defined by

$$X(\xi) := \big(L_{ss}\xi + \phi_s(\mathrm{id}_u, \xi)\big)\big(L_{uu} + \phi_u(\mathrm{id}_u, \xi)\big)^{-1}, \tag{34.2}$$

Finding a map ξ which solves (34.1) is equivalent to finding a fixed point of X.

2. In this step we aim to apply the Banach Fixed Point Theorem 30.17 to X and thus obtain our desired fixed point ξ. Note that we have not yet specified the domain of X. This requires a bit of care: choosing the domain of X in such a way that X is a contraction is the most subtle part of the proof. First, let us extend the operator norm to non-linear maps: given a continuous map $\psi\colon F \to G$ between two linear spaces, define

$$\|\psi\|^* := \sup_{v \neq 0} \frac{\|\psi(v)\|_G}{\|v\|_F}.$$

If ψ is linear, this is just the operator norm.

Now set

$$\Sigma := \left\{\xi \in C^0(E^u, E^s) \mid \xi(0) = 0,\ \|\xi\|^* < \infty\right\}.$$

The space Σ equipped with the norm $\|\cdot\|^*$ is a Banach space, as you will enjoy proving on Problem Sheet P. If $\psi \in C^0(E, E)$ is Lipschitz continuous, then clearly $\psi \in \Sigma$ with

$$\|\psi\|^* \leq \mathrm{lip}(\psi).$$

However there exist non-Lipschitz functions that belong to Σ. Given $r > 0$ let

$$\Sigma(r) := \{\xi \in \Sigma \mid \xi \text{ is Lipschitz with } \mathrm{lip}(\xi) \leq r\}.$$

The set $\Sigma(1)$ is a closed subset of the unit ball of Σ, and hence is itself a Banach space. We consider the map X from (34.2) as a map[1]

$$X\colon \Sigma(1) \to \Sigma.$$

To apply the Banach Fixed Point Theorem 30.17, we must show that $X(\Sigma(1)) \subseteq \Sigma(1)$ and that X is a strict contraction. The fact that X maps $\Sigma(1)$ into itself is easy: if $\xi \in \Sigma(1)$ then $X(\xi)(0) = 0$ and $X(\xi)$ is Lipschitz with

$$\operatorname{lip}(X(\xi)) \leq \frac{\tau + 2\operatorname{lip}(\phi)}{\frac{1}{\tau} - 2\operatorname{lip}(\phi)} < 1, \tag{34.3}$$

and thus $X(\xi) \in \Sigma(1)$. The proof that X is a strict contraction is rather trickier. Fix $\xi, \zeta \in \Sigma(1)$. Let us abbreviate

$$\begin{aligned} A &:= L_{uu} + \phi_u(\mathrm{id}_u, \xi)\colon E^u \to E^u, \\ B &:= L_{uu} + \phi_u(\mathrm{id}_u, \zeta)\colon E^u \to E^u. \end{aligned} \tag{34.4}$$

Then as already mentioned, A and B are both bi-Lipschitz, thanks to the Lipschitz Inverse Function Theorem 31.7, and hence in particular both are homeomorphisms. This means that we can compute the $\|\cdot\|^*$ norm using $A(v)$ or $B(v)$ instead of v:

$$\|\psi\|^* = \sup_{v\neq 0} \frac{\|\psi(v)\|}{\|v\|} = \sup_{v\neq 0} \frac{\|\psi(A(v))\|}{\|A(v)\|} = \sup_{v\neq 0} \frac{\|\psi(B(v))\|}{\|B(v)\|}.$$

Fix $v \in E$. We compute:

$$\begin{aligned} \big\|X(\xi)(A(v)) - X(\zeta)(A(v))\big\| \leq& \big\|X(\xi)(A(v)) - X(\zeta)(B(v))\big\| \\ &+ \big\|X(\zeta)(B(v)) - X(\zeta)(A(v))\big\| \\ \leq& \big\|L_{ss}(\xi(v) - \zeta(v))\big\| + \big\|\phi_s(v, \xi(v)) - \phi_s(v, \zeta(s))\big\| \\ &+ \operatorname{lip}(X(\zeta))\|B(v) - A(v)\| \\ \leq& \tau\|\xi(v) - \zeta(v)\| + \operatorname{lip}(\phi)\|\xi(v) - \zeta(v)\| \\ &+ \operatorname{lip}(\phi)\|\xi(v) - \zeta(v)\| \\ \leq& \big(\tau + 2\operatorname{lip}(\phi)\big)\|\xi(v) - \zeta(v)\|, \end{aligned}$$

where the penultimate inequality used that $\operatorname{lip}(X(\zeta)) < 1$. Next, since $\phi(0) = 0$ and $\xi(0) = 0$, we have

$$\begin{aligned} \|A(v)\| &= \big\|L_{uu}v + \phi_u(v, \xi(v)) - \phi_u(0, \xi(0))\big\| \\ &\geq \frac{1}{\tau}\|v\| - \operatorname{lip}(\phi)\big(\|v\| + \operatorname{lip}(\xi)\|v\|\big) \\ &\geq \left(\frac{1}{\tau} - 2\operatorname{lip}(\phi)\right)\|v\|. \end{aligned}$$

Combining these last three statements, we see that

$$\begin{aligned} \big\|X(\xi) - X(\zeta)\big\|^* \leq& \frac{\tau + 2\operatorname{lip}(\phi)}{\frac{1}{\tau} - 2\operatorname{lip}(\phi)} \cdot \sup_{v\neq 0} \frac{\|\xi(v) - \zeta(v)\|}{\|v\|} \\ =& \frac{\tau + 2\operatorname{lip}(\phi)}{\frac{1}{\tau} - 2\operatorname{lip}(\phi)}\|\xi - \zeta\|^*. \end{aligned}$$

[1] If you are concerned about whether $X(\xi)$ belongs to Σ for $\xi \in \Sigma(r)$, see (34.3) below.

Since $\frac{\tau+2\,\mathrm{lip}(\phi)}{\frac{1}{\tau}-2\,\mathrm{lip}(\phi)} < 1$, it follows that X is a contraction, as required.

3. The Banach Fixed Point Theorem 30.17 therefore gives us a unique $\xi = \xi_\phi$ such that (34.1) holds. Note that for any $(v, \xi(v)) \in \mathrm{gr}(\xi)$, letting $u = A^{-1}v$, where A is as in (34.4), gives

$$f(u, \xi(u)) = (v, \xi(v)).$$

Thus we actually have

$$f(\mathrm{gr}(\xi)) = \mathrm{gr}(\xi).$$

Finally we prove that $\mathrm{gr}(\xi) = W^u(0, f)$. Since $\xi(0) = 0$ and $\mathrm{lip}(\xi) \le 1$, one has $\mathrm{gr}(\xi) \subset \mathrm{cone}_1(E^u)$. We will use Corollary 33.7 but applied to f^{-1} (recall last lecture we used stable manifolds). Set $\psi := f^{-1} - L^{-1}$. Since $\mathrm{lip}(\phi)$ is small, the Lipschitz Inverse Function Theorem 31.7 tells us that $\mathrm{lip}(\psi)$ is also small, and therefore after possibly shrinking δ, we may therefore assume that $L^{-1} + \psi$ satisfies the hypotheses of Corollary 33.7. Since $\mathrm{gr}(\xi)$ is invariant under f^{-1} and is contained in $\mathrm{cone}_1(E^u)$ (note the stable subspace of L^{-1} is E^u!), it follows from Corollary 33.7 that

$$\mathrm{gr}(\xi) \subseteq W^u(0, f).$$

Now suppose there exists $v \in W^u(0, f) \setminus \mathrm{gr}(\xi)$. Let $w := (v_u, \xi(v_u))$, so that $w \in \mathrm{gr}(\xi)$ and $v_u = w_u$. Then $v - w \notin \mathrm{cone}_1(E^u)$. By equation (33.5) from the last lecture (but again using $L^{-1} + \psi$ instead), it follows that

$$\left\| f^{-k}(v) - f^{-k}(w) \right\| \to \infty.$$

But since both v and w belong to $W^u(0, f)$, we also have

$$\left\| f^{-k}(v) - f^{-k}(w) \right\| \to 0.$$

This contradiction shows that $\mathrm{gr}(\xi) = W^u(0, f)$, and thus completes the proof[2]. ■

REMARK 34.2. If $L\colon E \to E$ is a hyperbolic linear dynamical system and $\phi\colon E \to E$ is C^1 and has a sufficiently small Lipschitz constant, then both $L + D\phi(0)$ and $(L + D\phi(0))^{-1}$ will satisfy the hypotheses of Proposition 31.2 and thus $L + D\phi(0)$ is hyperbolic. In this case not only is the map ξ continuously differentiable, but one can also show that

$$\mathrm{gr}(D\xi(0)) = E^u(L + D\phi(0)). \tag{34.5}$$

In words, this is saying that the unstable manifold of $L + \phi$ is *tangent* at zero to the unstable subspace of the hyperbolic linear dynamical system $L + D\phi(0)$. We will use this in the proof of Theorem 34.3 below.

We now prove the following theorem, which is one of the cornerstones of hyperbolic dynamics. Again, for variety this time we will use the stable manifold.

THEOREM 34.3 (The Local Stable Manifold Theorem). *Let $f\colon \Omega \to E$ be a dynamical system and suppose $u \in \Omega$ is a hyperbolic fixed point of f. Then for $r > 0$ sufficiently small the stable manifold $W^s_{\mathrm{loc},r}(u, f)$ is an embedded C^1 submanifold of E which is diffeomorphic to a ball in E^s.*

[2] Apart from the red statement, which we are skipping...

If you are not yet familiar with submanifolds, don't worry: the proof will make it clear exactly what exactly we mean by an "embedded C^1 submanifold". The following proof is non-examinable, since it is rather involved.

(♣) *Proof.* Denote by $\|\cdot\|$ the norm on E. Without loss of generality we may assume that $u = 0 \in \Omega$. Let $E = E^s \oplus E^u$ denote the hyperbolic splitting corresponding to $Df(0)$. Since the local stable manifold depends on the choice of norm, this time we *cannot* assume our norm $\|\cdot\|$ is already adapted to $Df(0)$ and of box-type with respect to the hyperbolic splitting. Thus let us denote by $\|\cdot\|_{\mathrm{ab}}$ the norm obtained from $\|\cdot\|$ via the procedure from Lemma 29.14, and let $0 < \tau < 1$ denote the skewness of $Df(0)$ with respect to $\|\cdot\|_{\mathrm{ab}}$. We will first prove the result using the norm $\|\cdot\|_{\mathrm{ab}}$, and then explain how to deduce the same result for the original norm $\|\cdot\|$. To help distinguish the two norms, let us denote by

$$\begin{aligned} E(r) &:= \text{ the closed ball of radius } r \text{ about } 0 \text{ with respect to } \|\cdot\|, \\ \widehat{E}(r) &:= \text{ the closed ball of radius } r \text{ about } 0 \text{ with respect to } \|\cdot\|_{\mathrm{ab}}, \end{aligned}$$

and

$$\begin{aligned} W^s_{\mathrm{loc},r}(u,f) &:= \text{ the local stable manifold with respect to } \|\cdot\|. \\ \widehat{W}^s_{\mathrm{loc},r}(u,f) &:= \text{ the local stable manifold with respect to } \|\cdot\|_{\mathrm{ab}}. \end{aligned}$$

Fix a C^∞ function $\beta \colon E \to [0,1]$ such that:

$$\beta(v) = \begin{cases} 1, & \|v\|_{\mathrm{ab}} \le \frac{1}{3}, \\ 0, & \|v\|_{\mathrm{ab}} \ge \frac{2}{3} \end{cases}$$

Let $\phi \colon = f - Df(0) \colon \Omega \to E$. Then $\phi(0) = 0$ and $D\phi(0) = 0$. We now want to extend ϕ to a function $\phi_r \colon E \to E$. Choose $r > 0$ small enough so that $\widehat{E}(3r) \subset \Omega$. We will later shrink r further. Define

$$\phi_r(v) := \beta\left(\frac{v}{3r}\right)\phi(v).$$

Then ϕ_r is C^1 and agrees with ϕ on $\widehat{E}(r)$. Just like in the proof of the Hartman-Grobman Theorem 32.2, if $r > 0$ is sufficiently small, then $\mathrm{lip}(\phi_r)$ will be small enough so that the hypotheses of both Corollary 33.7 and Theorem 34.1 are satisfied (with respect to $\|\cdot\|_{\mathrm{ab}}$). This means there is a C^1 map $\xi_r \colon E^s \to E^u$ with $\xi_r(0) = 0$ and $\mathrm{lip}(\xi_r) \le 1$ such that

$$\widehat{W}^s\big(0, Df(0) + \phi_r\big) = \mathrm{gr}(\xi_r).$$

Since $D\phi_r(0) = D\phi(0) = 0$, the stable subspace of the hyperbolic linear dynamical system $Df(0) + D\phi_r(0)$ is just E^s, and hence it follows from (34.5) that

$$D\xi_r(0) = 0.$$

Since the norm $\|\cdot\|_{\mathrm{ab}}$ is of box type, we can write[3]

$$\widehat{E}(r) = \widehat{E}^s(r) \times \widehat{E}^u(r),$$

[3]Here we are performing the harmless (ab)use of notation $E^s \oplus E^u = E^s \times E^u$.

where $\widehat{E}^s(r)$ is the closed ball in E^s with respect to $\|\cdot\|_{\mathrm{ab}}$ etc. Consider the C^1 embedding $i_r \colon E^s \to E$ given by

$$i_r(v) = (v, \xi_r(v)).$$

Then i maps E^s onto $\mathrm{gr}(\xi_r)$. Since $\mathrm{lip}(\xi_r) \leq 1$, we have

$$i_r\big(\widehat{E}^s(r)\big) = \widehat{W}^s\big(0, Df(0) + \phi_r\big) \cap \widehat{E}(r).$$

We claim that

$$\widehat{W}^s_{\mathrm{loc},r}(0, f) = \widehat{W}^s\big(0, Df(0) + \phi_r\big) \cap \widehat{E}(r). \tag{34.6}$$

Since $\phi_r = \phi$ on $E(r)$, it is clear that

$$\widehat{W}^s_{\mathrm{loc},r}(0, f) \subseteq \widehat{W}^s\big(0, Df(0) + \phi_r\big) \cap \widehat{E}(r)$$

To prove the other direction it suffices to show that if $v \in \widehat{W}^s\big(0, Df(0) + \phi_r\big) \cap \widehat{E}(r)$ then $(Df(0) + \phi_r)^k(v) \in \widehat{E}(r)$ for all $k \geq 0$, since then $(Df(0) + \phi_r)^k(v) = f^k(v)$ for all $k \geq 0$. The fact that $(Df(0) + \phi_r)^k(v) \in \widehat{E}(r)$ for all $k \geq 0$ is clear from Corollary 33.7, which gave us the alternative description of $\widehat{W}^s\big(0, Df(0) + \phi_r\big)$ as:

$$\widehat{W}^s\big(0, Df(0) + \phi_r\big) = \left\{ v \in E \mid \left\|(Df(0) + \phi_r)^k(v)\right\|_{\mathrm{ab}} \leq (\tau + \mathrm{lip}(\phi_r))^k \, \|v\|_{\mathrm{ab}} \right\}.$$

Thus

$$\widehat{W}^s_{\mathrm{loc},r}(0, f) = i_r\big(\widehat{E}^s(r)\big),$$

which shows that $\widehat{W}^s_{\mathrm{loc},r}(0, f)$ is an embedded submanifold diffeomorphic to the ball $\widehat{E}^s(r)$.

It remains to deduce the same result for the local stable manifold $W^s_{\mathrm{loc},r}(u, f)$ with respect to the original norm $\|\cdot\|$. Since $\|\cdot\|$ and $\|\cdot\|_{\mathrm{ab}}$ are equivalent, we may choose $0 < r_2 < r_1 < r$ such that

$$\widehat{W}^s_{\mathrm{loc},r_2}(0, f) \subset W^s_{\mathrm{loc},r_1}(0, f) \subset \widehat{W}^s_{\mathrm{loc},r}(0, f).$$

Set

$$B := i_r^{-1}\left(W^s_{\mathrm{loc},r_1}(0, f)\right).$$

Then B is a neighbourhood of 0 in E^s which is diffeomorphic[4] to a ball (it is a star-shaped set with respect to the origin in E^s, which is squeezed in between the two balls $\widehat{E}^s(r_2)$ and $\widehat{E}^s(r)$). Moreover $i|_B$ is a C^1 embedding from this ball onto $W^s_{\mathrm{loc},r_1}(0, f)$. This completes the proof. ∎

[4]We are concealing some details here...

LECTURE 35

Introduction to Differential Geometry

In the next two lectures we will give a brief introduction to the aspects of differential and Riemannian geometry that we will need during the rest of the course. Most of this should be familiar to anyone who has attended an introductory course on Differential Geometry, although we seize the opportunity to define infinite-dimensional **Banach manifolds**, which you are probably less likely to have seen before.

None of the material in the next two lectures is *directly* examinable.

To motivate the definition of a manifold, let us first take a step back.

- Suppose A and B are sets. Let
$$\mathrm{Maps}(A, B)$$
denote the set of all maps $f\colon A \to B$. Since A and B are just sets, it doesn't make sense to ask whether a given element f of $\mathrm{Maps}(A, B)$ is continuous.
- A metric space (or more generally, any topological space) consists of a set X, equipped with an extra structure—the metric—which allows us make sense of continuity. If X and Y are metric spaces, we can therefore speak of the subset
$$C^0(X, Y) \subseteq \mathrm{Maps}(X, Y)$$
of all continuous functions.
- Roughly speaking, a **smooth manifold** consists of a metric space M, equipped with an extra structure—an **atlas**—which allows us make sense of differentiability. Thus if M and N are manifolds, we can speak of the subset
$$C^p(M, N) \subset C^0(M, N), \qquad p = 1, \dots, \infty,$$
of functions on M that are p times differentiable.

We already know one special type of metric space on which it makes sense to differentiate things: namely, normed vector spaces $\big(E, \|\cdot\|\big)$. In fact, normed vector spaces are the prototypical examples of manifolds, and in a sense all manifolds are obtained by piecing together normed vector spaces.

Definition 35.1. Let E denote a finite-dimensional normed vector space, and let M be a separable metric space. An E-**atlas** on M is a collection $\mathcal{E} := \{\sigma_i \colon U_i \to E\}$ of functions on open subsets of M, such that:

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

- $\{U_i\}$ is an open cover of M,
- the maps $\sigma_i \colon U_i \to E$ are homeomorphisms onto their images,
- for each pair i, j, the set $\sigma_i(U_i \cap U_j)$ is an open (possibly empty) subset of E. If it is non-empty then the map

$$\sigma_{ij} := \sigma_j \circ \sigma_i^{-1} \colon \sigma_i(U_i \cap U_j) \to E \tag{35.1}$$

  must be a smooth (i.e. of class C^∞) map in the sense of Remark 30.5, which moreover is a diffeomorphism onto its image.

We call the individual functions σ_i the **charts** of the atlas $\mathcal{E}$. If an E-atlas exists, we call the pair $(M, \mathcal{E})$ a **smooth manifold modelled on** E.

In practice, we refer to a manifold $(M, \mathcal{E})$ simply by M, leaving the E-atlas to be understood. Here are some comments on the definition.

Remarks 35.2.

(i) One could start with weaker point-set topological hypotheses: namely, one could require merely that M was a separable Hausdorff space. The existence of an E-atlas implies that the topology on M is metrisable, and thus a posteriori, M admits the structure of a metric space.

(ii) It may very well be the case that two different E-atlases $\mathcal{E}_1$ and $\mathcal{E}_2$ on M define the "same" smooth manifold (this is analogous to the fact that there are often many ways to specify the topology on a given topological space). This can be formally rectified by working with equivalence classes of E-atlases: say that $\mathcal{E}_1$ and $\mathcal{E}_2$ are **equivalent** if $\mathcal{E}_1 \cup \mathcal{E}_2$ is another E-atlas. Then define a smooth manifold modelled on E to be a separable metric space equipped with an equivalence class of $\mathcal{E}$-atlases. Nevertheless, we will avoid the pedantry and work solely with honest E-atlases.

(iii) Since all normed vector spaces of a given finite dimension n are isomorphic, the only thing that matters is the dimension of E. So instead it is more common to say that M is an **n-dimensional smooth manifold** if M admits an E-structure for some (and hence any) n-dimensional normed vector space. Normally, one just takes $E = \mathbb{R}^n$. Nevertheless, it can often be insightful to keep track of E: for example, the manifold version of Stable Manifold Theorem states that W^s is a manifold modelled on E^s.

(iv) The "smooth" in the name "smooth manifold" refers to the fact that the functions σ_{ij} in (35.1) are of class C^∞. It is sometimes useful to work with less regularity. Given $1 \le p < \infty$, a **C^p-manifold modelled on** E consists of a pair $(M, \mathcal{E})$, where $\mathcal{E}$ is an **atlas of class** C^p, i.e. an atlas with the property that the maps σ_{ij} from (35.1) are of class C^p.

(v) Exactly the same definition works if E is a Banach (or Hilbert) space. This gives rise to the notion of a **Banach manifold**. In this case, however, one really does need to keep track of E—the "dimension" alone is not enough!

Moreover when E is infinite-dimensional, the requirement that the topology on M is metrisable *cannot* be replaced with “Hausdorff” (cf. (i) above). An example of an infinite-dimensional manifold is given in Example 35.18 below.

CONVENTION. For convenience, all manifolds we consider are implicitly assumed to be *connected*. From a dynamical systems point of view, this is harmless—if a given manifold has two connected components, then we simply treat it as two manifolds.

EXAMPLE 35.3. Let $\big(E, \|\cdot\|\big)$ be a normed vector space. Then E admits an E-atlas consisting of exactly one element: $\mathcal{E} = \{\mathrm{id}\colon E \to E\}$.

DEFINITION 35.4. Suppose M is a manifold modelled on E and N is a manifold modelled on F. A continuous map $f\colon M \to N$ is said to be of **class** C^p if for every $x \in M$, every chart $\sigma\colon U \to E$ with $x \in U$, and every chart $\tau\colon V \to F$ with $f(x) \in V$, the corresponding map

$$f_{\sigma,\tau} \coloneqq \tau \circ f \circ \sigma^{-1}\colon \sigma(U \cap f^{-1}(V)) \to F \tag{35.2}$$

is of class C^p. We say f is a **diffeomorphism** of class C^p (for $p \geq 1$) if f is a homeomorphism and $f^{-1}\colon N \to M$ is also of class C^p.

We call the maps $f_{\sigma,\tau}$ the **local representations** of f. When the precise choices of charts σ and τ are not important, we will refer to the local representation using the simpler notation $\hat{f}$ instead of $f_{\sigma,\tau}$.

DEFINITION 35.5. We denote by $\mathrm{Diff}^p(M,N)$ the group of all diffeomorphisms from M to N of class C^p, and abbreviate $\mathrm{Diff}^p(M) = \mathrm{Diff}^p(M,M)$.

Now let us introduce the *tangent space* to a manifold at a point x. The tangent space T_xM will be a finite-dimensional normed vector space which is non-canonically isomorphic to E. There are various ways this can be done, here is one:

DEFINITION 35.6. Let $(M, \mathcal{E})$ be a smooth manifold modelled on E and fix $x \in M$. Let $\mathcal{E}_x \subseteq \mathcal{E}$ denote the subset consisting of those charts $\sigma_i\colon U_i \to E$ for which x belongs to U_i. Note that $\mathcal{E}_x$ is non-empty, since by assumption the $\{U_i\}$ are an open cover of M.

If σ_i and σ_j are two elements of $\mathcal{E}_x$ then the differential of the map σ_{ij} from (35.1) at $\sigma_i(x)$ is a linear isomorphism

$$D\sigma_{ij}(\sigma_i(x))\colon E \to E.$$

Thus we can define an equivalence relation $\sim$ on $\mathcal{E}_x \times E$ by declaring that

$$(\sigma_i, u) \sim (\sigma_j, v) \qquad \Leftrightarrow \qquad D\sigma_{ij}(\sigma_i(x))u = v.$$

Let $[\sigma_i, u]$ denote the equivalence class containing (σ_i, u). The set of equivalence classes is denoted by T_xM and is called the **tangent space to** M **at** x. Any chart $\sigma_i \in \mathcal{E}_x$ determines an isomorphism $\mathcal{J}_{\sigma_i,x}\colon T_xM \to E$ given by

$$\mathcal{J}_{\sigma_i,x}\colon T_xM \to E, \qquad [\sigma_i, v] \mapsto v. \tag{35.3}$$

It is important to understand that the isomorphism $T_xM \cong E$ in (35.3) is *not* canonical if $\mathcal{E}_x$ contains more than one chart.

Example 35.7. Let $(E, \|\cdot\|)$ be a normed vector space. Think of E as a manifold, equipped with the E atlas consisting of the identity only (cf. Example 35.3). Then for any $u \in E$, the set $\mathcal{E}_u$ contains exactly one element (the identity), and thus in this case the identification $T_uE \cong E$ *is* canonical. In this case we write simply $\mathcal{J}_u$ instead of $\mathcal{J}_{\mathrm{id},u}$:

$$\mathcal{J}_u \colon T_uE \to E, \qquad [\mathrm{id}, v] \mapsto v.$$

Convention. From now on we write a point in T_xM simply as a vector v, rather than the more cumbersome notation $[\sigma_i, v]$.

Definition 35.8. Suppose M and N are two manifolds modelled on E and F respectively. Suppose $f \colon M \to N$ is a map of class C^p. The **differential** of f at a point x is a map $Df(x) \colon T_xM \to T_{f(x)}N$. This is the unique linear map with the following property: if σ is any chart about x on M and τ is any chart about $f(x)$ on N then the following diagram should commute:

$$\begin{tikzcd} T_xM \arrow[r, "\mathcal{J}_{\sigma,x}"] \arrow[d, "Df(x)"'] & E \arrow[d, "Df_{\sigma,\tau}(\sigma(x))"] \\ T_{f(x)}N \arrow[r, "\mathcal{J}_{\tau,f(x)}"'] & F \end{tikzcd}$$

Here the map on the right-hand side is the differential $Df_{\sigma,\tau}(\sigma(x)) \colon E \to F$ of the map (35.2).

The tangent spaces all fit together to define the *tangent bundle*. This is a special case of a more general notion of a *vector bundle*.

Definition 35.9. Suppose M is a manifold modelled on E, and P is a topological space. Suppose $\pi \colon P \to M$ is a surjective continuous map. Write $P(x) := \pi^{-1}(x)$. Let F denote another finite-dimensional normed vector space. An F**-bundle atlas** of $\pi \colon P \to M$ is a collection $\mathcal{F} = \{\lambda_i \colon \pi^{-1}(V_i) \to V_i \times F\}$ of functions defined on open subsets of P such that:

- $\{V_i\}$ is an open cover of M and the functions $\lambda_i \colon \pi^{-1}(V_i) \to V_i \times F$ are homeomorphisms,
- the homeomorphisms λ_i are required to make the following diagram commute:

$$\begin{tikzcd} \pi^{-1}(V_i) \arrow[rr, "\lambda_i"] \arrow[dr, "\pi"'] & & V_i \times F \arrow[dl, "\mathrm{proj}"] \\ & V_i & \end{tikzcd}$$

  where $\mathrm{proj} \colon V_i \times F \to V_i$ the projection onto the first factor. Thus there is a well-defined map $\lambda_{i,x} \colon P(x) \to F$ given by restricting λ_i to $P(x)$,

- if $V_i \cap V_j \neq \emptyset$ then for each $x \in V_i \cap V_j$ the map

$$\lambda_{ij,x} := \lambda_{j,x} \circ \lambda_{i,x}^{-1} \colon F \to F \tag{35.4}$$

  is a **linear** isomorphism.

The maps λ_i are called **trivialisations**. If such an F-trivialising structure exists, we say that $\pi \colon P \to M$ is a **vector bundle over** M **with fibre** F. We call P the **total space** and M the **base space**.

Remarks 35.10.

(i) As with atlases on manifolds (cf. part (i) of Remarks 35.2), we should really work with equivalence classes of bundle atlases. We won't bother, though.

(ii) If $\pi \colon P \to M$ is a vector bundle over M with fibre F then each set $P(x)$ is non-canonically isomorphic to F via the map $\lambda_{i,x}$. The total space P can then be given the structure of a manifold modelled on $E \times F$, in such a way that the map $\pi \colon P \to M$ is of class C^∞.

Definition 35.11. Suppose $\pi \colon P \to M$ and $\theta \colon Q \to N$ are two vector bundles. Let $f \colon M \to N$ be a of class C^p. A **fibre-preserving map over** f of class C^q (for $1 \leq q \leq p$) is a map $\Phi \colon P \to Q$ of class C^q with the property that the following diagram commutes:

$$\begin{tikzcd} P \arrow[r, "\Phi"] \arrow[d, "\pi"'] & Q \arrow[d, "\theta"] \\ M \arrow[r, "f"'] & N \end{tikzcd}$$

This means that Φ restricts to define a map

$$\Phi_x \colon P(x) \to Q(f(x)).$$

If this map is a **linear** map for each $x \in M$ then we say that the pair (f, Φ) is a **vector bundle map**. If f is a diffeomorphism and Φ_x is a linear isomorphism for each $x \in M$ then we say that pair (f, Φ) is a **vector bundle isomorphism**.

Definition 35.12. Suppose $\pi \colon P \to M$ is a vector bundle. A **section** γ of π of class C^p is a function $\gamma \colon M \to P$ such that $\pi \circ \gamma = \mathrm{Id}$. We denote by $\Gamma^p(M, P)$ the space of sections of π. The space $\Gamma^p(M, P)$ is a vector space, where addition is defined pointwise:

$$(\gamma_1 + \gamma_2)(x) := \gamma_1(x) + \gamma_2(x).$$

The addition makes sense as it takes place in the vector space $P(x)$. If $p < \infty$, this defines a Banach space structure on $\Gamma^p(M, P)$.

If $U \subseteq M$ is an open set we denote by $\Gamma^p(U, P)$ the space of sections which are only defined on U. These are referred to as **local sections** of P.

Definition 35.13. Suppose $\pi\colon P \to M$ is a vector bundle with fibre F. Let $U \subset M$ be open. A **local frame of class** C^p for P over U is a collection $\{\gamma_1, \dots, \gamma_n\}$ of elements of $\Gamma^p(U, P)$ which form a basis for $\Gamma^p(U, P)$ (thought of as a module over $C^\infty(U)$).

If a local frame exists for $U = M$ then we call the vector bundle **trivial**. In this case P is isomorphic as a vector bundle to $M \times F$. Most vector bundles are not trivial, however one can at least always find a basis of local sections.

Example 35.14. Suppose $\pi\colon P \to M$ is a vector bundle with fibre F, and suppose $\lambda\colon \pi^{-1}(V) \to V \times F$ is a trivialisation. Let $\{v_1, \dots, v_n\}$ be a basis of F. Define elements $\gamma_i \in \Gamma^\infty(V, P)$ by

$$\gamma_i(x) := \lambda_i^{-1}(x, v_i).$$

The sections $\{\gamma_1, \dots, \gamma_n\}$ form a local frame of class C^∞ for P over V.

We now define the tangent bundle.

Definition 35.15. Now suppose M is a smooth manifold

$$TM := \bigcup_{x \in M} T_x M$$

and let $\pi\colon TM \to M$ denote the map that sends T_xM to x. We call TM the **tangent bundle** of M. We write an element of TM as a pair (x, v)—this is shorthand for saying that v belongs to T_xM.

Proposition 35.16. *Suppose M is a smooth manifold modelled on E. Then the tangent bundle $\pi\colon TM \to M$ is a vector bundle over M with fibre E.*

Proof (Sketch). Let $\mathcal{E} = \{\sigma_i\colon U_i \to E\}$ be an E-atlas for M. We use $\mathcal{E}$ to define a E-bundle atlas. Define a function $\lambda_i\colon \pi^{-1}(U_i) \to U_i \times E$ by

$$\lambda_i(x, v) := \big(x, \mathcal{J}_{\sigma_i, x}(v)\big), \tag{35.5}$$

where $\mathcal{J}_{\sigma_i,x}$ was defined in (35.3). We endow TM with a topology by declaring that the λ_i are homeomorphisms. We claim that $\mathcal{F} := \{\lambda_i\colon \pi^{-1}(U_i) \to U_i \times E\}$ is an E-bundle atlas. For this we need to check that the functions $\lambda_{ij,x}\colon E \to E$ from (35.4) are linear isomorphisms. But this is clear, since unravelling the definitions shows that.

$$\lambda_{ij,x} = D\sigma_{ij}(\sigma_i(x))$$

This completes the proof. ■

Definition 35.17. Let M be a smooth manifold. A **vector field** on M is an element of $\Gamma^\infty(M, TM)$.

Example 35.18. Here is an example of an infinite-dimensional manifold: if $1 \le p < \infty$ then $\mathrm{Diff}^p(M)$ is a Banach manifold (cf. part (v) of Remarks 35.2). It is modelled on the Banach space $\Gamma^p(M, TM)$ (i.e. vector fields of class C^p). Note however that $\mathrm{Diff}^\infty(M)$ is not a Banach manifold, since $\Gamma^\infty(M, TM)$ is not a Banach space.

Definition 35.19. Let $f\colon M \to N$ be of class C^p. Define a map $Df\colon TM \to TN$ by requiring that $Df|_{T_xM} = Df(x)$. Then Df is of class C^{p-1}, and the pair (f, Df) is a vector bundle morphism.

We next define submanifolds of manifolds, which are smaller manifolds sitting inside larger manifolds.

Definition 35.20. Let N be a smooth manifold and suppose $M \subseteq N$ is a subset. We say that M is an **immersed submanifold** of N if there exists a topology on M such that M can made into a smooth manifold. This topology need *not* coincide with the subspace topology inherited from N. If it does, however, we call M a **embedded submanifold** of N.

In a similar way we can speak of submanifolds of class C^p.

Definition 35.21. Suppose $f\colon M \to N$ is a C^p map. We say that f is:

- an **immersion** if $Df(x)\colon T_xM \to T_{f(x)}N$ is injective for all $x \in M$,
- a **submersion** if $Df(x)\colon T_xM \to T_{f(x)}N$ is surjective for all $x \in M$.

Immersions can only exist when $\dim M \leq \dim N$, and submersions can only exist when $\dim M \geq \dim N$.

The manifold version of the Implicit Function Theorem that you no doubt remember from calculus shows how immersions and submersions can be used to create new manifolds. We conclude today's lecture by stating it.

Theorem 35.22 (The Implicit Function Theorem). *Suppose $f\colon M \to N$ is a C^p map.*

(i) *If f is an injective immersion then $f(M)$ is an immersed submanifold of N, with $\dim f(M) = \dim M$.*

(ii) *If f is an injective immersion which in addition is a homeomorphism onto its image then $f(M)$ is an embedded submanifold of N.*

(iii) *If f is a submersion then for every point $y \in f(M)$, the preimage $f^{-1}(y)$ is an embedded submanifold of M of dimension $\dim M - \dim N$.*

LECTURE 36

Introduction to Riemannian Geometry

Today we focus on the elements of Riemannian geometry that we will need throughout the remainder of the course.

DEFINITION 36.1. Suppose M is a manifold modelled on E. A **Riemannian metric**[1] of class C^p on M is an assignment of an inner product

$$m_x \colon T_x M \times T_x M \to \mathbb{R}, \qquad \forall\, x \in M,$$

such that the map $x \mapsto m_x$ is of class C^p.

The study of smooth manifolds equipped with Riemannian metrics is (unsurprisingly) referred to as *Riemannian Geometry*. This is a rich and interesting field of modern mathematics, but not one we will touch upon. For us, however, the Riemannian metric should simply be regarded as part of the "background".

Here are some remarks on this definition.

REMARKS 36.2.

(i) The word "metric" is slightly unfortunate, since a Riemannian metric is *not* the same as a metric in the sense of point-set topology (i.e. metric spaces). Nevertheless, the two concepts are closely linked (see Theorem 36.7 below).

(ii) Let us make explicit exactly what it means to say that $x \mapsto m_x$ is of class C^p. Fix an inner product $\langle \cdot, \cdot \rangle$ on E, and fix $x_0 \in M$. Let $\sigma \colon U \to E$ denote a chart on M containing x_0. The chart σ determines an isomorphism

$$\mathcal{J}_{\sigma,x} \colon T_x M \to E, \qquad \forall\, x \in U,$$

see Definition 35.6. The Riemannian metric m determines a map $x \mapsto A_x$, where $A_x \colon E \to E$ is a symmetric positive definite matrix, by requiring

$$m_x(v, w) = \langle A_x \mathcal{J}_{\sigma,x} v, \mathcal{J}_{\sigma,x} v \rangle\,, \qquad v, w \in T_x M. \tag{36.1}$$

We can think of A as a map $U \to \mathcal{L}(E, E)$. Saying that m is of class C^p near x_0 is equivalent to saying that the map A is of class C^p.

(iii) It can be shown that every smooth manifold admits (many) Riemannian metrics of any desired class C^p. Indeed, it is obvious that they exist locally (i.e. on chart domains)—simply read equation (36.1) from right to left. To go from local to global existence one then uses a partition of unity.

[1] It is common in the literature to use the symbol "g" to denote a Riemannian metric. We have elected to use "m" instead (m for metric), since g is typically used to denote a dynamical system.

(iv) Most of the time we will be interested in smooth (i.e. of class C^∞) Riemannian metrics. To this end, we adopt the convention that a **Riemannian manifold** consists of a pair (M, m), where M is a smooth manifold and m is a smooth Riemannian metric on M.

(v) Nevertheless, we will sometimes be forced to work with C^0 metrics. These metrics are not very useful (for example, Theorem 36.16 below doesn't hold for them), however they crop up naturally in hyperbolic dynamics. Roughly speaking, this is because we will primarily work with C^1 diffeomorphisms. Next lecture we will introduce the notion of a **hyperbolic set** of such a dynamical system f. In analogy to Proposition 29.11, we will then show that starting from a smooth Riemannian metric, one can construct a new metric that is "adapted" with respect to the hyperbolic splitting. Since f is only of class C^1, this new metric is only of class C^0.

(vi) Just as smooth functions are dense in the set of continuous functions, smooth Riemannian metrics are dense in the set of C^0 metrics. Thus we can always approximate a C^0 metric arbitrarily well with a smooth metric. Therefore in practice, when working with a hyperbolic set of a diffeomorphism f, we first adapt the metric for f, and then approximate this new adapted metric with a smooth one. See Proposition 37.15 for the details.

Example 36.3. Let $\big(E, \langle\cdot,\cdot\rangle\big)$ denote a vector space equipped with an inner product Since on a vector space there is a canonical identification $T_uE \cong E$ for every point $u \in E$ (cf. Example 35.7), we can view $\langle\cdot,\cdot\rangle$ as defining a smooth Riemannian metric on E. Thus any vector space can be given the structure of a Riemannian manifold.

Definition 36.4. A **smooth curve** on M is a smooth map $\alpha\colon \mathbb{R} \to M$, (here we think of $\mathbb{R}$ as a 1-dimensional manifold, cf. Example 35.3. It is convenient to abbreviate $\dot\alpha(t) \colon = D\alpha(t)1$, for $1 \in T_t\mathbb{R} \cong \mathbb{R}$. We call $\dot\alpha(t)$ the **velocity vector** of the curve α at time t.

Using a Riemannian metric we can define the *length* of a smooth curve. Here and elsewhere, we write $\|\cdot\|_m$ to denote the norm associated to a Riemannian metric m.

Definition 36.5. Suppose $\alpha\colon [a,b] \to M$ is a smooth curve[2] and m is a Riemannian metric on M. The **length** of α with respect to m is given by

$$\mathrm{length}_m(\alpha) := \int_a^b \|\dot\alpha(t)\|_m \, dt.$$

This allows us to endow any smooth manifold with a convenient choice of metric (in the sense of topology).

[2]Strictly speaking we should introduce the notion of a *manifold with boundary* to make sense of this, since $[a,b]$ has boundary. If you really don't like this, think of α as being defined on $(a-\varepsilon, b+\varepsilon)$ for some small $\varepsilon > 0$.

Definition 36.6. Suppose M is a manifold and m is a Riemannian metric on M. Define

$$d_m(x,y) := \inf \operatorname{length}_m(\alpha),$$

where the infimum is taken over all (piecewise) smooth curves $\alpha\colon [a,b] \to M$ (for any interval $[a,b]$) such that $\alpha(a) = x$ and $\alpha(b) = y$. Due to our standing implicit assumption that all manifolds are connected, $d_m(x,y)$ is a well-defined finite number for every pair $x, y \in M$.

As the notation suggests, d_m is a metric.

Theorem 36.7. *Let (M,m) be a Riemannian manifold. The function $d_m\colon M \times M \to [0,\infty)$ is a metric on M. Moreover the topology induced by d_m coincides with the original manifold topology on M.*

Proof. If α is a smooth curve from x to y, then by traversing along α backwards we get a new smooth curve from y to x. Denoting this curve by $\bar{\alpha}$, one has

$$\operatorname{length}_m(\alpha) = \operatorname{length}_m(\bar{\alpha}).$$

This shows that d_m is symmetric. Next, suppose α is a smooth curve from x to y, and β is a smooth curve from y to z. Then the concatenation of α and β, denoted by $\alpha * \beta$, is a piecewise smooth curve from x to z. Moreover directly from the definition,

$$\operatorname{length}_m(\alpha * \beta) = \operatorname{length}_m(\alpha) + \operatorname{length}_m(\beta).$$

We now prove the triangle inequality. Fix $x, y, z \in M$ and $\varepsilon > 0$ By definition of d_m as an infimum, there exists a smooth curve α from x to y, and a smooth curve β from y to z, such that

$$d_m(x,y) < \operatorname{length}_m(\alpha) + \frac{\varepsilon}{2}, d_m(y,z) < \operatorname{length}_m(\beta) + \frac{\varepsilon}{2}.$$

Then

$$d_m(x,z) \le \operatorname{length}_m(\alpha * \beta) < d_m(x,y) + d_m(y,z) + \varepsilon.$$

Since ε was arbitrary, the triangle inequality follows.

The hardest part of the proof is to show that

$$d_m(x,y) = 0 \qquad \Rightarrow \qquad x = y. \tag{36.2}$$

To prove this, first suppose we are given two points $x, y \in M$ with the property that there exists a chart $\sigma\colon U \to E$ on M whose domain contains both x and y. If α is a smooth curve from x to y whose image is entirely contained in U, then $\sigma \circ \alpha$ is a smooth curve in E from $\sigma(x)$ to $\sigma(y)$. Moreover there exists a constant $\delta > 0$ depending only on σ such that

$$\operatorname{length}_m(\alpha) \ge \delta \operatorname{length}(\sigma \circ \alpha), \tag{36.3}$$

where the length on the right-hand side is measured with respect to some fixed norm on E. Now we use the "geometrically obvious" fact that the claim is true in E. Indeed, the shortest path in E is a straight line, and thus

$$\operatorname{length}(\beta) \ge \|\sigma(x) - \sigma(y)\|, \tag{36.4}$$

for any smooth curve β from $\sigma(x)$ to $\sigma(y)$. Combining (36.3) and (36.4), we see that

$$\text{length}_m(\alpha) \geq \delta \|\sigma(x) - \sigma(y)\|,$$

for any smooth curve α from x to y whose image is entirely contained in U. Moreover the same argument shows that if α is any smooth curve from x to y, then

$$\begin{aligned} \text{length}_m(\alpha) &\geq \text{length}_m(\text{portion of } \alpha \text{ contained in } U) \\ &\geq \delta \inf_{v \in \partial(\sigma(U))} \|\sigma(x) - v\| \\ &> 0. \end{aligned}$$

This shows that $d_m(x, y) > 0$ for two points x, y contained in a chart domain. For the general case, suppose x, y are any two points in M. Let $\sigma\colon U \to E$ denote any chart about x. If $y \in U$ we are done. If y does not belong to E then every smooth curve α from x to y must pass through ∂U, and thus

$$\text{length}_m(\alpha) \geq d_m(x, \partial U) > 0$$

by the argument above. This proves (36.2).

