Notes on Entropy

Sayantan Khan
University of Michigan

Fall 2018

Note: These notes are incomplete, and have plenty of errors. Read at your own peril.

1 Ergodic transformations

Definition 1.1. A measure-preserving transformation T mapping a probability space (X, μ) to itself is called ergodic if the only T-invariant subsets are of measure 1 or measure 0.

We only care about ergodic transformations on probability spaces because if a transformation T acting on a space is not ergodic, we can write the space X as a union of two smaller subsets X_1 and X_2, on which T acts invariantly. In that sense, ergodicity is a an irreducibility condition.

Functions which are invariant under ergodic transformations are quite special: they are in fact constant. To be more precise, if $f \in L^2(X)$ is T invariant, i.e. $f(x) = f(Tx)$, then f is constant almost everywhere. This is in fact equivalent to ergodicity, so one way of showing that a transformation is ergodic is to show that the only L^2 functions invariant under the transformation are the constant functions.

We also have the pointwise ergodic theorem, due to Birkhoff.

Theorem 1 (Birkhoff, 1931)**.** *Suppose T is an ergodic transformation on the probability space (X, μ). Then the following sequence of functions converges pointwise almost everywhere to the given limit*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i x) = \int_X f d\mu$$

In the above result, if we set f to be the indicator function of a measurable subset, then the left hand side indicates the fraction of time the orbit of any point spends inside the set is the same as the measure of the set. This is what equidistribution is, so the pointwise ergodic theorem tells us that ergodicity implies equidistribution.

1.1 Examples (and non examples) of ergodic transformations

There are three ways of showing a transformation is ergodic.

(1) Harmonic analysis

(2) The Hopf Argument

(3) Assorted tricks

1.1.1 Rotations of S^1

Consider S^1 with the Lebesgue measure, and let T_α be the map sending the angle θ to $\theta + \alpha$. For which α is the transformation ergodic? To answer this question, we'll use the first technique, i.e. harmonic analysis.

To determine when T_α is ergodic, it's useful to see what the action of T on L^2 functions is. If $f \in L^2(S^1)$, and $f_\alpha = f \circ T_\alpha$, then the relation between the Fourier coefficients of f and f_α is given by the following identity.

$$\widehat{f_\alpha}(n) = e^{2\pi i n \alpha} \cdot \widehat{f}(n)$$

If α is rational, of the form $\frac{p}{q}$, then for all multiples of q, $e^{2\pi i n q x}$ is a non-constant L^2 function invariant under T_α, so T_α is clearly not ergodic. On the other hand, if α is irrational, then only n for which $\widehat{f_\alpha}(n)$ equals $\widehat{f}(n)$ is $n = 0$, i.e. the constant L^2 functions. This shows for irrational α, T_α is ergodic.

We can't really apply the Hopf argument here, because we'll see later that the Hopf argument relies on the transformation shrinking some part of the space, and here the transformation is an isometry.

1.1.2 Cat map on $S^1 \times S^1$

Consider the torus as a quotient of $\mathbb{R}^2$ by $\mathbb{Z}^2$ along with the induced Lebesgue measure, and the following transformation to itself.

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The fact that this map is measure preserving is a consequence of the fact that the determinant of the map is 1. There are multiple ways of showing this map is ergodic. One way is to use harmonic analysis, and look at what this map does to Fourier coefficients, much like the previous example.

Another way is to use the Hopf argument. To use the Hopf argument, observe the map has two eigenvectors, with eigenvalues $\frac{3-\sqrt{5}}{2}$, and $\frac{3+\sqrt{5}}{2}$. One of these eigenvalues is less than 1, and another is greater than 1. Let the eigenvector corresponding to the first eigenvalue be v_1, and the vector corresponding to the second eigenvalue be v_2. What the first eigenvalue being less than 1 means is that any two points lying on the same leaf of a foliation defined by v_1 will get closer and closer as you apply the map T. We call this leaf a leaf of the stable foliation. Similarly, any two points lying on the same leaf of a foliation defined by v_2 will get closer as you apply the map T^{-1}. This is the foliation we'll call the unstable foliation. Consider any continuous function f which is T-invariant. Because of the fact that points lying on the same leaf of a foliation get closer either forward or backwards in time means that the function f is constant on a leaf.