We now know that d_m is a metric. It remains to shows that the topology induced by d_m coincides with the original topology on M. This argument however is a little more involved, and we omit the details. ■

Definition 36.8. Suppose (M, m) is a Riemannian manifold. A **geodesic** on M is a curve α which is locally length minimising. That is, for any $s < t$ in the domain of α with $t - s$ sufficiently small,

$$d_m(\alpha(s), \alpha(t)) = \text{length}_m\left(\alpha|_{[s,t]}\right).$$

One can characterise geodesics as solutions to a certain second order ordinary differential equation on M. Therefore the usual existence and uniqueness theorems for solutions to ordinary differential equations proves the following statement:

Lemma 36.9. *Let (M, m) be a Riemannian manifold. For every $x \in M$ and every $v \in T_xM$ there exists a unique geodesic $\alpha_{x,v}$ whose maximal interval of existence is an open interval containing* 0, *such that*

$$\alpha_{x,v}(0) = x, \qquad \dot{\alpha}_{x,v}(0) = v$$

Geodesics behave nicely with respect to scaling. That is, for any $x \in M$, $v \in T_xM$ and $t > 0$,

$$\alpha_{x,v}(t) = \alpha_{x,tv}(1). \tag{36.5}$$

Here are some examples.

Example 36.10. Let $\left(E, \langle\cdot,\cdot\rangle\right)$ denote a vector space with an inner product. If we regard $\left(E, \langle\cdot,\cdot\rangle\right)$ as a Riemannian manifold as in Example 36.3, then geodesics are straight lines. In this case all geodesics are injective and defined on all of $\mathbb{R}$. Moreover geodesics are globally length minimising.

Example 36.11. As a slightly more interesting example, consider $E \setminus \{0\}$. This is an open subset of E, and hence we can consider it as a Riemannian manifold. Geodesics are still straight lines, but now not all of them are defined on all of $\mathbb{R}$. Indeed, any straight line passing through the origin determines a pair of geodesics that are only defined for finite time in one direction (we introduced a "hole" in our manifold).

Example 36.12. If we think of S^2 as the unit sphere in $\mathbb{R}^3$, then the Euclidean inner product on $\mathbb{R}^3$ restricts to define a Riemannian metric on S^2. Geodesics in this metric are great circles. In this case, all geodesics are defined for all $t \in \mathbb{R}$, but no geodesics are injective—they are all periodic with period 2π. Moreover geodesics are only length minimising up to the antipodal point, after which going the other way round the great circle gives a shorter curve.

In fact, when M is compact geodesics are always defined for all time.

Proposition 36.13. *Let (M, m) denote a compact Riemannian manifold. Then every geodesic $\alpha_{x,v}$ is defined for all $t \in \mathbb{R}$.*

From now on we will restrict our attention to compact Riemannian manifolds.

Definition 36.14. Let (M, m) be a compact Riemannian manifold. The **exponential map** of m is the map $\exp = \exp^m \colon TM \to M$ whose restriction $\exp_x$ to T_xM is given by

$$\exp_x \colon T_xM \to M, \qquad \exp_x(v) := \alpha_{x,v}(1).$$

One can show that the map $\exp \colon TM \to M$ is smooth. Note that $\exp_x(0_x) = x$, where $0_x \in T_xM$ is the origin.

Lemma 36.15. *Under the canonical identification $T_{0_x}T_xM \cong T_xM$,*

$$D\exp_x(0_x) = \mathrm{id}\,.$$

Proof. Let $v \in T_xM$. Then

$$\begin{aligned} D\exp_x(0_x)v &= \frac{d}{dt}\Big|_{t=0} \exp_x(tv) \\ &= \frac{d}{dt}\Big|_{t=0} \alpha_{x,tv}(1) \\ &\overset{(\heartsuit)}{=} \frac{d}{dt}\Big|_{t=0} \alpha_{x,v}(t) \\ &= \dot{\alpha}_{x,v}(0) \\ &= v, \end{aligned}$$

where $(\heartsuit)$ used (36.5). ∎

It follows from Lemma 36.15 and the Inverse Function Theorem that $\exp_x$ is a diffeomorphism in a neighbourhood of 0_x. In fact, a much stronger result is true. To state this, set

$$T_xM(r) := \{v \in T_xM \mid \|v\|_m \leq r\}\,, \qquad TM(r) := \bigcup_{x \in M} T_xM(r),$$

and write $B_m(x, r)$ for the open ball of radius r about x in the d_m-metric on M.

THEOREM 36.16. *Suppose* (M, m) *is a compact Riemannian manifold. Then there exists* $r_m > 0$ *such that for each* $0 < r < r_m$*, the restriction of* $\exp_x$ *to* $T_xM(r)$ *is a diffeomorphism onto its image, which is exactly* $\overline{B}_m(x, r)$*. In particular,*

$$d_m\big(x, \exp_x(v)\big) = \|v\|_m, \qquad \forall\, \|v\|_m < r_m. \tag{36.6}$$

We call r_m the **injectivity radius** of m. Equation (36.6) can be interpreted as saying that $\exp_x$ maps straight lines in T_xM through 0_x to geodesics in M.

EXAMPLE 36.17. The injectivity radius of S^2, equipped with the Riemannian structure from Example 36.12) is π. (The distance from a point to its antipodal point).

A consequence of Theorem 36.16 is that for $r < r_m$ we can use the inverse of the exponential map as a chart on M. This concludes all the background material we need from differential and Riemannian geometry. In the next lecture we will commence our study of differentiable dynamical systems on smooth manifolds.

LECTURE 37

Hyperbolic Sets

Throughout this lecture, M denotes a compact smooth manifold modelled on a d-dimensional normed vector space E, and m denotes a fixed (smooth) Riemannian metric on M. To keep the notation uncluttered, when there is no possibility of confusion we will write the induced norm on the tangent spaces simply as $\|\cdot\|$ instead of $\|\cdot\|_m$. If A is a subset of M we abbreviate

$$T_A M := \bigcup_{x \in A} T_x M.$$

DEFINITION 37.1. Let M be a compact smooth manifold. A **differentiable dynamical system** on M is an element $f \in \mathrm{Diff}^1(M)$.

As usual when the context is clear we will simply call f a "dynamical system".

REMARK 37.2. If f is a differentiable dynamical system on M, then any local representation of f (cf. Definition 35.4) is a local differentiable dynamical system in the sense of Definition 30.8.

The following key definition generalises the notion of a hyperbolic fixed point to a manifold setting.

DEFINITION 37.3. Let f be a dynamical system on M, and suppose $\Lambda \subseteq M$ is an *completely invariant* set for f (i.e. $f(\Lambda) = \Lambda$.) We say that Λ is a **hyperbolic set** for f if for each $x \in \Lambda$ the tangent space $T_x M$ splits as a direct sum

$$T_x M = E^s(x) \oplus E^u(x),$$

which is invariant for Df, i.e.

$$Df(x)E^s(x) = E^s(f(x)), \qquad Df(x)E^u(x) = E^u(f(x)),$$

and such that there exists constants $C \geq 1$ and $0 < \mu < 1$ such that

$$\|Df^k(x)v\| \leq C\mu^k \|v\|, \qquad \forall\, x \in \Lambda,\ \forall\, v \in E^s(x),\ \forall\, k \geq 0,$$

and

$$\|Df^{-k}(x)v\| \leq C\mu^k \|v\|, \qquad \forall\, x \in \Lambda,\ \forall\, v \in E^u(x),\ \forall\, k \geq 0$$

(note the notation Df^k is unambiguous, since $D(f^k) = (Df)^k$).

If Λ is a single point x then x is called a **hyperbolic fixed point**. If Λ is a periodic orbit $\mathcal{O}_f^{\mathrm{total}}(x)$ then $\mathcal{O}_f^{\mathrm{total}}(x)$ is called a **hyperbolic orbit**.

REMARKS 37.4.

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

(i) Since M is compact, the hyperbolicity of Λ is independent of the choice of Riemannian metric m. (The precise values of the constants C and μ do depend on the choice of metric m though).

(ii) If Λ is a hyperbolic set for f then it is also a hyperbolic set for f^{-1}. Any completely invariant subset of a hyperbolic set is hyperbolic. Conversely, any *finite* union of hyperbolic sets is hyperbolic (just take the maximum of the constants C, μ).

(iii) The same argument as in Remark 29.6 shows that the inequalities in the definition of $E^s(x)$ and $E^u(x)$ actually hold for all iterates $k \in \mathbb{Z}$ (and not just $k \geq 0$). Thus in particular

$$\|Df(x)^{-k}v\| \to \infty \qquad \text{for } v \in E^s(x) \text{ as } k \to \infty,$$

$$\|Df(x)^{k}v\| \to \infty \qquad \text{for } v \in E^u(x) \text{ as } k \to \infty,$$

(iv) If $x \in \Lambda$ then as $Df(x)$ is an isomorphism, the dimensions of $E^s(x)$ and $E^u(x)$ are constant along the orbit of x. In fact, the dimensions of $E^s(x)$ and $E^u(x)$ are locally constant across all of Λ. This is not immediate, and we will prove in Proposition 37.12 below.

Nothing in the definition of a hyperbolic set prevents either E^s or E^u being 0. Nevertheless, these hyperbolic sets are not very interesting, as the next proposition shows.

Proposition 37.5. *Let $\Lambda \subset M$ be a hyperbolic set for a dynamical system f. Suppose that $E^s(x) = \{0_x\}$ for each $x \in \Lambda$. Then Λ consists of finitely many periodic orbits of f. The same is true if instead $E^u(x) = \{0_x\}$ for each $x \in \Lambda$.*

The proof of Proposition 37.5 is on Problem Sheet Q. We say a periodic orbit is **expanding** if $E^s = \{0\}$ along the orbit, and **contracting** if $E^u = \{0\}$ along the orbit.

Example 37.6. If $f \colon E \to E$ is a differentiable dynamical system on a normed vector space, with $u \in E$ a hyperbolic fixed point in the sense of Definition 30.13, then if we regard E as a smooth manifold then $\{u\}$ is a hyperbolic set in the sense of Definition 37.3. Indeed, the only difference between Definition 30.13 and Definition 37.3 in this case is that in Definition 30.13 the hyperbolic splitting is of E itself, whereas in Definition 37.3 the hyperbolic splitting takes place in T_uE. However after we perform the canonical identification $T_uE \cong E$ (cf. Example 35.7), the two definitions become identical.

A much more interesting class of examples is:

Example 37.7. A hyperbolic toral automorphism $f_L \colon \mathbb{T}^2 \to \mathbb{T}^2$ is differentiable, and the entire manifold $\mathbb{T}^2$ is a hyperbolic set. On Problem Sheet Q you are asked to verify this.

Dynamical systems with the property that the entire manifold is hyperbolic get their own special name.

DEFINITION 37.8. Let $f\colon M \to M$ be a dynamical system. We say that f is **Anosov**[1] if the entire manifold M is a hyperbolic set for f.

Thus hyperbolic toral automorphisms are examples of Anosov dynamical systems.

Here is the analogue of Definition 29.9. As in the linear case, given $v \in T_x M$ for $x \in \Lambda$, we write $v = (v_s, v_u)$ to indicate the components in the stable and unstable spaces respectively.

DEFINITION 37.9. Let $\Lambda \subset M$ be a hyperbolic set for f with splitting $T_\Lambda M = E^s \oplus E^u$. Given $x \in \Lambda$ and $\varepsilon > 0$, we define the **ε-cones** about $E^s(x)$ and $E^u(x)$ by:

$$\operatorname{cone}_\varepsilon(E^s(x)) := \{v \in T_x M \mid \|v_u\| \le \varepsilon \|v_s\|\},$$

and

$$\operatorname{cone}_\varepsilon(E^u(x)) := \{v \in T_x M \mid \|v_s\| \le \varepsilon \|v_u\|\},$$

Just as in Proposition 29.10 we can alternatively characterise the splitting as follows.

PROPOSITION 37.10. *Let $\Lambda \subset M$ be a hyperbolic set for f with splitting $T_\Lambda M = E^s \oplus E^u$. For any $x \in \Lambda$, $E^s(x)$ can be alternatively characterised as:*

$$\begin{aligned} E^s(x) &= \left\{v \in T_x M \mid \|Df^k(x)v\| \to 0 \text{ as } k \to \infty\right\} \\ &= \left\{v \in T_x M \mid \exists\, r > 0,\ \|Df^k(x)v\| \le r,\ \forall k \ge 0\right\} \\ &= \left\{v \in T_x M \mid \exists\, \varepsilon > 0,\ Df^k(x)v \in \operatorname{cone}_\varepsilon(E^s(f^k(x)),\ \forall k \ge 0\right\} \end{aligned}$$

Similarly

$$\begin{aligned} E^u(x) &= \left\{v \in T_x M \mid \|Df^{-k}(x)v\| \to 0 \text{ as } k \to \infty\right\} \\ &= \left\{v \in T_x M \mid \exists\, r > 0,\ \|Df^{-k}(x)v\| \le r,\ \forall k \ge 0\right\} \\ &= \left\{v \in T_x M \mid \exists\, \varepsilon > 0,\ Df^{-k}(x)v \in \operatorname{cone}_\varepsilon(E^u(f^{-k}(x)),\ \forall k \ge 0\right\} \end{aligned}$$

In particular, the hyperbolic splitting is unique: if $T_x M = F^s(x) \oplus F^u(x)$, $x \in \Lambda$, is another hyperbolic splitting for f, then $E^s(x) = F^s(x)$ and $E^u(x) = F^u(x)$ for each $x \in \Lambda$.

Proof. We discuss E^s only. The proof is essentially the same as Proposition 29.10—one just needs to remember which tangent space the relevant vectors live in. For instance, to show that the second set of the right-hand side is contained in the third, suppose

$$v \in T_x M \setminus \left\{v \in T_x M \mid \exists\, \varepsilon > 0,\ Df^k(x)v \in \operatorname{cone}_\varepsilon(E^s(f^k(x)),\ \forall k \ge 0\right\}.$$

Then there exists $k \ge 0$ such that $w := Df^k(x)v \in T_{f^k(x)}M$ does not belong to $\operatorname{cone}_1(E^s(f^k(x))$. Thus in particular $w_u \ne 0$, and thus by part (iii) of Remark 37.4

$$\|Df^n(f^k(x))w_u\| \to \infty, \qquad |Df^n(f^k(x))w_s| \to 0,$$

[1]Named after the Russian mathematician Anosov.

as $n \to \infty$. Thus

$$\|Df^n(f^k(x))w\| \geq \|Df^n(f^k(x))w_u\| - \|Df^n(f^k(x))w_s\| \to \infty,$$

as $n \to \infty$. Thus $\{Df^k(x)v\}_{k \geq 0}$ is an unbounded sequence of numbers, and thus v does not belong to the second set on the right-hand side. ■

We will now show that $E^s(x)$ varies continuously with x (and hence also $E^u(x)$). For this to make sense, we need to introduce a topology on the set of linear subspaces.

DEFINITION 37.11. Let $1 \leq l \leq d = \dim M$. The **l-Grassmannian** space is the set

$$\mathrm{Grass}(M; l) := \{F \mid F \text{ is an } l\text{-dimensional subspace of } T_xM, \ x \in M\}.$$

Let us introduce a topology on $\mathrm{Grass}(M;l)$. Suppose (F_k) is a sequence in the Grassmannian $\mathrm{Grass}(M;l)$ and F is another element of $\mathrm{Grass}(M;l)$. We specify what it means for $F_k \to F$. Suppose $F_k \subset T_{x_k}M$ and $F \subset T_xM$. Then we first require that $x_k \to x$. Thus for k sufficiently large we may assume that x_k and x all belong to a chart $\sigma \colon U \to E$ of M. Let $\lambda \colon \pi^{-1}(U) \to U \times E$ denote the associated trivialisation of the tangent bundle over U (cf. the proof of Proposition 35.16). Via the trivialisation we can see all the subspaces F_k and F as subspaces of E. We then require that there exists a basis $(e_1, \dots, e_l)$ of F and bases $(e_1^k, \dots, e_k^k)$ of F_k such that $e_i^k \to e_i$ for each $i = 1, \dots, l$.

Since M is compact, the space $\mathrm{Grass}(M;l)$ is also compact. Moreover by definition of the topology on TM, if $F_k \to F$ then for any $(x,v) \in F$ there exists $(x_k, v_k) \in F_k$ such that $(x_k, v_k) \to (x, v)$.

PROPOSITION 37.12. *If Λ is a hyperbolic set for f then $E^s(x)$ and $E^u(x)$ vary continuously in $x \in \Lambda$. In particular, $\dim E^s(x)$ and $\dim E^u(x)$ are locally constant.*

Proof. Let $x \in \Lambda$. We prove that E^s is continuous at x. Since the Grassmannian is compact, it suffices to show that if $x_k \in \Lambda$ is any sequence such that $E^s(x_k)$ converges to some subspace $F \subset T_xM$ then $F = E^s(x)$.

Suppose $v \in F$. Then there exists $v_k \in E^s(x_k)$ such that $(x_k, v_k) \to (x, v)$. By definition of E^s, there exists $C \geq 1$ and $0 < \mu < 1$ such that

$$\|Df^n(x_k)v_k\| \leq C\mu^n \|v_k\|, \qquad \forall\, n \geq 0, \ k \geq 1.$$

Fixing n and letting $k \to \infty$ tell us that

$$\|Df^n(x)v\| \leq C\mu^n \|v\|, \qquad \forall\, n \geq 0.$$

Thus $F \subset E^s(x)$. Next, after passing to a subsequence, we may assume that $E^u(x_k) \to G$. The same proof shows that $G \subset E^u(x)$. Since $E^s(x_k) \oplus E^u(x_k) = T_{x_k}M$, it follows that $F \oplus G = T_xM$, and hence the fact that $F \subseteq E^s(x)$ and $G \subseteq E^u(x)$ implies that $F = E^s(x)$ and $G = E^u(x)$. This completes the proof. ■

The proof prompts the following definition.

Definition 37.13. Let $A \subseteq M$ be a set. A C^0 **subbundle of rank** l of $T_A M$ consists of a choice of l-dimensional subspace $S(x) \subseteq T_x M$ for each $x \in A$ such that $S(x)$ depends continuously on x in the sense of Definition 37.11.

With this definition the continuity of E^s and E^u proved in Proposition 37.12 can be succinctly stated as saying that E^s and E^u are C^0 subbundles of $T_\Lambda M$.

Proposition 37.14. *If Λ is a hyperbolic set for f then its closure $\overline{\Lambda}$ is also a hyperbolic set.*

Proposition 37.14 implies that, taking the closure if necessary, we may always assume our hyperbolic sets are compact.

Proof. Let us first check that $\overline{\Lambda}$ is completely invariant. Indeed, if $x \in \overline{\Lambda}$ then there exists $x_k \in \Lambda$ such that $x_k \to x$. Since $f(x_k) \in \Lambda$ and $f(x_k) \to f(x)$ we have $f(x) \in \overline{\Lambda}$. Thus $f(\overline{\Lambda}) \subseteq \overline{\Lambda}$. The same argument with f replaced by f^{-1} shows complete invariance.

Now let us prove that $\overline{\Lambda}$ is hyperbolic. It suffices to show that $\overline{\Lambda} \setminus \Lambda$ is hyperbolic. Let $x \in \overline{\Lambda} \setminus \Lambda$. Take a sequence $x_k \in \Lambda$ such that $E^s(x_k) \to F \subset T_x M$ and $E^u(x_k) \to G \subset T_x M$. The same argument as in the proof of Proposition 37.12 tells us that

$$\|Df^k(x)v\| \le C\mu^k \|v\|, \qquad \forall\, v \in F,\ \forall\, k \ge 0,$$

and

$$\|Df^k(x)v\| \ge \frac{1}{C}\frac{1}{\mu^k}\|v\|, \qquad \forall\, v \in G,\ \forall\, k \ge 0,$$

which implies that $F \cap G = \{0\}$. But as before, since $E^s(x_k) \oplus E^u(x_k) = T_{x_k}M$, we have $F + G = T_x M$, and hence $F \oplus G = T_x M$ is a direct sum. We can therefore *define*

$$E^s(x) := F, \qquad E^u(x) := G.$$

Since Λ is completely invariant, if $x \in \overline{\Lambda} \setminus \Lambda$ then so is the entire orbit $\mathcal{O}_f^{\mathrm{total}}(x)$. We now construct a hyperbolic splitting at each point in $\mathcal{O}_f^{\mathrm{total}}(x) \subseteq \overline{\Lambda} \setminus \Lambda$. For this we *define*

$$E^s(f^k(x)) := Df^k(x)E^s(x), \qquad E^u(f^k(x)) := Df^k(x)E^u(x).$$

Since a linear isomorphism preserves direct sum, we have $T_{f^k(x)}M = E^s(f^k(x)) \oplus E^u(f^k(x))$ for each k, and since the constants C, μ were independent of $x \in \Lambda$, the same argument as above shows that vectors in $E^s(f^k(x))$ and $E^u(f^k(x))$ satisfy the required growth/decay conditions. Thus we have constructed a hyperbolic splitting for every orbit in $\overline{\Lambda} \setminus \Lambda$. This completes the proof. ∎

We now prove that analogue of Proposition 29.11, which says that up to changing the Riemannian metric, we may always assume that $C = 1$.

Proposition 37.15. *Let f be a dynamical system on M, and suppose Λ is a hyperbolic set with splitting $T_\Lambda M = E^s \oplus E^u$. There exists a Riemannian metric m_{a} on M and $0 < \tau < 1$ such that if $\|\cdot\|_{\mathrm{a}}$ denotes the norm associated to this metric then*

$$\|Df(x)v\|_{\mathrm{a}} \le \tau \|v\|_{\mathrm{a}}, \qquad \forall\, x \in \Lambda,\ v \in E^s(x), \tag{37.1}$$

$$\|Df^{-1}(x)v\|_{\mathrm{a}} \le \tau \|v\|_{\mathrm{a}}, \qquad \forall\, x \in \Lambda,\ v \in E^u(x), \tag{37.2}$$

Such a Riemannian metric m_{a} is said to be **adapted** to f and Λ.

Proof. Let $C \geq 1$ and $0 < \mu < 1$ denote the original constants. Choose n large enough so that $C\mu^n < 1$. We define a C^0 Riemannian metric $\widehat{m}$ on M via

$$\widehat{m}_x(v, w) \coloneqq \sum_{k=0}^{n-1} m_{f^k(x)}\big(Df^k(x)v, Df^k(x)w\big), \qquad \forall\, v, w \in T_xM,\ \forall\, x \in M.$$

Let us first check that $\widehat{m}$ satisfies the requirements of the Proposition. The proof is essentially identical to that of Proposition 29.11, apart from the fact that there are squares everywhere because we are working with inner products instead of norms. Then $\|\cdot\|_{\mathrm{a}}$ is obviously a norm on E. Setting $\alpha \coloneqq \sum_{k=0}^{n-1} C^2\mu^{2k}$, one has

$$\|v\|^2_{\widehat{m}} \leq \alpha \|v\|^2_m, \qquad \forall\, v \in E^s(x),$$

and similarly

$$\|v\|^2_{\widehat{m}} \leq \alpha \|Df^{n-1}(x)v\|^2_m, \qquad \forall\, v \in E^u(x).$$

Now suppose $v \in E^s(x)$. Then

$$\begin{aligned}
\|Df(x)v\|^2_{\widehat{m}} &= \|v\|^2_{\widehat{m}} - \|v\|^2_m + \|Df^n(x)v\|^2_m \\
&\leq \|v\|^2_{\widehat{m}} - \left(1 - C^2\mu^{2n}\right)\|v\|^2_m \\
&\leq \left(1 - \frac{1}{\alpha}\left(1 - C^2\mu^{2n}\right)\right)\|v\|^2_{\widehat{m}}.
\end{aligned}$$

Similarly if $v \in E^u(x)$ one has

$$\begin{aligned}
\|Df^{-1}(x)v\|^2_{\widehat{m}} &= \|v\|^2_{\widehat{m}} + \|Df^{-1}(x)v\|^2_m - \|Df^{n-1}(x)v\|^2_m \\
&\leq \|v\|^2 - \left(1 - C^2\mu^{2n}\right)\|Df^{n-1}(x)v\|^2_m \\
&\leq \left(1 - \frac{1}{\alpha}\left(1 - C^2\mu^{2n}\right)\right)\|v\|^2_{\widehat{m}}.
\end{aligned}$$

Set

$$\tau' \coloneqq \left(1 - \frac{1}{\alpha}\left(1 - C^2\mu^{2m}\right)\right).$$

Since $\alpha \geq 1$ one has $0 < \tau' < 1$, and (37.1) and (37.2) hold for $\|\cdot\|_{\widehat{m}}$ and τ'.

Finally, we approximate $\widehat{m}$ by a smooth Riemannian metric m_{a} (cf. part (vi) of Remarks 36.2.) This new metric will still satisfy (37.1) and (37.2), but now for a slightly larger $\tau > \tau'$. Nevertheless for a sufficiently good approximation we will still have $\tau < 1$. This completes the proof. ∎

As in the linear case, going forward we will always assume that the process outlined in Proposition 37.15 has already been carried out, and write simply m instead of m_{a} for the metric.

Definition 37.16. If m is an adapted metric to f and Λ, and $\|\cdot\|$ denotes the associated norm, then we can speak of the **skewness** of m with respect to f and Λ.

$$\tau(f, \Lambda) \coloneqq \sup_{x \in \Lambda} \left\{ \left\|Df(x)|_{E^s(x)}\right\|^{\mathrm{op}}, \left\|Df^{-1}(x)|_{E^u(x)}\right\|^{\mathrm{op}} \right\}.$$

Persistence of Hyperbolic Sets

In this lecture we extend the persistence results proved in Lecture 31 to a manifold setting. Let us first introduce the notion of a norm of **box type** in this setting.

DEFINITION 38.1. Suppose $\Lambda \subseteq M$ is any set and $T_\Lambda M$ has a splitting into two C^0 subbundles F and G, so that

$$T_x M = F(x) \oplus G(x), \qquad \forall\, x \in \Lambda.$$

A norm $\|\cdot\|$ on $T_\Lambda M$ is said to be of **box type** with respect to the C^0 subbundles F and G if

$$\|v\| = \max\big\{\|v_F\|, \|v_G\|\big\}, \qquad \forall\, v \in T_\Lambda M,$$

where v_F and v_G are the components of v in this splitting. As in the linear case, it is easy to make a box-type norm: if $\|\cdot\|$ is any norm then clearly the norm $\|\cdot\|_{\mathrm{b}}$ defined by

$$\|v\|_{\mathrm{b}} \coloneqq \max\big\{\|v_F\|, \|v_G\|\big\}, \qquad \forall\, v \in T_\Lambda M,$$

is of box-type. On calls $\|\cdot\|_{\mathrm{b}}$ the **box-adjusted norm** of $\|\cdot\|$.

REMARK 38.2. The box-adjusted norm $\|\cdot\|_{\mathrm{b}}$ will only be of class C^0 (since the subbundles are only assumed to be of class C^0). It is important to realise this trick only works for norms, not metrics: a box-adjusted norm cannot be induced from any (even C^0) metric, since it violates the parallelogram law (unless either F or G is zero). Moreover such a box-adjusted norm typically cannot be extended to be defined on the entire manifold. This means that using box type norms in this setting is rather less useful than in the linear setting. We will only use them as tools in the "middle" of proofs.

If f is a dynamical system and Λ is a hyperbolic set, then if m is a Riemannian metric adapted to f, we can form a box-adjusted norm from the norm determined by m. This will again satisfies equations (37.1) and (37.2) from the previous lecture, and will have the same skewness.

DEFINITION 38.3. Suppose f is a dynamical system and Λ is a compact invariant set. Suppose moreover that we are given two C^0 subbundles F and G of $T_\Lambda M$. Let us abbreviate by $\mathcal{V}(F, G; f)$ the set of bounded continuous maps $\Phi \colon F \to G$ with the property that (f, Φ) is a C^0 vector bundle morphism[1] from F to G. We make $\mathcal{V}(F, G; f)$ into a Banach space with the operator norm

$$\|\Phi\|_0 \coloneqq \sup_{x\in\Lambda} \|\Phi_x\|^{\mathrm{op}} = \sup_{x\in\Lambda} \Big\{ \|\Phi_x v\| \mid v \in F(x),\ \|v\| = 1 \Big\},$$

[1] Thus $\Phi_x \colon F(x) \to G(f(x))$ is a linear map for each $x \in \Lambda$.

note by assumption $\|\Phi\|^{\mathrm{op}} < \infty$ since Φ is bounded. Given $r > 0$ we set[2]

$$\mathcal{V}_r(F, G; f) := \{\Phi \in \mathcal{V}(F, G; f) \mid \|\Phi\|_0 \leq r\}.$$

REMARK 38.4. If $\Psi \in \mathcal{V}(F, G; \mathrm{id})$ then the graph of Ψ is another C^0 subbundle:

$$\mathrm{gr}(\Psi) := \bigcup_{x \in \Lambda} \mathrm{gr}(\Psi_x) = \bigcup_{x \in \Lambda} \{(v, \Psi_x v) \mid v \in F(x)\}.$$

The next result is essentially a duplicate of Proposition 31.9.

PROPOSITION 38.5. *Suppose $f\colon M \to M$ is a dynamical system on a smooth compact manifold M and $\Lambda \subseteq M$ is a compact invariant set for f. Assume we are given:*

- *A continuous map $\Phi\colon T_\Lambda M \to T_\Lambda M$ such that (f, Φ) is a vector bundle isomorphism.*
- *Two C^0 subbundles F, G of $T_\Lambda M$ such that $T_\Lambda M = F \oplus G$.*

Let $\|\cdot\|$ be a C^0 norm on $T_\Lambda M$ which is of box type with respect to $F \oplus G$, and write Φ in matrix form as

$$\Phi = \begin{pmatrix} A & B \\ C & D \end{pmatrix} : F \oplus G \to F \oplus G.$$

Suppose there exist two constants $\lambda, \varepsilon > 0$ such that

$$\lambda + \varepsilon < 1$$

and

$$\max\left\{\|A^{-1}\|_0, \|D\|_0\right\} < \lambda,$$

and

$$\max\{\|B\|_0, \|C\|_0\} < \varepsilon,$$

Then there is a unique linear map $\Psi = \Psi_\Phi \in \mathcal{V}_1(F, G; \mathrm{id})$ such that the C^0 subbundle $\mathrm{gr}(\Psi)$ is Φ-invariant:

$$\Phi_x\big(\mathrm{gr}(\Psi_x)\big) = \mathrm{gr}(\Psi_{f(x)}), \qquad \forall\, x \in \Lambda.$$

Moreover for all $x \in \Lambda$ and $u \in \mathrm{gr}(\Psi_x)$ one has

$$\|\Phi_x u\| \geq \left(\frac{1}{\lambda} - \varepsilon\right)\|u\|.$$

Finally, Ψ_Φ—and hence also $\mathrm{gr}(\Psi_\Phi)$—depend continuously on Φ.

[2]It would be more notationally consistent to write this $\mathcal{V}(F, G; f)(r)$, but a double pair of parentheses is visually unpleasant.

(♣) *Proof.* The proof is literally word-for-word identical to that of Proposition 31.9, apart from the fact that everything has a subscript "x" appended to it to indicate the base point. Thus for instance the first step in the proof is to show that the desired map Ψ can be found as a fixed point of the map

$$\mathcal{X} = \mathcal{X}_\Phi \colon \mathcal{V}_1(F, G; \mathrm{id}) \to \mathcal{V}(F, G; \mathrm{id})$$

defined as follows: suppose $\Psi \in \mathcal{V}_1(F, G; \mathrm{id})$, $x \in \Lambda$ and $u \in F(f(x))$. Then $\mathcal{X}(\Psi)_{f(x)}u$ is the vector in $G(f(x))$ given by

$$\mathcal{X}(\Psi)_{f(x)}u = (C_x + D_x\Psi_x) \circ (A_x + B_x\Psi_x)^{-1}(u).$$

If we agree to omit the basepoints from our notation (which is harmless, since by assumption all our maps are fibre-preserving), the above formula would read

$$\mathcal{X}(\Psi) = (C + D\Psi)(A + B\Psi)^{-1},$$

which is then identical to the formula (31.8) in the proof of Proposition 31.9. As usual, such a fixed point is found by applying the Banach Fixed Point Theorem, after verifying that $\mathcal{X}$ is a contraction. The rest of the proof is formally identical to that of Proposition 31.9, and is thus omitted. ■

Assume M is a compact smooth manifold. Let us now explain how to define a topology[3] on $\mathrm{Diff}^1(M)$. We proceed in three steps.

(i) Fix an arbitrary metric d on M that defines the topology on M. For instance, d could be a metric induced from a Riemannian metric m on M, cf. Theorem 36.7). This induces a metric d_0 on $C^0(M, M)$ via

$$d_0(f, g) := \sup_{x \in M} d(f(x), g(x)).$$

(ii) Now cover M by finitely many charts (this is possible as M is compact). Choose a function $\delta \colon M \to (0, \infty)$ such that for every $x \in M$ the ball $B_d(x, 2\delta(x))$ of radius $2\delta(x)$ is contained in the domain of a chart. Compactness implies there exist points $x_1, \dots, x_l$ in M such that the balls $B_d(x_i, \delta(x_i))$ for $i = 1, \dots, l$ cover M. Set

$$\delta := \min_{i=1,\dots,l} \delta(x_i).$$

(iii) Then if $f, g \in \mathrm{Diff}^1(M)$ satisfy $d_0(f, g) < \delta$, then for every $x \in M$ there is a chart containing both $f(x)$ and $g(x)$. We can therefore define the C^1 distance $d_1(f, g)$ between f and g by taking the maximum C^1 distance in the sense of (30.2) between the corresponding local representations of f and g in these charts (35.2).

[3] A similar definition gives a topology on $\mathrm{Diff}^p(M)$ for all $p \geq 1$, although we won't need this.

Since M is compact, a different set of choices would yield an equivalent metric[4]. As hinted at in Example 35.18, the space $\mathrm{Diff}^1(M)$ with this topology can be given the structure of a Banach manifold modelled on $\Gamma^1(M, TM)$. We won't need this, though.

Here is the analogue of Proposition 31.2 in this setting.

PROPOSITION 38.6. *Let f be a dynamical system on M and suppose Λ is a compact invariant hyperbolic set. Then there exists a C^1 neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and a number $a > 0$ such that for any $g \in \mathcal{U}$ and any compact g-invariant set Δ with*

$$\Delta \subset B(\Lambda, a) := \{x \in M \mid d(x, \Lambda) < a\},$$

the set Δ is a hyperbolic set for g. Moreover as $g \xrightarrow{C^1} f$ and $x \in \Delta$ approaches $y \in \Lambda$, the stable subspace $E^s(x, g)$ approaches the stable subspace $E^s(y, f)$ in the sense of Definition 37.11, and similarly for the unstable subspaces.

REMARK 38.7. Note that Proposition 38.6 is *not* asserting the *existence* of a hyperbolic invariant set, merely that *if* any such set exists, it is necessarily hyperbolic. We will show later in Corollary 41.8 on that for g sufficiently close to f, such a set always does in fact always exist, but this is much harder to prove. This should be contrasted to the linear case, where we first proved the existence of a fixed point (Proposition 30.14) and only later on proved that the fixed point is hyperbolic (Proposition 31.2). Here it is the other way round.

Proof of Proposition 38.6. Let $T_\Lambda M = E^s \oplus E^u$ denote the hyperbolic splitting of Λ. We may assume that the Riemannian metric m is adapted to f and Λ. Since Λ is compact, the C^0 splitting of $T_\Lambda M$ extends to define a C^0 splitting of $T_U M$ for some neighbourhood U of Λ. Call this splitting $F^s \oplus F^u$, so that $E^s(x) = F^s(x)$ if $x \in \Lambda$, and similarly for the unstable spaces. Note however that if $x \in U \setminus \Lambda$, then this splitting is not preserved by Df.

Now let $\|\cdot\|_{\mathrm{b}}$ denote the box-adjusted norm of $\|\cdot\|$. Thus $\|\cdot\|_{\mathrm{b}}$ is defined on $T_U M$ and is both adapted and of box type with respect to the hyperbolic splitting of $T_\Lambda M$. Let $0 < \tau < 1$ denote the skewness of f and Λ with respect to $\|\cdot\|$ (and hence also $\|\cdot\|_{\mathrm{b}}$.) Then on $T_\Lambda M$, Df is represented[5] by

$$Df = \begin{pmatrix} (Df)_{uu} & 0 \\ 0 & (Df)_{ss} \end{pmatrix}$$

with $\left\|(Df)_{uu}^{-1}\right\|_{\mathrm{b}} \leq \tau$ and $\left\|(Df)_{ss}\right\|_{\mathrm{b}} \leq \tau$.

Take $\tau < \lambda < 1$ and $0 < \varepsilon < 1 - \lambda$. If $\mathcal{U}$ and $a > 0$ are sufficiently small, then for any invariant set $\Delta \subset B(\Lambda, a) \subset U$ of any $g \in \mathcal{U}$, the four block bundle homomorphisms of Dg and Dg^{-1} represented under $T_\Delta M = F^s \oplus F^u|_\Delta$, with respect to $\|\cdot\|_{\mathrm{b}}$, satisfy the hypotheses of Proposition 38.5. The proof is now completed in exactly the same way as the proof of Proposition 31.2 was, and thus the details are omitted. ■

[4]We used compactness of M at every stage in the definition of d_1. With a bit more work the definition can be extended to non-compact manifolds as well, however in this case the metric d_1 is not independent of the choices made.

[5]Here we write the splitting as $E^u \oplus E^s$ in order to fit in with the notation from Proposition 38.5

As remarked at the beginning of this lecture, using norms of box type can sometimes be a little unfortunate, since these cannot come from Riemannian metrics. It is therefore useful to be able to control the difference between a norm of box type and one induced from the Riemannian metric.

DEFINITION 38.8. Let (M, m) be a Riemannian manifold. Suppose $\Lambda \subseteq M$ is a subset and F, G are two C^0 subbundles of $T_\Lambda M$ such that $T_\Lambda M = F \oplus G$. For $x \in \Lambda$, set

$$\measuredangle_m\big(F(x), G(x)\big) := \inf\left\{\measuredangle_{m_x}(u, v) \mid u \in F(x) \setminus \{0_x\},\ v \in G(x) \setminus \{0_x\}\right\},$$

where $\measuredangle_{m_x}(u, v)$ is measured using the inner product m_x. We call $\measuredangle_m\big(F(x), G(x)\big)$ the **angle** between the subspaces $F(x)$ and $G(x)$. We set

$$\measuredangle_m(F, G) := \inf_{x \in \Lambda} \measuredangle_m\big(F(x), G(x)\big),$$

and call $\measuredangle_m(F, G)$ the **angle** between the subbundles F and G.

Since F and G are continuous subbundles, the function $x \mapsto \measuredangle_m\big(F(x), G(x)\big)$ is also continuous. Thus if Λ is compact and $\measuredangle_m(F(x), G(x)) > 0$ for all $x \in \Lambda$ then so is $\measuredangle_m(F, G)$. This implies:

LEMMA 38.9. *Let $\Lambda \subseteq M$ be a compact hyperbolic set of a dynamical system f, with hyperbolic splitting $T_\Lambda M = E^s \oplus E^u$. Then for any Riemannian metric m on M, one has $\measuredangle_m(E^s, E^u) > 0$.*

Now we use the following piece of Euclidean geometry.

PROPOSITION 38.10. *Suppose E is a finite-dimensional normed vector space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$. For any $\delta > 0$, there exists $c \geq 1$ such that if $E = F \oplus G$ is a direct sum and*[6] *$\measuredangle(F, G) > \delta$, then if $\| \cdot \|_{\mathrm{b}}$ denotes the box-adjusted norm from $\| \cdot \|$ with respect to $F \oplus G$ then*

$$\frac{1}{c}\|v\| \leq \|v\|_{\mathrm{b}} \leq c\|v\|, \qquad \forall v \in E.$$

The content of the Proposition is that the constant c depends only on δ. The proof of Proposition 38.10 is on Problem Sheet R.

Now suppose Λ is a compact hyperbolic set of f. By a slight abuse of notation, let us write $\| \cdot \|_\Lambda$ for the box-adjusted norm for the splitting of $T_\Lambda M$. Thus $\| \cdot \|_\Lambda$ is only defined on $T_\Lambda M$. We now prove the following enhancement of Proposition 38.6.

PROPOSITION 38.11. *Let f be a dynamical system on M and suppose Λ is a compact invariant hyperbolic set. Let m denote an arbitrary Riemannian metric on M (not necessarily adapted to f and Λ). There exists a C^1 neighbourhood $\mathcal{U}$ of f in* $\mathrm{Diff}^1(M)$ *and numbers $a > 0$ and $c \geq 1$ such that for any $g \in \mathcal{U}$ and any compact*

[6] Here $\measuredangle$ is measured with respect to the inner product $\langle \cdot, \cdot \rangle$.

g-invariant set Δ with $\Delta \subset B(\Lambda, a)$, not only is Δ hyperbolic but the box-adjusted norm $\|\cdot\|_\Delta$ is equivalent to $\|\cdot\|$ with constant c, i.e.

$$\frac{1}{c}\|v\| \leq \|v\|_\Delta \leq c\|v\|, \qquad \forall\, v \in T_\Delta M. \tag{38.1}$$

Moreover if the Riemannian metric m is adapted to f, then for any $\varepsilon > 0$ there exists a neighbourhood $\mathcal{U}_\varepsilon \subset \mathcal{U}$ and a number $0 < a_\varepsilon \leq a$ such that if $\Delta \subset B(\Lambda, a_\varepsilon)$ is a compact invariant set for $g \in \mathcal{U}_\varepsilon$, then m is also adapted to g and

$$\tau(g, \Delta) \leq \tau(f, \Lambda) + \varepsilon.$$

Proof. By Proposition 38.6, if g is C^1 close enough to f and Δ is close enough to Λ, then for every $x \in \Delta$ there exists $y \in \Lambda$ such that $E^s(x, g) \oplus E^u(x, g)$ is close to $E^s(y, f) \oplus E^u(y, f)$. Thus $\measuredangle_m(E^s(x, f), E^u(x, g))$ is close to $\measuredangle_m(E^s(y, f), E^u(y, f))$, and hence is positive (c.f. the discussion before Proposition 38.10.) Thus by Proposition 38.10 there exists a constant $c \geq 1$, independent of Δ, such that (38.1) holds.

For the last statement, it is clear that if g is close enough to f then m will also be adapted to g and Δ. The statement about the skewness then follows immediately from the continuity statement in Proposition 38.6 about the stable and unstable spaces as g approaches f. This completes the proof. ∎

Lifting Dynamical Systems

Let M be a compact manifold and let m be a Riemannian metric on M with exponential map $\exp\colon TM \to M$. We denote by r_m the injectivity radius of m (cf. Theorem 36.16). Given a dynamical system f on M, there exists $r_* = r_*(f, m) > 0$ such that

$$d(x,y) \le r_* \qquad \Rightarrow \qquad d(f(x), f(y)) < r_m, \qquad \forall x, y \in M. \tag{39.1}$$

Definition 39.1. We define the **lifting** of f, written $\widehat{f}$, to be the map

$$\widehat{f}\colon TM(r_*) \to TM, \qquad \widehat{f}(x,v) := \exp^{-1}_{f(x)}\Big(f\big(\exp_x(v)\big)\Big).$$

This map is well defined thanks to the choice of r_* and Theorem 36.16, and of class C^1, being the composition of C^1 maps. Thus the following commutes (where defined)

$$\begin{tikzcd} TM \arrow[r, "\widehat{f}"] \arrow[d, "\exp"'] & TM \arrow[d, "\exp"] \\ M \arrow[r, "f"'] & M \end{tikzcd}$$

We should really write $\widehat{f}_m$ since the map $\widehat{f}$ depends on the choice of Riemannian metric m. Nevertheless, in keeping with the rest of the notation, we will typically not do this.

The map $\widehat{f}$ is fibre-preserving over f. If one thinks of $\exp_x^{-1}(y)$ as being the "vector from x to y" inside T_xM, then $\widehat{f}$ carries the vector from x to y to the vector from $f(x)$ to $f(y)$:

$$\widehat{f}(\exp_x^{-1}(y)) = \exp^{-1}_{f(x)}(f(y)). \tag{39.2}$$

It follows from (39.2) that

$$\big\|\widehat{f}(\exp_x^{-1}(y))\big\| = d(f(x), f(y)). \tag{39.3}$$

Note that if

$$d(f^i(x), f^i(y)) \le r_*, \qquad \forall\, i = 1, \dots, p,$$

then $\widehat{f}^p(\exp_x^{-1}(y))$ is defined, and satisfies

$$\big\|\widehat{f}^p(\exp_x^{-1} y)\big\| = d(f^p(x), f^p(y)). \tag{39.4}$$

Remark 39.2. **Warning:** The map $\widehat{f}$ is typically *not* linear on the fibres. Thus $(f, \widehat{f})$ is *not* a vector bundle morphism.

Definition 39.3. Suppose $\Phi\colon TM(r) \to TM$ is a (not necessarily linear) fibre-preserving map over f. If we restrict Φ to one fibre, we get a map between two linear spaces:

$$\Phi_x\colon T_xM(r) \to T_{f(x)}M.$$

Assume this map is differentiable: then we can take the derivative of this map at any point[1] $u \in T_xM(r)$:

$$D\Phi_x(u)\colon T_xM \to T_{f(x)}M.$$

We call this map the **fibre derivative** of Φ at the point (x, u), and write

$$D_{\text{fib}}\Phi(x, u) := D\Phi_x(u).$$

If the map $(x, u) \mapsto D_{\text{fib}}\Phi(x, u)$ is continuous, we say that Φ is **continuously differentiable on the fibres**, or C^1 on the fibres.

Where possible we will omit the base-point x from the notation and just write $D_{\text{fib}}\Phi(u)$.

Remarks 39.4.

(i) If Φ is C^1 then Φ is also C^1 on the fibres. The converse is not true: any C^0 vector bundle morphism is automatically C^1 on the fibres (actually C^∞), since a linear map is always differentiable.

(ii) Moreover if Φ is a vector bundle morphism then since the derivative of a linear map is the linear map itself, we have $D_{\text{fib}}\Phi = \Phi$.

(iii) If Φ is C^1 on the fibres then $D_{\text{fib}}\Phi$ is a C^0 vector bundle morphism (even when Φ is not itself a vector bundle morphism).

The fibre derivative of $\widehat{f}$ at 0_x is not hard to guess:

Lemma 39.5. *Let $f\colon M \to M$ be a dynamical system. Then $\widehat{f}(0_x) = 0_{f(x)}$ and*

$$D_{\text{fib}}\widehat{f}(0_x) = Df(x).$$

Proof. We have

$$\begin{aligned}\widehat{f}(0_x) &= \exp_{f(x)}^{-1}\left(f(\exp_x(0_x))\right)\\ &= \exp_{f(x)}^{-1}(f(x))\\ &= 0_{f(x)}.\end{aligned}$$

Differentiating via the chain rule gives

$$\begin{aligned}D_{\text{fib}}\widehat{f}(x, v) &= D\left(\exp_{f(x)}^{-1} f \exp_x\right)(v)\\ &= D(\exp_{f(x)}^{-1})(f(\exp_x(v)) \circ Df(\exp_x(v)) \circ D\exp_x(v).\end{aligned}$$

[1] To fit in with the manifold formalism, we are implicitly using the canonical identification $T_uT_xM \cong T_xM$ from Example 35.7.

Since $D\exp_x(0_x) = \mathrm{id}$ by Lemma 36.15, we have

$$\begin{aligned} D_{\mathrm{fib}}\widehat{f}(0_x) &= D\left(\exp_{f(x)}^{-1} f \exp_x\right)(0_x) \\ &= \mathrm{id}\,|_{T_{f(x)}M} \circ Df(x) \circ \mathrm{id}\,|_{T_xM} \\ &= Df(x). \end{aligned}$$

This completes the proof. ∎

Recall in Lecture 30 we looked for fixed points of maps of the form $L+\phi$ where L had a hyperbolic fixed point and ϕ was Lipschitz small. In the Hartman-Grobman Theorem 32.2 we applied this with $L = Df$ and $\phi = f - Df$ (so that $L + \phi = f$). Here we will do something similar, only with $\widehat{f}$ in place of f. Set

$$\Phi_f := \widehat{f} - Df \colon TM(r_*) \to TM.$$

The map Φ_f is fibre-preserving over f and satisfies

$$\Phi_f(0_x) = 0_{f(x)}, \qquad D_{\mathrm{fib}}\Phi_f(0_x) = 0, \qquad \forall\, x \in M. \tag{39.5}$$

The map Φ_f is only C^0, even though $\widehat{f}$ is C^1, since Df is only C^0. However the restriction of Φ_f to each fibre is C^1, and the fibre derivative $D_{\mathrm{fib}}\Phi_f$ is continuous on $TM(r_*)$. Indeed, this is because Df restricted to a fibre is a linear map, and hence of class C^∞. In particular, Φ_f is Lipschitz on each fibre.

Definition 39.6. Suppose $\Phi\colon TM(r) \to TM$ is a (not necessarily linear) fibre-preserving map over f. Assume that the map

$$\Phi_x \colon T_xM(r) \to T_{f(x)}M$$

is Lipschitz for each x. We define the **fibre-Lipschitz constant** of Φ, written as $\mathrm{lip}_{\mathrm{fib}}(\Phi)$, to be the number

$$\mathrm{lip}_{\mathrm{fib}}(\Phi) := \sup_{x\in M} \mathrm{lip}(\Phi_x).$$

The next result is the analogue of Proposition 30.12 in this setting.

Proposition 39.7. *Let $f\colon M \to M$ be a smooth dynamical system. For any $\varepsilon > 0$ there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and $r > 0$ such that for any $g \in \mathcal{U}$ one has $\mathrm{lip}_{\mathrm{fib}}(\Phi_g) \leq \varepsilon$ on $TM(r)$, where $\Phi_g = \widehat{g} - Dg$.*

Proof. Since $D_{\mathrm{fib}}\widehat{f}$ is continuous on $TM(r_*)$ and since $D_{\mathrm{fib}}\Phi_f(0_x) = 0$ and M is compact, for any $\varepsilon > 0$ there exists $r > 0$ such that for any $v \in TM(r)$ one has $\left\| D_{\mathrm{fib}}\Phi_f(v) \right\|^{\mathrm{op}} \leq \frac{\varepsilon}{2}$. Using compactness of $TM(r)$ again, there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ such that for any $g \in \mathcal{U}$ and $v \in TM(r)$, one has $\left\| D_{\mathrm{fib}}\Phi_g(v) \right\|^{\mathrm{op}} \leq \varepsilon$. Applying the Mean Value Theorem 30.11 to the fibres, we obtain $\mathrm{lip}_{\mathrm{fib}}(\Phi_g) \leq \varepsilon$. ∎

Now suppose f is a dynamical system on M and $\Lambda \subseteq M$ is a compact invariant set. As usual we write $\Gamma^0(\Lambda, T_\Lambda M)$ for continuous vector fields on Λ, which is a Banach space when endowed with the C^0 norm

$$\|\gamma\|_0 := \sup_{x \in \Lambda} \|\gamma(x)\|.$$

If $\Phi \colon T_\Lambda M \to T_\Lambda M$ is a fibre-preserving map over f, we say a section γ is **Φ-invariant** if

$$\Phi_x(\gamma(x)) = \gamma(f(x)), \qquad \forall\, x \in \Lambda, \tag{39.6}$$

i.e. that the following commutes

$$\begin{tikzcd} T_\Lambda M \arrow[r, "\Phi"] & T_\Lambda M \\ \Lambda \arrow[u, "\gamma"] \arrow[r, "f"'] & \Lambda \arrow[u, "\gamma"'] \end{tikzcd}$$

We denote by $0_\Lambda \in \Gamma^0(\Lambda, T_\Lambda M)$ the zero section:

$$0_\Lambda(x) = 0_x, \qquad \forall\, x \in \Lambda.$$

If we omit basepoints from the notation the invariance condition becomes simply

$$\Phi(\gamma) = \gamma$$

which explains the name "invariant". The following result is the analogue of Proposition 30.18 in this setting. This result will be crucial in our proof of the stability of hyperbolic sets in a few lecture's time.

Proposition 39.8. *Let f be a dynamical system on M, and let Λ be a compact hyperbolic set of f with splitting $T_\Lambda M = E^s \oplus E^u$. Assume m is adapted to f and Λ, and let $\|\cdot\|$ denote a C^0 box adjusted norm on $T_\Lambda M$, with skewness $\tau = \tau(f, \Lambda)$. Fix $r > 0$, and suppose $\Phi \colon T_\Lambda M(r) \to T_\Lambda M$ is a continuous fibre-preserving map over f which is fibrewise Lipschitz with*

$$\operatorname{lip}_{\mathrm{fib}}(\Phi) < 1 - \tau. \tag{39.7}$$

Then $Df + \Phi$ has at most one invariant section. If in addition

$$\sup_{x \in \Lambda} \|\Phi_x(0_x)\| \le \big(1 - \tau - \operatorname{lip}_{\mathrm{fib}}(\Phi)\big) r \tag{39.8}$$

then $Df + \Phi$ has at least one invariant section (which is thus unique). Denoting this section by γ_Φ, one has

$$\|\gamma_\Phi\|_0 \le \frac{\sup_{x \in \Lambda} \|\Phi_x(0_x)\|}{1 - \tau - \operatorname{lip}_{\mathrm{fib}}(\Phi)}. \tag{39.9}$$

Proof. We wish to solve the equation

$$(Df + \Phi) \circ \gamma = \gamma \circ f$$

for $\gamma \in \Gamma^0(\Lambda, T_\Lambda M(r))$, that is,

$$(Df + \Phi)\gamma(x) = \gamma(f(x)), \qquad \forall x \in \Lambda.$$

To keep the notation simple we will write L instead of Df and omit the base-point x from the notation. Writing the previous equation in components using the hyperbolic splitting this becomes

$$L_{ss}\gamma(x) + \Phi_s(\gamma(x)) = \gamma_s(f(x)), \qquad L_{uu}\gamma(x) + \Phi_u\gamma(x) = \gamma_u(f(x)),$$

which can be rewritten as

$$L_{ss}\gamma_s(f^{-1}(x)) + \Phi_s\gamma(f^{-1}(x)) = \gamma_s(x), \qquad L_{uu}^{-1}\gamma_u(f(x)) - L_{uu}^{-1}\Phi_u\gamma(x) = \gamma_u(x).$$

Thus we consider the map $\mathcal{X} = \mathcal{X}_\Phi \colon \Gamma^0(\Lambda, T_\Lambda M(r)) \to \Gamma^0(\Lambda, T_\Lambda M)$ given by

$$\mathcal{X}(\gamma)(x) = \Big(L_{ss}\gamma_s(f^{-1}(x)) + \Phi_s\gamma(f^{-1}(x)), L_{uu}^{-1}\gamma_u(f(x)) - T_{uu}^{-1}\Phi_u\gamma(x)\Big)$$

for $x \in \Lambda$. Thus γ is an invariant section of $L + \Phi$ if and only if γ is a fixed point of $\mathcal{X}$. Just as in the proof of Proposition 30.18 the trick now is to show that $\mathcal{X}$ is a strict contraction if (39.7) holds. By definition

$$\|\mathcal{X}(\gamma) - \mathcal{X}(\zeta)\|_0 = \sup_{x\in\Lambda} \|\mathcal{X}(\gamma)(x) - \mathcal{X}(\zeta)(x)\|.$$

Arguing as in the proof of Proposition 30.18, we see that the E^s component is less than or equal to

$$\sup_{x\in\Lambda} \Big(\tau\big\|\gamma_s(f^{-1}(x)) - \zeta_s(f^{-1}(x))\big\| + \mathrm{lip}_{\mathrm{fib}}(\Phi)\big\|\gamma(f^{-1}(x)) - \zeta(f^{-1}(x))\big\|\Big),$$

and similarly the E^u component is less than or equal to

$$\sup_{x\in\Lambda} \Big(\tau\big\|\gamma_u(f(x)) - \zeta_u(f(x))\big\| + \tau\,\mathrm{lip}_{\mathrm{fib}}(\Phi)\big\|\gamma(x) - \zeta(x)\big\|\Big),$$

and hence

$$\|\mathcal{X}(\gamma) - \mathcal{X}(\zeta)\|_0 \le \big(\tau + \mathrm{lip}_{\mathrm{fib}}(\Phi)\big)\|\gamma - \zeta\|_0,$$

which on account of (39.7) shows that $\mathcal{X}$ is indeed a contraction.

Now assume (39.8). We will show that $\mathcal{X}$ maps $\Gamma^0(\Lambda, T_\Lambda M(r))$ into itself, whence the desired fixed point follows from the Banach Fixed Point Theorem 30.17. For this we note that

$$\begin{aligned}\|\mathcal{X}(0_\Lambda)(x)\| &= \big\|\big(\Phi_s(0_{f^{-1}(x)}), -L_{uu}^{-1}\Phi_u(0_x)\big)\big\| \\ &\le \sup_{x\in\Lambda} \|\Phi(0_x)\|.\end{aligned}$$

Now fix $\gamma \in \Gamma^0(\Lambda, T_\Lambda M(r))$ and argue:

$$\begin{aligned}\|\mathcal{X}(\gamma)\|_0 &\le \|\mathcal{X}(0_\Lambda)\|_0 + \|\mathcal{X}(\gamma) - \mathcal{X}(0_\Lambda)\|_0 \\ &\le \sup_{x \in \Lambda} \|\Phi(0_x)\| + \big(\tau + \mathrm{lip}_{\mathrm{fib}}(\Phi)\big)\|\gamma\|_0 \\ &\le r.\end{aligned}$$

This proves the existence of a unique fixed point γ_Φ of $\mathcal{X}$. Moreover the calculation above tells us that

$$\|\gamma_\Phi\|_0 \le \sup_{x \in \Lambda} \|\Phi(0_x)\| + \big(\tau + \mathrm{lip}_{\mathrm{fib}}(\Phi)\big)\|\gamma_\Phi\|_0,$$

and hence (39.9) holds. This completes the proof. ■

The attentive reader will have noticed this proof was formally identical to the proof of Proposition 30.18. In fact, Proposition 39.8 can be deduced from the (infinite dimensional analogue of) Proposition 30.18, thanks to the following construction.

DEFINITION 39.9. Let $f \colon M \to M$ be a dynamical system, and let $\Lambda \subseteq M$ denote a compact completely invariant set. Define a linear operator L_f on the Banach space $\Gamma^0(\Lambda, T_\Lambda M)$ by

$$(L_f \gamma)(x) \coloneqq Df(f^{-1}(x))\gamma(f^{-1}(x)).$$

Then we have the following result, which was originally due to Mather.

PROPOSITION 39.10. *Let $f \colon M \to M$ be a dynamical system, and let $\Lambda \subseteq M$ denote a compact completely invariant set. Then Λ is hyperbolic if and only if L_f is a hyperbolic linear dynamical system.*

The proof of Proposition 39.10 is on Problem Sheet R. The importance of this result should not be understated: it tells us that the study of hyperbolic sets can be reduced to the study of linear hyperbolic dynamics, at the expense of passing from a finite dimensional manifold to the infinite dimensional Banach space of vector fields. However we will not make use Proposition 39.10 in the rest of the course (even though doing so would speed up some of the proofs), since doing so would require us to assume more background in functional analysis. Nevertheless, let us note that Proposition 39.8 is an immediate corollary of Proposition 39.10 and Proposition 30.18.

The Stable Manifold Theorem for Hyperbolic Sets

In this lecture we study (un)stable manifolds for hyperbolic sets. Let f be a dynamical system on M, and let Λ be a compact completely invariant set. Let $\|\cdot\|$ denote a C^0 norm on $T_\Lambda M$ (possibly only of class C^0). The following definition is the analogue of Definition 33.1.

DEFINITION 40.1. Fix $r_0 > 0$, and suppose $\Phi\colon T_\Lambda M(r_0) \to T_\Lambda M$ is a continuous fibre-preserving map over f which is fibrewise Lipschitz. We define the **local fibre stable manifold** $\mathbb{W}^s_{\mathrm{loc},r}(0_x, Df+\Phi)$ at $x \in \Lambda$ of size $0 < r \le r_0$ to be the set[1]

$$\mathbb{W}^s_{\mathrm{loc},r}(0_x, Df+\Phi) := \Big\{ v \in T_xM \;\Big|\; \big\|(Df+\Phi)^k(v)\big\| \le r,\ \forall\, k \ge 0 \text{ and } \lim_{k\to\infty} \big\|(Df+\Phi)^k(v)\big\| = 0 \Big\}.$$

Similarly the **local fibre unstable manifold** $\mathbb{W}^u_{\mathrm{loc},r}(0_x, Df+\Phi)$ of size $0 < r \le r_0$ is the set

$$\mathbb{W}^u_{\mathrm{loc},r}(0_x, Df+\Phi) := \Big\{ v \in T_xM \;\Big|\; \big\|(Df+\Phi)^{-k}(v)\big\| \le r,\ \forall\, k \ge 0 \text{ and } \lim_{k\to\infty} \big\|(Df+\Phi)^{-k}(v)\big\| = 0 \Big\}.$$

Just as in Definition 33.2, the definition also works when $r_0 = \infty$, only then one drops the word "local" from the notation.

DEFINITION 40.2. Suppose $\Phi\colon T_\Lambda M \to T_\Lambda M$ is a continuous and fibre-preserving map over f which is fibrewise Lipschitz. We define the **global fibre stable manifold** associated to $Df + \Phi$ at $0_x \in T_xM$ to be the set

$$\mathbb{W}^s(0_x, Df+\Phi) := \Big\{ v \in T_xM \;\Big|\; \lim_{k\to\infty} \big\|(Df+\Phi)^k(v)\big\| = 0 \Big\}.$$

and similarly we define the **global fibre unstable manifold** associated to $Df+\Phi$ at $0_x \in T_xM$ to be the set

$$\mathbb{W}^u(0_x, Df+\Phi) := \Big\{ v \in T_xM \;\Big|\; \lim_{k\to\infty} \big\|(Df+\Phi)^{-k}(v)\big\| = 0 \Big\}.$$

[1]It would be more logical to write this as $W^s_{\mathrm{fib;loc},r}$ but such notation should not be written in polite company.

Remark 40.3. The fibre stable manifolds are subsets of the *tangent space* to the manifold. In Definitions 40.5 and 40.6 below we will introduce yet another notion of a stable manifold, written with a normal W^s instead of a blackboard $\mathbb{W}^s$. This will be a subset of the *manifold* itself. It is important you do not confuse the two!

The proof of the following result is essentially identical to the proof of Proposition 33.6, and hence we will omit it.

Proposition 40.4. *Let f be a dynamical system on M, and let Λ be a compact hyperbolic set of f with splitting $T_\Lambda M = E^s \oplus E^u$. Let $\|\cdot\|$ denote a C^0 norm on $T_\Lambda M$ which is adapted to f and Λ and of box type with respect to the splitting, and let $\tau = \tau(f, \Lambda)$ denote the skewness of f and Λ with respect to $\|\cdot\|$. Fix $r_0 > 0$, and suppose $\Phi \colon T_\Lambda M(r_0) \to T_\Lambda M$ is a continuous fibre-preserving map over f which is fibrewise Lipschitz with*

$$\operatorname{lip}_{\mathrm{fib}}(\Phi) < 1 - \tau, \qquad \Phi(0_x) = 0_{f(x)}, \qquad \forall\, x \in \Lambda.$$

Then for any $0 < r \le r_0$ and any $x \in \Lambda$, one has

$$\begin{aligned}
\mathbb{W}^s_{\mathrm{loc},r}(0_x, Df + \Phi) &= \Big\{ v \in T_x M(r) \;\Big|\; \big\|(Df + \Phi)^k(v)\big\| \le r, \ \forall\, k \ge 0 \Big\} \\
&= \big\{ v \in T_x M(r) \;\big|\; (Df + \Phi)^k(v) \in \operatorname{cone}_1(E^s(f^k(x))), \ \forall\, k \ge 0 \big\} \\
&= \Big\{ v \in T_x M(r) \;\Big|\; \big\|(Df + \Phi)^k(v)\big\| \le (\tau + \operatorname{lip}_{\mathrm{fib}}(\Phi))^k \, \|v\|, \ \forall\, k \ge 0 \Big\}.
\end{aligned}$$

Now we define the local stable manifolds on M itself. We caution the reader again that without additional hypotheses, the stable manifolds are *not* necessarily manifolds. The Stable Manifold Theorem 40.11 below tells us that hyperbolicity is one such set of hypotheses.

Definition 40.5. Let f be a dynamical system on M. Let m denote a Riemannian metric on M, and let $d = d_m$ denote the metric in the sense of topology on M induced by m (Theorem 36.7). Given $x \in M$ and $r > 0$, we define the **local stable manifold** $W^s_{\mathrm{loc},r}(x, f)$ at x of size r to be the set

$$\begin{aligned}
W^s_{\mathrm{loc},r}(x, f) := \Big\{ y \in M \;\Big|\; & d\big(f^k(x), f^k(y)\big) \le r, \ \forall\, k \ge 0 \\
& \text{and } \lim_{k \to \infty} d\big(f^k(y), f^k(x)\big) = 0 \Big\}.
\end{aligned}$$

Similarly the **local unstable manifold** at x of size r is given by

$$\begin{aligned}
W^u_{\mathrm{loc},r}(x, f) := \Big\{ y \in M \;\Big|\; & d\big(f^{-k}(x), f^{-k}(y)\big) \le r, \ \forall\, k \ge 0 \\
& \text{and } \lim_{k \to \infty} d\big(f^{-k}(y), f^{-k}(x)\big) = 0 \Big\}.
\end{aligned}$$

Finally we have the global stable manifolds.

Definition 40.6. Let f be a dynamical system on M. Let m denote a Riemannian metric on M, and let d denote the Riemannian distance. Given $x \in M$ we define the **global stable manifold** $W^s(x,f)$ at x to be the set

$$W^s(x,f) := \left\{ y \in M \;\middle|\; \lim_{k\to\infty} d\big(f^k(y), f^k(x)\big) = 0 \right\}.$$

Similarly the **global stable manifold** at x is the set

$$W^u(x,f) := \left\{ y \in M \;\middle|\; \lim_{k\to\infty} d\big(f^{-k}(y), f^{-k}(x)\big) = 0 \right\}.$$

Remark 40.7. The local stable manifolds depend on the choice of Riemannian metric m. The global stable manifolds do not.

On Problem Sheet S you will show that for any $r > 0$, one has

$$W^s(x,f) = \bigcup_{k\geq 0} f^{-k}\left(W^s_{\mathrm{loc},r}(f^k(x), f)\right). \tag{40.1}$$

The following result is the analogue of Proposition 33.8.

Proposition 40.8. *Let Λ denote a hyperbolic set for f. Then for all $r > 0$ sufficiently small, there are constants $C \geq 1$ and $0 < \mu < 1$ such that for any $x \in \Lambda$,*

$$\begin{aligned} W^s_{\mathrm{loc},r}(x,f) &= \left\{ y \in M \mid d\big(f^k(x), f^k(y)\big) \leq r,\ \forall\, k \geq 0 \right\} \\ &= \left\{ y \in M \mid d\big(f^k(x), f^k(y)\big) \leq \min\left\{r, C\mu^k d(x,y)\right\},\ \forall k \geq 0 \right\}, \end{aligned}$$

and similarly

$$\begin{aligned} W^u_{\mathrm{loc},r}(x,f) &= \left\{ y \in M \mid d\big(f^{-k}(x), f^{-k}(y)\big) \leq r,\ \forall\, k \geq 0 \right\} \\ &= \left\{ y \in M \mid d\big(f^{-k}(x), f^{-k}(y)\big) \leq \min\left\{r, C\mu^k d(x,y)\right\},\ \forall k \geq 0 \right\}, \end{aligned}$$

Proof. We prove the result for the stable manifold only. We may assume that the Riemannian metric m is adapted to f and Λ. It suffices to show that there exist $r_0 > 0$, $C \geq 1$ and $0 < \mu < 1$ such that for all $0 < r \leq r_0$, the first set on the right-hand side is contained in the second. For $x \in M$ and y close to x, setting $v = \exp_x^{-1}(y)$ and applying (39.3), we have

$$d\big(f^k(y), f^k(x)\big) = \big\|\widehat{f}^k(v)\big\| = \big\|(Df + \Phi_f)^k(v)\big\|,$$

where $\Phi_f := \widehat{f} - Df$, as long as the iterates make sense. Thus the statement reduces to showing that there exist $r_0 > 0$, $C \geq 1$ and $0 < \mu < 1$ such that if $0 < r \leq r_0$ and $v \in T_xM(r)$ then

$$\big\|(Df + \Phi_f)^k(v)\big\| \leq r,\ \forall\, k \geq 0 \qquad \Rightarrow \qquad \big\|(Df + \Phi_f)^k(v)\big\| \leq C\mu^k\|v\|,\ \forall\, k \geq 0.$$

We wish to apply Proposition 40.4, but this requires us to use a norm of box type. So let $\|\cdot\|_\Lambda$ denote the box-adjusted norm, which is only of class C^0 and only defined on $T_\Lambda M$. Then $\|\cdot\|_\Lambda$ is equivalent to the norm $\|\cdot\|$ coming from m with a constant c, say:

$$\frac{1}{c}\|v\| \leq \|v\|_\Lambda \leq c\|v\|, \qquad \forall\, v \in T_\Lambda M.$$

(this is a special case of Proposition 38.11). Now let $C = 1$ and choose $\tau < \mu < 1$. By Proposition 39.7, there is $r > 0$ sufficiently small such that

$$\operatorname{lip}_{\text{fib}}\left(\Phi_f; \|\cdot\|\right) \leq \frac{\mu - \tau}{c^2},$$

on[2] $\{v \in T_\Lambda M \mid \|v\| \leq cr\}$. Thus

$$\operatorname{lip}_{\text{fib}}\left(\Phi_f; \|\cdot\|_\Lambda\right) \leq \mu - \tau$$

on $\{v \in T_\Lambda M \mid \|v\|_\Lambda \leq r\}$. The result now directly follows from Proposition 40.4. ∎

Let us now compare the local stable manifold $W^s_{\text{loc},r}(x, f)$ with the local fibre unstable manifold $\mathbb{W}^s_{\text{loc},r}(0_x, \widehat{f})$.

Corollary 40.9. *Let f be a dynamical system on M and $\Lambda \subseteq M$ a compact invariant hyperbolic set. Then for all $x \in \Lambda$ and r small enough, one has*

$$W^s_{\text{loc},r}(x, f) = \exp_x\left(\mathbb{W}^s_{\text{loc},r}(0_x, \widehat{f})\right),$$

where $\widehat{f}$ denotes the lift of f.

Proof. Combining Proposition 40.4 and Proposition 40.8, we have for all $x \in \Lambda$ and $r > 0$ sufficiently small:

$$W^s_{\text{loc},r}(x, f) = \left\{y \in M \,\middle|\, d\big(f^k(x), f^k(y)\big) \leq r,\ \forall\, k \geq 0\right\}$$
$$\mathbb{W}^s_{\text{loc},r}(0_x, \widehat{f}) = \left\{v \in T_x M(r) \,\middle|\, \big\|\widehat{f}^k(v)\big\| \leq r,\ \forall\, k \geq 0\right\}.$$

By (39.3), for x and y close enough,

$$\|\widehat{f}^k(\exp_x^{-1}(y)\| = d\big(f^k(x), f^k(y)\big).$$

The result follows. ∎

Suppose $\xi\colon E^u \to E^s$ is a fibre-preserving map over id, i.e. $\xi(E^u(x)) \subseteq E^s(x)$. If the map is Lipschitz when restricted to the fibres, then just as before we can speak of the **fibre-Lipschitz constant**

$$\operatorname{lip}_{\text{fib}}(\xi) := \sup_{x \in \Lambda} \operatorname{lip}(\xi|_{E^u(x)}).$$

As before, we also say that ξ is **continuously differentiable on the fibres**, if the **fibre derivative**

$$D_{\text{fib}}\xi\colon E^u \to \mathcal{V}(E^u, E^s; \text{id})$$

defined by

$$D_{\text{fib}}\xi(v) = D\big(\xi|_{E^u(x)}\big)(v)\colon E^u(x) \to E^s(x), \qquad v \in E^u(x)$$

exists and is continuous. The next result is the analogue of Theorem 34.1. We state it for the fibre unstable manifold for variety.

[2]Since there are now two norms in play, we won't use the $TM(r)$ notation here ...

Theorem 40.10 (The Fibre Stable Manifold Theorem). *Let $\Lambda \subseteq M$ be a compact hyperbolic set for f with splitting $T_\Lambda M = E^s \oplus E^u$. Let $\|\cdot\|$ be a C^0 norm on $T_\Lambda M$ which is adapted to f and Λ and of box type with respect to the splitting. There is a $\delta > 0$ with the following property: if $\Phi\colon T_\Lambda M \to T_\Lambda M$ is a continuous fibre-preserving map over f satisfying*

$$\operatorname{lip}_{\mathrm{fib}}(\Phi) < \delta, \qquad \Phi(0_x) = 0_{f(x)}, \qquad \forall\, x \in \Lambda,$$

then there is a continuous fibre-preserving map $\xi\colon E^u \to E^s$ over id *such that*

$$\operatorname{lip}_{\mathrm{fib}}(\xi) \le 1, \qquad \xi(0_x) = 0_x, \qquad \forall\, x \in \Lambda,$$

and such that for any $x \in \Lambda$ the global fibre unstable manifold of $Df + \Phi$ at x is the graph of ξ:

$$\mathbb{W}^u(0_x, Df + \Phi) = \operatorname{gr}(\xi|_{E^u(x)}).$$

Moreover if Φ is continuously differentiable on the fibres then so is ξ.

The proof of the Lipschitz case of Theorem 40.10 is very similar to the proof of Theorem 34.1, and as such is omitted. The proof of the differentiable case is also omitted, since we skipped this part in the linear case.

The following theorem is the generalisation of Local Stable Manifold Theorem 34.3 to the setting of hyperbolic sets. It is arguably the single most important result in hyperbolic dynamics.

Theorem 40.11 (The Stable Manifold Theorem for Hyperbolic Sets). *Let f be a dynamical system on M, and let $\Lambda \subseteq M$ be a compact hyperbolic set with splitting $T_\Lambda M = E^s \oplus E^u$.*

(i) *For all $r > 0$ sufficiently small and for every $x \in \Lambda$, the local unstable manifold $W^s_{\mathrm{loc},r}(x, f)$ is a C^1 embedded submanifold of M of dimension $\dim E^s(x)$, which is diffeomorphic to a ball in $E^s(x)$. Similarly the unstable manifold $W^u_{\mathrm{loc},r}(x, f)$ is a C^1 embedded submanifold of M of dimension $\dim E^u(x)$, which is diffeomorphic to a ball in $E^u(x)$.*

(ii) *The global stable manifold $W^s(x, f)$ is an C^1 immersed submanifold of M of dimension $\dim E^s(x)$, and the global unstable manifold $W^u(x, f)$ is an C^1 immersed submanifold of M of dimension $\dim E^u(x)$.*

(♣) *Proof.* We consider the stable manifold only. Since $\exp_x$ is a diffeomorphism, it suffices by Corollary 40.9 to show that $\mathbb{W}^s_{\mathrm{loc},r}(0_x, \widehat{f})$ is a C^1 embedded submanifold of T_xM. The strategy is similar to how we proved Theorem 34.3. Let $\Phi_f := \widehat{f} - Df\colon TM(r_*) \to TM$, where r_* was defined in (39.1). Then Φ_f is a C^0 map which is continuously differentiable on the fibres and satisfies

$$\Phi_f(0_x) = 0_{f(x)}, \qquad D_{\mathrm{fib}}\Phi_f(0_x) = 0, \qquad \forall\, x \in \Lambda$$

(cf. (39.5)). Fix a C^∞ function $\beta\colon \mathbb{R} \to [0, 1]$ such that:

$$\beta(t) = \begin{cases} 1, & t \le \frac{1}{3}, \\ 0, & t \ge \frac{2}{3} \end{cases}$$

Choose $r > 0$ small enough so that $TM(3r) \subset TM(r_*)$. Define

$$\Phi_r(v) := \beta\left(\frac{\|v\|}{3r}\right)\Phi_f(v).$$

After shrinking r if necessary, we may assume that $\mathrm{lip}_{\mathrm{fib}}(\Phi_r)$ is small enough to satisfy the hypotheses of Proposition 40.4 and Theorem 40.10. This gives us a continuous fibre-preserving map $\xi \colon E^s \to E^u$ which is differentiable in the fibres, fixes the zero section, and has fibre-Lipschitz constant at most 1, such that for any $x \in \Lambda$,

$$\mathbb{W}^s(0_x, Df + \Phi_r) = \mathrm{gr}(\xi|_{E^s(x)}).$$

Just like in the proof of Theorem 34.3, the next step is to show that the C^1 embedding $i \colon E^s(r) \to T_\Lambda M$ given by

$$i(v) = (v, \xi(v)),$$

satisfies

$$i(E^s(x,r)) = \mathbb{W}^s(0_x, Df + \Phi_r) \cap T_x M(r).$$

This is done in exactly the same way as in the proof of Theorem 34.3. Our choice of r then tells us that

$$i(E^s(x,r)) = \mathbb{W}^s_{\mathrm{loc},r}(0_x, \widehat{f}),$$

and hence $\mathbb{W}^s_{\mathrm{loc},r}(0_x, \widehat{f})$ is a C^1 embedded submanifold of $T_x M$. Actually, strictly speaking we are using the wrong norm here, and so as in the proof of Theorem 34.3, we must switch back to the original Riemannian norm, rather than the box-adjusted norm. The details are left to you.

It remains to prove part (ii). This is a formal consequence of (40.1). Indeed, since Λ is invariant, we know that $W^s_{\mathrm{loc},r}(f^k(x), f)$ is an embedded C^1 submanifold of M for each k. Thus the same is true of $f^{-k}\left(W^s_{\mathrm{loc},r}(f^k(x), f)\right)$. Thus by (40.1), $W^s(x,f)$ is union of a increasing sequence of C^1 embedded submanifolds, whence $W^s(x,f)$ is itself a C^1 immersed submanifold. This completes the proof. ■

LECTURE 41

Structural Stability

Suppose $f\colon M \to M$ and $g\colon N \to N$ are two differentiable dynamical systems. Recall we say that f and g are *conjugate* if there exists a homeomorphism $H\colon M \to N$ such that

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "H"'] & M \arrow[d, "H"] \\ N \arrow[r, "g"'] & N \end{tikzcd}$$

In today's lecture we introduce three notions of *stability*.

DEFINITION 41.1. Fix $p \geq 1$. A dynamical system $f \in \mathrm{Diff}^p(M)$ is C^p **structurally stable** if there exists a neighbourhood[1] $\mathcal{U}$ of f in $\mathrm{Diff}^p(M)$ such that any $g \in \mathcal{U}$ is topologically conjugate to f.

REMARKS 41.2.

(i) It doesn't make sense to talk about C^0-structural stability. This is because a C^0 perturbation can change almost anything about a dynamical system. For instance, if x is an isolated fixed point of f then a C^0-small perturbation g of f may fix an entire neighbourhood of x. Thus it is hopeless to expect all C^0 perturbations to yield conjugate dynamics—C^0 structural stability is never satisfied. Thus structural stability is only interesting in the differentiable category, and there is no analogue in topological dynamics.

(ii) On the other hand, it is crucial that we only require the conjugacy H to be a homeomorphism and not a diffeomorphism. This may at first seem somewhat unnatural, given that we are working in the differentiable category. However it is the correct notion to study. This is because asking f and g to be conjugate by a diffeomorphism is too restrictive to be useful. Indeed, if f and g are conjugate via a diffeomorphism H then the matrices $Df(x)$ and $Dg(H(x))$ are similar for all $x \in M$, as they are conjugated by the linear isomorphism $DH(x)$. But this already fails for linear maps: a generic perturbation of an invertible linear map changes its eigenvalues. Thus structural stability is *never* satisfied for differentiable conjugacies.

In fact, in this course we will only ever be interested in C^1 structural stability, and as such we will abbreviate "C^1 structurally stable" as simply "structurally

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]Strictly speaking, we didn't formally define the topology on $\mathrm{Diff}^p(M)$ for $p > 1$ (cf. the discussion before Proposition 38.6). This is no big deal, for two reasons: (a) the construction is entirely analogous for $p > 1$, and (b), in this course, however, we will only ever work with C^1 structural stability.

simple" from now on.

To get a feel for the definition, on Problem Sheet S you will prove:

Example 41.3. Suppose $f\colon [0,1] \to [0,1]$ is an orientation preserving diffeomorphism. Then f is structurally stable if and only if $f'(0) \neq 1$ and $f'(1) \neq 1$.

There is also a quantitative version of structural stability.

Definition 41.4. Let f be a dynamical system of M, and fix $\varepsilon > 0$. We say that f is **structurally ε-stable** if there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^p(M)$ such that any $g \in \mathcal{U}$ is topologically conjugate to f via a homeomorphism H_g satisfying $d_0(H_g, \mathrm{id}) \leq \varepsilon$.

Structural stability guarantees that invariant sets persist.

Definition 41.5. Suppose $\Lambda \subseteq M$ is a compact completely invariant set for a structurally stable dynamical system f. Then for g sufficiently close to f, there exists a homeomorphism $H = H_g$ conjugating f and g. The image $\Lambda_g := H_g(\Lambda)$ is thus a compact completely invariant set for g. We call Λ_g the **continuation** of Λ to g. If f is structurally ε-stable then the continuation Λ_g is contained in the ball $B(\Lambda, \varepsilon)$ of radius ε about Λ.

Why is this useful? In Lecture 38 we set ourselves the goal of showing that hyperbolicity "persists". That is, if Λ is a compact hyperbolic set for f then for g close enough to f, we want to prove that g has a compact hyperbolic set close to Λ. We already proved in Proposition 38.6 that for g close enough to f, any compact completely invariant set Δ for g is necessarily hyperbolic, and thus it remains to show such Δ exists. The notion of structural ε-stability provides a mechanism for ensuring the existence of such Δ—namely, one could take $\Delta = \Lambda_g$.

Thus one might hope that hyperbolicity implies structural ε-stability. Unfortunately it turns out this is *not* quite true. Nevertheless, a slightly weaker notion is true, and this is still enough for our purposes. In the following, we denote by $i_\Lambda\colon \Lambda \hookrightarrow M$ the inclusion, and we continue to use d_0 for the supremum metric on $C^0(\Lambda, M)$.

Definition 41.6. Suppose f is a dynamical system on M and Λ is a compact completely invariant set. We say that f is **weakly structurally stable on** Λ if there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ such that if $g \in \mathcal{U}$ then there exists a continuous injective map $H_g\colon \Lambda \to M$ such that $H_g \circ f = g \circ H_g$ on Λ. Similarly[2] we say that f is **weakly structurally ε-stable** if such H_g can be chosen to satisfy $d_0(H_g, i_\Lambda) \leq \varepsilon$.

Since Λ is compact, $H_g\colon \Lambda \to H_g(\Lambda)$ is a homeomorphism, and thus the continuation $\Lambda_g := H_g(\Lambda)$ of Λ to g from Definition 41.5 is still a well-defined compact completely invariant set for g.

[2] As with the linguistic monstrosity surrounding the stable manifolds (eg. Definition 40.1), no definition in dynamical systems is complete without at least three adjectives...

With these preliminaries out of the way, we are ready to state the following fundamental result, which states, roughly speaking, that hyperbolicity implies weak structural stability.

THEOREM 41.7. *Let f be a dynamical system on a manifold M, and let $\Lambda \subset M$ be a compact hyperbolic set of f. Then there exists ε such that f is uniquely weakly structurally ε-stable on Λ. That is, there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ such that for $g \in \mathcal{U}$ there exists a unique continuous injection $H_g \colon \Lambda \to M$ such that $H_g \circ f = g \circ H_g$ on Λ. Finally, $d_0(H_g, i_\Lambda) \to 0$ as $d_1(g, f) \to 0$.*

The proof of Theorem 41.7 will come next lecture. We will spend the rest of this lecture working on auxiliary results that will be needed in the proof of Theorem 41.7. For now, however, let us note that Theorem 41.7 does indeed show that hyperbolicity persists.

COROLLARY 41.8 (Persistence of Hyperbolicity). *Let Λ be a compact hyperbolic set for f. There exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and a number $a > 0$ such that if $g \in \mathcal{U}$ then g has a compact hyperbolic set $\Delta \subset B(\Lambda, a)$.*

Proof. Immediate from Proposition 38.6 and Theorem 41.7. ∎

Before continuing with the applications of Theorem 41.7, let us recall the following famous result.

THEOREM 41.9 (Brouwer's Invariance of Domain Theorem). *Let $H \colon M \to N$ be a continuous map between manifolds of the same dimension. If H is injective then $H(M)$ is an open subset of N.*

This beautiful proof is due to W. Kulpa.

(♣) *Proof.* The statement is local in nature, and thus it suffices to show that if B denotes the closed unit ball in $\mathbb{R}^n$ and $H \colon B \to \mathbb{R}^n$ is continuous, then $H(0)$ lies in the interior of $C := H(B)$.

Suppose for contradiction this is not the case. Without loss of generality we may assume that $H(0) = 0$. Since B is compact, $H \colon B \to C$ is a homeomorphism, and hence by the Tietze Extension Theorem, there exists a continuous map $G \colon \mathbb{R}^n \to \mathbb{R}^n$ such that $G = H^{-1}$ on C. Then $G(0) = 0$. Since 0 is not an interior point of C, we may perturb[3] G slightly to a new function $F \colon C \to \mathbb{R}^n$ such that

$$F(x) \neq 0, \qquad \forall\, x \in C. \tag{41.1}$$

and

$$\|F(x) - G(x)\| \leq 1, \qquad \forall x \in C. \tag{41.2}$$

Now consider the continuous map

$$f \colon B \to \mathbb{R}^n, \qquad f(x) := x - F(H(x)).$$

By (41.2), $f(B) \subseteq B$. Moreover by (41.1), f has no fixed points. This contradicts the Brouwer Fixed Point Theorem[4], which states that any continuous map $B \to B$ has at least one fixed point. ∎

[3]This is reasonably believable if you draw a picture. The proof is much less so, and we refer the reader to Kulpa's original paper for the details.

[4]The Brouwer Fixed Point Theorem is easiest to prove using Algebraic Topology. If one accepts the existence of (singular) homology, the proof is one line. See for instance here.

Another immediate consequence of Theorem 41.7 is the following theorem of Anosov.

Corollary 41.10. *Anosov diffeomorphisms on compact manifolds are structurally stable.*

Proof. Let $f\colon M \to M$ be Anosov. By Theorem 41.7, there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ such that if $g \in \mathcal{U}$ then there exists a continuous injective map $H = H_g\colon M \to M$ such that $H \circ f = g \circ H$. We need only prove that H is surjective. By Theorem 41.9, $H(M)$ is an open set. Since $H(M)$ is also closed, it follows that $H(M) = M$ (recall we always assume our manifolds are connected). ∎

In order to prove Theorem 41.7, we will need the following result which is of interest in its own right. Recall Definition 9.8: a reversible topological dynamical system f on a compact metric space X is **weakly expansive** if there exists a constant $\delta > 0$ such that

$$d\big(f^k(x), f^k(y)\big) \le \delta, \quad \forall\, k \in \mathbb{Z} \qquad \Rightarrow \qquad x = y.$$

The constant δ (which is not unique) is called a **weak expansivity constant**. We now prove that the restriction of a differentiable dynamical system to a hyperbolic set is weakly expansive, as the next result shows. This result is the analogue of Corollary 33.9 in this setting. In fact, we will prove a stronger statement that works for nearby maps too, since this will be needed in the proof of Theorem 41.7 in the next lecture.

Proposition 41.11. *Let f be a dynamical system on M and let $\Lambda \subseteq M$ be a compact hyperbolic set. Then $f|_\Lambda$ is weakly expansive. Moreover there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and $\delta, a > 0$ such that if $g \in \mathcal{U}$ and Δ is a compact completely invariant set of g with $\Delta \subset B(\Lambda, a)$ then $g|_\Delta$ is weakly expansive with weak expansivity constant δ.*

Proof. We may assume that the Riemannian norm m on M is adapted to f and Λ. Let $d = d_m$ denote the induced point-set topology metric on M from Theorem 36.7. Let us first reformulate what it means for $g|_\Delta$ to be weakly expansive in terms of the lifting $\widehat{g}$ of g. Assume that δ is chosen smaller than the constant $r_*(g,\rho)$ defined in (39.1). Then if $d(g^k(x), g^k(y)) < \delta$ for all $k \in \mathbb{Z}$, setting $v := \exp_x^{-1}(y)$, we have from equation (39.4) that

$$d(g^k(x), g^k(y)) = \|\widehat{g}^k(v)\| \le \delta$$

for all $k \in \mathbb{Z}$. Set $\Phi_g = \widehat{g} - Dg$. It thus suffices to show that there exists $\delta > 0$ such that if a vector $v \in T_xM$ satisfies

$$\big\|(Dg + \Phi_g)^k(v)\big\| \le \delta, \qquad \forall\, k \in \mathbb{Z},$$

then $v = 0_x$.