But now observe that on this torus we can get from any point to any other point by travelling along a leaf of the stable foliation, and then along the leaf of the unstable foliation. This means the function is constant on all of the torus. This tells us that continuous T-invariant functions are constant. But the constant functions are dense in L^2, which means the T-invariant L^2 functions are also constant.

The Hopf argument relies breaking up a space into stable and unstable sets, and if one can get from any point to any other point by travelling along stable and unstable sets, the associated transformation is ergodic.

1.1.3 The Bernoulli shift on $\{0, 1\}^{\mathbb{Z}}$

Consider the space $\{0, 1\}$ with probability measure that assigns $\{0\}$ with probability p, where $0 < p < 1$. Then we can define the product probability measure on $\{0, 1\}^{\mathbb{Z}}$. The Bernoulli shift takes a sequence of 0 and 1 in the space, and shifts it to the left. It's clear that this is a measure preserving transformation. To show that this transformation is ergodic, we can use the Hopf argument, or "assorted tricks". We'll first do use the Hopf argument technique to highlight similarities with the cat map example.

To apply the Hopf argument, we also need a metric on this space. We define the metric in the following manner. The distance between sequences $\{a_n\}$ and $\{b_n\}$ is given by the following expression.

$$d(a, b) = \sum_{i=\infty}^{\infty} \frac{|a_i - b_i|}{2^{|i|}}$$

To draw an analog with the cat map example, we need to describe what would correspond to a leaf of the stable foliation. Consider any sequence $\{a_n\}$. Now consider the set of all sequences $\{b_i\}$ such that $a_i = b_i$ for $i > 0$. It's clear that applying the Bernoulli map on any two elements of this set halves their distance. This is what we'll call the stable set of a. Similarly, the analog of the leaf of an unstable foliation will be the set of all sequences $\{c_n\}$ such that $c_i = a_i$ for $i < 0$. It's clear that applying the inverse of the Bernoulli shift halves the distance between any two pair of points in this set. This is the unstable set of a. It's also not too hard to see that any sequence can be reached from any sequence by going from stable to unstable sets. That means continuous Bernoulli-shift invariant functions are constant, i.e. Bernoulli shift is ergodic.

We can also use a more direct approach in this example. What we can do is show that any positive measure set that is shift invariant has to have measure 1. Suppose we have some positive measure shift invariant set. We can pick a positive measure subset which is a cylinder set, i.e. it's of the following form.

$$\cdots \times \{0, 1\} \times \{0, 1\} \times \{a_i\} \times \cdots \times \{a_{i+n}\} \times \{0, 1\} \times \{0, 1\} \times \cdots$$

We now claim that the shift orbit of such a set has measure 0. An easy way to see that is to see that if some element is not in the shift orbit of this set, the sequence $a_i \dots a_{i+n}$ never appears in the sequence. But it's not too hard to see that the measure of the set of such sequences is 0, which means the orbit has full-measure, i.e. the shift is an ergodic map.

2 Measure theoretic entropy

Consider you have a probability space (X, μ), with finitely many partitions $\{P_i\}$. Suppose you randomly sample a point, and instead of knowing what point you happened to pick, you only know which partition that point lies in. If the partition is just the single partition with measure 1, knowing what partition you landed in gave you no more information than you already had. On the other hand, if you had n partitions, for $n \geqslant 2$, then knowing what partition the point came from tells you a lot. The entropy of a partition quantifies this notion: a partition has high entropy if knowing which element of the partition you land in gives you a lot of information, and a partition has low entropy, if knowing the partition you land in doesn't give you too much information. In our example, the former partition has low entropy (in fact 0), and the latter partition has high entropy (the entropy is $\log n$). This notion was quantified by several people, including Boltzmann, Kolmogorov, and Shannon[1].