Let $\tau = \tau(f, \Lambda)$ denote the skewness of f and Λ with respect to $\|\cdot\|$, and choose $\tau < \tau_0 < \mu < 1$. Let

$$\Psi_g := (\widehat{g})^{-1} - (Dg)^{-1},$$

so that

$$(Dg)^{-1} + \Psi_g = (Dg + \Phi_g)^{-1}.$$

Note that as $g \to f$ in C^1, one also has $g^{-1} \to f^{-1}$. Now by Proposition 38.11, there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and two numbers $a > 0$ and $c \geq 1$ such that for any $g \in \mathcal{U}$ and any compact g-invariant set Δ with $\Delta \subset B(\Lambda, a)$, the set Δ is hyperbolic and the box-adjusted norm $\|\cdot\|_\Delta$ is equivalent to $\|\cdot\|$ with constant c, i.e.

$$\frac{1}{c}\|v\| \leq \|v\|_\Delta \leq c\|v\|, \qquad \forall v \in T_\Delta M,$$

and finally the skewness of g and Λ with respect to $\|\cdot\|$ satisfies

$$\tau(g, \Delta) \leq \tau_0.$$

Next, by Proposition 39.7, up to shrinking $\mathcal{U}$, there exists a number $\delta > 0$ such that for any $g \in \mathcal{U}$, one has

$$\mathrm{lip}_{\mathrm{fib}}\left(\Phi_g; \|\cdot\|\right) \leq \frac{\mu - \tau_0}{c^2} \qquad \text{and} \qquad \mathrm{lip}_{\mathrm{fib}}\left(\Psi_g; \|\cdot\|\right) \leq \frac{\mu - \tau_0}{c^2},$$

on[5]

$$\left\{v \in TM \mid \|v\| \leq c^2\delta\right\},$$

Thus also

$$\mathrm{lip}_{\mathrm{fib}}\left(\Phi_g; \|\cdot\|_\Delta\right) \leq \mu - \tau_0, \qquad \text{and} \qquad \mathrm{lip}_{\mathrm{fib}}\left(\Psi_g; \|\cdot\|_\Delta\right) \leq \mu - \tau_0$$

on

$$\{v \in T_\Delta M \mid \|v\|_\Delta \leq c\delta\}.$$

Now suppose a vector $v \in T_x M$ for $x \in \Delta$ has

$$\left\|(Dg + \Phi_g)^k(v)\right\| < \delta, \qquad \forall\, k \in \mathbb{Z}.$$

Then also

$$\left\|(Dg + \Phi_g)^k(v)\right\|_\Delta < c\delta, \qquad \forall\, k \in \mathbb{Z}.$$

In other words,

$$v \in \mathbb{W}^u_{\mathrm{loc},r}(0_x, Dg + \Phi_g) \cap \mathbb{W}^s_{\mathrm{loc},r}(0_x, Dg + \Phi_g).$$

Since $\|\cdot\|_\Delta$ is of box type, Proposition 40.4 is applicable. Thus as in the proof of Corollary 33.9, we have

$$\begin{aligned}
\|v\|_\Delta &= \left\|((Dg)^{-1} + \Psi_g)(Dg + \Phi_g)(v)\right\|_\Delta \\
&\leq \Big(\tau_0 + \mathrm{lip}_{\mathrm{fib}}\left(\Psi_g; \|\cdot\|_\Delta\right)\Big)\left\|(Dg + \Phi_g)(v)\right\|_\Delta \\
&\leq \Big(\tau_0 + \mathrm{lip}_{\mathrm{fib}}\left(\Psi_g; \|\cdot\|_\Delta\right)\Big)\Big(\tau_0 + \mathrm{lip}_{\mathrm{fib}}\left(\Phi_g; \|\cdot\|_\Delta\right)\Big)\|v\|_\Delta \\
&\leq \mu^2\|v\|_\Delta.
\end{aligned}$$

Since $\mu^2 < 1$, it follows that $v = 0$. This completes the proof. ■

[5] Since there are now multiple norms in play, we refrain from using the $TM(r)$ notation for the remainder of this proof.

LECTURE 42

Hyperbolicity Implies Weak Structural Stability

The aim of this lecture is to prove Theorem 41.7. In fact, we will prove a somewhat stronger statement, which will be useful in future applications. Let us first explain the strategy in the proof.

Let (M, m) be a compact Riemannian manifold. We are given a smooth dynamical system f with a hyperbolic set Λ, and a nearby smooth dynamical system g. We wish to find a continuous function $H \colon \Lambda \to M$ which is close to the inclusion $i_\Lambda \colon \Lambda \hookrightarrow M$ such that

$$g \circ H = H \circ f, \qquad \text{on } \Lambda. \tag{42.1}$$

As usual, we would like to recast this in such a way that it becomes a fixed point problem. Unfortunately, $C^0(\Lambda, M)$ does not have a Banach space structure[1], which complicates matters. So we pass to the exponential map. Since H is meant to be close to i_Λ, it will lie in the image of the exponential map. Thus if such H exists, there also exists a continuous section $\gamma \in \Gamma^0(\Lambda, T_\Lambda M)$ such that

$$H(x) = \exp_x(\gamma(x)), \qquad \forall\, x \in \Lambda.$$

Indeed, one can simply define $\gamma(x) := \exp_x^{-1}(H(x))$. In terms of γ, our desired equation (42.1) becomes

$$g\big(\exp_x(\gamma(x)\big) = \exp_{f(x)}(\gamma(f(x))), \qquad \forall x \in \Lambda,$$

or equivalently,

$$\exp_{f(x)}^{-1}\big(g(\exp_x(\gamma(x))\big) = \gamma(f(x)), \qquad \forall x \in \Lambda. \tag{42.2}$$

The expression on the left-hand side of (42.2) is similar to the lifting $\widehat{f}$, apart from g is in the middle, not f. This motivates the following definition.

Definition 42.1. Let $f, g \colon M \to M$ denote two smooth dynamical systems which are close to each other. Define $r_* = r_*(f, g, m) > 0$ by requiring that if

$$d(x, y) < r_* \qquad \Rightarrow \qquad d(f(x), g(y)) < r_m,$$

where r_m is the injectivity radius of our metric m. We then define the **dual lift**

$$\widehat{f}_g \colon TM(r_*) \to TM, \qquad \widehat{f}_g(x, v) = \exp_{f(x)}^{-1}\big(\exp_x(v))\big).$$

[1]You cannot add together functions that take values in a manifold!

Just as with the normal lift $\widehat{f}$, the map $\widehat{f}_g$ is a C^1 fibre-preserving map over f. Note that $\widehat{f}_f = \widehat{f}$. The equation (42.2) is thus equivalent to asking that γ is an *invariant* section of $\widehat{f}_g$:

$$\widehat{f}_g(\gamma) = \gamma(f), \tag{42.3}$$

in the sense of (39.6). The existence of such an invariant section γ will follow from Proposition 39.8, if we can show that $\widehat{f}_g$ is a Lipschitz-small perturbation of Df if g is sufficiently close to f.

Let us now state the improved version of Theorem 41.7 that we will prove in this lecture.

THEOREM 42.2. *Let f be a dynamical system on a compact manifold M and let Λ be a compact hyperbolic set for f.*

(i) *There is a C^1 neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and two numbers $a > 0$ and $\varepsilon_0 > 0$ such that if $g, h \in \mathcal{U}$ and $\Delta \subset B(\Lambda, a)$ is a compact completely invariant set for g then there is at most one continuous map $H \colon \Delta \to M$ such that $H \circ g = h \circ H$ and $d_0(H, i_\Delta) \le \varepsilon_0$.*

(ii) *Moreover, for any $0 < \varepsilon \le \varepsilon_0$ there is a neighbourhood $\mathcal{U}_\varepsilon \subset \mathcal{U}$ such that if $g, h \in \mathcal{U}_\varepsilon$ and $\Delta \subset B(\Lambda, a)$ is a compact completely invariant set for g then there is at least one (and hence exactly one) continuous map $H \colon \Delta \to M$ such that $H \circ g = h \circ H$ and $d_0(H, i_\Delta) \le \varepsilon$. Moreover this map H is injective.*

Theorem 41.7 follows immediately from Theorem 42.2 by taking "g" to be equal to f and "h" to be equal to g.

Proof. We may assume our Riemannian metric m is adapted to f and Λ. We prove the result in five steps.

1. First take $\mathcal{V}$ and r_* small enough so that the dual lifting

$$\widehat{g}_h \colon TM(r_*) \to TM$$

is well defined for each pair $g, h \in \mathcal{V}$. Now set

$$\Phi_{g,h} := \widehat{g}_h - Dg \colon TM(r_*) \to TM,$$

so that $\Phi_{g,h}$ is a fibre-preserving map over g. We claim that for any $\delta > 0$ there exists a neighbourhood $\mathcal{V}_\delta \subset \mathcal{V}$ of f and $0 < r_\delta < r_*$ such that for any $g, h \in \mathcal{V}_\delta$, one has

$$\mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}) \le \delta, \qquad \text{on} \qquad TM(r_\delta).$$

Indeed, from the proof of Proposition 39.7 given $\delta > 0$ there is a C^1 neighbourhood $\mathcal{V}_\delta$ of f and $r_\delta > 0$ such that for $g \in \mathcal{V}_\delta$,

$$\big\| D_{\mathrm{fib}}\widehat{g}(v) - Dg(v) \big\|^{\mathrm{op}} \le \frac{\delta}{2}, \qquad \text{on} \qquad TM(r_\delta).$$

Since

$$D_{\mathrm{fib}}\widehat{g}(v) = D(\exp_{g(x)}^{-1})(g(\exp_x(v)) \circ Dg(\exp_x(v)) \circ D\exp_x(v),$$

and

$$D_{\mathrm{fib}}\widehat{g}_h(v) = D(\exp_{g(x)}^{-1})(h(\exp_x(v)) \circ Dh(\exp_x(v)) \circ D\exp_x(v),$$

up to shrinking $\mathcal{V}_\delta$, we may assume that

$$\left\| D_{\mathrm{fib}}\widehat{g}_h(v) - D_{\mathrm{fib}}\widehat{g}(v)\right\|^{\mathrm{op}} \le \frac{\delta}{2}, \qquad \text{on} \qquad TM(r_\delta).$$

Then

$$\begin{aligned} \left\| D_{\mathrm{fib}}\Phi_{g,h}(v)\right\|^{\mathrm{op}} &= \left\| D_{\mathrm{fib}}\widehat{g}_h(v) - Dg(v)\right\|^{\mathrm{op}} \\ &\le \left\| D_{\mathrm{fib}}\widehat{g}_h(v) - D_{\mathrm{fib}}\widehat{g}(v)\right\|^{\mathrm{op}} + \left\| D_{\mathrm{fib}}\widehat{g}(v) - Dg(v)\right\|^{\mathrm{op}} \\ &\le \frac{\delta}{2} + \frac{\delta}{2} = \delta. \end{aligned}$$

Applying the Mean Value Theorem 30.11 on the fibres, we obtain $\mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}) \le \delta$ on $TM(r_\delta)$.

2. Now let $\tau = \tau(f, \Lambda)$ denote the skewness of f with respect to m, and choose $\tau < \tau_0 < \mu < 1$. By Proposition 38.11 there exists a neighbourhood $\mathcal{W}$ of f and $a_0 > 0$ and $c \ge 1$ such that if $g \in \mathcal{W}$ and $\Delta \subset B(\Lambda, a_0)$ is a compact invariant set of g then Δ is hyperbolic with skewness

$$\tau(g, \Delta) \le \tau_0, \tag{42.4}$$

and the box-adjusted norm $\|\cdot\|_\Delta$ is equivalent to the Riemannian norm $\|\cdot\|$ with constant c. Now take

$$\delta := \frac{\mu - \tau_0}{c^2}.$$

Then if $g, h \in \mathcal{W} \cap \mathcal{V}_\delta$, one has

$$\mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}; \|\cdot\|) \le \delta \tag{42.5}$$

on[2] $\{v \in TM \mid \|v\| \le r_\delta\}$. The "norm switching" argument we used in the proof of Proposition 40.8 then tells us that

$$\mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}; \|\cdot\|_\Delta) \le \mu - \tau_0$$

on $\left\{v \in T_\Delta M \mid \|v\|_\Delta \le \frac{r_\delta}{c}\right\}$. Finally, by Proposition 41.11 there exists $\mathcal{U} \subset \mathcal{W} \cap \mathcal{V}_\delta$, $0 < a \le a_0$ and $0 < \varepsilon_0 \le \frac{r_\delta}{2c^2}$ such that if $g \in \mathcal{U}$ and Δ is any compact completely invariant set for g contained in $B(\Lambda, a_0)$, then $g|_\Delta$ is weakly expansive with weak expansivity constant $2\varepsilon_0$.

3. We claim that this choice of $\mathcal{U}$, a and ε_0 satisfies part (i) of the Theorem. Indeed, if $g, h \in \mathcal{U}$ and $\Delta \subset B(\Lambda, a)$ is a compact completely invariant set for g then by (42.4) and (42.5)

$$\mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}; \|\cdot\|_\Delta) \le \mu - \tau(g, \Delta) \qquad \text{on} \qquad \left\{v \in T_\Delta M \mid \|v\|_\Delta \le c\,\varepsilon_0\right\}.$$

Since $\|\cdot\|_\Delta$ is of box type, Proposition 39.8 is applicable. We conclude that $\widehat{g}_h = Dg + \Phi_{g,h}$ has at most one invariant section γ with $\|\gamma(x)\|_\Delta \le c\,\varepsilon_0$ for all $x \in \Delta$.

[2]As before, now there are multiple norms in play we will drop the $TM(r)$ notation.

As explained at the start of the lecture, this implies that there exists at most one continuous function $H\colon \Delta \to M$ such that $d_0(H, i_\Lambda) \le \varepsilon_0$ and $h \circ H = H \circ g$. Indeed, given such a H, if we define

$$\gamma(x) = \exp_x^{-1}(H(x))$$

then γ is $\widehat{g}_h$ invariant and

$$\begin{aligned}\sup_{x\in\Delta} \|\gamma(x)\|_\Delta &\le c\,\|\gamma(x)\| \\ &\le c\,\sup_{x\in\Delta} d(x, H(x)) \\ &\le c\,\varepsilon_0.\end{aligned}$$

This finishes the proof of part (i).

4. Now suppose $0 < \varepsilon < \varepsilon_0$ is given. We choose $\mathcal{U}_\varepsilon \subset \mathcal{U}$ such that for every $g \in \mathcal{U}_\varepsilon$, one has

$$d_0(f, g) \le \frac{(1-\mu)\varepsilon}{2c^2}. \tag{42.6}$$

Now suppose $g, h \in \mathcal{U}_\varepsilon$ and $\Delta \subset B(\Lambda, a)$ is a compact completely invariant set for g. Then

$$\begin{aligned}\sup_{x\in\Delta} \big\|\Phi_{g,h}(0_x)\big\|_\Delta &= \sup_{x\in\Delta} \big\|\widehat{g}_h(0_x)\big\|_\Delta \\ &= \sup_{x\in\Delta} \big\| \exp_{g(x)}^{-1}(h(x))\big\|_\Delta \\ &\le c\,\sup_{x\in\Delta} \big\| \exp_{g(x)}^{-1}(h(x))\big\| \\ &\le c\,d_0(g,h) \\ &\le \frac{(1-\mu)\varepsilon}{c},\end{aligned}$$

where the last inequality used (42.6) and the triangle inequality. Thus by Proposition 39.8, the dual lift $\widehat{g}_h$ at least one invariant section $\gamma_{g,h} \in \Gamma^0(\Delta, T_\Delta M)$ satisfying

$$\begin{aligned}\sup_{x\in\Delta} \|\gamma_{g,h}(x)\|_\Delta &\le \frac{(1-\mu)\frac{\varepsilon}{c}}{1 - \tau(g,\Delta) - \mathrm{lip}_{\mathrm{fib}}(\Phi_{g,h}; \|\cdot\|_\Delta)} \\ &\le \frac{\varepsilon}{c}.\end{aligned}$$

Switching back to the Riemannian norm, this $\gamma_{g,h}$ (which is necessarily unique) satisfies $\sup_{x\in\Delta} \|\gamma_{g,h}(x)\| \le \varepsilon$. Setting $H(x) = \exp_x(\gamma_{g,h}(x))$, the continuous map $H\colon \Delta \to M$ satisfies $d_0(H, i_\Delta) \le \varepsilon$ and $h \circ H = H \circ g$ on Δ.

5. It remains to prove that H is injective. Suppose $x, y \in \Delta$ satisfy $H(x) = H(y)$. Then for any $k \in \mathbb{Z}$, by the triangle equality, we have

$$\begin{aligned}d\big(g^k(x), g^k(y)\big) &\le d\big(g^k(x), H(g^k(x))\big) + d\big(H(g^k(x)), H(g^k(y))\big) + d\big(H(g^k(y)), g^k(y)\big) \\ &\le \varepsilon + d\big(h^k(H(x)), h^k(H(y))\big) + \varepsilon \\ &= \varepsilon + 0 + \varepsilon \\ &< 2\varepsilon_0.\end{aligned}$$

Since $g|_\Delta$ is expansive with weak expansivity constant $2\varepsilon_0$, it follows that $x = y$. This finally completes the proof. ■

LECTURE 43

Isolated Hyperbolic Sets

In this lecture we look at isolated hyperbolic sets. Roughly speaking, an invariant set is *isolated* if there are no other invariants sets sufficiently close to it. The notion of an isolated invariant set makes sense in the topological category, so we begin there.

DEFINITION 43.1. Let $f\colon X \to X$ be a dynamical system on a compact[1] metric space. Given any set $A \subseteq X$, the **maximal invariant set** of f in A, written $\mathrm{inv}(A, f)$, is defined to be

$$\mathrm{inv}(A, f) := \bigcap_{k \in \mathbb{Z}} f^k(A).$$

Thus $\mathrm{inv}(A, f)$ consists of points whose total orbit never goes out of A.

LEMMA 43.2. *Let $f\colon X \to X$ be a dynamical system on a compact metric space, and let $A \subseteq X$ be either open or closed. Then:*

(i) $\mathrm{inv}(A, f) \subseteq A$ *is (possibly empty) completely invariant set, with* $\mathrm{inv}(A, f) = A$ *if and only if A is completely invariant.*

(ii) *if* $f(A) \subsetneq A$ *then* $\mathrm{inv}(A, f)$ *is compact.*

Proof. Part (i) is clear. Assume $f(A) \subsetneq A$, and let K be a compact set such that

$$\mathrm{inv}(A, f) \subseteq K \subsetneq A.$$

Then $\mathrm{inv}(A, f) = \mathrm{inv}(K, f)$. Since K is compact and f is continuous, $\mathrm{inv}(K, f)$ is compact. This proves part (ii). ∎

DEFINITION 43.3. A compact completely invariant set $Y \subseteq X$ of f is said to be **isolated** if there is a neighbourhood[2] U of Y such that

$$Y = \mathrm{inv}(U, f).$$

In this case we call U an **isolating neighbourhood** for Y.

REMARK 43.4. As the proof of Lemma 43.2 shows, if Y is an isolated invariant set with isolating neighbourhood U, then if K is any compact set with

$$Y \subset K^\circ \subset K \subset U$$

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]All of this material works in the non-compact case too, albeit with minor changes. However for the sake of a uniform presentation we stick to the compact setting.

[2]Recall our convention is that neighbourhoods are always open sets.

then we also have $Y = \operatorname{inv}(K, f)$. We then call K a **isolating compactum**[3] for Y. Conversely if Y is a compact completely invariant set with the property that there exists a compact set K such that $Y \subset K^\circ$ and $Y = \operatorname{inv}(K, f)$, then we also have $Y = \operatorname{inv}(K^\circ, f)$, and thus Y is isolated.

In order to cut down on the number of adjectives, in later lectures we will adopt the terminology that an **isolating set** for a compact completely invariant set is either an isolating neighbourhood or an isolating compactum. Thus isolating sets are rather malleable: once we have one we can shrink or enlarge it as required, and we can choose it to be open or compact, depending on which is more convenient.

The entire space X is trivially always isolated. A more interesting example is:

Example 43.5. Let $f\colon M \to M$ be a differentiable dynamical system, and suppose $x \in M$ is a hyperbolic fixed point. Then by Corollary 33.7, the set $\{x\}$ is always isolated.

Not all hyperbolic sets are isolated however. This is an important difference between hyperbolic fixed points and more general hyperbolic sets, and explains why hyperbolic dynamics can be more interesting than those of a fixed point.

Our key example of a non-isolated hyperbolic set—which we will cover in detail in Lecture 47—is the closure of the orbit of a **transverse homoclinic point**. The dynamics near a non-isolated hyperbolic set are typically extremely complicated. For instance, in Lecture 47 we will prove that a dynamical system is always chaotic in a neighbourhood of a transverse homoclinic point (cf. Corollary 47.10).

Remark 43.6. Despite the fact that both Theorem 41.7 and its big brother, Theorem 42.2, come with uniqueness clauses attached, they do *not* imply that all hyperbolic sets are isolated (which is good, since they're not!) As an instructive exercise, try to fallaciously use Theorem 42.2 to prove that all hyperbolic sets are isolated. You will no doubt quickly see why the proof breaks down.

Our aim in this lecture is to improve Theorem 42.2 in the case where the hyperbolic set is isolated. We begin with the following piece of topological dynamics.

Proposition 43.7. *Let X be a compact metric space and $f\colon X \to X$ a reversible dynamical system. Assume $Y \subseteq X$ is an isolated completely invariant set with isolating neighbourhood U. For any $a > 0$ there exists a neighbourhood $\mathcal{U}$ of f in* $\operatorname{Hom}(X)$ *such that if $g \in \mathcal{U}$ then the maximal invariant set of g in U is contained in $B(Z, a)$:*

$$\operatorname{inv}(U, g) \subset B(Z, a).$$

Proof. For any $a > 0$ there is $n \geq 1$ such that

$$\bigcap_{k=-n}^{n} f^k(U) \subset B(Z, a/2).$$

[3] A *compactum* is a fancy, if somewhat quaint, name for a compact set. The plural of compactum is *compacta*.

Since this is a finite intersection, there exists a neighbourhood $\mathcal{U}$ of f in $\mathrm{Hom}(X)$ such that for $g \in \mathcal{U}$ one also has

$$\bigcap_{k=-n}^{n} g^k(U) \subset B(Z, a).$$

Then by definition one also has $\mathrm{inv}(U, g) \subset B(Z, a)$. ■

Before stating the improvement of Theorem 41.7 to isolated hyperbolic sets, let us first introduce some more terminology, so as to make the forthcoming proof less clunky.

Definition 43.8. Suppose f and g are two reversible dynamical systems on a compact metric space (X, d). Let Y be a compact completely invariant set for f and Z be a compact completely invariant set for g. We say that $f|_Y$ and $g|_Z$ are **ε-conjugate** if there exists a homeomorphism $H\colon Y \to Z$ such that the following commutes:

$$\begin{tikzcd} Y \arrow[r, "f|_Y"] \arrow[d, "H"'] & Y \arrow[d, "H"] \\ Z \arrow[r, "g|_Z"'] & Z \end{tikzcd}$$

and that

$$d(H(x), x) \leq \varepsilon, \qquad \forall\, x \in Y.$$

Note such an H can exist only when

$$Y \cup Z \subset \big(B(Y, \varepsilon) \cap B(Z, \varepsilon)\big).$$

This definition is essentially just a rephrasing of the notion of weak structural ε-stability, and hence we have:

Example 43.9. Suppose Λ is a compact completely invariant set for f, and assume f is weakly structurally ε-stable on Λ. Then for g close enough to f, the continuation Λ_g of Λ to g is well-defined, and $f|_\Lambda$ and $g|_{\Lambda_g}$ are ε-conjugate.

We now present a strengthening of Theorem 41.7 in the isolated case.

Theorem 43.10. *Let f be a dynamical system on a compact smooth manifold M. Let Λ be an isolated hyperbolic set with isolating neighbourhood U. For any $\varepsilon > 0$, there is a neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ such that if $g \in \mathcal{U}$ then $\mathrm{inv}(U, g)$ is an isolated hyperbolic set for g, and $g|_{\mathrm{inv}(U,g)}$ is ε-conjugate to $f|_\Lambda$.*

Remark 43.11. Theorem 43.10 is rather similar to Theorem 41.7. Both of them assert that if Λ is hyperbolic for f then any sufficiently near g will also have a hyperbolic set Δ, near to Λ, and moreover $f|_\Lambda$ and $g|_\Delta$ are ε-conjugate. However Theorem 43.10 is stronger than Theorem 41.7 for three reasons.

- Firstly, Theorem 43.10 tells us that if original hyperbolic set Λ was isolated for f, then the new hyperbolic set Δ is isolated for g.

- Secondly, Theorem 43.10 actually gives us a way to *construct* the new invariant set Δ, namely,
$$\Delta = \mathrm{inv}(U, g).$$
In contrast, Theorem 41.7 merely proves the abstract existence of some ε-conjugacy H. It is not at all easy to extract from the proof of Theorem 41.7 a recipe for writing H down explicitly.

- Finally, Theorem 43.10 tells us that the new isolated set Δ only depends on U and g. This is surprising, since the ε-conjugacy H from Theorem 41.7 manifestly depends on both f and g.

The proof of Theorem 43.10 should remind you of the proof of the uniqueness part of Proposition 32.4.

Proof of Theorem 43.10. To keep the notation uncluttered during the proof, we will denote all inclusions simply by i. By Theorem 42.2 there is a neighbourhood $\mathcal{U}_0$ of f in $\mathrm{Diff}^1(M)$ and two numbers $a_0, \varepsilon_0 > 0$ such that for any $g, h \in \mathcal{U}_0$ and any compact completely invariant set $\Delta \subset B(\Lambda, a_0)$ of g, there is at *most* one continuous map $H \colon \Delta \to M$ such that

$$\begin{cases} H \circ g = h \circ H \text{ on } \Delta, \\ d_0(H, i) \leq \varepsilon_0. \end{cases} \tag{43.1}$$

Without loss of generality we may assume that $B(\Lambda, a_0 + \varepsilon_0) \subset U$. By Proposition 43.7 there is a C^0 neighbourhood $\mathcal{U}_1$ of f in $\mathrm{Hom}(M)$ such that if $g \in \mathcal{U}_1$ then

$$\mathrm{inv}(U, g) \subset B(\Lambda, a_0).$$

Now let $\varepsilon > 0$ be given. We may assume that $2\varepsilon < \varepsilon_0$. We apply Theorem 42.2 again to find another neighbourhood $\mathcal{U} \subset \mathcal{U}_0 \cap (\mathrm{Diff}^1(M) \cap \mathcal{U}_1)$ such that if $g, h \in \mathcal{U}$ and $\Delta \subset B(\Lambda, a_0)$ is a compact completely invariant set for g then there is at *least* one continuous injective map such that (43.1) holds (with ε_0 replaced with ε).

In particular, since $\mathrm{inv}(U, g)$ is one such set (cf. part (i) of Lemma 43.2), there is a continuous injective map $H \colon \mathrm{inv}(U, g) \to M$ with

$$\begin{cases} H \circ g = h \circ H \text{ on } \mathrm{inv}(U, g), \\ d_0(H, i) \leq \varepsilon. \end{cases}$$

Reversing the roles of g and h gives another continuous injection $G \colon \mathrm{inv}(U, h) \to M$ such that

$$\begin{cases} G \circ h = g \circ G \text{ on } \mathrm{inv}(U, h), \\ d_0(G, i) \leq \varepsilon. \end{cases}$$

Observe that

$$H\big(\mathrm{inv}(U, g)\big) \subset B\big(\mathrm{inv}(U, g), \varepsilon\big) \subset B(\Lambda, a_0 + \varepsilon_0) \subset U.$$

Since $H\big(\mathrm{inv}(U, g)\big)$ is h-invariant and $\mathrm{inv}(U, h)$ is the maximal invariant set of h in U, it follows that

$$H\big(\mathrm{inv}(U, g)\big) \subseteq \mathrm{inv}(U, h).$$

Thus the composition $F := G \circ H\colon \operatorname{inv}(U, g) \to M$ is therefore a well-defined continuous injective map which satisfies

$$\begin{cases} F \circ g = g \circ F \text{ on } \operatorname{inv}(U, g), \\ d_0(F, i) \leq \varepsilon + \varepsilon \leq \varepsilon_0. \end{cases} \tag{43.2}$$

However the inclusion $i\colon \operatorname{inv}(U, g) \hookrightarrow M$ is another solution of (43.2), and thus by the uniqueness of solutions to (43.1), we deduce that $F = i$. Similarly $H \circ G$ agrees with the inclusion $\operatorname{inv}(U, h) \hookrightarrow M$. Thus we must have

$$H\big(\operatorname{inv}(U, g)\big) = \operatorname{inv}(U, h)$$

and $g|_{\operatorname{inv}(U,g)}$ and $h|_{\operatorname{inv}(U,h)}$ are ε-conjugate. Now take $h = f$ to complete the proof. ■

We only used the hypothesis that Λ was isolated right at the very end of the proof. Thus the same argument proves:

Corollary 43.12. *Let f be a dynamical system on a compact smooth manifold M, and let Λ be a compact hyperbolic set. There exists a compact set K such that $\Lambda \subset K^\circ$ with the following property: For any $\varepsilon > 0$, there is a neighbourhood $\mathcal{U}$ of f in $\operatorname{Diff}^1(M)$ such that if $g \in \mathcal{U}$ then $\operatorname{inv}(K, g)$ is a hyperbolic set for g, and $g|_{\operatorname{inv}(K,g)}$ is ε-conjugate to $f|_\Lambda$.*

Applying Corollary 43.12 with $g = f$ shows that $\operatorname{inv}(K, f)$ is a hyperbolic set for f with $\Lambda \subseteq \operatorname{inv}(K, f)$. If Λ is not isolated then this inclusion is necessarily strict $\Lambda \subsetneq \operatorname{inv}(\overline{U}, f)$. This proves:

Corollary 43.13. *Let f be a dynamical system on a compact smooth manifold M, and let Λ be a compact hyperbolic set. There exists a compact set K with $\Lambda \subset K^\circ$ such that $\operatorname{inv}(K, f)$ is a compact hyperbolic set containing Λ.*

Remark 43.14. **Warning:** Corollary 43.13 does *not* assert that every compact hyperbolic set can be extended to an isolated hyperbolic set (since it may not hold that $\operatorname{inv}(K, f) \subset K^\circ$). In fact, there exist compact hyperbolic sets that cannot be extended to isolated hyperbolic sets. The first examples were found by Crovisier in 2001, and ten years later Fisher constructed various robust such examples.

LECTURE 44

The Shadowing Theorem

Let $f\colon X \to X$ be a reversible dynamical system on a compact metric space. Recall from Definition 3.13 that a **δ-chain** is a sequence (x_k), $k \in \mathbb{Z}$ such that

$$d\big(f(x_k), x_{k+1}\big) \le \delta, \qquad \forall\, k \in \mathbb{Z}.$$

A δ-chain (x_k) is **periodic** if there exists $p > 0$ such that $x_{k+p} = x_k$ for all $k \in \mathbb{Z}$. As explained in Remark 3.14, a (periodic) δ-chain can be thought of as a sequence of points in X that are indistinguishable from a (periodic) orbit of f by a "measuring device" which is only accurate up to the nearest δ.

Definition 44.1. Let $f\colon X \to X$ denote a reversible dynamical system on a compact metric space. Let (x_k) denote a δ-chain for f. A point $y \in X$ is said to **ε-shadow** the δ-chain (x_k) if

$$d\big(f^k(y), x_k\big) \le \varepsilon, \qquad \forall\, k \in \mathbb{Z}.$$

Roughly speaking, the main results of the next two lectures—the Shadowing Theorem 44.3 and the Anosov Closing Lemma 45.1—assert that, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that any δ-chain (x_k) is ε-shadowed by exactly one point y. Moreover if the δ-chain (x_k) is periodic then y is a periodic point of f.

Remark 44.2. In terms of our measuring device analogy, this tells us that an accuracy of δ is "good enough". From a real-world point of view, the importance of this statement cannot be overstated. No real-world measuring device can ever be perfectly accurate, and therefore in practice it is *never* possible to detect with 100% certainty a periodic orbit of a real-world dynamical system. The Shadowing Theorem tells us it suffices to use a device which is only accurate up to δ in order to prove the existence of a true orbit. The price to pay is that we can only claim that the true orbit lives somewhere within an ε-neighbourhood of where our measuring device thinks it does (and typically $\delta \ll \varepsilon$). Nevertheless, this is still a significant conceptual improvement—up to now, we had no way of proving the existence of *any* periodic orbits at all!

Here is the precise statement of the main result we will prove today.

Theorem 44.3 (The Shadowing Theorem). *Let f be a dynamical system on a compact manifold M and let $\Lambda \subset M$ be a compact hyperbolic set. There is $\varepsilon_0 > 0$ and $\delta_0 > 0$ such that every δ_0-chain contained in Λ is ε_0-shadowed by at most one point. Moreover for any $0 < \varepsilon < \varepsilon_0$ there is a $0 < \delta < \delta_0$ such that every δ-chain contained in Λ is ε-shadowed by at least (and hence exactly one) point.*

Remark 44.4. Note that Theorem 44.3 is *not* asserting that the unique point that ε-shadows the δ-chain actually belongs to Λ. However this is indeed the case if Λ is isolated, as you will prove on Problem Sheet S.

In order to prove Theorem 44.3 we will introduce a new concept: Our treatment will be somewhat ad hoc, as treating this in full generality would lead to unnecessarily complicated notation.

Definition 44.5. Suppose $\mathtt{x} = (x_k)$ is any sequence in M. Let

$$\mathbb{T}_{\mathtt{x}}M := \big\{(k, v) \in \mathbb{Z} \times TM \mid v \in T_{x_k}M\big\},$$

and let $\omega\colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{Z}$ denote the map that sends (k, v) to k. If we think of $\mathbb{Z}$ as a (disconnected!) zero-dimensional manifold, then $\omega\colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{Z}$ is a vector bundle over $\mathbb{Z}$. We call $\mathbb{T}_{\mathtt{x}}M$ the **tangent bundle over** $\mathtt{x}$.

Definition 44.6. $\|\cdot\|$

A **norm** $|\!|\!|\cdot|\!|\!|$ on $\mathbb{T}_{\mathtt{x}}M$ is a choice $\big(\|\cdot\|_k\big)_{k\in\mathbb{Z}}$ of norms on each tangent space $T_{x_k}M$. Given such a norm $|\!|\!|\cdot|\!|\!|$, we denote by

$$\mathbb{T}_{\mathtt{x}}M(r) := \big\{(k, v) \in \mathbb{Z} \times TM \mid v \in T_{x_k}M, \|v\|_k \le r\big\}.$$

Remark 44.7. If $U \subseteq M$ is an open set containing every point x_k, then any C^0 norm on $T_U M$ induces a norm $|\!|\!|\cdot|\!|\!|$ on $\mathbb{T}_{\mathtt{x}}M$ by restriction. However it is important to realise that not every norm $|\!|\!|\cdot|\!|\!|$ on $\mathbb{T}_{\mathtt{x}}M$ comes from a norm on TM. As an extreme example, if $\mathtt{x}$ is the constant sequence $x_k \equiv x$ then a norm $|\!|\!|\cdot|\!|\!|$ on $\mathbb{T}_{\mathtt{x}}M$ is a simply a collection $\big(\|\cdot\|_k\big)$ of norms on T_xM. Unless $\|\cdot\|_k$ is also a constant sequence of norms, the norm $|\!|\!|\cdot|\!|\!|$ cannot be induced from a norm on TM.

Definition 44.8. We denote by $\Gamma(\mathtt{x})$ the space of sections of $\mathbb{T}_{\mathtt{x}}M$. Such a section γ takes the form

$$\gamma\colon \mathbb{Z} \to \mathbb{T}_{\mathtt{x}}M, \qquad \gamma(k) = (k, \tilde{\gamma}(k)), \tag{44.1}$$

where $\tilde{\gamma}\colon \mathbb{Z} \to TM$ is a map such that $\tilde{\gamma}(k) \in T_{x_k}M$.

Equip $\Gamma(\mathtt{x})$ with the norm

$$|\!|\!|\gamma|\!|\!| := \sup_{k\in\mathbb{Z}} \|\tilde{\gamma}(k)\|_k.$$

Let $\Gamma_b(\mathtt{x}) \subset \Gamma(\mathtt{x})$ denote the bounded sections

$$\Gamma_b(\mathtt{x}) := \big\{\gamma \in \Gamma(\mathtt{x}) \mid |\!|\!|\gamma|\!|\!| < \infty\big\}.$$

Then $\big(\Gamma_b(\mathtt{x}), \|\cdot\|\big)$ is a Banach space. Given $r \ge 0$, let $\Gamma_r(\mathtt{x}) \subset \Gamma_b(\mathtt{x})$ denote those sections γ with $|\!|\!|\gamma|\!|\!| \le r$.

Notation. Denote by $+\colon \mathbb{Z} \to \mathbb{Z}$ the map $k \mapsto k + 1$.

We will mainly be interested continuous fibre preserving maps $\boldsymbol{\Phi}$ over $+$.

$$\begin{tikzcd}
\mathbb{T}_{\mathtt{x}}M \arrow[r, "\boldsymbol{\Phi}"] \arrow[d, "\omega"'] & \mathbb{T}_{\mathtt{x}}M \arrow[d, "\omega"] \\
\mathbb{Z} \arrow[r, "+"'] & \mathbb{Z}
\end{tikzcd}$$

Explicitly, this means $\mathbf{\Phi}$ is of the form

$$\mathbf{\Phi}(k,v) = \big(k+1, \Phi_k(v)\big), \tag{44.2}$$

where $\Phi_k \colon T_{x_k}M \to T_{x_{k+1}}M$ is a continuous map. If each Φ_k is a linear map, we say that $\mathbf{\Phi}$ is a **vector bundle morphism** over $+$.

(♣) REMARK 44.9. The preceding definitions may all seem somewhat contrived. However they fit into a standard differential geometry construction of a **pullback bundle**. For completeness, let us briefly recall how pullback bundles are defined. Suppose $f \colon M \to N$ is a smooth between two smooth manifolds, and $\pi \colon P \to N$ is a vector bundle over N with fibre F. We define the **pullback bundle** $f^*P \to M$ as follows: Set

$$f^*E \coloneqq \big\{(x,p) \in M \times P \mid f(x) = \pi(p)\big\},$$

and define $\omega \colon f^*E \to M$ by

$$\omega(x,p) \coloneqq x.$$

It follows straightforwardly from the definitions that $\omega \colon f^*E \to M$ is a vector bundle over M with fibre F, with

$$f^*E(x) \cong P(f(x)), \qquad \forall\, x \in M.$$

To fit Definition 44.5 into this framework, observe that we can regard a sequence $\mathtt{x} = (x_k)$ in M as a smooth map $\mathtt{x} \colon \mathbb{Z} \to M$, where $\mathbb{Z}$ is thought of as (disconnected!) 0-dimensional manifold. Then by definition, we have

$$\mathtt{x}^*TM = \mathbb{T}_{\mathtt{x}}M.$$

DEFINITION 44.10. Fix a sequence $\mathtt{x} = (x_k)$ in M, and fix a norm $|\!|\!|\cdot|\!|\!| = \big(\|\cdot\|\big)_{k\in\mathbb{Z}}$ on $\mathbb{T}_{\mathtt{x}}M$. A **hyperbolic linear operator over** $\mathtt{x}$ is vector bundle morphism

$$\mathbb{L} \colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{T}_{\mathtt{x}}M, \qquad (k,v) \mapsto (k+1, L_k v)$$

over $+$ which is uniformly hyperbolic, in the sense that each

$$L_k \colon T_{x_k}M \to T_{x_{k+1}}M$$

is a hyperbolic linear operator[1], such that the same constants work for all L_k. That is, each space $T_{x_k}M$ has a splitting

$$T_{x_k}M = E^s_k \oplus E^u_k,$$

such that $L_k E^s_k \subseteq E^s_{k+1}$ and $L_k E^u_k \subseteq E^u_{k+1}$, and that there exist constants $C \geq 1$ and $0 < \mu < 1$ such that for all $k \in \mathbb{Z}$,

$$\begin{aligned} \|L_{k+i} \circ \cdots \circ L_{k+1} \circ L_k v\|_{k+i+1} &\leq C\mu^i \|v\|_k, \qquad \forall\, v \in E^s_k,\ \forall\, i \geq 0, \\ \|L^{-1}_{k-i} \circ \cdots \circ L^{-1}_{k-1} \circ L^{-1}_k v\|_{k-i} &\leq C\mu^i \|v\|_k, \qquad \forall\, v \in E^u_{k+1},\ \forall\, i \geq 0. \end{aligned}$$

[1] This is a slight generalisation of our previous notion of a hyperbolic linear operator, as the domain and range are not the same normed vector space.

Since the constants C, μ are independent of k, one can apply Proposition 29.11 to produce a norm $\lvert\cdot\rvert_{\mathrm{a}}$ on the bundle $\mathbb{T}_{\mathtt{x}}M$ which is adapted to each L_k (i.e. for which $C = 1$) and for which the skewness of $\mathbb{L}$ with respect to this norm:

$$\tau(\mathbb{L}) := \sup_{k \in \mathbb{Z}} \tau(L_k) \tag{44.3}$$

is strictly less than one[2]. One can then also perform the process described in Lemma 29.14 to produce a norm $\lvert\cdot\rvert_{\mathrm{ab}}$ which is both adapted and of box type for $\mathbb{L}$. Note that even if $\lvert\cdot\rvert$ is induced from a norm on M (cf. Remark 44.7), $\lvert\cdot\rvert_{\mathrm{ab}}$ may not be.

The notion of a Lipschitz perturbation of a hyperbolic linear operator is defined as you would guess:

Definition 44.11. Let $\mathbf{\Phi} \colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{T}_{\mathtt{x}}M$ be a continuous fibre preserving map over $+$. Write $\mathbf{\Phi}$ as in (44.2). We say that $\mathbf{\Phi}$ is **Lipschitz continuous** if each map

$$\Phi_k \colon \big(T_{x_k}M, \|\cdot\|_k\big) \to \big(T_{x_{k+1}}M, \|\cdot\|_{k+1}\big)$$

is Lipschitz continuous, and moreover if the Lipschitz constants are bounded uniformly, that is,

$$\operatorname{lip}(\mathbf{\Phi}) := \sup_{k\in\mathbb{Z}} \operatorname{lip}(\Phi_k) < \infty.$$

Definition 44.12. Let $\mathbf{\Phi} \colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{T}_{\mathtt{x}}M$ be a continuous fibre preserving map over $+$. A section γ is said to be **invariant** if

$$\mathbf{\Phi}(\gamma(k)) = \gamma(k+1), \qquad \forall k \in \mathbb{Z}.$$

that is, if

$$\Phi_k(\tilde{\gamma}(k)) = \tilde{\gamma}(k+1), \qquad \forall k \in \mathbb{Z},$$

where $\tilde{\gamma}$ and Φ_k are as in (44.1) and (44.2) respectively.

Let $\mathbf{0}$ denote the trivial section

$$\mathbf{0}(k) = (k, 0_{x_k}).$$

The following result is very similar to (but simpler than) Proposition 39.8.

Proposition 44.13. *Let $\mathtt{x} = (x_k)$ denote any sequence in M. Suppose $\mathbb{L} \colon \mathbb{T}_{\mathtt{x}}M \to \mathbb{T}_{\mathtt{x}}M$ is a hyperbolic linear operator. Let $\lvert\cdot\rvert$ denote a norm on $\mathbb{T}_{\mathtt{x}}M$ which is adapted and of box type with respect to $\mathbb{L}$, and let $0 < \tau < 1$ denote the skewness of $\mathbb{L}$ with respect to $\lvert\cdot\rvert$. Fix $r > 0$, and assume $\mathbf{\Phi} \colon \mathbb{T}_{\mathtt{x}}M(r) \to \mathbb{T}_{\mathtt{x}}M$ is a Lipschitz continuous map satisfying*

$$\operatorname{lip}(\mathbf{\Phi}) < 1 - \tau.$$

Then $\mathbb{L} + \mathbf{\Phi}$ has at most one invariant section $\gamma \in \Gamma_r(\mathtt{x})$. If in addition one has

$$\lvert\mathbf{\Phi}(\mathbf{0})\rvert \le \big(1 - \tau - \operatorname{lip}(\mathbf{\Phi})\big)r,$$

then $\mathbb{L} + \mathbf{\Phi}$ has at least one (and hence exactly one) invariant section $\gamma_{\mathbf{\Phi}} \in \Gamma_r(\mathtt{x})$, which moreover satisfies

$$\lvert\gamma_{\mathbf{\Phi}}\rvert \le \frac{\lvert\mathbf{\Phi}(\mathbf{0})\rvert}{1 - \tau - \operatorname{lip}(\mathbf{\Phi})}.$$

[2]If the constants C, μ were not bounded independently of k then one could still produce an adapted norm $\lvert\cdot\rvert_{\mathrm{a}}$ on the bundle $\mathbb{T}_{\mathtt{x}}M$, but then the supremum in (44.3) could be equal to 1.

Proof. Argue as in Proposition 39.8, replacing all "x" subscripts with "k". ■

We now prove the Shadowing Theorem 44.3.

Proof of the Shadowing Theorem 44.3. Let $\|\cdot\|$ denote a C^0 norm on $T_\Lambda M$ which is adapted and of box type with respect to f and Λ. Take $\varepsilon_0 > 0$ and $\delta_0 > 0$ small enough such that for every $0 < \delta < \delta_0$ and every δ-chain $\mathtt{x} = (x_k)$, the map

$$\mathbb{F} \colon \mathbb{T}_{\mathtt{x}} M(\varepsilon_0) \to \mathbb{T}_{\mathtt{x}} M$$

given by

$$\mathbb{F}(k, v) := \left(k+1, \exp^{-1}_{x_{k+1}} \circ f \circ \exp_{x_k}(v)\right)$$

is well defined. Let $0 < \varepsilon < \varepsilon_0$. If a point $y \in M$ ε-shadows a δ-chain $\mathtt{x} = (x_k)$ then

$$\gamma(k) := \left(k, \exp^{-1}_{x_k}(f^k(y))\right)$$

is an $\mathbb{F}$-invariant section belonging to $\Gamma_\varepsilon(\mathtt{x})$. Thus the existence of a point y that ε-shadows a δ-chain is equivalent to the existence of an invariant section of $\mathbb{F}$.

First note that there exists a constant $r_0 > 0$ such that if $d(f(x), y) < r_0$ then the operator

$$D_{x,y} := D(\exp_y)^{-1}(f(x)) \circ Df(x) \colon T_x M \to T_y M$$

is a well-defined linear operator which depends continuously on both x and y and agrees with $Df(x)$ for $y = f(x)$. Thus for $r < r_0$ sufficiently small, if $x, y \in \Lambda$ satisfy $d(f(x), y)) \leq r$ then the operator $D_{x,y}$ satisfies the hypotheses of Proposition 31.9 with respect to the hyperbolic splitting of $T_x M = E^s(x) \oplus E^u(x)$ and $T_y M = E^s(y) \oplus E^u(y)$. Thus by Proposition 31.9, for $r > 0$ sufficiently small, if $x, y \in \Lambda$ satisfy $d(f(x), y) \leq r$ then $D_{x,y}$ is a hyperbolic linear operator from $T_x M$ to $T_y M$.

Thus up to shrinking δ_0, the operator

$$\mathbb{D}f \colon \mathbb{T}_{\mathtt{x}} M \to \mathbb{T}_{\mathtt{x}} M$$

defined by

$$\mathbb{D}f(k, v) = \left(k+1, D_{x_k, x_{k+1}} v\right)$$

is a hyperbolic linear operator over $\mathtt{x}$. Moreover using the continuity statement from Proposition 31.9 together with Proposition 38.10, we can produce from $\|\cdot\|$ a new norm $|\!|\!|\cdot|\!|\!| = \left(\|\cdot\|_k\right)_{k\in\mathbb{Z}}$ on $\mathbb{T}_{\mathtt{x}} M$ which is both adapted and of box type with respect to $\mathbb{D}f$, and which is uniformly equivalent to $\|\cdot\|$ in the sense that there exists $c \geq 1$ such that

$$\frac{1}{c}\|v\| \leq \|v\|_k \leq c\|v\|, \qquad \forall\, v \in T_{x_k} M, \qquad \forall\, k \in \mathbb{Z}.$$

Since

$$\mathbb{F}(k, v) = \left(k+1, \exp^{-1}_{x_{k+1}} \circ \exp_{f(x_k)} \circ \widehat{f}(x_k, v)\right),$$

it follows from the same argument as Proposition 39.7 that for any $r > 0$, there exists $\varepsilon > 0$ such that

$$\mathbf{\Phi} := \mathbb{F} - \mathbb{D}f \colon \mathbb{T}_{\mathtt{x}} M \to \mathbb{T}_{\mathtt{x}} M$$

satisfies $\operatorname{lip}(\mathbf{\Phi}; \|\cdot\|) < r$ on a small ball about the zero section in $\mathbb{T}_{\mathtt{x}}M$. Since $\left|\!\left|\cdot\right|\!\right|$ is uniformly equivalent to $\|\cdot\|$, the "norm-switching" argument we saw in the proofs of Proposition 40.8 and Theorem 42.2 tells us that $\operatorname{lip}(\mathbf{\Phi}; \left|\!\left|\cdot\right|\!\right|)$ can also be made arbitrarily small on a small ball in $\mathbb{T}_{\mathtt{x}}M$.

Thus after shrinking δ_0 and ε_0, we can apply Proposition 44.13 to $\mathbb{F}$. Both statements of the Theorem follow immediately from Proposition 44.13, and so the proof is complete. ∎

Axiom A and the Anosov Closing Lemma

We begin this lecture with an extension of the Shadowing Theorem that allows us to "detect" periodic orbits (cf. Remark 44.2). This famous result is due to Anosov. Recall we say that a δ-chain (x_k) is **periodic** if there exists $p > 0$ such that $x_{k+p} = x_k$ for all $k \in \mathbb{Z}$. The minimal such p is called the **period** of (x_k).

THEOREM 45.1 (The Anosov Closing Lemma). *Let f be a dynamical system on a compact manifold M and let $\Lambda \subset M$ be a compact hyperbolic set. For every $\varepsilon > 0$ there exists $\delta > 0$ such that every periodic δ-chain in Λ is ε-shadowed by a periodic point of f.*

The name "Closing Lemma" stems from the fact that the Theorem allows us to "close" a periodic chain up into a true orbit.

Proof. By the Shadowing Theorem 44.3 there are $\varepsilon_0 > 0$ and $\delta_0 > 0$ such that every δ_0-chain can be ε_0-shadowed by at most one point, and moreover if $0 < \varepsilon \le \varepsilon_0$ then there exists $0 < \delta = \delta(\varepsilon) \le \delta_0$ such that every δ-chain is ε-shadowed by at least one point.

Now let $\varepsilon > 0$ and let $\delta = \delta(\varepsilon)$ be as above. Let (x_k) be a periodic δ-chain of period p. Then there exists a point y that ε-shadows the (x_k):

$$d\big(f^k(y), x_k\big) \le \varepsilon, \qquad \forall\, k \in \mathbb{Z}.$$

Since $x_{k+p} = x_k$, it follows that

$$d\big(f^k(f^p(y)), x_k\big) = d(f^{k+p}(y), x_{k+p}) < \varepsilon, \qquad \forall\, k \in \mathbb{Z}.$$

Thus $f^p(y)$ also ε-shadows (x_k). By uniqueness, this implies that $f^p(y) = y$. Since f is reversible, y is a periodic point[1]. This completes the proof. ∎

We now prove an stronger version of the Shadowing Theorem 44.3 and the Anosov Closing Lemma 45.1. This result allows the chain to slightly "escape" the hyperbolic set. The improvement is mild and somewhat technical, but it will be useful later on in the course.

PROPOSITION 45.2. *Let f be a dynamical system on a compact manifold M and let $\Lambda \subset M$ be a compact hyperbolic set. For every $\varepsilon > 0$ sufficiently small there exists $\rho > 0$ such that every ρ-chain contained in $B(\Lambda, \rho)$ is ε-shadowed by a unique point in M. Similarly for every $\varepsilon > 0$ sufficiently small there exists $\rho > 0$ such that every periodic ρ-chain contained in $B(\Lambda, \rho)$ in ε-shadowed by a unique periodic point of f.*

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]Note however that p does not have to be the minimal period of y.

Proof. We prove the first statement only, since the second follows in exactly the same way as the proof of of the Anosov Closing Lemma 45.1. Given $\varepsilon > 0$ sufficiently small, the Shadowing Theorem 44.3 gives us $\delta > 0$ such that every δ-chain contained in Λ itself is $\frac{\varepsilon}{2}$-shadowed by a point, and moreover that there is at most one point that $\frac{3\varepsilon}{2}$-shadows this point. Let $r > 0$ be such that

$$d(w, z) \leq r \qquad \Rightarrow \qquad d(f(w), f(z)) \leq \frac{\delta}{3}.$$

Set

$$\rho := \min\left\{ r, \frac{\delta}{3}, \frac{\varepsilon}{2} \right\}.$$

Then if (x_k) is any r-chain and (y_k) is any collection of points in M such that

$$d(y_k, x_k) \leq r, \qquad \forall\, k \in \mathbb{Z},$$

then (y_k) is actually a δ-chain. Now suppose (x_k) is a ρ-chain contained in $B(\Lambda, \rho)$. Then there exists $y_k \in \Lambda$ such that $d(x_k, y_k) \leq \rho$. We claim that (y_k) is itself a δ-chain. Indeed:

$$\begin{aligned} d\big(f(y_k), y_{k+1}\big) &\leq d\big(f(y_k), f(x_k)\big) + d\big(f(x_k), x_{k+1}\big) + d\big(x_{k+1}, y_{k+1}\big) \\ &\leq \frac{\delta}{3} + \rho + \rho \\ &\leq \delta. \end{aligned}$$

Thus $\frac{\varepsilon}{2}$-shadowed by a point $z \in M$. Then since $\rho \leq \frac{\varepsilon}{2}$, (x_k) is also ε-shadowed by z. Finally, this point z is unique, since if w is any point that ε-shadows (x_k), then w also $\frac{3\varepsilon}{2}$-shadows the y_k, and by assumption there is at most such point. ■

Next, let us recall the notion of the chain recurrent set from Definition 3.13. If $f\colon X \to X$ is a reversible dynamical system on a compact metric space then a point $x \in X$ is said to be **chain recurrent** if for any $\delta > 0$ there is a finite δ-chain that starts and ends at x. That is, for any $\delta > 0$ there exists a tuple (x_k), $k = 0, \dots, p$, such that $x_0 = x_p = x$ and $d\big(f(x_k), x_{k+1}\big) \leq \delta$ for each $k = 0, \dots, p-1$. The set of chain recurrent points is denoted[2] by $\mathsf{cha}(f)$. By Proposition 3.15 the chain recurrent set is a non-empty compact completely invariant set which contains the non-wandering set $\mathsf{nw}(f)$, and hence also

$$\overline{\mathsf{per}(f)} \subseteq \mathsf{cha}(f). \tag{45.1}$$

We will need the following slight extension of Proposition 3.15.

PROPOSITION 45.3. *Let $f\colon X \to X$ be a reversible dynamical system on a compact metric space. For any $\varepsilon > 0$ there is $\delta > 0$ such that if $(x_k)_{k\in\mathbb{Z}}$ is any periodic δ-chain passing through a point in $\mathsf{cha}(f)$ then $(x_k) \subset B(\mathsf{cha}(f), \varepsilon)$.*

[2]In Definition 3.13 we wrote $\mathsf{cha}_d(f)$ to indicate the dependence of the chain recurrent set on the metric. However by Proposition 3.16, in the compact case the chain recurrent set is independent of the metric. Since we will only be interested in compact spaces, we omit the d in our notation.

Proof. If the result is false we can find numbers $(p_k) \in \mathbb{N}$ and sequences (x_i^k) of points in X for $k \in \mathbb{Z}$ and $0 \le i \le p_k$ such that for each $k \in \mathbb{Z}$:

- $(x_i^k)_{0 \le i \le p_k}$ is a finite periodic $\frac{1}{k}$-chain that starts in $\mathsf{cha}(f)$,
- $(x_i^k)_{0 \le i \le p_k}$ is not contained in the ball of radius ε about $\mathsf{cha}(f)$, i.e. there exists $1 \le l_k \le p_k - 1$ such that $x_k^{l_k} \notin B(\mathsf{cha}(f), \varepsilon)$.

Up to passing to a subsequence, we may assume that $x_{l_k}^k \to y$ for some point $y \in X$. This point y does not belong to $\mathsf{cha}(f)$, since it is a limit of points that lie at least ε from $\mathsf{cha}(f)$.

This means that there exists $\delta > 0$ such that there is no finite periodic δ-chain that starts and ends at y. Let $r > 0$ be such that

$$d(w, z) \le r \qquad \Rightarrow \qquad d(f(w), f(z)) \le \frac{\delta}{2},$$

and choose $k\frac{2}{\delta}$ such that $d\big(y, x_{l_k}^k\big) \le r$. Then

$$\begin{aligned} d\big(f(y), x_{l_k+1}^k\big) &\le d\big(f(y), f(x_{l_k}^k)\big) + d\big(f(x_{l_k}^k), x_{l_k+1}^k\big) \\ &\le \frac{1}{k} + \frac{\delta}{2} + \frac{1}{k} \\ &\le \delta, \end{aligned}$$

and similarly

$$\begin{aligned} d\big(f(x_{l_k-1}^k), y\big) &\le d\big(f(x_{l_k-1}^k), x_{l_k}^k\big) + d\big(x_{l_k}^k, y)\big) \\ &\le \frac{\delta}{2} + \frac{1}{k} \\ &\le \delta. \end{aligned}$$

Therefore

$$y, x_{l_k+1}^k, \ldots, x_{p_k-1}^k, x_{p_k}^k = x_0^k, x_1^k, \ldots, x_{l_k-1}^k, y$$

is a finite δ-chain from y to itself, which contradicts the choice of δ. The proof is complete. ∎

We will use Proposition 45.3 to show that if the chain recurrent set is hyperbolic then (45.1) is actually an equality.

Proposition 45.4. *Let f be a dynamical system on a compact manifold M. If* $\mathsf{cha}(f)$ *is hyperbolic then* $\mathsf{cha}(f) = \overline{\mathsf{per}(f)}$.

Proof. Let $x \in \mathsf{cha}(f)$ and take $\varepsilon > 0$ small. By Proposition 45.2 there exists $\rho > 0$ such that every periodic ρ-chain contained in $B(\mathsf{cha}(f), \rho)$ in ε-shadowed by a periodic point of f. Next, by the last statement of Proposition 45.3 there exists $0 < \delta < \rho$ such that any periodic δ-chain through x is contained in $B(\mathsf{cha}(f), \rho)$. Since $x \in \mathsf{cha}(f)$ such a δ-chain certainly exists. Thus there exists a periodic point y of f that ε-shadows this chain. In particular, $d(x, f^i(y)) \le \varepsilon$ for some iterate $f^i(y)$ of y. Since ε was arbitrary, this shows that $x \in \overline{\mathsf{per}(f)}$, and since x was arbitrary it offlows that $\mathsf{cha}(f) \subseteq \overline{\mathsf{per}(f)}$. Since the reverse inclusion always holds by Proposition 3.15, the proof is complete. ∎

If $\mathsf{cha}(f)$ is hyperbolic then

$$\mathsf{cha}(f) = \mathsf{nw}(f),$$

since the non-wandering set is sandwiched between $\overline{\mathsf{per}(f)}$ and $\mathsf{cha}(f)$. Thus

$$\mathsf{cha}(f) \text{ hyperbolic} \qquad \Rightarrow \qquad \mathsf{nw}(f) \text{ hyperbolic.}$$

The converse direction is not true. It is clear from Proposition 45.3 that a necessary condition for the chain recurrent set to be hyperbolic is that $\mathsf{nw}(f)$ is hyperbolic and $\mathsf{nw}(f) = \overline{\mathsf{per}(f)}$. Such a class of dynamical systems gets its own (rather unhelpful) name.

Definition 45.5. Let $f \colon M \to M$ be a dynamical system on a compact manifold. We say that f satisfies **Axiom A** if the non-wandering set $\mathsf{nw}(f)$ is hyperbolic and

$$\mathsf{nw}(f) = \overline{\mathsf{per}(f)}.$$

The importance of the Axiom A condition will become clear in Lecture 50 when we discuss **omega stability**. For now let us just note that:

Corollary 45.6. *Let $f \colon M \to M$ be a dynamical system on a compact manifold. If f does not satisfy Axiom A then $\mathsf{cha}(f)$ is not hyperbolic.*

In fact, an equivalent formulation of the main theorem from Lecture 49 is that the chain recurrent set is hyperbolic *if and only if* f satisfies Axiom A and has **no basic cycles**. The latter condition will be defined in Lecture 49.

We conclude this lecture by discussing (yet another) type of stable manifold.

Definition 45.7. Let $f \colon X \to X$ be a reversible dynamical system on a compact metric space X and let $A \subseteq X$ be a compact completely invariant set. We define the **global stable manifold of** A to be the set

$$W^s(A, f) := \big\{x \in X \mid d\big(f^k(x), A\big) \to 0 \text{ as } k \to \infty\big\},$$

and the **global unstable manifold of** A to be the set

$$W^u(A, f) := \big\{x \in X \mid d\big(f^{-k}(x), A\big) \to 0 \text{ as } k \to \infty\big\},$$

The name is a bit of a misnomer, as these sets are typically *not* manifolds. The next lemma is on Problem Sheet S.

Lemma 45.8. *Let $f \colon X \to X$ be a reversible dynamical system on a compact metric space X. Let $A \subset X$ be a compact completely invariant set. Then*

$$x \in W^s(A, f) \qquad \Leftrightarrow \qquad \omega_f(x) \subset A,$$

and

$$x \in W^u(A, f) \qquad \Leftrightarrow \qquad \alpha_f(x) \subset A,$$

If $x \in A$ and $y \in W^s(x, f)$ then clearly $y \in W^s(A, f)$, and similarly for the unstable case. But what about the other direction?

Definition 45.9. Let $f\colon X \to X$ be a reversible dynamical system on a compact metric space X. Let $A \subseteq X$ be a compact completely invariant set. We say that a point $y \in W^s(A, f)$ is **in phase** with a point $x \in A$ if $y \in W^s(x, f)$. Similarly we say that a point $y \in W^u(A, f)$ is **negatively in phase** with a point $x \in A$ if $y \in W^u(x, f)$.

In general there is no reason why every point must be in phase. However for an isolated hyperbolic set, this is always the case.

Theorem 45.10 (The In Phase Theorem). *Let f be a dynamical system on a compact manifold M and let $\Lambda \subset M$ be an isolated compact hyperbolic set. Then every point is in phase:*

$$W^s(\Lambda, f) = \bigcup_{x \in \Lambda} W^s(x, f),$$

and

$$W^u(\Lambda, f) = \bigcup_{x \in \Lambda} W^u(x, f).$$

Proof. We give the proof for the stable sets only. It is clear that the right-hand side is contained in $W^s(\Lambda, f)$, and thus if suffices to show that if $y \in W^s(\Lambda, f)$ then $y \in W^s(x, f)$ for some $x \in \Lambda$.

Let $r > 0$ be the number from Proposition 40.8 so that for all $x \in \Lambda$,

$$W^s_{\mathrm{loc},r}(x, f) = \big\{ z \in M \mid d\big(f^k(z), f^k(x)\big) \le r, \ \forall\, k \ge 0 \big\}.$$

Shrinking r if necessary, we may assume that $B(\Lambda, r)$ is contained in an isolating neighbourhood U of Λ. By the Shadowing Lemma 44.3, there is a $0 < \delta < \frac{r}{2}$ such that every δ-chain in Λ is $\frac{r}{2}$-shadowed by some point in M. Using continuity of f, take $0 < \varepsilon < \delta$ such that if $(x_k) \subset \Lambda$ is any sequence of points such that

$$d(f^k(y), x_k) \le \varepsilon, \qquad \forall\, k \ge 0,$$

then (x_k) is actually part of a δ-chain. Since $y \in W^s(\Lambda, f)$, there exists $p \ge 1$ such that $d\big(f^k(y), \Lambda\big) \le \varepsilon$ for all $k \ge p$. This means that for all $k \ge p$ there are points $x_k \in \Lambda$ such that $d\big(f^k(y), x_k\big) \le \varepsilon$. Thus $(x_k)_{k \ge p}$, is part of a δ-chain. If we set $x_k := f^{k-p}(x_p)$ for $k < p$ then $(x_k)_{k \in \mathbb{Z}}$ is a true δ-chain.

Thus (x_k) is $r/2$-shadowed by some point x. This implies that

$$\mathcal{O}_f^{\mathrm{total}}(x) \subset B(\Lambda, r/2) \subset U.$$

Since U is an isolating neighbourhood for Λ, it follows that $x \in \Lambda$. Then for $k \ge p$,

$$d\big(f^k(y), f^k(x)\big) \le d\big(f^k(y), x_k\big) + d\big(x_k, f^k(x)\big) \le \varepsilon + \frac{r}{2} \le r.$$

By choice of r, it follows that $f^p(y) \in W^s_{\mathrm{loc},r}(f^p(x), f)$. Thus $y \in W^s(x, f)$ by (40.1) (Problem S.1). This completes the proof. ∎

We will use the In Phase Theorem 45.10 in Lecture 49 (cf. Corollary 49.4 and Proposition 49.11.)

LECTURE 46

The Inclination Lemma

The aim of this lecture is to prove the following theorem of Palis, which is one of the cornerstones of chaotic dynamics. This result is sometimes also called[1] the "Inclination Lemma".

THEOREM 46.1 (The Inclination Lemma). *Let f be a dynamical system on a compact manifold M and suppose x is a hyperbolic fixed point. Set $d := \dim E^u(x)$ and fix a point $y \in W^s(x, f)$. Suppose B and D are two C^1 embedded discs of dimension d in M with*

$$B \subset W^u(x, f), \qquad y \in D,$$

and such that D is transverse to $W^s(x, f)$ at y:

$$T_y D + T_y W^s(x, f) = T_y M.$$

Then for any $\varepsilon > 0$ there is a $p \geq 1$ such that for all $k \geq p$, $f^k(D)$ contains a C^1 embedded disc of dimension d which is ε-close to B in the C^1-topology.

We will see next lecture why this theorem is powerful. There are three "surprising" things about the statement: D could be very small, B could be very large, and the angle between D and $W^s(x, f)$ could be very small. See Figure 46.1. The difficult bit is, of course, the C^1 statement: the C^0 statement is (almost) obvious from the picture.

REMARK 46.2. The Inclination Lemma 46.1 has a straightforward extension to hyperbolic periodic points. Indeed, if x is a hyperbolic periodic point of f of period p, then x is a hyperbolic fixed point of f^p. Moreover from the definition we have

$$W^u(x, f) = W^u(x, f^p), \qquad W^s(x, f) = W^s(x, f^p).$$

Thus in the statement of Theorem 46.1 the phrase "hyperbolic fixed point" can be replaced with "hyperbolic periodic point".

Let us first consider a local version. Suppose $f\colon \Omega \to E$ is a local dynamical system, and suppose $0 \in \Omega$ is a hyperbolic fixed point. We endow E with a norm which is adapted to $Df(0)$ and of box type with respect to the splitting $E = E^s \oplus E^u$. Instead of the direct sum notation, it is more useful today to think of $E = E^s \times E^u$ as a product. Since the norm is of box type, one has $E(r) = E^s(r) \times E^u(r)$. To keep the notation in this lecture as simple as possible, we will always use the letter u for a point in E, v for a point in E^s, and w for a point in E^u. Thus the notation $u = (v, w)$ means that $v = u_s$ and $w = u_u$.

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]The reason for the name "λ-Lemma" comes from the fact that Palis denoted a certain constant—defined in (46.2) below— by λ. (You might think I'm joking. I'm not.)

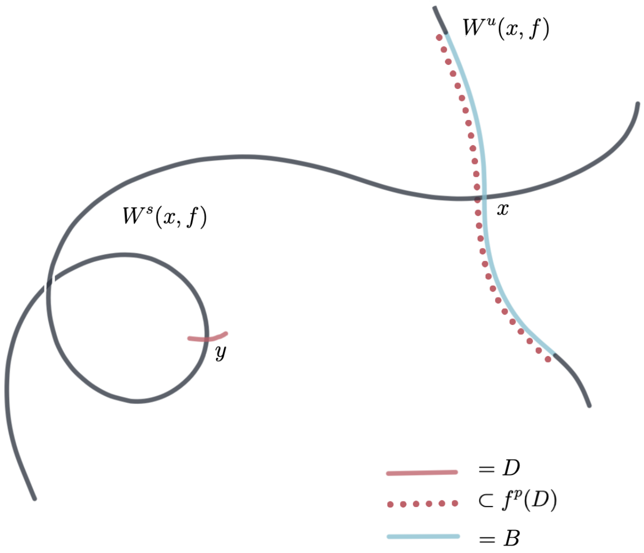

Figure 46.1: The Inclination Lemma.

Now fix $r > 0$ such that the conclusion of the Local Unstable Manifold Theorem 34.3 holds. Thus there are C^1 maps

$$\xi_s\colon E^s(r) \to E^u(r), \qquad \xi_u\colon E^u(r) \to E^s(r),$$

with

$$\xi_s(0) = 0, \qquad D\xi_s(0) = 0, \qquad \xi_u(0) = 0, \qquad D\xi_u(0) = 0,$$

and such that

$$W^s_{\mathrm{loc},r}(0,f) = \mathrm{gr}(\xi_s), \qquad W^u_{\mathrm{loc},r}(0,f) = \mathrm{gr}(\xi_u).$$

Consider the map

$$\sigma\colon E^s(r) \times E^u(r) \to E, \qquad \sigma(v,w) = \big(v - \xi_u(w), w - \xi_s(v)\big).$$

Then

$$\sigma(0,0) = (0,0), \qquad D\sigma(0,0) = \mathrm{id},$$

and hence by the Inverse Function Theorem 30.7, after possibly shrinking r, the map σ is a C^1 diffeomorphism onto its image. We can therefore view σ as a chart on E, and look at the representation $\hat{f}$ of f in this chart:

$$\hat{f} := \sigma \circ f \circ \sigma^{-1}\colon E(r) \to E.$$

Then

$$\hat{f}(0) = 0, \qquad D\hat{f}(0) = Df(0),$$

and hence 0 is also a hyperbolic fixed point of $\hat{f}$. This coordinate change has the nice effect of putting the stable manifolds "on the axes". See Figure 46.2.

Lemma 46.3. *The local stable manifold of $\hat{f}$ is of the form $B^s \times \{0\}$ for B^s a ball in E^s, and the local unstable manifold of $\hat{f}$ is of the form $\{0\} \times B^u$ for B^u a ball in E^u,*

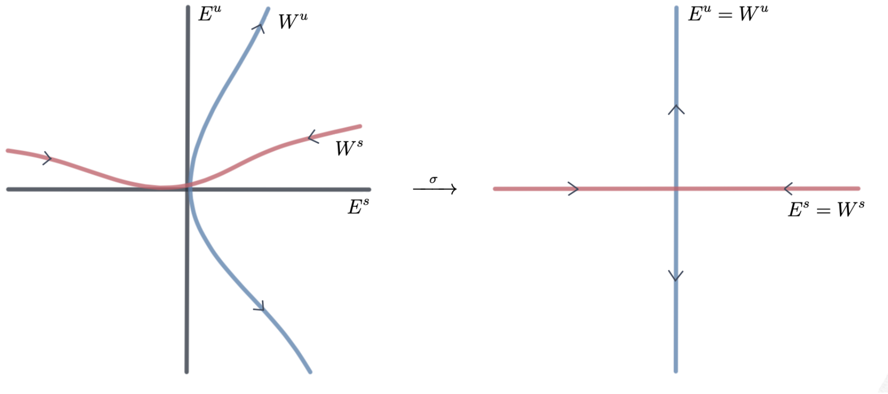

Figure 46.2: The coordinate change σ.

Proof. Observe that $\sigma(v,w) = (v',0)$ for some v' if and only if $w = \xi_s(v)$. Thus for $r > 0$ sufficiently small,

$$\sigma^{-1}\big(E^s(r) \times \{0\}\big) \subset \mathrm{gr}(\xi_s).$$

The lemma follows. ■

From now on we suppress the $\hat{f}$ notation and simply write f again. Choose $r > 0$ small enough that $E^s(r) \subset B^s$ and $E^u(r) \subset B^u$. Set $V = E^s(r) \times E^u(r)$. Then every point in V of the form $(v,0)$ belongs to the local stable manifold and every point in V of the form $(0,w)$ belongs to the local unstable manifold.

Consider a point $(q,0) \in V$ and let D denote a disc of dimension $d := \dim E^u$ which is transversal to the local stable manifold at $(q,0)$. For $k \geq 0$ let D_k denote the connected component of $f^k(D) \cap V$ containing $f^k(q,0)$. The difficult part of the proof of Theorem 46.1 is contained in the following local version. See Figure 46.3.

Theorem 46.4 (The Local Inclination Lemma). *Given $\varepsilon > 0$ there exists $p \geq 1$ such that for all $k \geq p$, D_k contains a C^1 embedded disc of dimension d which is ε-close to $\{0\} \times E^u(r)$ in the C^1 topology.*

Let us show how the (global) Inclination Lemma follows from the local one.

Proof of the Inclination Lemma 46.1. We may assume that y is in $W^s_{\mathrm{loc},r}(x,f)$ for r small by replacing y with $f^k(y)$ for k large. Similarly by shrinking D if necessary we may assume that D is contained in small ball about x. It then suffices to show that we can make D close to a small disc B inside $W^u_{\mathrm{loc},r}(x,f)$, since a small disc in $W^u(x,f)$ gets mapped onto a larger disc by successive applications of f. This means that it is enough to prove a local statement. Moreover by choosing a chart σ on M about x appropriately, we may assume that we are in the special situation prescribed in the Local Inclination Lemma 46.4. Thus Theorem 46.1 is a direct consequence of the Local Inclination Lemma. ■

The rest of the lecture is devoted to the proof of Theorem 46.4. This proof is non-examinable.

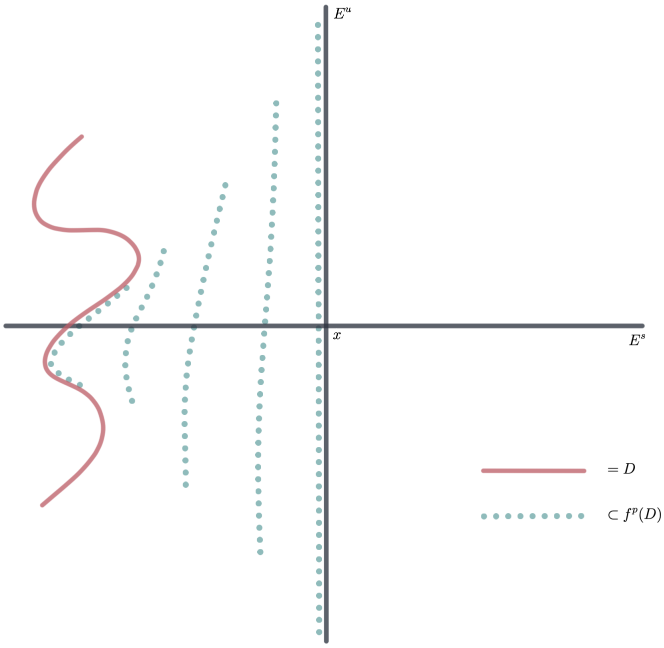

Figure 46.3: The Local Inclination Lemma.

Proof of Theorem 46.4. We proceed in five steps.

1. In this first step we set up notation. Let us abbreviate $L \coloneqq Df(0,0)$ and $A \coloneqq L_{ss}$ and $B \coloneqq L_{uu}$. Then we can write

$$f(v,w) \overset{\text{def}}{=} \big(v + \varphi(v,w), Bw + \psi(v,w)\big),$$

for two C^1 functions φ, ψ. Let $\tau < 1$ denote the skewness of L, so that

$$\|A\|^{\mathrm{op}} \le \tau, \qquad \|B^{-1}\|^{\mathrm{op}} \le \tau.$$

Note also that since $L_{us} = L_{su} = 0$, one has

$$\frac{\partial \varphi}{\partial w}\Big|_{\{0\}\times E^u(r)} = \frac{\partial \psi}{\partial v}\Big|_{E^s(r)\times\{0\}} = 0.$$

Thus by continuity of these partial derivatives there exists $0 < \mu < 1$ and $0 < r' < r$ such that

$$\tau + \mu < 1, \qquad c \coloneqq \frac{1}{\tau} - \mu > 1, \qquad \mu < \frac{(c-1)^2}{4}, \tag{46.1}$$

and such that if $V' \coloneqq E^s(r') \times E^u(r')$ then

$$\max_{V'} \left\{ \left\| \frac{\partial \varphi}{\partial w} \right\|^{\mathrm{op}}, \left\| \frac{\partial \psi}{\partial v} \right\|^{\mathrm{op}} \right\} \le \mu.$$

We may assume that $q \in E^s(r')$ and that $D \subset V'$. Now suppose u_0 is a unit length vector in $T_{(q,0)}D$. Write $u_0 = (v_0, w_0)$. Note that $w_0 \neq 0$ as D is transversal to E^s at $(q,0)$. Consider the **inclination** of u_0:

$$\frac{\|v_0\|}{\|w_0\|}.$$

The main goal of the proof is to show that under iteration by Df, the inclination gets smaller in a uniform manner, which will imply that TD approaches $\{0\} \times E^u$ under Df. This is essentially just a long, delicate, and rather tedious computation.

2. In this step we show we can control the inclination of u_0 under iteration. Set

$$(q_k, 0) \coloneqq f^k(q, 0), \qquad u_k = (v_k, w_k) \coloneqq Df(q_{k-1}, 0)u_{k-1}, \qquad k \geq 0.$$

Let λ_k denote the inclination of u_k:

$$\lambda_k \coloneqq \frac{\|v_k\|}{\|w_k\|}. \tag{46.2}$$

Since the unit ball in $T_{(q,0)}D$ is compact, we may assume that we have chosen u_0 so as to maximise λ_0. Let us now estimate λ_k. We have

$$\begin{aligned} Df(q,0)u_0 &= \begin{pmatrix} A + \frac{\partial \varphi}{\partial v}(q,0) & \frac{\partial \varphi}{\partial w}(q,0) \\ 0 & B + \frac{\partial \psi}{\partial w}(q,0) \end{pmatrix} \begin{pmatrix} v_0 \\ w_0 \end{pmatrix} \\ &= \begin{pmatrix} Av_0 + \frac{\partial \varphi}{\partial v}(q,0)v_0 + \frac{\partial \varphi}{\partial w}(q,0)w \\ Bw_0 + \frac{\partial \psi}{\partial w}(q,0)w_0 \end{pmatrix}. \end{aligned}$$

Thus

$$\begin{aligned} \|v_1\| &\leq \|Av_0\| + \left\| \frac{\partial \varphi}{\partial v}(q,0)v_0 \right\| + \left\| \frac{\partial \varphi}{\partial w}(q,0)w_0 \right\| \\ &\leq \tau\|v_0\| + \mu\|v_0\| + \mu\|w_0\|. \end{aligned}$$

Similarly

$$\begin{aligned} \|w_1\| &\geq \|Bw_0\| - \left\| \frac{\partial \psi}{\partial w}(q,0)v_0 \right\| \\ &\geq \frac{1}{\tau}\|w_0\| - \mu\|w_0\|, \end{aligned}$$

and hence

$$\begin{aligned} \lambda_1 &\leq \frac{\tau\lambda_0 + \mu\lambda_0 + \mu}{\frac{1}{\tau} - \mu} \\ &\leq \frac{\lambda_0 + \mu}{c} \\ &= \frac{\lambda_0}{c} + \frac{\mu}{c}. \end{aligned}$$

Similarly we have

$$\begin{aligned} \lambda_2 &\leq \frac{\lambda_1 + \mu}{c} \\ &\leq \frac{\lambda_0}{c^2} + \mu\left(\frac{1}{c} + \frac{1}{c^2}\right), \end{aligned}$$

and more generally

$$\lambda_k \le \frac{\lambda_0}{c^k} + \mu \sum_{i=1}^{k} \frac{1}{c^i}$$
$$\le \frac{\lambda_0}{c^k} + \frac{\mu}{c-1}.$$

Since $\frac{\lambda_0}{c^k} \to 0$ as $k \to \infty$ and

$$\frac{\mu}{c-1} < \frac{c-1}{4},$$

by (46.1), there exists $p \ge 1$ such that for all $k \ge p$,

$$\lambda_k \le \frac{c-1}{4}.$$

3. Since f is a diffeomorphism, by the choice of u_0 we know that know that any non-zero vector in $T_{(q_p,0)}D_p$ has inclination less than or equal to $\frac{c-1}{4}$. But what about vectors in $T_z D_k$ for points $z \ne q_k$? In this short step we address this point. Since D is a C^1 embedded submanifold and f is of class C^1, the tangent planes $T_z D_p$ depend continuously on z. Thus there exists a small C^1 embedded d-dimensional disc $D' \subset D_p$ with centre $(q_p, 0)$ such that for any $z \in D'$ and any unit vector $u \in T_z D'$, the inclination of u is at worst $\frac{c-1}{2}$.

4. We have shown that we can bound the inclinations uniformly. But this is not good enough—in order to show that the discs approach $\{0\} \times E^u$ we need to prove that the inclinations converge to zero. We prove this now.

Let $0 < \delta < \mu$ be arbitrary. Since

$$\frac{\partial \varphi}{\partial w}\Big|_{\{0\}\times E^u(r)} = 0,$$

we may choose $r'' < r'$ such that on $V'' := E^s(r'') \times E^u(r')$ we have

$$\max_{V'} \left\| \frac{\partial \varphi}{\partial w} \right\|^{\mathrm{op}} \le \delta.$$

Up to increasing p and shrinking D', we may assume that $D' \subset V''$ and $(q_k, 0) \in V''$ for all $k \ge p$. Fix $z \in D'$ and let $u' = (v', w')$ denote a unit vector in $T_z D'$. Let λ' denote the inclination of u'. As before, we may assume that u' is chosen in such a way to maximise the inclination λ'. We again compute the inclination of the iterates of u' under Df. This is the same computation as before, only the bottom left-hand term no longer vanishes:

$$Df(z)u' = \begin{pmatrix} A + \frac{\partial \varphi}{\partial v}(z) & \frac{\partial \varphi}{\partial w}(z) \\ \frac{\partial \psi}{\partial v}(z) & B + \frac{\partial \psi}{\partial w}(z) \end{pmatrix} \begin{pmatrix} v' \\ w' \end{pmatrix}.$$

Thus the inclination is

$$\begin{aligned}
\frac{\left\|Av' + \frac{\partial\varphi}{\partial v}(z)v' + \frac{\partial\psi}{\partial w}(z)v'\right\|}{\left\|\frac{\partial\psi}{\partial v}(z)v' + Bw' + \frac{\partial\psi}{\partial w}(z)w'\right\|} &\leq \frac{\tau\|v'\| + \mu\|v'\| + \delta\|w'\|}{\frac{1}{\tau}\|w'\| - \mu\|w'\| - \mu\|v'\|} \\
&\leq \frac{\tau\lambda' + \mu\lambda' + \delta}{\frac{1}{\tau} - \mu - \mu\lambda'} \\
&\leq \frac{\lambda' + \delta}{c - \mu\lambda'} \\
&\overset{(\heartsuit)}{\leq} \frac{\lambda' + \delta}{c - \frac{1}{2}\mu(c-1)} \\
&\leq \frac{\lambda' + \delta}{\frac{1}{2}(c+1)},
\end{aligned}$$

where $(\heartsuit)$ used Step 3. Set $a := \frac{1}{2}(c+1) > 1$. Denoting by λ'_k the inclination of the kth iterate $u'_k = Df^k u'$, as before we obtain

$$\lambda'_k \leq \frac{\lambda'}{a^k} + \frac{\delta}{a-1}.$$

Thus there exists l such that for all $k \geq l$, one has

$$\lambda'_k \leq \delta\left(1 + \frac{1}{a-1}\right).$$

By choice of u', this shows that for any $k \geq l$, every non-zero vector tangent to $f^k(D') \cap V''$ has inclination at most

$$\delta\left(1 + \frac{1}{a-1}\right).$$

Since δ was arbitrary, this shows we can makes the inclinations uniformly small.

5. We are almost done. The last step is to estimate the length of a vector tangent to $f^k(D') \cap V''$, compared to that of its iterate. But this is easy: with $u'_k = (v'_k, w'_k)$ and λ'_k as above,

$$\begin{aligned}
\|u'_k\| &= \max\left\{\|v'_k\|, \|w'_k\|\right\} \\
&= \|w'_k\|,
\end{aligned}$$

provided $\lambda'_k < 1$. Moreover from the computation above,

$$\frac{\|w'_{k+1}\|}{\|w'_k\|} \geq c - \mu\lambda'_k.$$

Since $\lambda'_k \to 0$ and $c > 1$, we see that $\frac{\|u'_{k+1}\|}{\|u'_k\|} > 1$ for k sufficiently large. Thus the length grows.

Hence the diameter of $f^k(D') \cap V''$ increases. Combining this with Step 4 tells us that for any $\varepsilon > 0$, there exists $p = p(\varepsilon) \geq 1$ such that, for all $k \geq p$, the set $f^k(D') \cap V''$ contains a C^1 embedded disc which is ε-close to $\{0\} \times E^u$ in the C^1 topology. This finally completes the proof. ■

LECTURE 47

Homoclinic Tangles

These intersections form a sort of trellis, web, or infinitely tight mesh...
One is struck by the complexity of this figure, which I shall not even attempt to draw.

HENRI POINCARÉ, 1889.

In this lecture we explore some applications of the Inclination Lemma 46.1 to chaotic dynamics. Recall the definition of homoclinic points and heteroclinic points from Definition 16.8. A *homoclinic tangle* arises from a special type of homoclinic point. Here is the definition.

DEFINITION 47.1. Let f be a dynamical system on a compact manifold M, and suppose $x \neq y$ are hyperbolic periodic points. A point $z \in M$ is called a **transverse heteroclinic point** if $z \in W^s(x, f) \cap W^u(y, f)$ and the intersection is transverse at z:

$$T_z W^s(x, f) + T_z W^u(y, f) = T_z M.$$

Similarly if x is a hyperbolic periodic point then a point $z \neq x \in W^s(x, f) \cap W^u(x, f)$ is called a **transverse homoclinic point** for x if the intersection is transverse at z:

$$T_z W^s(x, f) + T_z W^u(x, f) = T_z M.$$

The transversality condition by itself is nothing special, as the following result shows.

LEMMA 47.2. *Let f be a dynamical system on a compact manifold M, and let $\Lambda \subseteq M$ be a compact hyperbolic set. Then there exists $r_0 > 0$ such that for any $0 < r \leq r_0$ there exists a $\delta > 0$ such that if $x, y \in \Lambda$ satisfy $d(x, y) \leq \delta$ then*

$$W^s_{\mathrm{loc},r}(x, f) \cap W^u_{\mathrm{loc},r}(y, f) = \{z\} \tag{47.1}$$

for a unique point $z \in M$, and moreover the intersection is always transverse.

The proof of Lemma 47.2 is deferred to Problem Sheet T.

REMARK 47.3. In fact, more is true. Let $\Delta \subseteq \Lambda \times \Lambda$ denote the diagonal. Lemma 47.2 tells us that there is a well-defined map $\varphi \colon B(\Delta, \delta) \to M$ that sends a pair (x, y) to the unique point z from the right-hand side of (47.1). In Problem T.1 you will show that φ is itself a continuous function, and that Λ is isolated if and only if[1] $\operatorname{im} \varphi \subseteq \Lambda$.

[1]Actually in Problem T.1 you are only asked to prove $\Rightarrow$. The converse direction—whilst true—is sadly too hard to set as an exercise.

Nevertheless, the existence of a transverse homoclinic point has *profound* consequences for the nearby dynamics. It was Poincaré who first realised this, during his work on the Three Body Problem in the late 19th century. His investigations led him to imagine a figure of quite breathtaking complexity. The famous quote[2] at the start of the lecture sums up his bewilderment at the implications of this discovery. This was arguably humanity's first taste of (mathematical!) chaos, and the development of the entire modern theory of chaotic and hyperbolic dynamics can be traced back to Poincaré's observations.

So what is so special about a transverse homoclinic point? The starting observation is that if $z \in W^s(x, f) \cap W^u(x, f)$ is a transverse homoclinic point then so is $f^k(z)$ for any $k \in \mathbb{Z}$. Thus the existence of one implies the existence of infinitely many. Now as an exercise (do this before turning the page!), try to draw what the unstable stable and stable manifolds of x must look like. The intersection is transverse at z. In order for it to be also transverse at $f(z)$, the unstable manifold and the stable manifold must "double back" on themselves. Then in order for them to be transverse at $f^2(z)$, they have to double back on themselves again. By the time you have reached $f^4(z)$, your picture will no doubt look rather messy... Since $f^k(z) \to x$ for $|k| \to \infty$, the upshot is that the unstable and stable manifolds are forced to seesaw back on themselves infinitely often in a neighbourhood of x, creating a "mesh" that Poincaré named a **homoclinic tangle**.

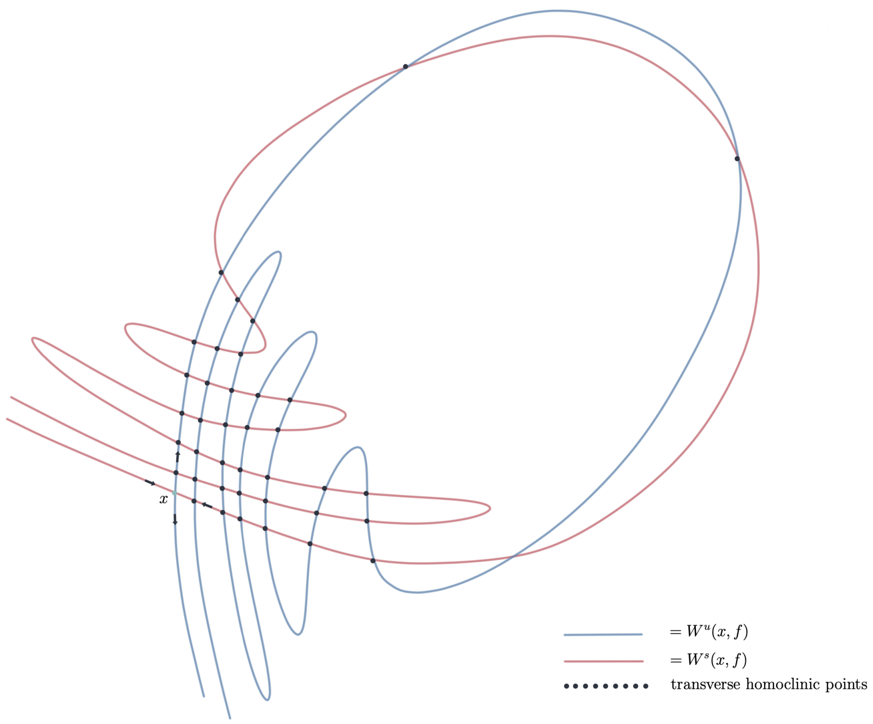

Figure 47.1: A homoclinic tangle.

[2]The original French version reads: *"Ces intersections forment une sorte de treillis, de tissu, de réseau à maille infiniment serrées... On sera frappé de la complexité de cette figure, que je ne cherche même pas à tracer."* This is from Volume 3 of Poincaré's series of papers *"Les Méthodes nouvelles de la mécanique céleste"*, published in 1899. The original text is available online here.

We shall shortly see how homoclinic tangles give rise to chaos. First, however, we use the Inclination Lemma to extend hyperbolic sets over the orbits of transverse heteroclinic (or homoclinic points).

PROPOSITION 47.4. *Suppose Λ is a hyperbolic set, and $x, y \in \Lambda$ are fixed points. Suppose there exists a point $z \in M$ such that $z \in W^s(x, f) \cap W^u(y, f)$, and assume that the intersection is transverse at z. Then $\Lambda' := \Lambda \cup \mathcal{O}_f^{\text{total}}(z)$ is another hyperbolic set.*

Proof. Let the hyperbolic splitting of Λ be given by $T_\Lambda M = E^s \oplus E^u$. To extend this to Λ', set

$$E^s(f^k(z)) := T_{f^k(z)} W^s(x, f)$$

and

$$E^u(f^k(z)) := T_{f^k(z)} W^u(y, f).$$

We first need to check this splitting is continuous. This means we need to show that:

(i) $E^s(f^k(z)) \to E^s(x)$ as $k \to \infty$,

(ii) $E^u(f^{-k}(z)) \to E^u(y)$ as $k \to \infty$,

(iii) $E^s(f^{-k}(z)) \to E^s(y)$ as $k \to \infty$,

(iv) $E^u(f^k(z)) \to E^u(x)$ as $k \to \infty$.

Of these (i) and (ii) are obvious. However (iii) and (iv) are not. However the Inclination Lemma comes to the rescue. Indeed, since the intersection is transverse, (iv) follows directly from the Theorem 46.1. Similarly (iii) can be deduced from applying Theorem 46.1 to f^{-1}.

We still need to prove the splitting is hyperbolic. We may assume our norm is adapted to f and Λ. Let τ denote the skewness and choose $\tau < \mu < 1$. Since both the splitting and Df are continuous, there exists a neighbourhood U of x and a neighbourhood V of y and such that

$$f^k(z) \in U \cup V \qquad \Rightarrow \qquad \max\left\{ \left\| Df|_{E^s(f^k(z))} \right\|^{\text{op}}, \left\| Df^{-1}|_{E^u(f^{-k}(z))} \right\|^{\text{op}} \right\} \le \mu.$$

Since $z \in W^s(x, f) \cap W^u(y, f)$, by Proposition 40.8 there exists $p \ge 1$ such that $f^k(z) \in U$ for all $k \ge p$ and $f^{-k}(z) \in V$ for all $k \ge p$. Thus there are only finitely many k such that $f^k(z) \notin U \cup V$, and thereforethere exists a constant $C \ge 1$ such that

$$\max\left\{ \left\| Df|_{E^s(f^k(z))} \right\|^{\text{op}}, \left\| Df^{-1}|_{E^u(f^{-k}(z))} \right\|^{\text{op}} \right\} \le C\mu, \qquad \forall\, k \in \mathbb{Z}.$$

This completes the proof. ■

As a special case, we can now give an important example of a hyperbolic set that is *not* isolated (cf. the discussion after Example 43.5.

COROLLARY 47.5. *If y is a transverse homoclinic point for a hyperbolic fixed point x then $\mathcal{O}_f^{\text{total}}(y) \cup \{x\}$ is a hyperbolic set which is not isolated.*

Proof. Proposition 47.4 tells us that $\Lambda := \mathcal{O}_f^{\text{total}}(y) \cup \{x\}$ is hyperbolic. Let r_0 be as in Lemma 47.2 and choose $0 < r \leq \min\{d(x,y), r_0\}$. Let $\delta > 0$ be the constant associated to r from Lemma 47.2. Choose k large enough so that both $f^k(y)$ and $f^{-k}(y)$ belongs to the ball of radius $\frac{\delta}{2}$ about x. Then Lemma 47.2 tells local (un)stable manifolds $W^s_{\text{loc},r}(f^{-k}(y), f)$ and $W^u_{\text{loc},r}(f^k(y), f)$ intersect at a unique point z. See Figure 47.2. Then $\mathcal{O}_f^{\text{total}}(z)$ is a completely invariant set which is disjoint from Λ but contained in $B(\Lambda, r)$. Since r was arbitrary it follows that Λ is not isolated. ■

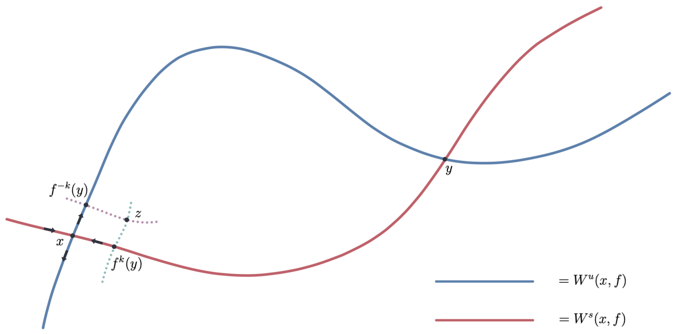

Figure 47.2: A non-isolated hyperbolic set.

Here is another application of the Inclination Lemma.

LEMMA 47.6. *Let f be a dynamical system on a compact 2-dimensional manifold M. Suppose x is a hyperbolic fixed point of f with a* **homoclinic loop** *(see Figure 47.3). Then every point on this loop is non-wandering.*

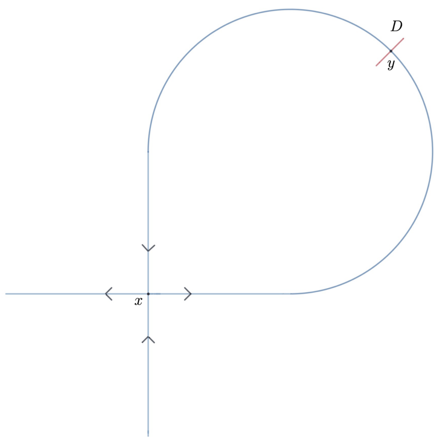

Figure 47.3: A homoclinic loop.

Proof. Fix a point y on this loop, and let D denote a small disc transversal to the loop passing through y. The Inclination Lemma 46.1 tells for k large enough, $f^k(D)$ will approach the stable manifold, and thus in particular for k large one has $f^k(D) \cap D \neq \emptyset$. Since D is arbitrary, this shows that $y \in \mathsf{nw}(f)$. ■

We will now state a famous theorem which explains how transverse homoclinic points gives rise to chaotic dynamics. We first recall the shift map from Definition 4.15.

Definition 47.7. Let Σ_2 denote the space of infinite sequences[3] $\mathtt{x} = (x_k)$ for $k \in \mathbb{Z}$, where each x_k is either 0 or 1. Endow Σ_2 with the metric

$$d(\mathtt{x}, \mathtt{y}) := \sum_{k \in \mathbb{Z}} \frac{|x_k - y_k|}{2^{|k|}}.$$

Let $\sigma \colon \Sigma_2 \to \Sigma_2$ denote the map that "shifts" a sequence along one position to the right, i.e. $\sigma(\mathtt{x})_k = x_{k+1}$.

The next result is the reversible reversion of Propositions 4.16 and 4.20 and Problem F.1.

Proposition 47.8. *The space Σ_2 is a compact metric space without isolated points which is totally disconnected*[4]*. Moreover σ is a chaotic reversible dynamical system on Σ_2 with $\mathsf{h}_{\mathrm{top}}(\sigma) = \log 2$.*

Here is the main result of today's lecture.

Theorem 47.9 (Birkhoff-Smale). *Let f be a dynamical system on a compact manifold M. Let $x \in M$ be a hyperbolic periodic point, and suppose $y \in M$ is a transverse homoclinic point. For any neighbourhood U of $\{x, y\}$, there exists $p \geq 1$ and a compact invariant set $\Lambda \subset U$ of f^p containing x and y such that $f^p|_\Lambda \colon \Lambda \to \Lambda$ is topologically conjugate to the shift map σ.*

Corollary 47.10. *Let f be a dynamical system on a compact manifold M. Let $x \in M$ be a hyperbolic periodic point, and suppose $y \in M$ is a transverse homoclinic point. Then the topological entropy of f is positive, and in a neighbourhood of $\{x, y\}$ an iterate of f displays chaotic behaviour.*

Proof. Using the notation from the statement of Theorem 47.9, we see that $f^p|_\Lambda$ is chaotic and has positive topological entropy. Thus the topological entropy of f^p on all of M is positive, since the topological entropy is bounded below by the entropy of the restriction to any invariant set (Proposition 8.5). Since $\mathsf{h}_{\mathrm{top}}(f^p) = p\,\mathsf{h}_{\mathrm{top}}(f)$ by Problem D.2, we also see that the topological entropy of f is positive. ■

The full proof of Theorem 47.9 is a bit too involved for us, but we will give a fairly compelling sketch. To do so we first introduce a special dynamical system called[5] the **horseshoe map**. The horseshoe map can be thought of as an abstraction of the system near a homoclinic tangle.

Definition 47.11. Let $Q \subset \mathbb{R}^2$ be a square of size 1. Define a diffeomorphism h so that Q gets contracted horizontally and expanded vertically and folded into a horse-shoe shape and then put back across itself. We call h the **horseshoe map**. See Figure 47.4.

[3] Strictly speaking in Definition 4.15 we focused on the non-reversible case and considered only half-infinite sequences. However the extension to the reversible case is straightforward.

[4] And hence is homeomorphic to a Cantor set, by Remark 4.17.

[5] See Remark 13.10 for a comment on the terminology.

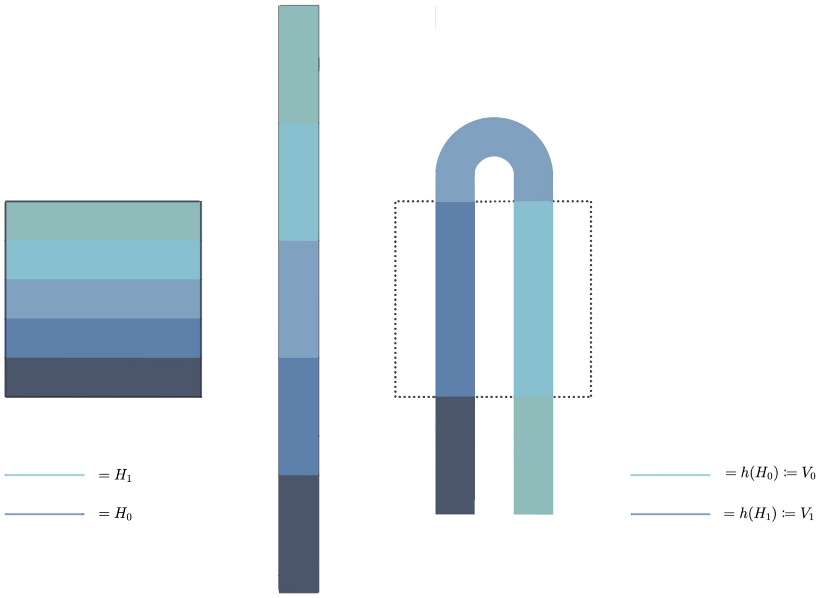

Figure 47.4: The horseshoe map h.

Think of Q as sitting inside S^2, as in Figure 47.5. We extend h to a global diffeomorphism of S^2 such that the south pole becomes an attracting fixed point and the northern hemisphere gets mapped into itself. Nevertheless, we will focus our attention on Q only.

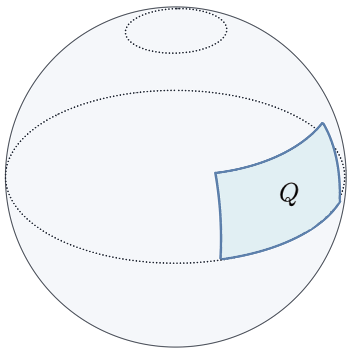

Figure 47.5: Q inside S^2.

Let us fix two horizontal strips H_0 and H_1 and two vertical strips V_0 and V_1 such that $h(H_0) = V_0$ and $h(H_1) = V_1$. Thus a point $z \in Q$ has $h(z) \in Q$ if and only if $z \in H_0 \cup H_1$. We assume that h is affine on the H_i with contraction rate $1/5$ and expansion rate 5 respectively. In Figure 47.4, H_0 and V_0 are dark blue and H_1 and V_1 are light blue.

Convention. Let us say temporarily say that a **horizontal strip** in Q is a rectangle that is contained in either H_0 or H_1 and which runs parallel from the left edge of Q to the right. See Figure 47.6. Similarly a rectangle that is contained in V_0 or V_1 and runs from the bottom of Q to the top of Q is a called **vertical strip**.

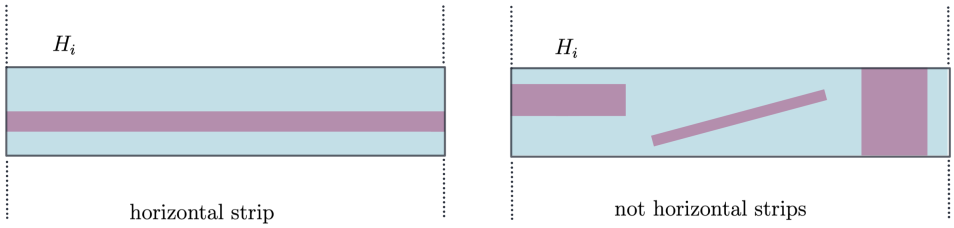

Figure 47.6: Horizontal strips.

LEMMA 47.12. *If V is a vertical strip then both $h(V) \cap V_0$ and $h(V) \cap V_1$ are vertical strips, and moreover their width is exactly $1/5$ of the width of V. Similarly if H is a horizontal strip then both $h^{-1}(H) \cap H_0$ and $h^{-1}(H) \cap H_1$ are horizontal strips, with height exactly $1/5$ of the height of H.*

Proof. The case of a vertical strip is obvious (draw a picture!) If H is a horizontal strip then

$$h^{-1}(H) \cap H_i = h^{-1}(H \cap h(H_i)) = h^{-1}(H \cap V_i).$$

Since $H \cap V_i$ is a rectangle contained in V_i that crosses V_i horizontally, its h^{-1} image is a horizontal strip. ∎

DEFINITION 47.13. Now set

$$\Lambda := \bigcap_{k \in \mathbb{Z}} h^k(Q).$$

Thus Λ is a compact h-invariant set. We call Λ the **Smale Horseshoe**.

PROPOSITION 47.14. *The horseshoe map $h\colon \Lambda \to \Lambda$ is conjugate to the shift map $\sigma\colon \Sigma_2 \to \Sigma_2$.*

Thus in particular Λ is homeomorphic to a Cantor set.

Proof. If $z \notin H_0 \cup H_1$ then $h(z) \notin Q$. Thus

$$\Lambda = \bigcap_{k \in \mathbb{Z}} h^k(H_0 \cup H_1).$$

Since $H_0 \cap H_1 = \emptyset$, for any $z \in \Lambda$ there is a unique $\mathtt{x} \in \Sigma_2$ such that $h^k(z) \in H_{x_k}$. We define map $F\colon \Lambda \to \Sigma_2$ by $F(z) = \mathtt{x}$. It is clear that $F \circ h = \sigma \circ F$ from the definition. The difficult part is showing that F is a homeomorphism. Let us first show F is a bijection. Suppose $\mathtt{x} \in \Sigma_2$. We need to find a unique $z \in \Lambda$ such that $h^k(z) \in H_{x_k}$ for each $k \in \mathbb{Z}$. Equivalently, for each $\mathtt{x} \in \Sigma_2$, we need to show that the intersection

$$\bigcap_{k \in \mathbb{Z}} h^{-k}(H_{x_k})$$

is a single point. Since $h(H_i) = V_i$, this is the same as

$$\cdots \cap h^2(V_{x_{-3}}) \cap h(V_{x_{-2}}) \cap V_{x_{-1}} \cap H_{x_0} \cap h^{-1}(H_{x_1}) \cap h^{-2}(H_{x_2}) \cap \ldots.$$

Let

$$I_k := H_{x_0} \cap h^{-1}\big(H_{x_1}\big) \cap \dots h^{-k}\big(H_{x_k}\big)$$

and

$$J_k := V_{x_{-1}} \cap h\big(V_{x_{-2}}\big) \cap \dots h^k\big(V_{x_{-(k+1)}}\big).$$

Then $I_{k+1} \subset I_k$ and $J_{k+1} \subset J_k$ for each k. By Lemma 47.12, since J_0 is a vertical strip, so is J_1. Then J_2 is also a vertical strip, since

$$J_2 = h\Big(h\big(V_{x_{-3}}\big) \cap V_{x_{-2}}\Big) \cap V_{x_{-1}}.$$

Moreover J_1 has width at most $1/5$ and thus J_2 has width at most $1/25$. Inductively, we see that J_k is a vertical strip with width at most 5^{-k}. Similarly I_k is a horizontal strip with height at most 5^{-k}. This means that the intersection

$$\bigcap_{k \in \mathbb{Z}} h^{-k}\big(H_{x_k}\big)$$

is the intersection of a vertical line and a horizontal line. This is a single point.

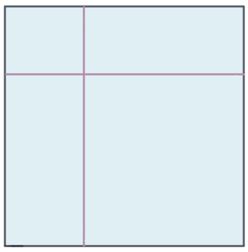

Figure 47.7: The intersection is a point.

Now let us show that F is continuous. Fix $n \geq 1$. If z and w are two points in Λ that are sufficiently close together then $h^k(z)$ and $h^k(w)$ will be within $1/10$ of each other for each $-n \leq k \leq n$. Thus for each $-n \leq k \leq n$, either $h^k(z)$ and $h^k(w)$ are both in H_0 or both in H_1. Thus $F(z)$ and $F(w)$ have the same entries x_k for $-n \leq k \leq n$. By definition of the topology on Σ_2 (cf. Proposition 4.20), this shows that F is continuous. Finally, since Λ is compact, F is a homeomorphism. This completes the proof. ∎

We conclude with our "picture proof" of Theorem 47.9.

Picture Proof of Theorem 47.9. We will illustrate the proof of Theorem 47.9 via a series of pictures. This argument is made rigorous by using the Inclination Lemma 46.1. The idea is to "embed" the horseshoe inside the homoclinic tangle. Consider again Figure 47.1. Here we have introduced a rectangular patch S inside a two-dimensional cross section. One side of S is bounded by a portion on the stable manifold, and the other three are transverse to the stable and unstable manifolds. We have divided S into three coloured strips. See Figure 47.8.

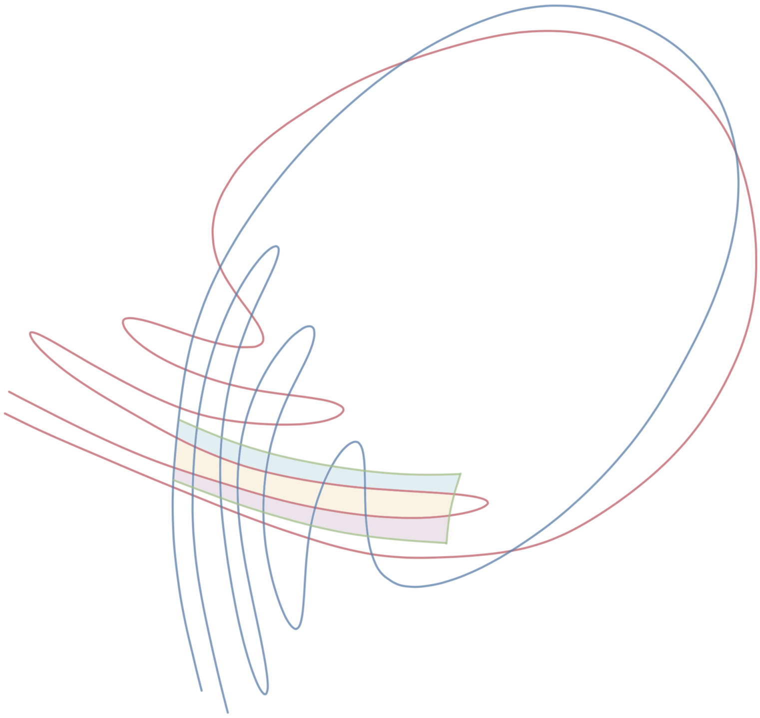

Figure 47.8: The rectangular patch S.

Since we are very close to a hyperbolic fixed point, by the Hartman-Grobman Theorem 32.2 the dynamics are conjugate to that of the linearised system Df. This means that the patch is squeezed in one direction and lengthened in the other. In Figure 47.9 we have illustrated the images $f(S)$, $f^2(S)$, and $f^3(S)$.

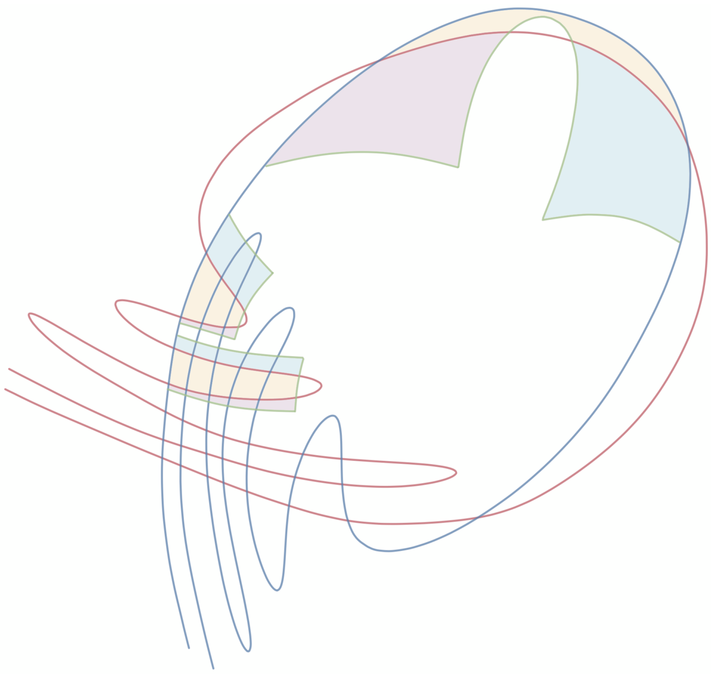

Figure 47.9: The images $f(S)$, $f^2(S)$, and $f^3(S)$.

By the Inclination Lemma, after sufficiently many iterations the patch is brought back onto itself (compare Lemma 47.6). Thus there exists p such that $f^p(S) \cap S \neq \emptyset$. In Figure 47.10 we have drawn the next three iterations $f^4(S)$, $f^5(S)$ and $f^6(S)$.

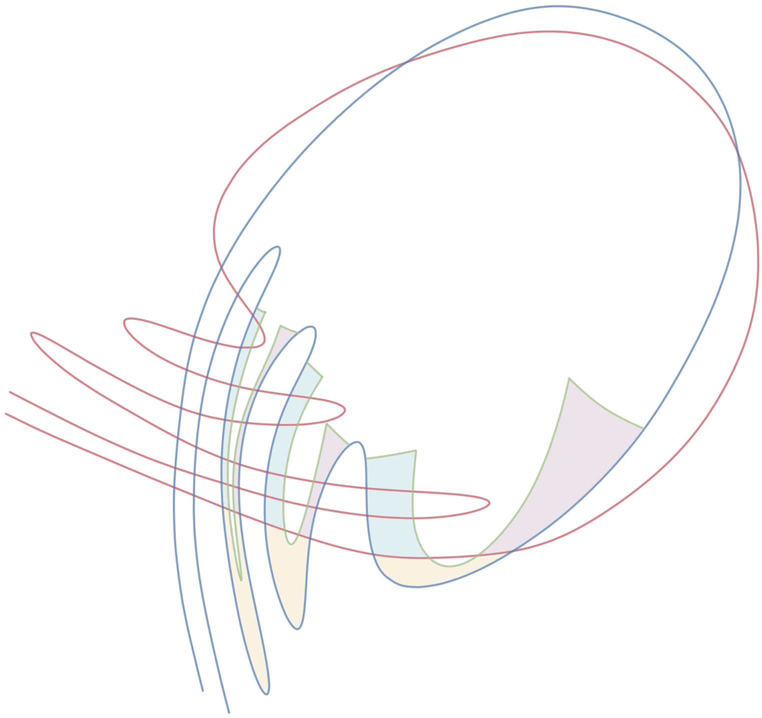

Figure 47.10: The images $f^4(S)$, $f^5(S)$, and $f^6(S)$.

Thus in our picture we may take $p = 6$. In the next two Figures 47.11 and 47.12 we focus our attention on the intersection $f^6(S) \cap S$—this is the square bounded by the dotted lines. We see that the image of the yellow strip under f^6 does not intersect S anymore, and the purple and blue strips are have been squeezed and rotated.

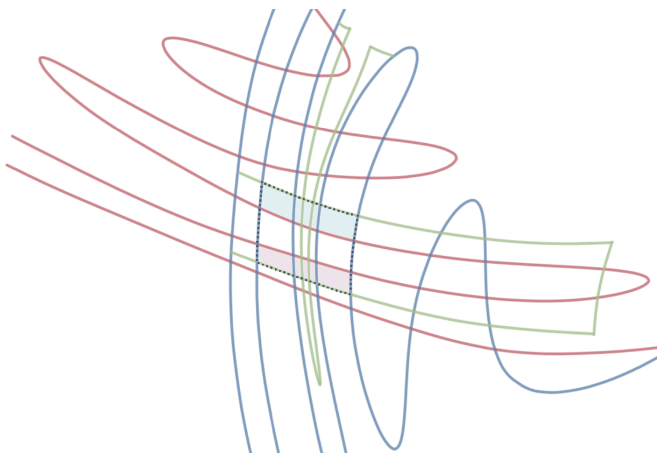

Figure 47.11: The purple and blue strips at the start.

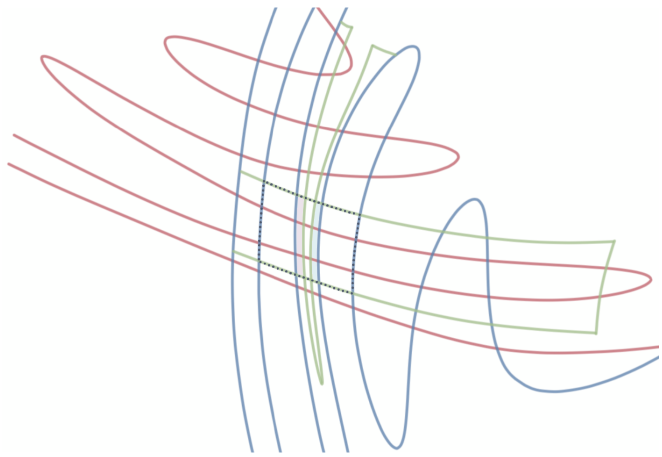

Figure 47.12: The purple and blue strips at the end.

Compare this to Figure 47.4, where the purple strip plays the role of H_0 and V_0, and the blue strip plays the role of V_0 and V_1. This shows that we have found a horseshoe inside the homoclinic tangle. The Theorem now follows from Proposition 47.14. ■

LECTURE 48

The Spectral Decomposition Theorem

Let f be a dynamical system on a compact manifold M. In this lecture we investigate the case where $\overline{\mathrm{per}(f)}$ is hyperbolic. Our main result is the Spectral Decomposition Theorem 48.5, of Smale, which states that if $\overline{\mathsf{per}(f)}$ is hyperbolic, then $\overline{\mathsf{per}(f)}$ can be split into finitely many compact completely invariant sets such that the restriction of f to each set is transitive. We begin with some definitions.

Definition 48.1. Let f be a dynamical system on a compact manifold M, and suppose $x \in M$ is a hyperbolic periodic point. We define the **index**[1] of x to be the integer

$$\imath(x) \coloneqq \dim\, E^u(x).$$

Lemma 48.2. *Let f be a dynamical system on a compact manifold M. There are at most finitely many hyperbolic periodic points with $\imath(x) = 0$ or $\imath(x) = \dim M$.*

Proof. This is an immediate consequence of Problem Q.4. ■

The following simple lemma is the key step behind the Spectral Decomposition Theorem 48.5.

Lemma 48.3. *Let f be a dynamical system on a compact manifold M, and suppose $x \neq y$ are two hyperbolic fixed points of f. Assume that*[2]

$$z \in W^s(x, f) \pitchfork W^u(y, f), \qquad w \in W^u(x, f) \pitchfork W^s(y, f).$$

Then both z and w both belong to $\mathsf{nw}(f)$.

Remark 48.4. If $W^s(x, f)$ intersects transversely with $W^u(y, f)$ then in particular we must have

$$\dim W^s(x, f) + \dim W^u(y, f) \geq \dim M,$$

and hence

$$\imath(y) \geq \imath(x).$$

This shows that the hypotheses of Lemma 48.2 imply that $0 < \imath(x) = \imath(y) < \dim M$.

We will give two proofs of Lemma 48.3, one using the Inclination Lemma 46.1 and one using the Anosov Closing Lemma 45.1. The Spectral Decomposition Theorem 48.5 will make use of both proofs.

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]The terminology comes from Morse theory.

[2]The $\pitchfork$ notation is shorthand for "transverse intersection".

Proof of Lemma 48.3 using the Inclination Lemma. Replacing f with f^p where p is the product of the periods of x and y, we may assume that both x and y are fixed points of f. This does not change the (un)stable manifolds, cf. Remark 46.2. Moreover if z and w are non-wandering with respect to f^p then they are also non-wandering with respect to f. Thus without loss of generality we may assume both x and y are fixed points of f.

Now let $D \subset W^u(y, f)$ be a small disc of dimension $\imath(x) = \imath(y)$ which contains z and is transverse to $W^s(x, f)$. See Figure 48.1. Applying the Inclination Lemma to D and x, we see that D will eventually accumulate on $W^u(x, f)$, and hence eventually cross $W^s(y, f)$. Thus there exists $w' \in W^s(y, f)$ arbitrarily close to w and a small disc B centred about w of dimension $\imath(x)$ which is transverse to $W^s(y, f)$. Then applying the Inclination Lemma to B and y, we see that B will eventually accumulate on $W^u(y, f)$, and hence cross $W^s(x, f)$ at a point z' arbitrarily close to z. Since any open neighbourhood of z intersects D, it follows that z is non-wandering. Reversing the roles of z and w shows that w is non-wandering. This completes the proof. ■

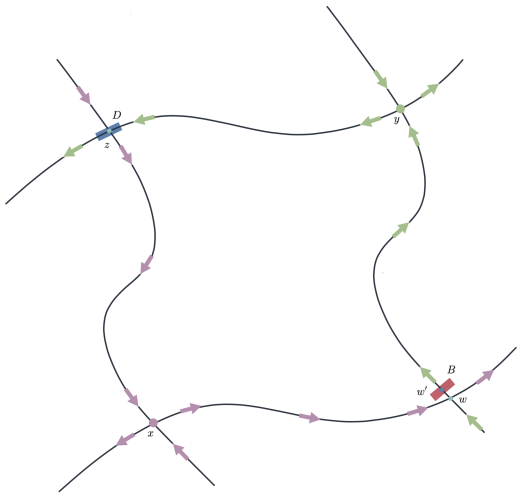

Figure 48.1: The discs B and D.

Proof of Lemma 48.3 using the Anosov Closing Lemma. As before we may without loss of generality assume that x and y are fixed points of f. Let

$$\Lambda \coloneqq \{x, y\} \cup \mathcal{O}_f^{\text{total}}(z) \cup \mathcal{O}_f^{\text{total}}(w).$$

Then Λ is hyperbolic by Proposition 47.4. Now fix $\varepsilon > 0$. By the Anosov Closing Lemma 45.1 there exists $\delta > 0$ such that every periodic δ-chain in Λ is ε-shadowed by a periodic point of f. Choose $k \geq 1$ such that

$$f^k(z), f^{-k}(w) \in B\big(x, \tfrac{\delta}{2}\big) \qquad \text{and} \qquad f^k(w), f^{-k}(z) \in B\big(y, \tfrac{\delta}{2}\big).$$

Then

$$\begin{aligned} z, f(z), \dots, f^{k-1}(z), &f^{-k}(w), f^{-(k-1)}(w), \dots, f^{-1}(w), w \\ &f(w), \dots, f^{(k-1)}(w), f^{-k}(z), f^{-(k-1)}(z), \dots, f^{-1}(z), z \end{aligned}$$

is a periodic δ-chain. See Figure 48.2. The Anosov Closing Lemma then tells us that there exists a periodic point x' of f and $i, j > 0$ such that

$$d\big(f^i(x'), z\big) \leq \varepsilon, \qquad d\big(f^j(x'), w\big) \leq \varepsilon.$$

Since ε was arbitrary, it follows that both z and w belong to $\overline{\mathsf{per}(f)}$. Since $\overline{\mathsf{per}(f)} \subseteq \mathsf{nw}(f)$ by Proposition 3.11, the proof is complete. ∎

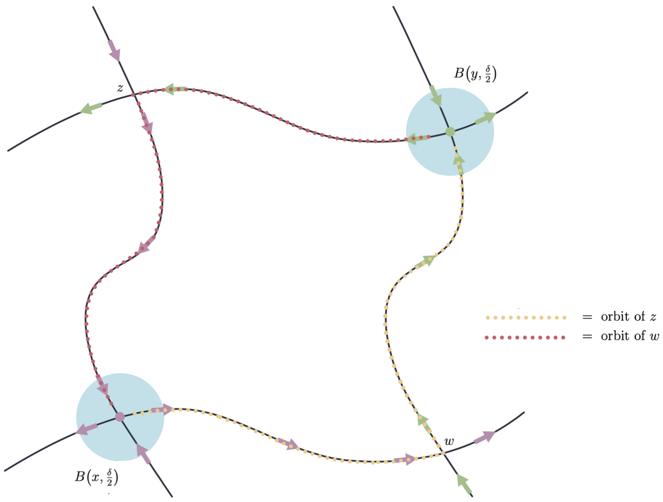

Figure 48.2: The periodic δ-chain.

We can now state Smale's famous Spectral Decomposition Theorem.

Theorem 48.5 (The Spectral Decomposition Theorem). *Let f be a dynamical system on a compact manifold M. Assume that $\overline{\mathsf{per}(f)}$ is hyperbolic. Then there exists a unique (up to relabelling) decomposition*

$$\overline{\mathsf{per}(f)} = P_1 \cup \cdots \cup P_q$$

into finitely many disjoint completely invariant closed sets such that $f|_{P_i}$ is transitive for each $i = 1, \dots, q$.

Despite the name, Theorem 48.5 does not refer to the spectra of anything[3] .

Proof. We prove the result in three steps.

1. In this first step we prove uniqueness. Suppose we had two such decompositions:

$$\overline{\mathsf{per}(f)} = P_1 \cup \cdots \cup P_q = Q_1 \cup \cdots \cup Q_p.$$

[3] In Smale's own words, the name "Spectral Decomposition Theorem" is used because *"the decomposition of the manifold into invariant sets of the diffeomorphism is quite analogous to the decomposition of a finite dimensional vector space into eigenspaces of a linear map. In one case we are considering automorphisms in the category of differential topology, in the other, finite dimensional vector spaces."*. Fair enough.

Since $f|_{P_i}$ is topologically transitive, there exists $x_i \in P_i$ such that $\overline{\mathcal{O}_f(x_i)} = P_i$ by Proposition 2.9. This implies that we cannot write any P_i as a finite disjoint union of closed invariant sets. But since

$$P_1 = \big(P_1 \cap Q_1\big) \cup \cdots \cup \big(P_1 \cap Q_p\big)$$

is such a decomposition, all but one of these sets must be empty. Thus each P_i is necessarily contained in some Q_j. Interchanging the roles of the P_i and the Q_j we also see that each Q_j is necessarily contained in some P_i. Thus $q = p$ and (up to relabelling the indices) we have $P_i = Q_i$ for each i.

2. In this step we construct the desired sets P_i. Lemma 48.2 tells us that there are at most finitely many orbits whose index is either zero or equal to the dimension of M. To each such orbit x we assign it the set $P_x := \mathcal{O}_f(x)$. Then $f|_{P_x}$ is clearly transitive. Set

$$\mathsf{per}_*(f) := \{x \in \mathsf{per}(f) \mid 0 < \imath(x) < \dim M\}\,.$$

It suffices to prove the theorem with $\mathsf{per}_*(f)$ in place of $\mathsf{per}(f)$. The idea of the proof is to use Lemma 48.3 to introduce an equivalence relation on $\mathsf{per}_*(f)$. Namely, we let us say that $x \sim y$ for two periodic orbits x and y if the hypotheses of Lemma 48.2 hold for x and y, that is:

$$x \sim y \qquad \Leftrightarrow \qquad \begin{cases} W^s(x,f) \pitchfork W^u(y,f) \neq \emptyset, & \text{and} \\ W^s(y,f) \pitchfork W^u(x,f) \neq \emptyset. \end{cases}$$

We claim this is an equivalence relation. It is obviously symmetric and reflexive. Transitivity follows from a similar argument to the first proof of Lemma 48.3. Suppose $x \sim y$ and $y \sim z$. Replace f with f^p, where p is the product of the periods of x, y and z. Now consider a disc $B \subset W^u(x,f)$ transverse to $W^s(y,f)$, and a disc $D \subset W^s(z,f)$ transverse to $W^u(z,f)$. As $k \to \infty$, $f^k(B)$ and $f^{-k}(D)$ accumulate on $W^u(y,f)$ and $W^s(y,f)$ respectively. See Figure 48.1. In particular $f^k(B) \pitchfork f^{-k}(D) \neq \emptyset$ for k large enough.

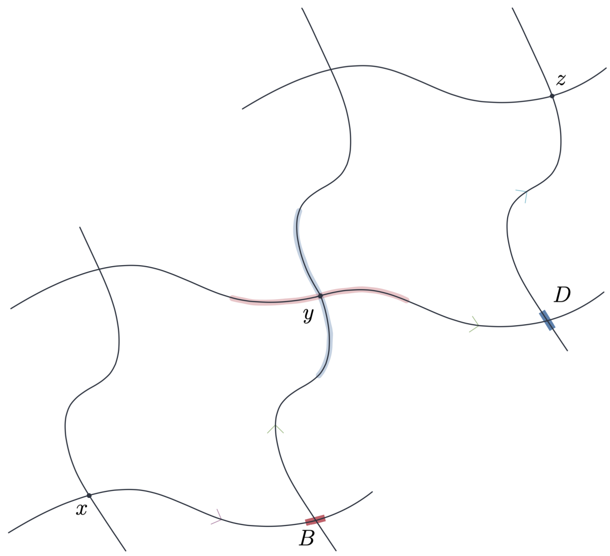

Figure 48.3: Proving $\sim$ is transitive.

Thus we can decompose $\mathsf{per}_*(f)$ into the equivalence classes of $\sim$. Moreover it follows from Lemma 47.2 that there exists $\delta > 0$ such that if $x, y \in \mathsf{per}_*(f)$ satisfy $d(x, y) \leq \delta$ then $x \sim y$. Thus there are only finitely many equivalence classes, say $C_1, \dots, C_m$. One has

$$\overline{C}_i \cap \overline{C}_j = \emptyset, \qquad i \neq j.$$

Each C_i may not be invariant, but since f clearly preserves equivalence classes, for any i there is a unique j such that $f(C_i) = C_j$ and also $f(\overline{C}_i) = \overline{C}_j$. The map $i \mapsto j$ is a bijection on $\{1, 2, \dots, m\}$ and is thus a product of cyclic permutations. This means we can decompose $\overline{\mathsf{per}_*(f)}$ into closed completely invariant sets

$$\overline{\mathsf{per}_*(f)} = P_1 \cup \cdots P_q,$$

where each P_i is a cyclic union of the some of the $\overline{C}_j$'s.

3. It remains to prove that the restriction of f to each P_i is topologically transitive. For definiteness we will prove that $f|_{P_1}$ is topologically transitive. The proof for the others is the same. After relabelling, we may assume that

$$P_1 = \overline{C}_1 \cup \cdots \cup \overline{C}_r,$$

and $f(\overline{C}_i) = \overline{C}_{i+1}$ for each $i = 1, \dots, r-1$ and $f(\overline{C}_r) = \overline{C}_1$. Suppose U and V are open sets of P_1 (note they may not be open in M). We may assume $U \subset \overline{C}_a$ and $V \subset \overline{C}_b$, where $1 \leq a \leq b \leq r$. Set

$$W := f^{b-a}(U),$$

so that $W \subset \overline{C}_b$. It suffices to show that there exists $l \geq 1$ such that $f^l(W) \cap V \neq \emptyset$. Suppose $x \in V \cap C_b$ and $y \in W \cap C_b$. Then $x \sim y$. Thus there are points

$$z \in W^s(x, f) \cap W^u(y, f), \qquad w \in W^s(y, f) \cap W^u(x, f).$$

The second proof of Lemma 48.3 shows that $z \in \overline{\mathsf{per}(f)}$. Thus z belongs to one of the P_i. Since the P_i are mutually disjoint and $d(f^k(z), x) \to 0$ as $k \to \infty$, we must have $z \in P_1$. Thus there are $k_1, k_2 \geq 1$ such that $f^{-k_1}(z) \in W$ and $f^{k_2}(z) \in V$. Thus if $l := k_1 + k_2$ then $f^l(W) \cap V \neq \emptyset$. This completes the proof. ■

DEFINITION 48.6. Let f be a dynamical system on a compact manifold M. Assume that $\overline{\mathsf{per}(f)}$ is hyperbolic. We call the sets P_i appearing in the statement of Theorem 48.5 the **basic sets** of f.

We conclude with the following result.

PROPOSITION 48.7. *If $\overline{\mathsf{per}(f)}$ is hyperbolic then it is isolated.*

Proof. Let $\overline{\mathsf{per}(f)} = P_1 \cup \cdots \cup P_q$ denote the spectral decomposition of f into basic sets. It suffices to show that each P_i is isolated. By Corollary 43.12 we can choose pairwise disjoint compact sets $K_1, \dots, K_q$ such that $P_i \subseteq K_i$ and such that $\Lambda_i := \mathrm{inv}(K_i, f)$ is hyperbolic. Then $P_i \subseteq \Lambda_i$, and it suffices to prove the reverse inequality. Take $x \in \Lambda_1$. We show that $x \in P_1$ in two steps.

1. In this step we show that $\omega_f(x) \subseteq P_1$. Since Λ_i is the maximal completely invariant set of f in K_i and $\omega_f(x)$ is itself completely invariant, we have $\omega_f(x) \subseteq \Lambda_1$, and thus $\omega_f(x)$ is a hyperbolic set. Fix $y \in \omega_f(x)$ and let $\varepsilon > 0$. By Proposition 45.2 there exists $0 < r \leq \varepsilon$ such that every periodic r-chain contained in $B(\omega_f(x), r)$ is $\frac{\varepsilon}{2}$-shadowed by a periodic point. Choose $l < m$ large such that both $f^l(x)$ and $f^m(x)$ belong to $B(y, r/2)$ and such that $f^k(x) \in B(\omega_f(x), r)$ for all $k = l, l+1, \dots, m$. Then

$$f^l(x), f^{l+1}(x), \dots, f^{m-1}(x), f^l(x)$$

is a periodic r-chain contained in $B(\omega_f(x), r)$ which passes through $B(y, r/2)$. Proposition 45.2 implies there exists a periodic point z of f which $\frac{\varepsilon}{2}$-shadows this chain. Replacing z with an iterate if necessary, we have

$$\begin{aligned} d(y,z) &\leq d(y, f^l(x)) + d(f^l(x), x) \\ &\leq \frac{r}{2} + \frac{\varepsilon}{2} \\ &\leq \varepsilon. \end{aligned}$$

Since ε was arbitrary we have $y \in \overline{\mathsf{per}(f)}$, and then as y was arbitrary, we have $\omega_f(x) \subseteq \overline{\mathsf{per}(f)}$. Since $\omega_f(x) \subseteq K_1$, we must have $\omega_f(x) \subseteq P_1$.

2. We now prove that $x \in P_1$ with a second application of the Anosov Closing Lemma. Fix again an arbitrary $\varepsilon > 0$, and let $\delta > 0$ be such that any periodic δ-chain is ε-shadowed by a periodic point. Since P_1 is transitive, by Proposition 2.9 there exists $w \in P_1$ with dense (forward) orbit. As $\omega_f(x)$ and $\alpha_f(x)$ are both contained in P_1 by Step 1, there exists $a, b, c, d \in \mathbb{N}$ with $b \leq d$ such that

$$d\big(f^a(x), f^b(w)\big) \leq \delta, \qquad d\big(f^{-c}(x), f^d(w)\big) \leq \delta.$$

Then

$$x, f(x), \dots, f^a(x), f^{b-1}(w), f^b(w), \dots f^{d-1}(w), f^{-c}(x), \dots, f^{-1}(x), x$$

is a periodic δ chain starting at x. See Figure 48.4.

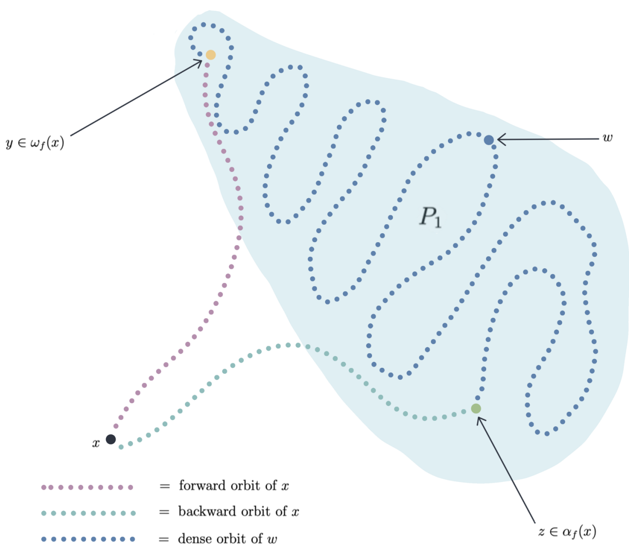

Figure 48.4: Travelling from x to $\omega_f(x)$, then to $\alpha_f(x)$, and then back to x again.

Thus by the Anosov Closing Lemma 45.1 we find a periodic point of f within ε of x. This shows that $x \in \overline{\mathsf{per}(f)}$. Thus $x \in P_1$, and the proof is complete. ■

LECTURE 49

No Cycles

The Axiom A condition (Definition 45.5) has two ingredients: that the non-wandering set should be hyperbolic, and that every non-wandering point should be the limit of a sequence of periodic points. However for a general dynamical system f, the difference between the set $\overline{\mathsf{per}(f)}$ and the non-wandering set $\mathsf{nw}(f)$ can be rather large. For example, the set $\overline{\mathsf{per}(f)}$ does not necessarily capture the limiting behaviour of all orbits, whereas the non-wandering set does.

In this lecture we examine this difference more closely. We show that if the set of all limit points admits a hyperbolic structure, then every limit point is reachable through periodic points. This in turn leads to a pleasing criterion for the chain recurrent set to be hyperbolic. Much of this lecture is valid in the topological category, and where possible we give the relevant definitions and statements for reversible topological dynamical systems.

DEFINITION 49.1. Let $f\colon X \to X$ denote a reversible dynamical system on a compact metric space. We define the **limit set** of f, written $\mathsf{lim}(f)$, to be:

$$\mathsf{lim}(f) := \overline{\bigcup_{x \in X} \omega_f(x) \cup \alpha_f(x)}.$$

Thus $\mathsf{lim}(f)$ is a compact, non-empty and completely invariant set for f.

It is clear that $\overline{\mathsf{per}(f)} \subseteq \mathsf{lim}(f)$. Meanwhile part (ii) of Proposition 3.11 shows that $\mathsf{nw}(f) \subseteq \mathsf{lim}(f)$. Thus we have

$$\overline{\mathsf{per}(f)} \subseteq \mathsf{lim}(f) \subseteq \mathsf{nw}(f) \subseteq \mathsf{cha}(f).$$

In general all of these inclusions can be strict. Nevertheless, under additional assumptions they agree. For instance, we already proved in Proposition 45.4 that if $\mathsf{cha}(f)$ is hyperbolic then $\mathsf{cha}(f) = \overline{\mathsf{per}(f)}$ (and hence all four sets agree). In this lecture we will investigate other implications. We begin with the following result, which is valid in the topological category.

PROPOSITION 49.2. *Suppose $f\colon X \to X$ is a reversible dynamical system on a compact metric space. Assume we can write*[1]

$$\mathsf{lim}(f) \subseteq L_1 \cup \cdots \cup L_q,$$

where the L_i are compact completely invariant sets that are pairwise disjoint. Then

$$X = \bigcup_{i=1}^{q} W^s(L_i, f) = \bigcup_{i=1}^{q} W^u(L_i, f).$$

[1]The "$\subseteq$" is not a typo: this will be useful in the proof of Proposition 50.6 next lecture.

Proof. Let K_i denote a compact set with $L_i \subset K_i^\circ$ such that if $i \neq j$ then

$$K_i \cap K_j = \emptyset \qquad f(K_i) \cap K_j = \emptyset. \tag{49.1}$$

Such neighbourhoods exist since the L_i are disjoint and invariant. Now let $x \in X$. Since $\omega_f(x) \subseteq \lim(f)$, there is a $n \geq 1$ such that for all $k \geq n$, one has $f^k(x) \in \bigcup_{i=1}^n K_i$. Without loss of generality, assume that $f^n(x) \in K_1$. Since $f^{n+1}(x)$ cannot belong to any of the other K_i by the second equality in (49.1), we must have $f^{n+1}(x) \in K_1$, and inductively $f^k(x) \in K_1$ for all $k \geq n$. Thus $\omega_f(x) \subseteq K_1$, and hence $\omega_f(x) \subseteq L_1$. Thus $x \in W^s(L_1, f)$ by Lemma 45.8. The argument for W^u is similar. This completes the proof. ■

Going back to the smooth case, we have:

Proposition 49.3. *Let f be a smooth dynamical system on M. If $\lim(f)$ is hyperbolic then $\lim(f) = \overline{\mathrm{per}(f)}$.*

Proof. This argument is word-for-word identical to the proof of Step 1 of Proposition 48.7 (replace Λ_1 with $\lim(f)$). ■

The In-Phase Theorem 45.10 yields the following corollary, which will be useful next lecture.

Corollary 49.4. *Let f be a smooth dynamical system on M. Assume that $\lim(f)$ is hyperbolic. Then*

$$M = \bigcup_{x \in \lim(f)} W^s(x, f) = \bigcup_{x \in \lim(f)} W^u(x, f).$$

Proof. Since $\lim(f)$ is hyperbolic, $\lim(f) = \overline{\mathrm{per}(f)}$ by Proposition 49.3, and hence $\lim(f)$ is isolated by Proposition 48.7. Thus by the In Phase Theorem 45.10 one has

$$W^s(\lim(f), f) = \bigcup_{x \in \lim(f)} W^s(x, f)$$

and

$$W^u(\lim(f), f) = \bigcup_{x \in \lim(f)} W^u(x, f)$$

By Proposition 49.2 (applied with $q = 1$ and $L_1 = \lim(f)$) one has

$$M = W^s(\lim(f), f) = W^u(\lim(f), f).$$

This completes the proof. ■

Now let us introduce the notion of a cycle.

Definition 49.5. Suppose $f \colon X \to X$ is a reversible dynamical system on a compact metric space. Let $Y \subseteq X$ be a compact completely invariant set, and suppose we are given a decomposition

$$Y = Y_1 \cup \cdots \cup Y_q$$

into disjoint compact completely invariant sets. We introduce a binary partial relation on $\{Y_i\}$ by declaring that

$$Y_i \prec Y_j \qquad \Leftrightarrow \qquad \Big(W^s(Y_i, f) \cap W^u(Y_j, f)\Big) \setminus Y \neq \emptyset.$$

Thus $Y_i \prec Y_j$ if there is an $x \in X \setminus Y$ that travels from Y_j to Y_i. See Figure 49.1. If $i \neq j$ then $W^s(Y_i, f) \cap W^u(Y_j, f) \neq \emptyset$ implies that $Y_i \prec Y_j$, but this is not necessarily the case if $i = j$. The binary relation $\prec$ is *not* reflexive, nor symmetric, nor transitive! The following construction should remind you of our treatment of the Sharkovsky Theorem 14.2.

Definition 49.6. We say that $Y_{i_1}, Y_{i_2} \dots, Y_{i_k}$ form a **cycle of length** k if

$$Y_{i_1} \prec Y_{i_2} \prec \cdots \prec Y_{i_k} \prec Y_{i_1}.$$

The case $k = 1$ is not excluded, and we call a cycle of length 1 a **trivial cycle**. We say that the decomposition $\{Y_i\}$ has **no cycles** if there do not exist any cycles (trivial or otherwise).

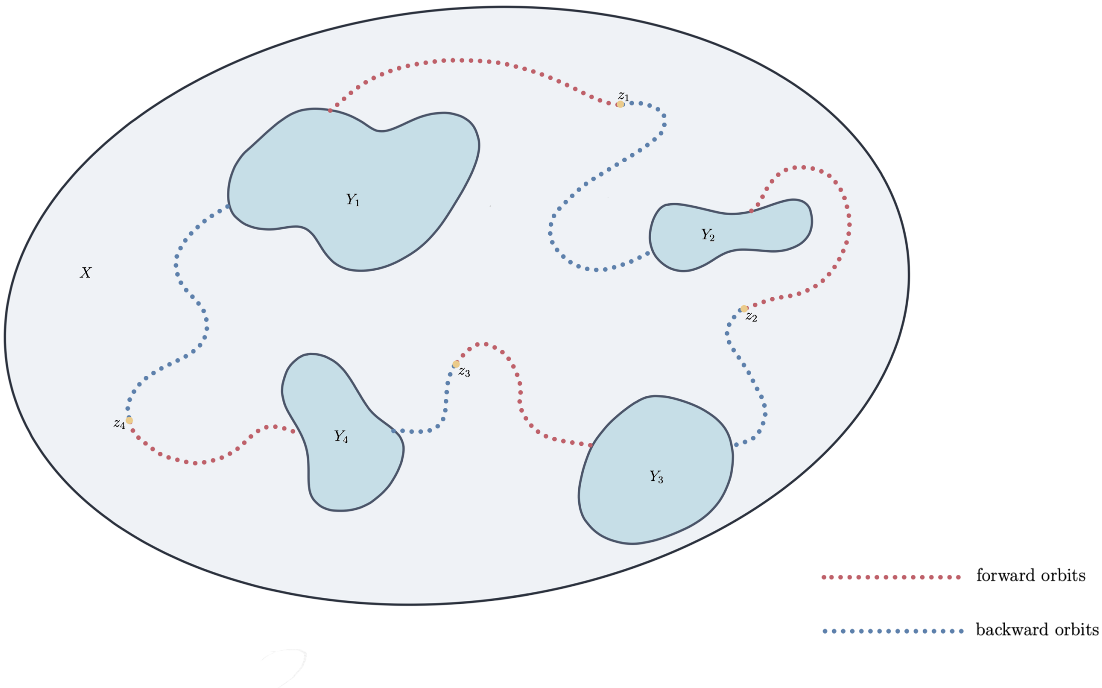

Figure 49.1: A cycle $Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_1$.

The following result is somewhat technical, but it will be key to all that follows. It tells us, roughly speaking, that the only way to produce "more" chain recurrent points than limit points is to exploit cycles in the limit set. This result is valid in the topological category.

Proposition 49.7. *Suppose $f\colon X \to X$ is a reversible dynamical system on a compact metric space. Assume we can write*

$$\mathsf{lim}(f) = L_1 \cup \cdots \cup L_q,$$

where the L_i are compact completely invariant sets that are pairwise disjoint. If the $\{L_i\}$ have no cycles then $\mathsf{lim}(f) = \mathsf{cha}(f)$.

Proof. Set
$$Z := \mathsf{cha}(f) \setminus \mathsf{lim}(f).$$
We prove that Z is empty in two steps.

1. Fix $1 \leq i \leq q$ and abbreviate $L := L_i$. In this step we show that
$$Z \cap W^s(L, f) \neq \emptyset \qquad \Rightarrow \qquad Z \cap W^u(L, f) \neq \emptyset. \tag{49.2}$$
Suppose $z \in Z \cap W^s(L, f)$. Take a compact set K such that
$$z \notin K, \qquad L \subset K^\circ, \tag{49.3}$$
and
$$K \cap L_j = \emptyset, \qquad f(K) \cap L_j = \emptyset, \qquad j \neq i. \tag{49.4}$$
Since $z \in \mathsf{cha}(f)$, for every k there is a $\frac{1}{k}$-chain from z to itself, say
$$z = z_0^k, z_1^k, \ldots, z_{p_k-1}^k, z_{p_k}^k = z.$$
Let $0 \leq l_k \leq p_k$ be such that $z_{l_k}^k$ is the closest point to L. Since $z \in W^s(L, f)$ one has $d\big(z_{l_k}^k, L\big) \to 0$ as $k \to \infty$. Thus for large k, there is a $1 \leq n_k \leq p_k - l_k$ such that
$$z_{l_k}^k, z_{l_k+1}^k, \ldots, z_{l_k+n_k-1}^k \in K^\circ, \tag{49.5}$$
$$w_k := z_{l_k+n_k}^k \notin K^\circ. \tag{49.6}$$
See Figure 49.2.

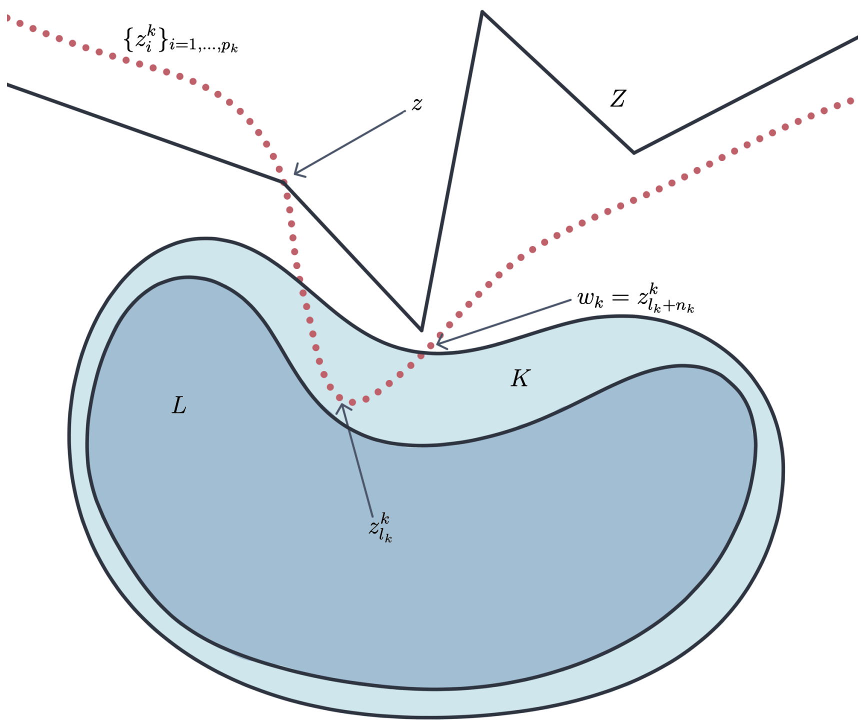

Figure 49.2: The points $z_{l_k}^k$ and $w_k := z_{l_k+n_k}^k$.

Then $w_k \in B(f(K), 1/k)$. Since $d\big(z^k_{l_k}, L_i\big) \to 0$ we must have $n_k \to \infty$. After passing to a further subsequence, we may assume that there exists a point $w \in X$ such that $w_k \to w$. Then $w \in f(K) \setminus L$ by (49.6) and (49.3), and hence also $w \notin \mathsf{lim}(f)$ by (49.4). However $w \in \mathsf{cha}(f)$, since given any $\varepsilon > 0$, for k large a periodic ε-chain is

$$w, z^k_{l_k+n_k+1}, \ldots, z^k_{p_k} = z = z^k_0, z^k_1, \ldots, z^k_{l_k+n_k-1}, w.$$

We claim that $w \in W^u(L, f)$. Indeed, since $n_k \to \infty$ if follows from (49.5) that $f^{-k}w) \in K$ for all $k \geq 1$, and hence also $\alpha_f(w) \subseteq K$. Thus $\alpha_f(w) \in L$, and hence $w \in W^u(L, f)$ by Lemma 45.8. This completes the proof of (49.2).

2. Now we prove the Proposition. If $\mathsf{lim}(f) \neq \mathsf{cha}(f)$, then choose $z_1 \in \mathsf{cha}(f) \setminus \mathsf{lim}(f)$. By Proposition 49.2, there exists i_1 such that $z_1 \in W^s(L_{i_1}, f)$. By equation (49.2) there exists $z_2 \in \mathsf{cha}(f) \setminus \mathsf{lim}(f)$ which belongs to $W^u(L_{i_1}, f)$. Then by Proposition 49.2 again, $z_2 \in W^s(L_{i_2}, f)$ for some i_2. Note that this implies $L_{i_2} \prec L_{i_1}$. If $i_2 = i_1$ we are done. If not, we keep going: By (49.2) there exists $z_3 \in \mathsf{cha}(f) \setminus \mathsf{lim}(f)$ such that $z_3 \in W^u(L_{i_2}, f)$, and then by Proposition 49.2 one has $z_3 \in W^s(L_{i_3}, f)$ for some i_3. Thus $L_{i_3} \prec L_{i_2}$. By induction, we find an infinite sequence i_k such that $L_{i_{k+1}} \prec L_{i_k}$. Since there are only finitely many L_i, we must have $i_k = i_1$ for some k. This means we have a cycle:

$$L_{i_1} \prec L_{i_{k-1}} \prec \cdots L_{i_2} \prec L_{i_1}.$$

The contradiction completes the proof. ∎

Definition 49.8. Let f be a dynamical system on a compact manifold M and assume $\overline{\mathsf{per}(f)}$ is hyperbolic. We say that f has **no basic cycles** if the basic sets $\{P_i\}$ from the Spectral Decomposition Theorem 48.5 have no cycles.

We use Proposition 49.7 to prove the following pleasing result.

Theorem 49.9. *Let f be a dynamical system on a compact manifold M. The following are equivalent:*

(i) *f satisfies Axiom A and has no basic cycles.*

(ii) *$\mathsf{lim}(f)$ is hyperbolic and f has no basic cycles.*

(iii) *$\mathsf{cha}(f)$ is hyperbolic.*

The most striking consequence of Theorem (i) will come next lecture, when we show that (i) is equivalent to f being *omega stable*.

Proof. The implication (i) $\Rightarrow$ (ii) follows from the fact that $\mathsf{lim}(f)$ is sandwiched between $\overline{\mathsf{per}(f)}$ and $\mathsf{nw}(f)$. The implication (ii) $\Rightarrow$ (iii) follows from Proposition 49.3 and Proposition 49.7.

Assume now that (iii) holds. Then by Proposition 45.4 we have that $\overline{\mathsf{per}(f)}$ is hyperbolic and equal to $\mathsf{cha}(f)$. Let $\overline{\mathsf{per}(f)} = P_1 \cup \cdots \cup P_q$ be the decomposition of $\mathsf{per}(\mathrm{f})$ from the Spectral Decomposition Theorem 48.5. Assume for contradiction that f has a basic cycle

$$P_{i_1} \prec \cdots \prec P_{i_k} \prec P_{i_1}.$$

Thus there exist points

$$z_1 \in \Big(W^s(P_{i_1}, f) \cap W^u(P_{i_2}, f) \Big) \setminus \mathsf{cha}(f),$$
$$z_2 \in \Big(W^s(P_{i_2}, f) \cap W^u(P_{i_3}, f) \Big) \setminus \mathsf{cha}(f),$$

and so on up to

$$z_k \in \Big(W^s(P_{i_k}, f) \cap W^u(P_{i_1}, f) \Big) \setminus \mathsf{cha}(f).$$

Then by definition $z_i \notin \mathsf{cha}(f)$. But since the restriction of f to each P_i is transitive, by gluing together dense orbits inside the P_{i_j}'s, for any $\delta > 0$ we can embed the z_i inside a periodic δ-chain: start at z_1 and travel via a positive orbit to P_{i_1}, then travel along a dense orbit of f to a point where we can leave P_{i_1} and head towards z_k via a negative orbit of f, and so on... See Figure 49.3. Thus actually each z_i belongs to $\mathsf{cha}(f)$, which is a contradiction. This completes the proof. ■

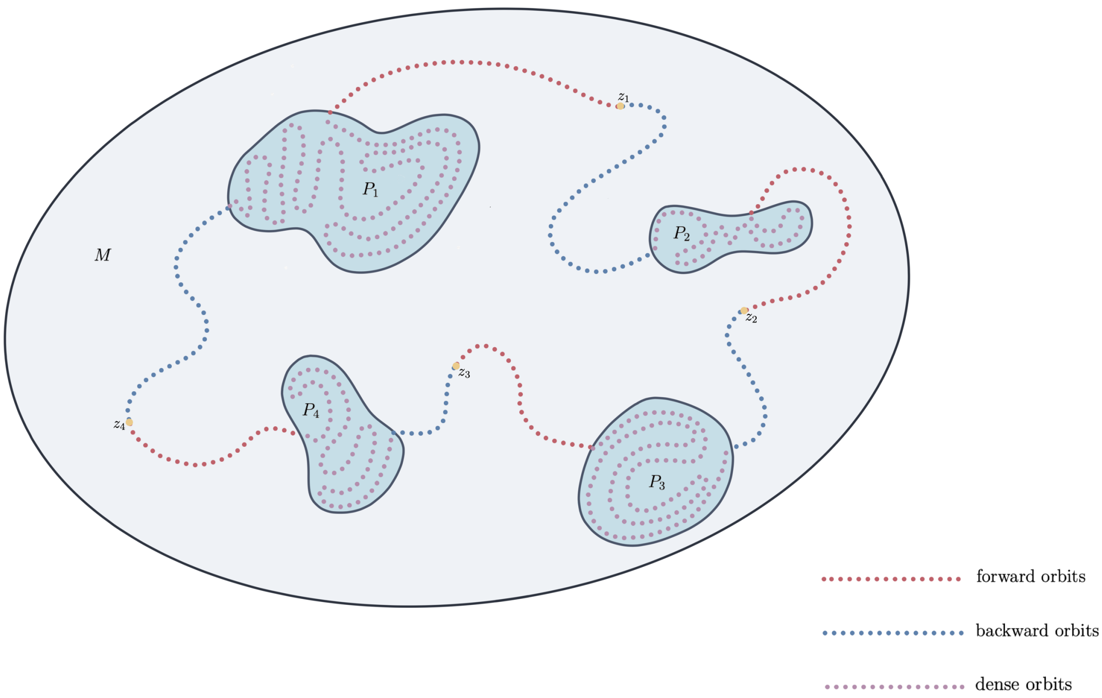

Figure 49.3: Building a chain from a cycle and dense orbits.

We conclude with the following interesting observation: an Axiom A dynamical system can never have a trivial cycle. First, a lemma.

LEMMA 49.10. *Let f be a dynamical system on a compact manifold M, and let x be a hyperbolic periodic point. Suppose U, V are open subsets of M such that*

$$U \cap W^s(x, f) \neq \emptyset, \qquad V \cap W^u(f^p(x), f) \neq \emptyset,$$

for some iterate $f^p(x)$ of x. Then there exists $k \geq 1$ such that $f^k(U) \cap V \neq \emptyset$.

Proof. Immediate[2] from the Inclincation Lemma 46.1. ■

Proposition 49.11. *A dynamical system satisfying Axiom A has no trivial basic cycles.*

Proof. Let $P \subseteq \overline{\mathsf{per}(f)}$ be one of the basic sets, and suppose $z \in W^s(P,f) \cap W^u(P,f)$. We will prove that $z \in \mathsf{nw}(f)$, so that $z \in P$ by the Axiom A hypothesis. Let U be a neighbourhood of z. We must find $k \geq 0$ such that $f^k(U) \cap U \neq \emptyset$. By the In Phase Theorem 45.10 (which is applicable by Proposition 48.7) there exists $x_0, y_0 \in P$ such that $z \in W^s(x_0, f) \cap W^u(y_0, f)$. Choose open sets $V, W \subset P$ containing x_0 and y_0 respectively such that

$$W^s(x,f) \cap U \neq \emptyset, \qquad \forall\, x \in V,$$

and

$$W^u(y,f) \cap U \neq \emptyset, \qquad \forall\, y \in W.$$

Since $f|_P$ is transitive, there exists a point $w \in P$ with dense forward orbit. Thus there exists $l_1, l_2 \geq 0$ such that $f^{l_1}(w) \in V$ and $f^{l_2}(w) \in W$. Since $w \in P$, there exists a sequence $w_k \in \mathsf{per}(f)$ of periodic points such that $w_k \to w$. Thus for k sufficiently large one has $f^{l_1}(w_k) \in U$ and $f^{l_2}(w_k) \in V$. The claim now follows from Lemma 49.10. ■

[2]Actually using the Inclination Lemma is overkill here: with a bit of work this follows directly from the Stable Manifold Theorem.

The Omega Stability Theorem

In this final lecture we return to the notion of structural stability from Lecture 41. Smale proved in 1966 that in dimensions and higher, structural stability is *not* a generic condition. That is, there exist dynamical systems that cannot be approximated by structurally stable ones. This led him to search for weaker conditions than structural stability that are still strong enough to capture the "interesting" dynamics. Suppose $f, g \colon M \to M$ are two dynamical systems on a compact manifold M. Instead of asking that f and g are globally conjugate, we can simply ask that their restrictions to the various invariant subsets we have introduced have to be conjugate.

DEFINITION 50.1. We say that $f, g \in \mathrm{Diff}^1(M)$ are:

(i) $\overline{\mathsf{per}}$ **equivalent** if $f|_{\overline{\mathsf{per}(f)}}$ is conjugate to $g|_{\overline{\mathsf{per}(g)}}$, that is, if there exists a homeomorphism $H \colon \overline{\mathsf{per}(f)} \to \overline{\mathsf{per}(g)}$ such that $H \circ f|_{\overline{\mathsf{per}(f)}} = g|_{\overline{\mathsf{per}(g)}} \circ H$.

(ii) lim **equivalent** if $f|_{\mathsf{lim}(f)}$ is conjugate to $g|_{\mathsf{lim}(g)}$, that is, if there exists a homeomorphism $H \colon \mathsf{lim}(f) \to \mathsf{lim}(g)$ such that $H \circ f|_{\mathsf{lim}(f)} = g|_{\mathsf{lim}(g)} \circ H$.

(iii) **omega equivalent**[1] if $f|_{\mathsf{nw}(f)}$ is conjugate to $g|_{\mathsf{nw}(g)}$, that is, if there exists a homeomorphism $H \colon \mathsf{nw}(f) \to \mathsf{nw}(g)$ such that $H \circ f|_{\mathsf{nw}(f)} = g|_{\mathsf{nw}(g)} \circ H$.

(iv) cha **equivalent** if $f|_{\mathsf{cha}(f)}$ is conjugate to $g|_{\mathsf{cha}(g)}$, that is, if there exists a homeomorphism $H \colon \mathsf{cha}(f) \to \mathsf{cha}(g)$ such that $H \circ f|_{\mathsf{cha}(f)} = g|_{\mathsf{cha}(g)} \circ H$.

This gives rise to four new notions of stability:

DEFINITION 50.2. Let f be a dynamical system on a compact manifold. We say that f is:

(i) $\overline{\mathsf{per}}$ **stable** if any sufficiently nearby system is $\overline{\mathsf{per}}$ equivalent to f.

(ii) lim **stable** if any sufficiently nearby system is lim equivalent to f.

(iii) **omega stable** if any sufficiently nearby system is omega equivalent to f.

(iv) cha **stable** if any sufficiently nearby system is cha equivalent to f.

Since

$$\overline{\mathsf{per}(f)} \subseteq \mathsf{lim}(f) \subseteq \mathsf{nw}(f) \subseteq \mathsf{cha}(f) \subseteq M$$

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1] Why the name "omega equivalent"? This is because the classical notation for the non-wandering set $\mathsf{nw}(f)$ is $\Omega(f)$—we elected not to use this notation due to the potential conflict with the omega limit sets $\omega_f(x)$.

one has:

$$\begin{aligned} \text{structural stability} &\Rightarrow \mathsf{cha} \text{ stability} \\ &\Rightarrow \text{omega stability} \\ &\Rightarrow \mathsf{lim} \text{ stability} \\ &\Rightarrow \overline{\mathsf{per}} \text{ stability} \end{aligned}$$

In this lecture we will consider only omega stability. This is mainly for historical reasons, but also because omega stability leads to the neatest statement[2]. Nevertheless, much of this lecture could be recast in terms of cha stability, with minor tweaks to the relevant assumptions.

It was originally hoped that omega stability would turn out to be a generic property. Unfortunately this too turned out to be false[3]. Nevertheless, the omega stability story has a particularly satisfying conclusion: as we will see there is a simple "if and only if" criteria for a dynamical system to be omega stable, namely:

THEOREM 50.3 (The Omega Stability Theorem). *A dynamical system f on a compact manifold is omega stable if and only if f satisfies Axiom A and has no basic cycles.*

We will only prove the "easier" of the two directions: that if f satisfies Axiom A and has no basic cycles then f is omega stable. This direction was proved by Smale in 1967, and is approachable using the machinery developed thus far. The converse direction was proved by Palis in 1988, building on work of Mañé (1988), and goes beyond the scope of this course.

REMARK 50.4. What we are really talking about here is C^1 **omega stability**. One can ask the analogous question in the C^p topology for any $p \geq 1$. However this is an open problem. It is conjectured that if $p > 1$, then a C^p dynamical system which is C^p omega stable necessarily satisfies Axiom A. In this sense the original problem of structural stability has an even more satisfactory solution than Theorem 50.3. We will return to this at the end of the lecture.

Using Theorem 49.9 we obtain the following slightly more compact form of the Omega Stability Theorem.

COROLLARY 50.5. *A dynamical system f on a compact manifold is omega stable if and only if $\mathsf{cha}(f)$ is hyperbolic.*

We have already done 99% of the work needed to prove the $\Leftarrow$ direction of Theorem 50.3. One last piece of the puzzle remains, which concerns cycles in the non-wandering set. As with Proposition 49.7, this statement is valid in the topological category.

PROPOSITION 50.6. *Let $f \colon X \to X$ denote a reversible dynamical system. Suppose*

$$\mathsf{nw}(f) = N_1 \cup \cdots \cup N_q$$

is a decomposition of $\mathsf{nw}(f)$ into pairwise disjoint compact completely invariant sets. If the $\{N_i\}$ have no cycles then for any neighbourhood V of $\mathsf{nw}(f)$ in X there is a neighbourhood $\mathcal{V}$ of f in $\mathrm{Hom}(X)$ such that if $g \in \mathcal{V}$ then $\mathsf{nw}(g) \subset V$.

[2]It also has the most catchy name.
[3]Proved by Abraham and Smale in 1970.

The proof of Proposition 50.6 is very similar to that of Proposition 49.7.

Proof. Let $(g_k) \in \mathrm{Hom}(X)$ be any sequence converging to f, and let Z denote the set of wandering points of f that can be approximated by non-wandering points of the g_k, i.e.

$$Z := \{z \in X \setminus \mathsf{nw}(f) \mid \exists\, z_k \in \mathsf{nw}(g_k) \text{ with } z_k \to z\}.$$

It suffices to prove that $Z = \emptyset$, which we do in two steps.

1. Fix $1 \le i \le q$ and abbreviate $N = N_i$. In this first step, we show that

$$Z \cap W^s(N, f) \neq \emptyset \qquad \Rightarrow \qquad Z \cap W^u(N, f) \neq \emptyset. \tag{50.1}$$

The proof of (50.1) is analogous to (49.2) last lecture. Let $z \in Z \cap W^s(N, f)$, and take a compact set K such that

$$z \notin K, \qquad N \subset K^\circ \tag{50.2}$$

and

$$K \cap N_j = \emptyset, \qquad f(K) \cap N_j = \emptyset, \qquad \forall j \neq i. \tag{50.3}$$

By definition of Z, there exists a sequence $z_k \to z$ and a sequence $p_k \to \infty$ such that $g_k^{p_k}(z_k) \to z_k$. Then also $g_k^{p_k}(z_k) \to z$. Let $0 \le l_k \le p_k$ be such that out of all points of the form $g_k^i(x_k)$ with $0 \le i \le p_k$, $w_k := g_k^{l_k}(x_k)$ is the point closest to N, i.e.

$$\min_{0 \le i \le p_k} d\big(g_k^i(x_k), N\big) = d\big(g_k^{l_k}(x_k), N\big).$$

Since $z \in W^s(N, f)$ and $g_k \to f$, up to passing to a subsequence we may assume that $d(w_k, N) \to 0$ as $k \to \infty$. Thus in particular one has $0 < l_k < p_k$.

Since $g_k^{p_k}(z_k) \to z \notin K$, for large k we can choose $1 \le n_k \le p_k - l_k$ such that

$$w_k, g_k(w_k), \dots, g_k^{n_k - 1}(w_k) \in K^\circ, \tag{50.4}$$

but

$$g_k^{n_k}(w_k) \notin K^\circ. \tag{50.5}$$

Since $w_k \to N$ we must have $n_k \to \infty$. Now set

$$m_k := l_k + n_k, \qquad \zeta_k := g_k^{m_k}(z_k) = g_k^{n_k}(w_k),$$

so that $\zeta_k \in g_k(K) \setminus K^\circ$. Note that $\zeta_k \in \mathsf{nw}(g_k)$, since $z_k \in \mathsf{nw}(g_k)$ by assumption, and the iterate of a non-wandering point is also non-wandering.

Up to passing to another sequence, we may assume that ζ_k converges to a point ζ. Then $\zeta \in f(K) \setminus K^\circ$, and hence by (50.2) and (50.3) we must have $\zeta \notin \mathsf{nw}(f)$. Thus $\zeta \in Z$. Moreover by (50.4) we have $f^{-k}(\zeta) \in K$ for all $k \ge 1$, and thus $\alpha_f(\zeta) \subseteq K$. Since $K \cap \mathsf{lim}(f) \subset N$, we must have $\alpha_f(z) \subseteq N$, and hence $z \in W^u(N, f)$ by Lemma 45.8. Thus $\zeta \in Z \cap W^u(N, f)$, and this completes the proof of (50.1).

2. We now prove that $Z = \emptyset$. This step is entirely analogous to that of Step 2 of Proposition 49.7. Assume for contradiction that $Z \neq \emptyset$, and fix $z \in Z$. Then since $\mathsf{lim}(f) \subseteq \mathsf{nw}(f)$, by Proposition 49.2 we have[4] $z \in W^s(N_i, f)$ for some i. Then

[4]Here it is important that in Proposition 49.2 we required only an inclusion "$\subseteq$" in the hypothesis rather than an equality.

by Step 1 there exists a point $\zeta \in W^u(N_i, f) \setminus \mathsf{nw}(f)$. By Proposition 49.2 again, we have $\zeta \in W^s(N_j, f)$ for some j. If $i = j$ we are done, since then $N_i \prec N_i$, contradicting the assumption there are no cycles. If $i \neq j$ then we keep going. As in the proof of Step 2 of Proposition 49.7, after at most q iterations we reach a contradiction. This completes the proof. ■

We can now prove half of the Omega Stability Theorem 50.3. The proof is straightforward, given what we have already accomplished, however it is worth pointing out that the argument uses practically all of the main theorems of the course.

Proof of the "easy" half of Theorem 50.3. Assume that f satisfies Axiom A and has no basic cycles. Let d denote a metric on M, and fix $\varepsilon > 0$. We will prove the formally stronger statement[5] that there exists a neighbourhood $\mathcal{W}$ of f in $\mathrm{Diff}^1(M)$ such that if $g \in \mathcal{W}$ then $f|_{\mathsf{nw}(f)}$ and $g|_{\mathsf{nw}(g)}$ are conjugate by a homeomorphism H such that $d(H(x), x) \leq \varepsilon$ for all $x \in \mathsf{nw}(f)$. By Axiom A and Proposition 48.7 $\mathsf{nw}(f)$ is isolated. Let $U \subset M$ be an open isolating neighbourhood. By Theorem 43.10 there is a C^1 neighbourhood $\mathcal{U}$ of f such that if $g \in \mathcal{U}$ then the maximal invariant set Δ_g of g in U is isolated in U and there is a homomorphism

$$H \colon \mathsf{nw}(f) \to \Delta_g, \qquad H \circ f|_{\mathsf{nw}(f)} = g|_{\Delta_g} \circ H, \qquad d(H(x), x) \leq \varepsilon, \ \forall\, x \in \mathsf{nw}(f).$$

Since f has no basic cycles, by Proposition 50.6 there is a C^0 neighbourhood $\mathcal{V}$ of f such that if $g \in \mathcal{V}$ then $\mathsf{nw}(g) \subset U$. Let

$$\mathcal{W} := \mathcal{U} \cap \big(\mathcal{V} \cap \mathrm{Diff}^1(M)\big).$$

We claim if $g \in \mathcal{W}$ then $\Delta_g = \mathsf{nw}(g)$. Indeed, certainly $\mathsf{nw}(g) \subseteq \Delta_g$, since Δ_g is the maximal invariant set in U. To see the other direction, observe that

$$\begin{aligned}
\Delta_g &= H(\mathsf{nw}(f)) \\
&= H\big(\overline{\mathsf{per}(f)}\big) && \text{by Axiom A,} \\
&= \overline{H(\mathsf{per}(f))} \\
&\subseteq \overline{\mathsf{per}(g)} && \text{by definition of a conjugacy,} \\
&\subseteq \mathsf{nw}(g).
\end{aligned}$$

This completes the proof. ■

We conclude the course by stating without proof the resolution of the structural stability question. This requires one last definition, which, in keeping with the rest of the course, has a rather bland name. Let f be a dynamical system such that $\mathsf{lim}(f)$ is hyperbolic. Then for all $x \in M$ the stable and unstable manifolds $W^s(x, f)$ and $W^u(x, f)$ are C^1 immersed submanifolds of M. For $x \in \mathsf{lim}(f)$ this is the content of (ii) of the Stable Manifold Theorem 40.11, and then for general x this follows from Corollary 49.4. Thus it makes sense to ask whether these submanifolds intersect transversely.

[5] In fact, this statement is equivalent to omega stability. This can be proved directly, or deduced from the $\Rightarrow$ direction of Theorem 50.3.

Definition 50.7. Let f be a dynamical system on a compact manifold such that $\mathsf{lim}(f)$ is hyperbolic. We say that f satisfies the **strong tranversality condition** if for every point $x \in M$, the stable and unstable manifolds at x intersect transversely: $W^s(x, f) \pitchfork W^u(x, f)$.

Remarks 50.8.

(i) The content of Definition 50.7 pertains to points in $M \setminus \mathsf{lim}(f)$. Indeed, if $\mathsf{lim}(f)$ is hyperbolic then it is automatically the case that $W^s(x, f) \pitchfork W^u(x, f)$ for $x \in M$, since in this case $T_x W^s(x, f) = E^s(x)$ and $T_x W^u(x, f) = E^u(x)$ by the Stable Manifold Theorem 40.11.

(ii) If f satisfies Axiom A and the strong transversality condition then f also has no basic cycles.

Theorem 50.9 (The Structural Stability Theorem). *A dynamical system on a compact manifold is structurally stable if and only if it satisfies Axiom A and the strong transversality condition.*

The $\Leftarrow$ direction of Theorem 50.9 was proved in the C^2 case by Robbin in 1971, and then in the C^1 case by Robinson in 1976. The $\Rightarrow$ direction was proved by Mañé in 1988. All of these results, however, are too difficult for us to cover, and therefore we will call it a day here.

Thank you all for attending, and enjoy your summer vacation!

Problem Sheet A

Problem A.1. Consider the restriction of the logistic map λ_4 to $[0,1]$.

(i) Prove that $\lambda_4|_{[0,1]}$ is conjugate to the tent map τ. *Hint:* $x \mapsto \sin^2(\frac{\pi}{2}x)$.

(ii) Prove that $\lambda_4|_{[0,1]}$ is a factor of the doubling map e_2. *Hint:* $x \mapsto \sin^2(\pi x)$.

Problem A.2. Consider the circle rotation $\rho_\theta \colon S^1 \to S^1$.

(i) Prove that

$$\mathsf{per}(\rho_\theta) = \begin{cases} S^1, & \theta \in \mathbb{Q}, \\ \emptyset, & \theta \notin \mathbb{Q}. \end{cases}$$

(ii) Prove that if $\theta \notin \mathbb{Q}$ then $\overline{\mathcal{O}_{\rho_\theta}(x)} = S^1$ for every $x \in S^1$.

(iii) Prove that ρ_θ is transitive if and only if $\theta \notin \mathbb{Q}$.

Problem A.3. Consider the dynamical system

$$f \colon (0,\infty) \to (0,\infty), \qquad f(x) := \frac{1}{2}\left(x + \frac{2}{x}\right).$$

Prove that $f^k(x) \to \sqrt{2}$ as $k \to \infty$ for all $x \geq 1$.

Problem A.4. Suppose X is a metric space with at least one isolated point and $f \colon X \to X$ is a transitive dynamical system. Show that X is necessarily finite, and $X = \mathcal{O}_f(x)$ for any point $x \in X$.

Problem A.5. Let $f \colon X \to X$ denote a transitive dynamical system. Prove that for any two open non-empty sets U, V, there are infinitely many $k \geq 0$ such that $f^k(U) \cap V \neq \emptyset$. *Hint:* If X has an isolated point, apply the previous problem.

Problem A.6. Recall that a continuous map $f \colon X \to X$ on a metric space is a **contraction** if $d(f(x), f(y)) \leq d(x,y)$ for all $x, y \in X$. Prove that a transitive dynamical system which is also a contraction is automatically minimal.

Problem A.7. Let $f \colon X \to X$ be a dynamical system on a metric space without isolated points. Prove that f is transitive if and only if for any $\varepsilon > 0$ and any two points $x, y \in X$ there exists $z \in X$ and $k, n \geq 0$ such that

$$d(f^k(z), x) < \varepsilon \qquad \text{and} \qquad d(f^n(z), y) < \varepsilon.$$

(♣) Problem A.8. Suppose Φ is a transitive flow on a metric space X. Prove that X is connected.

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

Problem Sheet B

Problem B.1. Let $f\colon X \to X$ be a dynamical system. Prove that the non-wandering set $\mathsf{nw}(f)$ can be characterised as the set of points $x \in X$ such that for every neighbourhood U of x and every $n \geq 1$ there exists $k \geq n$ such that $f^k(U) \cap U \neq \emptyset$.

Problem B.2. Let $f\colon X \to X$ be a dynamical system on a compact metric space. Prove that $\mathsf{cha}_{d_1}(f) = \mathsf{cha}_{d_2}(f)$ for any two metrics d_1 and d_2 defining the topology on X.

Problem B.3. Prove that the tent map has sensitive dependence on initial conditions.

Problem B.4. Prove that the doubling map $e_2\colon S^1 \to S^1$ is a factor of the shift map $\sigma\colon \Sigma_2 \to \Sigma_2$.

Problem B.5. Let $f\colon (1,\infty) \to (1,\infty)$ be the dynamical system defined by

$$f(x) = 2x.$$

Show that with respect to the standard metric $d(x,y) := |x-y|$ on $(1,\infty)$, the map f has sensitive dependence on initial conditions. Find an equivalent metric on $(1,\infty)$ for which f does not have sensitive dependence on initial conditions. *Hint:* Think of logarithms.

Problem B.6. Show that a dynamical system $f\colon X \to X$ is chaotic if and only if for each finite family $(U_1,\dots,U_n)$ of non-empty open sets there is a periodic point $x \in U_1$ of f and non-negative integers $k_2,\dots,k_n$ such that $f^{k_i}(x) \in U_i$ for $2 \leq i \leq n$.

(♣) Problem B.7. Let $\mathbb{T}^2 = S^1 \times S^1$ denote the torus. Given $\theta,\omega \in [0,1)$, consider the product dynamical system (cf. Example 1.24)

$$\rho_\theta \times \rho_\omega\colon \mathbb{T}^2 \to \mathbb{T}^2, \qquad (x,y) \mapsto (\rho_\theta(x), \rho_\omega(y)).$$

Prove that $\rho_\theta \times \rho_\omega$ is transitive if and only if the numbers $\{\theta,\omega,1\}$ are rationally independent[1].

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

[1]This means that the only solution to the equation $a\theta + b\omega = c$ for $a,b,c \in \mathbb{Z}$ is $a = b = c = 0$.

Problem Sheet C

Problem C.1. Let $f\colon X \to X$ be a dynamical system. Prove that the following statements are all equivalent (compare Proposition 5.13):

(i) f is weakly mixing.

(ii) For any three non-empty open subsets $U, V, W \subseteq X$, one has

$$\mathrm{ret}_f(U, W) \cap \mathrm{ret}_f(V, W) \neq \emptyset.$$

(iii) For any three non-empty open subsets $U, V, W \subseteq X$, one has

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(V, W) \neq \emptyset.$$

(iv) For any three non-empty open subsets $U, V, W \subseteq X$, one has

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(W, W) \neq \emptyset.$$

(v) For any two non-empty open subsets $U, V \subseteq X$, one has

$$\mathrm{ret}_f(U, V) \cap \mathrm{ret}_f(V, V) \neq \emptyset.$$

Problem C.2. Prove that a dynamical system $f\colon X \to X$ on a complete separable metric space is weakly mixing if and only if for every pair U, V of non-empty open subsets the set $\mathrm{ret}_f(U, V)$ contains two consecutive integers.

Problem C.3. Let us say that a dynamical system $f\colon X \to X$ is **totally transitive** if for every $k \in \mathbb{N}$ the system $f^k\colon X \to X$ is transitive.

(i) Prove that a weakly mixing system is totally transitive.

(ii) Prove that a chaotic totally transitive system is weakly mixing.

(♣) Problem C.4. Let X be a separable complete metric space and let Y be a metric space without isolated points. Suppose $f\colon X \to X$ and $g\colon Y \to Y$ are dynamical systems, and assume that f is mixing. Suppose $y \in Y$ has the property that $\mathcal{O}_g(y)$ is dense. Prove there exists $x \in X$ such that $\mathcal{O}_{f\times g}((x, y))$ is dense in $X \times Y$. *Remark:* This is harder than it looks. You cannot immediately deduce it from part (iv) of Proposition 5.4, since you don't get to choose y!

Problem Sheet D

Problem D.1. Let $f\colon X \to X$ and $g\colon Y \to Y$ be dynamical systems on compact metric spaces. Suppose that g is a factor of f. Prove that $\mathsf{h}_{\mathrm{top}}(g) \le \mathsf{h}_{\mathrm{top}}(f)$.

Problem D.2. Let $f\colon X \to X$ be a dynamical system on a compact metric space.

(i) Prove that $\mathsf{h}_{\mathrm{top}}(f^k) = k\,\mathsf{h}_{\mathrm{top}}(f)$ for $k \ge 1$.

(ii) Now assume that f is reversible. Prove that $\mathsf{h}_{\mathrm{top}}(f^{-1}) = \mathsf{h}_{\mathrm{top}}(f)$, and deduce that $\mathsf{h}_{\mathrm{top}}(f^k) = |k|\,\mathsf{h}_{\mathrm{top}}(f)$ for all $k \in \mathbb{Z}$.

Problem D.3. Let $f\colon [0,1] \to [0,1]$ be a reversible dynamical system. Prove that $\mathsf{h}_{\mathrm{top}}(f) = 0$. *Hint:* Use Proposition 8.8.

Problem D.4. Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic toral automorphism. Prove that f_L is mixing.

(♣) Problem D.5. Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ be a toral automorphism that is *not* hyperbolic.

(i) Suppose that $\det L = -1$. Prove that $\operatorname{tr} L = 0$.

(ii) Suppose that $\det L = 1$. Prove that $\operatorname{tr} L \in \{-2, -1, 0, 1, 2\}$.

(iii) Suppose that $|\operatorname{tr} L| < 2$. Prove that there exists $k \in \mathbb{N}$ such that $f_L^k = \mathrm{id}$.

(iv) Prove that there exist uncountably many non-empty open connected f_L-invariant subsets of $\mathbb{T}^2$. *Hint:* If $|\operatorname{tr} L| < 2$ then use part (iii). If $\det L = 1$ and $\operatorname{tr} L = \pm 2$ then you can write down an explicit formula for L.

Problem Sheet E

Problem E.1. Let $f \colon X \to X$ be a dynamical system on a compact metric space. Assume that either:

(i) f is expansive, or,

(ii) f is reversible and weakly expansive.

Recall the **minimal period** of a periodic point x of f is the minimal $k \geq 1$ such that $f^k(x) = x$. Define

$$\mathsf{per}_k(f) := \{x \in \mathsf{per}(f) \mid x \text{ has minimal period } k\}.$$

Prove that

$$\mathsf{h}_{\text{top}}(f) \geq \limsup_{k\to\infty} \frac{1}{k} \log \#\mathsf{per}_k(f).$$

Hint: Let δ be a (weak) expansivity constant for f. Show that $\mathsf{per}_k(f)$ is a (k,δ)-separated set.

Problem E.2. Let $f \colon X \to X$ and $g \colon Y \to Y$ be conjugate dynamical systems on compact metric spaces. Prove directly that

$$\mathsf{h}^*_{\text{top}}(f) = \mathsf{h}^*_{\text{top}}(g).$$

Problem E.3. Suppose $f \colon S^1 \to S^1$ is a reversible dynamical system with a periodic point. Prove that f is not weakly expansive. *Remark:* In fact the hypothesis that f has no periodic points is unnecessary. By the end of the course you should be able to prove this.

Problem E.4. Let $f \colon X \to X$ be a dynamical system on a compact metric space. Suppose there exist constants a, b such that

$$\mathsf{h}_{\text{top}}(f^k) \leq ak + b, \qquad \forall\, k \in \mathbb{N}.$$

Prove that $\mathsf{h}_{\text{top}}(f) \leq a$.

Problem E.5. Let $f, g \colon X \to X$ be two dynamical systems on a compact metric space. Suppose that

$$\left|\mathsf{h}_{\text{top}}(f^k) - \mathsf{h}_{\text{top}}(g^k)\right| < \sqrt{k}, \qquad \forall\, k \in \mathbb{N}.$$

Prove that $\mathsf{h}_{\text{top}}(f) = \mathsf{h}_{\text{top}}(g)$.

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

Problem Sheet F

Problem F.1. This problem computes the topological entropy of the shift map $\sigma\colon \Sigma_2 \to \Sigma_2$.

(i) Consider the following two open sets of Σ_2:

$$U := \{\mathtt{x} = (x_k)_{k\geq 0} \mid x_0 = 0\}\,, \qquad V := \{\mathtt{x} = (x_k)_{k\geq 0} \mid x_0 = 1\}\,,$$

Let $\mathcal{U} = \{U, V\}$. Prove that $\mathcal{U}$ is a generator for σ. Deduce that σ is expansive with respect to the metric d on Σ_2 from Definition 4.15.

(ii) Show that $\mathsf{h}_{\mathrm{top}}(\sigma) = \log 2$.

(♣) Problem F.2. This problem explores the ball dimension of two somewhat exotic spaces.

(i) Let C denote the Cantor ternary set obtained by iteratively deleting the open middle third from subintervals of $[0,1]$:

$$C := [0,1] \setminus \left(\bigcup_{k=0}^{\infty} \bigcup_{n=0}^{3^k-1} \left(\frac{3n+1}{3^{k+1}}, \frac{3n+2}{3^{k+1}} \right) \right).$$

Let d denote the metric inherited from $[0,1]$. Prove that

$$\text{ball-dim}_d(C) = \frac{\log 2}{\log 3}.$$

(ii) Let

$$X := \left\{ 0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\} \subset [0,1].$$

Let d denote the metric inherited from $[0,1]$. What is $\text{ball-dim}_d(X)$?

Problem F.3. This problem constructs an example of a dynamical system with infinite topological entropy.

(i) Let $f\colon [0,1] \to [0,1]$ denote the dynamical system defined by

$$f(x) := \begin{cases} 3x, & 0 \leq x \leq \frac{1}{3}, \\ 2 - 3x, & \frac{1}{3} \leq x \leq \frac{2}{3}, \\ 3x - 2, & \frac{2}{3} \leq x \leq 1. \end{cases}$$

Prove that $\mathsf{h}_{\mathrm{top}}(f) = \log 3$.

(ii) Let

$$I_k := \left[2^{-k}, 2^{-(k-1)}\right],$$

and define homeomorphisms

$$H_k \colon I_k \to [0,1], \qquad H_k(x) = 2^k x - 1.$$

Define a new dynamical system $g\colon [0,1] \to [0,1]$ by setting $g(0) := 0$ and

$$g(x) := H_k^{-1} \circ f^k \circ H_k, \qquad \text{for } x \in I_k.$$

Prove that g is continuous and that $\mathsf{h}_{\mathrm{top}}(g) = \infty$.

PROBLEM F.4. Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system.

(i) Prove there exists a fixed point a of f such that $0 < a < 1$.

(ii) Prove that the image of a non-trivial interval under f is another non-trivial interval.

(iii) Prove that f is surjective.

(♣) PROBLEM F.5. Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. Suppose there exists a fixed point $a \in (0,1)$ such that either a lies in the interior of $f\big([0,a]\big)$, or a lies in the interior of $f\big([a,1]\big)$. Prove that f is mixing.

Problem Sheet G

Problem G.1. Let $f\colon [0,1] \to [0,1]$ be a transitive dynamical system. Prove that f has a periodic point (which is not a fixed point) of period 6.

Problem G.2. Prove that a dynamical system $f\colon [0,1] \to [0,1]$ is turbulent if and only if there exists $a, b, c \in [0,1]$ such that[1]

$$c \in (\!(a, b)\!), \qquad f(a) = f(b) = a, \qquad f(c) = b.$$

Problem G.3. Let $f\colon [0,1] \to [0,1]$ be a mixing dynamical system. Suppose that $0 \notin f\big((0,1]\big)$. Then there exists a sequence $(x_k) \subset (0,1)$ of fixed points of f such that $x_k \to 0$.

Problem G.4. Let $f\colon [0,1] \to [0,1]$ be a dynamical system, and suppose $x \in \mathsf{per}(f)$ has minimal period $p > 1$. Let $\mathsf{G}\big(\mathcal{O}_f(x)\big)$ denote the graph of $\mathcal{O}_f(x)$ with vertices $I_1, \dots, I_{p-1}$.

(i) Prove that for every vertex I_k there is a vertex I_i such that $I_k \boldsymbol{\to} I_i$. Prove moreover that if $p \neq 2$ then it is always possible to choose $i \neq k$.

(ii) Prove that for every vertex I_k there is a vertex I_j such that $I_j \boldsymbol{\to} I_k$. Prove moreover that it is possible to choose $I_j \neq I_k$, unless p is even and $k = \frac{p}{2}$.

(♣) Problem G.5. Let $f\colon [0,1] \to [0,1]$ be a dynamical system. We say that f is **k-turbulent** if there exist k intervals $I_1, \dots, I_k$ with pairwise disjoint interiors such that

$$I_1 \cup \cdots \cup I_k \subseteq f(I_1) \cap \cdots \cap f(I_k).$$

We say f is **strictly k-turbulent** if the I_i can be chosen disjoint.

(i) Suppose f is k-turbulent. Prove that f^n is k^n-turbulent.

(ii) Suppose f is k-turbulent for $k \geq 3$. Prove that f is strictly $\lceil \frac{k}{2} \rceil$-turbulent.

(iii) Suppose f is k-turbulent. Prove that $\mathsf{h}_{\mathrm{top}}(f) \geq \log k$. *Hint:* First prove the result when f is strictly k-turbulent. To deduce the general case from this, apply parts (i) and (ii).

[1]The $(\!(\cdot,\cdot)\!)$ notation was introduced in Definition 13.1.

Problem Sheet H

Problem H.1. Let $F, G\colon \mathbb{R} \to \mathbb{R}$ be the maps

$$F(x) = x + \frac{1}{2}\sin(2\pi x), \qquad G(x) = x + \frac{1}{4\pi}\sin 2\pi x.$$

Decide whether there exist reversible dynamical systems on S^1 that lift to give either F or G. If so, are they orientation-preserving? And if yes, what is the rotation number?

Problem H.2. Suppose $f\colon S^1 \to S^1$ is an orientation-reversing reversible dynamical system. Prove that f has exactly two fixed points. Prove that f^2 is orientation-preserving and deduce that $\mathrm{rot}(f^2) = 0$.

Problem H.3. Give an example of an orientation-reversing reversible dynamical system on S^1 such that $\mathsf{per}(f) \neq \emptyset$ but such that the periodic points of f do not all have the same minimal period.

Problem H.4. Let $f\colon X \to X$ be a dynamical system on a compact metric space, and suppose $x \in X$ has the property that there exists $y \in \mathsf{per}(f)$ such that $\mathcal{O}_f(y) \subseteq \omega_f(x)$. Prove that $\omega_f(x) = \mathcal{O}_f(y)$ Deduce that if $\omega_f(x)$ is finite then there exists $y \in \mathsf{per}(f)$ such that $\omega_f(x) = \mathcal{O}_f(y)$.

Problem H.5. Let $f\colon S^1 \to S^1$ denote an orientation-preserving reversible dynamical system with rational rotation number $\mathrm{rot}(f) = \frac{p}{q}$ with p and q relatively prime. Suppose $z \in S^1$ is not a periodic point for f. Let $w_1, w_2 \in \mathsf{per}(f)$ denote the periodic points such that z is positively asymptotic to w_1 and negatively asymptotic to w_2 under f^q (these points exist by Proposition 16.10). Prove that for each $1 \leq i \leq q-1$, $f^i(z)$ is positively asymptotic to $f^i(w_1)$ and negatively asymptotic to $f^i(w_2)$ under f^q.

Problem H.6. Let $\mathrm{Hom}^+(S^1)$ denote the set of orientation-preserving reversible dynamical systems of S^1. Endow $\mathrm{Hom}^+(S^1)$ with the metric

$$\tilde{d}(f,g) := \max_{z \in S^1} d(f(z), g(z)),$$

where d is the standard metric on S^1 (cf. (8.1)). Prove that

$$\mathrm{rot}\colon \mathrm{Hom}^+(S^1) \to (S^1, d)$$

is a continuous function.

Problem H.7. Suppose $f\colon S^1 \to S^1$ is an orientation-preserving reversible dynamical system with rational rotation number. Prove that $\mathsf{nw}(f) = \mathsf{per}(f)$.

Will J. Merry, Dyn. Systems I, Autumn 2019, ETH Zürich. Last modified: June 08, 2020.

Problem Sheet I

(♣) Problem I.1. Consider the continuous function $g\colon S^1 \to \mathbb{R}$ defined by

$$g(x) \coloneqq \begin{cases} x \sin \frac{\pi}{x}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Prove that g does not have bounded variation.

Problem I.2. Let $f, g\colon S^1 \to S^1$ be two commuting orientation-preserving reversible dynamical systems. Prove that $\mathrm{rot}(f \circ g) = \mathrm{rot}(f) + \mathrm{rot}(g) \mod 1$. Find an example to show that the commuting hypothesis is necessary.

Problem I.3. Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with rational rotation number $\theta = \frac{p}{q}$. Suppose that ρ_θ is a factor of f. Prove that $\mathsf{per}(f)$ is uncountable. *Remark:* The converse of this statement is also true, but the proof is harder.

Problem I.4. Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system with irrational rotation number. Prove that the non-wandering set $\mathsf{nw}(f)$ is a minimal set. Prove that either $\mathsf{nw}(f)$ is a perfect nowhere dense set or $\mathsf{nw}(f) = S^1$.

(♣) Problem I.5. Let $f\colon S^1 \to S^1$ be an orientation-preserving reversible dynamical system. Prove that the chain recurrent[1] set $\mathsf{cha}_d(f)$ is either equal to all of S^1 or equal to $\overline{\mathsf{per}(f)}$. Deduce that if $\mathrm{rot}(f)$ is irrational then $\mathsf{cha}_d(f) = S^1$.

[1] Here d is any metric on S^1 inducing the usual topology. For example, d could be given by (8.1). It does not matter which metric we choose by Proposition 3.16.

Problem Sheet J

PROBLEM J.1. Show that the doubling map $e_2\colon S^1 \to S^1$ is an ergodic measure-preserving dynamical system with respect to the Lebesgue measure. *Remark:* The Lebesgue measure λ on S^1 is defined by identifying S^1 with $[0,1)$ and taking the Lebesgue measure on $[0,1)$.

PROBLEM J.2. Prove that the circle rotation ρ_θ is a measure-preserving dynamical system with respect to Lebesgue measure. Prove however that ρ_θ is ergodic if and only if θ is irrational.

PROBLEM J.3. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Prove that f is ergodic if and only if for every measurable function $u\colon X \to \mathbb{R}$, if $f^*(u) \geq u$ almost everywhere then u is constant almost everywhere.

(♣) PROBLEM J.4. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Given $A \in \mathscr{A}$, define the **first return time** $\tau_A\colon X \to \mathbb{N} \cup \{\infty\}$ by

$$\tau_A(x) := \inf \left\{ k \geq 1 \mid f^k(x) \in A \right\},$$

where by convention $\inf \emptyset := \infty$.

(i) Prove that for any $A \in \mathscr{A}$ such that $\mu(A) > 0$, the function τ_A is integrable with
$$\int_A \tau_A \, d\mu = \mu\Big(\{x \in X \mid \tau_A(x) < \infty\}\Big).$$

(ii) Now assume that f is ergodic. Prove that for any $A \in \mathscr{A}$ such that $\mu(A) > 0$ one has
$$\int_A \tau_A \, d\mu = 1.$$

PROBLEM J.5. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Let $A \in \mathscr{A}$ have $\mu(A) > 0$, and define τ_A for $A \in \mathscr{A}$ as in Problem J.4. Define the **Poincaré return map** $\sigma_A\colon A \to A$ by

$$\sigma_A(x) := f^{\tau_A(x)}(x) \qquad \text{if } \tau_A(x) < \infty.$$

(i) Prove that this is a well-defined measure-preserving dynamical system on the restricted space $(A, \mathscr{A}_A, \mu_A)$ (cf. Example 18.12). *Remark:* Your proof should include an explanation as to why it did not matter that we only bothered to define σ_A on points for which $\tau_A(x) < \infty$!

(ii) Now suppose that f is ergodic. Prove that σ_A is also ergodic.

Problem Sheet K

PROBLEM K.1. Prove that the circle rotation $\rho_\theta \colon S^1 \to S^1$ is never weakly mixing with respect to Lebesgue measure.

(♣) PROBLEM K.2. Let $(X, \mathscr{A}, \mu)$ be a probability space with a countable basis (Definition 18.43). Let f be a weakly mixing dynamical system on $(X, \mathscr{A}, \mu)$. Prove there exists a set $K \subset \{0, 1, 2, \dots\}$ of density zero such that

$$\lim_{k \notin K} \mu(f^{-k}A \cap B) = \mu(A)\mu(B), \qquad \forall\, A, B \in \mathscr{A}.$$

PROBLEM K.3. Let $(X, \mathscr{A}, \mu)$ be a probability space with a countable basis, and let f be a dynamical system on $(X, \mathscr{A}, \mu)$. Prove that f is weakly mixing if and only if there exists a set $K \subset \{0, 1, 2, \dots\}$ of density zero such that

$$\lim_{k \notin K} \langle\!\langle (f^*)^k(u), v \rangle\!\rangle = \int_X u \, d\mu \int_X \overline{v} \, d\mu, \qquad \forall\, u, v \in L^2(\mu; \mathbb{C}).$$

(♣) PROBLEM K.4. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Let $1 \leq p < \infty$ and suppose $u \in L^p(\mu)$. Prove there exists $v \in L^p(\mu)$ such that

$$f^*(v) = v \text{ almost everywhere}, \qquad \text{and} \qquad \lim_{k \to \infty} \left\| \left(\frac{1}{k} \sum_{i=0}^{k-1} u \circ f^i \right) - v \right\|_p = 0.$$

Hint: If u is bounded this follows from the Birkhoff Ergodic Theorem 20.2.

Problem Sheet L

Problem L.1. Let X be a compact metric space, and let $\mathcal{M}(X)$ denote the space of all Borel probabilty measures on X, equipped with the weak star topology. Prove that the map

$$\imath\colon X \to \mathcal{M}(X), \qquad \imath(x) := \delta_x$$

is a topological embedding (i.e. a homeomorphism onto its image).

Problem L.2. Let X be a compact metric space.

(i) Suppose $f, g\colon X \to X$ be two commuting topological dynamical systems. Prove that $\mathcal{M}(f) \cap \mathcal{M}(g)$ is not empty.

(ii) Now suppose $\{f_i \mid i \in I\}$ is an arbitrary (possibly uncountable) commuting family of topological dynamical systems on X (i.e. $f_i \circ f_j = f_j \circ f_i$ for all $i, j \in I$). Prove that

$$\bigcap_{i \in I} \mathcal{M}(f_i) \neq \emptyset.$$

Problem L.3. Let $f\colon X \to X$ be a topological dynamical system on a compact metric space. Let $p \in \mathbb{N}$ and $x \in X$. Suppose $f^p(x) = x$. Prove that the periodic orbit measure $\wp_{x,p}$ is ergodic.

Problem L.4. Let f be a topological dynamical system on a compact metric space.

(i) Suppose $\mu \in \mathcal{M}(f)$ is a purely atomic measure. Prove that μ is a (possibly countably infinite) convex combination of periodic orbit measures.

(ii) Suppose $\mu \in \mathcal{E}(f)$ is a purely atomic measure. Prove that μ is a periodic orbit measure.

Problem Sheet M

Problem M.1. Let $f\colon X \to X$ denote a topological dynamical system on a compact metric space X. Assume f is uniquely ergodic, with $\mathcal{M}(f) = \{\mu\}$. Prove that f is minimal if and only if $\mu(U) > 0$ for all non-empty open subsets U.

Problem M.2. Let $f\colon S^1 \to S^1$ be a reversible topological dynamical system with no periodic points. Prove that f is uniquely ergodic.

Problem M.3. Let $(X, \mathscr{A}, \mu)$ denote a probability space, and let ξ, η, ζ be three partitions of $(X, \mathscr{A}, \mu)$. Prove that:

(i) If $\eta \preceq \zeta$ then $\mathsf{H}(\xi|\zeta) \le \mathsf{H}(\xi|\eta)$.

(ii) $\mathsf{H}(\xi \vee \eta|\zeta) = \mathsf{H}(\xi|\zeta) + \mathsf{H}(\eta|\xi \vee \zeta)$.

(iii) $\mathsf{H}(\xi|\zeta) \le \mathsf{H}(\xi|\eta) + \mathsf{H}(\eta|\zeta)$.

Problem M.4. Let $(X, \mathscr{A}, \mu)$ denote a probability space, and let $\mathscr{P}_p \subseteq \mathscr{P}$ denote those partitions with exactly p elements. Define a function

$$\tilde{d}_p\colon \mathscr{P}_p \times \mathscr{P}_p \to \mathbb{R}^+$$

as follows. If $\xi = \{C_1, \dots, C_p\}$ and $\eta = \{D_1, \dots, D_p\}$, set

$$\tilde{d}_p(\xi, \eta) := \min_{\sigma \in \mathfrak{S}(p)} \sum_{i=1}^{p} \mu(C_i \mathbin{\triangle} D_{\sigma(i)}),$$

where the sum is over the symmetric group of permutations of $\{1, 2, \dots, p\}$. Prove that $\tilde{d}_p$ is a metric.

(♣) Problem M.5. Let $(X, \mathscr{A}, \mu)$ denote a probability space. Fix $p \in \mathbb{N}$ and $\varepsilon > 0$. Prove that there exists $\delta > 0$ such that if $\xi, \eta \in \mathscr{P}_p$ then

$$d_{\mathrm{R}}(\xi, \eta) < \delta \qquad \Rightarrow \qquad \tilde{d}_p(\xi, \eta) < \varepsilon.$$

Problem Sheet N

PROBLEM N.1. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$.

(i) Prove that
$$\mathsf{h}_\mu(f^p) = p\,\mathsf{h}_\mu(f), \qquad \forall\, p \in \mathbb{N}.$$

(ii) Now suppose that f is reversible. Prove that
$$\mathsf{h}_\mu(f) = \mathsf{h}_\mu(f^{-1}),$$
and deduce that $\mathsf{h}_\mu(f^p) = |p|\,\mathsf{h}_\mu(f)$ for any $p \in \mathbb{Z}$.

PROBLEM N.2. Let f be a dynamical system on a probability space $(X, \mathscr{A}, \mu)$, and let g be a dynamical system on another probability space $(Y, \mathscr{B}, \nu)$. Prove that
$$\mathsf{h}_{\mu\times\nu}(f \times g) = \mathsf{h}_\mu(f) + \mathsf{h}_\nu(g).$$

PROBLEM N.3. Let f be a reversible dynamical system on a probability space $(X, \mathscr{A}, \mu)$. Prove that if f has a generator ξ then $\mathsf{h}_\mu(f) = 0$.

PROBLEM N.4. Prove that a circle rotation satisfies $\mathsf{h}_\lambda(\rho_\theta) = 0$, where λ is the Lebesgue measure.

PROBLEM N.5. Let f be a topological dynamical system on a compact metric space X.

(i) Let $\mu \in \mathcal{M}(f)$. Prove that $\mu(\mathsf{nw}(f)) = 1$.

(ii) Prove that $\mathsf{h}_{\text{top}}(f) = \mathsf{h}_{\text{top}}\big(f|_{\mathsf{nw}(f)}\big)$. *Hint:* Use the Variational Principle.

Problem Sheet O

(♣) Problem O.1. Let $L\colon E \to E$ denote a linear dynamical system. Prove that L is hyperbolic if and only if every eigenvalue λ of L has absolute value different to 1.

Problem O.2. Let $L\colon E \to E$ be a reversible linear dynamical system. Define

$$F^s := \left\{ v \in E \;\Big|\; \sup_{k \geq 0} \|L^k v\| < \infty \right\},$$
$$F^u := \left\{ v \in E \;\Big|\; \sup_{k \geq 0} \|L^{-k} v\| < \infty \right\}.$$

Prove that L is hyperbolic if and only if $F^s \cap F^u = \{0\}$.

Problem O.3. Let $L\colon E \to E$ be a reversible linear dynamical system. Define

$$G^s := \left\{ v \in E \mid \|L^k v\| \to 0 \text{ as } k \to \infty \right\},$$
$$G^u := \left\{ v \in E \mid \|L^{-k} v\| \to 0 \text{ as } k \to \infty \right\}.$$

Prove that L is hyperbolic if and only if $G^s + G^u = E$.

Problem O.4. Let $L\colon E \to E$ be a reversible linear dynamical system. Suppose we can write $E = F \oplus G$ such that L has matrix form

$$L = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} : F \oplus G \to F \oplus G,$$

with

$$\max\left\{ \|A^{-1}\|^{\mathrm{op}}, \|D\|^{\mathrm{op}} \right\} < 1.$$

Prove that L is hyperbolic.

(♣) Problem O.5. Let $L\colon E \to E$ be a reversible linear dynamical system. Denote by $r(L)$ the **spectral radius** of L, that is, the maximal absolute value of the eigenvalues of L. Prove that for any $\varepsilon > 0$ there is a norm $\|\cdot\|$ on E with associated operator norm $\|\cdot\|^{\mathrm{op}}$ such that

$$\|L\|^{\mathrm{op}} < r(L) + \varepsilon.$$

Problem O.6. Given $\lambda \geq 0$, let $L_\lambda\colon \mathbb{R} \to \mathbb{R}$ denote the linear map

$$L_\lambda(x) = \lambda x.$$

Suppose $0 < \lambda < \mu < 1$. Show that L_λ and L_μ are conjugate. Prove however that it is not possible to choose the conjugating homeomorphism h in such a way that both h and h^{-1} are Lipschitz.

Problem Sheet P

PROBLEM P.1. Let $f\colon \Omega \subseteq E \to E$ be a dynamical system and suppose $u \in \Omega$ is a hyperbolic fixed point of f. There exists $r > 0$ such that if $v \in \Omega$ satisfies

$$\|f^k(v) - u\| \le r, \qquad \forall\, k \in \mathbb{Z},$$

then $v = u$.

PROBLEM P.2. Let $f\colon \Omega \subseteq E \to E$ be a dynamical system with a hyperbolic fixed point u. Prove that given any $n \in \mathbb{N}$, there exists a neighbourhood V_n of u such that any periodic point of f in $V_n \setminus \{u\}$ has period greater than n.

(♣) PROBLEM P.3. Let $\big(F, \|\cdot\|_F\big)$ and $\big(G, \|\cdot\|_G\big)$ be finite-dimensional normed vector spaces. Given a continuous map $\psi\colon F \to G$, define

$$\|\psi\|^* := \sup_{v \neq 0} \frac{\|\psi(v)\|_G}{\|v\|_F},$$

and let

$$\Sigma := \{\psi \mid \psi(0) = 0,\ \|\psi\|^* < \infty\}.$$

Prove that $\big(\Sigma, \|\cdot\|^*\big)$ is a Banach space, and that the inclusion

$$\Big(\mathcal{L}(F, G), \|\cdot\|^{\mathrm{op}}\Big) \hookrightarrow \Big(\Sigma, \|\cdot\|^*\Big)$$

is an isometry.

(♣) PROBLEM P.4. Let $L\colon E \to E$ denote a hyperbolic linear dynamical system, with corresponding hyperbolic splitting $E = E^s \oplus E^u$.

(i) Show that for any $v \in E$, the stable manifold[1] $W^s(v, L)$ is given by the affine space $v + E^s$, and similarly that $W^u(v, L) = v + E^u$.

(ii) Now let us specialise this to the case $E = \mathbb{R}^2$. Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ denote the associated hyperbolic toral automorphism (cf. Definition 8.13). Prove that for any $x \in \mathbb{T}^2$, the stable manifold $W^u(x, f_L)$ is given by

$$W^s(x, f_L) = \pi(W^s(v, L)),$$

where $\pi\colon \mathbb{R}^2 \to \mathbb{T}^2$ is the projection and $v \in \pi^{-1}(x)$.

PROBLEM P.5. Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ denote a hyperbolic toral automorphism. Prove that

$$\#\mathsf{fix}(f_L) = |\det(L - \mathrm{id})|.$$

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1] Recall from Remark 33.4 that the stable manifold can be defined for any point, not necessarily a fixed point. In general such a stable “manifold” is not actually a manifold—however in this case it is.

Problem Sheet Q

Problem Q.1. Let $f_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic toral automorphism (cf. Definition 8.13). Prove that f_L is Anosov.

Problem Q.2. Let (M, m) be a compact Riemannian manifold. Let $\Lambda \subseteq M$ be a compact invariant set of a dynamical system f. Let $E \subseteq T_\Lambda M$ be a Df-invariant C^0 subbundle. Prove that the following conditions are equivalent:

(i) There exists $C \geq 1$ and $0 < \mu < 1$ such that

$$\|Df^k(x)v\| \leq C\mu^k\|v\|, \qquad \forall\, x \in \Lambda,\ v \in E(x),\ k \geq 0.$$

(ii) There exists $0 < \lambda < 1$ and $n \geq 0$ such that

$$\|Df^k(x)v\| \leq \lambda\|v\|, \qquad \forall\, x \in \Lambda,\ v \in E(x),\ k \geq n.$$

(iii) For any $x \in \Lambda$ and $v \neq 0_x \in E(x)$, there exists a $n = n(v)$ such that

$$\|Df^n(x)v\| < \|v\|.$$

Problem Q.3. Let f be a dynamical system on a compact smooth manifold M. Let $x \in \mathsf{per}(f)$ be a periodic point, and assume that the orbit $\mathcal{O}_f(x)$ is hyperbolic with $E^u(y) = \{0_y\}$ for all $y \in \mathcal{O}_f(x)$. Suppose[1] $x \in \alpha_f(z)$ for some $z \in M$. Prove that $z \in \mathcal{O}_f(x)$.

Problem Q.4. Let $\Lambda \subset M$ be a hyperbolic set for a dynamical system f. Suppose that $E^s(x) = \{0_x\}$ for each $x \in \Lambda$. Prove that Λ consists of finitely many periodic orbits of f.

[1]Recall $\alpha_f(z) = \omega_{f^{-1}}(z)$—see Definition 3.5.

Problem Sheet R

Problem R.1. Suppose E is a finite-dimensional normed vector space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$. Prove that for any $\delta > 0$, there exists $c \geq 1$ such that if $E = F \oplus G$ is a direct sum and $\measuredangle(F, G) > \delta$, then if $\| \cdot \|_{\mathrm{b}}$ denotes the box-adjusted norm from $\| \cdot \|$ with respect to $F \oplus G$ then

$$\frac{1}{c}\|v\| \leq \|v\|_{\mathrm{b}} \leq c\|v\|, \qquad \forall v \in E.$$

(♣) Problem R.2. Let f denote a dynamical system on a compact Riemannian manifold (M, m) and suppose $\Lambda \subseteq M$ is an invariant set. Let us say that a continuous invariant splitting $T_\Lambda M = F^s \oplus F^u$ is **dominated** if there exist constants $C \geq 1$ and $0 < \mu < 1$ such that

$$\big\|Df^k|_{F^s(x)}\big\| \cdot \big\|Df^{-k}|_{F^u(f^k(x))}\big\| \leq C\mu^k, \qquad \forall\, k \in \mathbb{N},\ x \in \Lambda.$$

If such a setting exists, we say that Λ is a **weakly hyperbolic** set[1].

(i) Prove that a hyperbolic set is also weakly hyperbolic.

(ii) Suppose $F^s \oplus F^u$ and $G^s \oplus G^u$ are two dominated splittings of $T_\Lambda M$ such that $\dim F^s(x) = \dim G^s(x)$ for all $x \in \Lambda$. Prove that $F^s = G^s$ and $F^u = G^u$. Thus dominated splittings of a given dimension are unique.

(iii) Suppose Λ is weakly hyperbolic. Prove that $\overline{\Lambda}$ is also weakly hyperbolic. Thus as in the hyperbolic case, we can without loss of generality assume a weakly hyperbolic set is compact.

(♣) Problem R.3. Let f denote a dynamical system on a compact Riemannian manifold (M, m) and suppose $\Lambda \subseteq M$ is a compact invariant set with a dominated splitting. Prove that there exists a C^1 neighbourhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and a number $a > 0$ such that for any $g \in \mathcal{U}$, any compact g-invariant set Δ with $\Delta \subset B(\Lambda, a)$ has a dominated splitting with respect to g.

Will J. Merry, Dyn. Systems II, Spring 2020, ETH Zürich. Last modified: June 08, 2020.

[1]There is another variant on this definition that turns out to be more useful in practice. A dominated splitting $F^s \oplus F^u$ is said to be **strongly dominating** if in addition there exists constants $c \geq 1$ and $\lambda > 0$ such that at least one of

$$\big\|Df^k|_{F^s}\big\| \leq c\lambda^k, \qquad \forall\, k \in \mathbb{N},\ x \in \Lambda.$$

or

$$\big\|Df^{-k}|_{F^u}\big\| \leq c\lambda^k, \qquad \forall\, k \in \mathbb{N},\ x \in \Lambda$$

holds. If a strongly dominated splitting exists, then we say that Λ is **partially hyperbolic**. Partial hyperbolicity sits in between weak hyperbolicity and hyperbolicity:

$$\text{hyperbolic} \quad \Rightarrow \quad \text{partially hyperbolic} \quad \Rightarrow \quad \text{weakly hyperbolic.}$$

Many of the results proved in this course extend to partially hyperbolic systems. For example, if f is a dynamical system with the property that the entire manifold M is a partially hyperbolic set (i.e. the partially hyperbolic version of Anosov), then $\mathsf{h}_{\mathrm{top}}(f) > 0$.

(♣) Problem R.4. Let f denote a dynamical system on a compact Riemannian manifold (M, m) and suppose $\Lambda \subseteq M$ is a compact invariant set. Let $\mathcal{E}$ denote the Banach space of bounded sections of $T_\Lambda M$:

$$\mathcal{E} \coloneqq \Gamma^0(\Lambda, T_\Lambda M),$$

equipped with the norm

$$\|\gamma\|_0 := \sup_{x \in \Lambda} \|\gamma(x)\|.$$

Consider the linear operator

$$L_f \colon \mathcal{E} \to \mathcal{E}$$

defined by

$$L_f(\gamma)(x) \coloneqq Df(f^{-1}(x))\gamma(f^{-1}(x)).$$

(i) Suppose L_f is a hyperbolic[2] linear dynamical system on $\mathcal{E}$. Prove that Λ is a hyperbolic set for f.

(ii) **(Hard)** Prove the converse to the previous statement: if Λ is a hyperbolic set then L_f is hyperbolic.

This problem shows that hyperbolic sets on manifolds can be reduced to linear hyperbolic systems, at the expense of moving into infinite dimensions.

[2] In case you are confused, the definition of being a hyperbolic *linear* operator is formally identical in a Banach space setting. To recap: if $(\mathcal{E}, |\cdot|)$ is Banach space and $L \colon \mathcal{E} \to \mathcal{E}$ is a linear isomorphism, we say that L is a **hyperbolic linear dynamical system** if $\mathcal{E}$ splits into a direct sum

$$\mathcal{E} = \mathcal{E}^s \oplus \mathcal{E}^u$$

which is L-invariant in the sense such that there exist constants $C \geq 1$ and $0 < \mu < 1$ such that

$$\|L^k v\| \leq C\mu^k \|v\|, \qquad \forall\, v \in \mathcal{E}^s,\ \forall\, k \geq 0,$$

and such that

$$\|L^{-k} v\| \leq C\mu^k \|v\|, \qquad \forall\, v \in \mathcal{E}^u,\ \forall\, k \geq 0.$$

This is again equivalent to asking that the spectrum of L does not meet the unit circle in $\mathbb{C}$, although in infinite dimensions this requires the Spectral Decomposition Theorem to prove.

Problem Sheet S

Problem S.1. Let $f\colon X \to X$ be a reversible dynamical system on a compact metric space X.

(i) Show that for any $r > 0$ and any $x \in X$, one has

$$W^s(x, f) = \bigcup_{k \geq 0} f^{-k}\left(W^s_{\mathrm{loc},r}(f^k(x), f)\right).$$

(ii) Let $A \subseteq X$ be a compact completely invariant set. Prove that

$$W^s(A, f) = \left\{x \in X \mid \omega_f(x) \subseteq A\right\},$$

(♣) Problem S.2. Suppose f is a smooth dynamical system on a closed manifold M. Assume we are given a Df-invariant splitting $TM = F^s \oplus F^u$ of the entire tangent bundle. Assume that the restriction of this splitting to the non-wandering set $\mathsf{nw}(f)$ is hyperbolic. Prove that the entire splitting is hyperbolic, and hence that f is Anosov.

Problem S.3. Prove the following enhancement of the Shadowing Theorem 44.3: Let f be a dynamical system of a compact manifold M and let $\Lambda \subset M$ be an **isolated** compact hyperbolic set. Prove that there is $\varepsilon_0 > 0$ and $\delta_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ there exists $0 < \delta < \delta_0$ such that every δ-chain contained in Λ is ε-shadowed by exactly one point,*which in addition belongs to* Λ.

Problem S.4. Let $f : M \to M$ be an Anosov diffeomorphism of a closed manifold. Prove that if the unstable manifold $W^u(x, f)$ is dense in M for every point x then f is mixing.

Problem S.5. Suppose f is a smooth dynamical system on a closed manifold M. Assume that $\mathsf{nw}(f)$ is hyperbolic. Prove that

$$\mathsf{nw}\left(f|_{\mathsf{nw}(f)}\right) = \overline{\mathsf{per}(f)}.$$

Problem Sheet T

PROBLEM T.1. Let f be a dynamical system on a compact manifold M, and let $\Lambda \subseteq M$ be a compact hyperbolic set. Prove that there exists $r_0 > 0$ with the following property: for any $0 < r \le r_0$ there exists a $\delta > 0$ such that if $x, y \in \Lambda$ satisfy $d(x, y) \le \delta$ then

$$W^s_{\mathrm{loc},r}(x, f) \cap W^u_{\mathrm{loc},r}(y, f) = \{z\} \tag{T.1}$$

for a unique point $z \in M$, and moreover the intersection is always transverse.

PROBLEM T.2. Let $\Delta \subseteq \Lambda \times \Lambda$ denote the diagonal. The previous problem tells us that there is a well-defined map $\varphi \colon B(\Delta, \delta) \to M$ that sends a pair (x, y) to the unique point z from the right-hand side of (T.1).

(i) Prove that φ is continuous.

(ii) Assume that Λ is isolated. Prove that $\operatorname{im} \varphi \subseteq \Lambda$. *Remark:* The converse of this statement is also true—if $\operatorname{im} \varphi \subseteq \Lambda$ then Λ is isolated—but this is harder to prove.

(♣) PROBLEM T.3. Let $f \colon M \to M$ be a dynamical system on a compact manifold. Assume that f is Anosov and that[1] $\mathsf{nw}(f) = M$. Prove that for every point $x \in M$ the stable manifold $W^s(x, f)$ is dense in M. *Hint:* First prove this for the case $x \in \mathsf{per}(f)$.

PROBLEM T.4. Let $f \colon M \to M$ be a dynamical system on a compact manifold. Assume that f is Anosov and that $\mathsf{nw}(f) = M$. Prove that f is transitive.

PROBLEM T.5. Let $f \colon M \to M$ be an Anosov diffeomorphism on a compact manifold. Prove that f satisfies Axiom A.

[1]It is a long-standing conjecture that the non-wandering set of any Anosov diffeomorphism is always entire manifold (and thus that this assumption is superfluous). This has been proved in many situations (for example when M is a nilmanifold or infranilmanifold), but remains open in general. **Bonus Problem:** Prove this conjecture.