Definition 2.1 (Entropy of a partition)**.** The entropy h_μ of a partition $\mathcal{P} = \{P_i\}$ of a probability space (X, μ) is given by the following expression.

$$h_\mu(\mathcal{P}) = \sum_{P_i \in \mathcal{P}} -\mu(P_i) \log(\mu(P_i))$$

Now that we have a notion of entropy for a partition, we can also talk about entropy of a measure-preserving map $T : X \to X$ with respect to a partition. In terms of our previous analogy, we can not only look at what partition the element we sampled lies in, but also the partitions all the elements of its orbit lie in. If we have a partition $\mathcal{P}$, we look at the partition $\mathcal{P}_n T^{-n}\mathcal{P}$, and take the common refinement of all the $\mathcal{P}_n$ (which we'll denote by the symbol $\bigwedge$) for $n \geqslant 0$ (I'm not sure why we're only taking T^{-n} of the partitions). We can then look at the entropy of this partition as n goes to infinity.

[1]Kotak has a great survey on entropy [Kat07].

Definition 2.2 (Entropy of a transformation with respect to a partition)**.** Given a measure preserving transformation T, and a partition $\mathcal{P}$, we define the entropy of T with respect to $\mathcal{P}$ by the following expression.

$$h_\mu(T, P) = \limsup_{n\to\infty} \frac{h_\mu\left(\bigwedge_{i=0}^{n} T^{-i}\mathcal{P}\right)}{n}$$

The measure theoretic entropy of the transformation T is defined by taking the supremum of the above quantity over all possible partitions.

Definition 2.3 (Measure theoretic entropy)**.** The measure theoretic entropy of a transformation T is the supremum of $h_\mu(T, \mathcal{P})$ over all possible partitions $\mathcal{P}$.

$$h_\mu(T) = \sup_{\mathcal{P}} h_\mu(T, \mathcal{P})$$

From the definition, it's not obvious that it's actually feasible to calculate the measure theoretic entropy of a transformation; after all, taking a supremum over all partitions is not easy. It's quite fortunate that doing so is not necessary, thanks to a theorem due to Kolmogorov and Sinai. Before we can state the theorem, we need to define what a generating partition is.

Definition 2.4 (Generating partition)**.** Given a measure preserving transformation T, and a partition $\mathcal{P}$, the partition $\mathcal{P}$ is said to be generating if $\bigwedge_{i=0}^{\infty} T^{-i}\mathcal{P}$ can approximate any element of the Borel sigma algebra within arbitrarily small error.

With generating partitions defined, we can state the Kolmogorov-Sinai theorem.

Theorem 2 (Kolmogorov-Sinai)**.** *The measure theoretic entropy of a transformation* T *can be computed by computing it with respect to a generating partition* $\mathcal{P}$*.*

$$h_\mu(T) = h_\mu(T, \mathcal{P})$$

The utility of this theorem is that in practice, it's not too hard to find generating partitions, and computing entropy with respect to them is fairly easy. The proof of this theorem is reasonably elementary, and has a well written proof in the lecture notes by Will Merry [Mer]. The notes also outline how to find generating partitions when working with probability spaces which are also compact metric spaces.

There's another way to think about the measure theoretic entropy of a partition. Suppose you partition up your probability space into a partition $\{P_1, \ldots, P_k\}$. Think of these partitions as the best possible resolution of the space you're looking at, i.e. the best you can say about any particular point is which element of the partition it lies in. If you have a measure preserving transformation T, and you apply it repeatedly on a point x, all you can see are the partition $T^i x$ lies in. That means applying T to a point n times gives you a length n word in the alphabet $\{P_1, \ldots, P_k\}$. The only way you can distinguish two points in this setting is apply the map T repeatedly, and hope that the orbits end up in different partitions eventually. The means the number of different n-length words one gets is a measure of how well the transformation T mixes things, or its entropy. A way of quantifying that is to compute the following quantity.

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\mu(\mathcal{A}^n(x))\right)$$

Here, $\mathcal{A}^n(x)$ denotes the length n word corresponding to the orbit of x, and the measure on this space is the pullback measure. This seems like a reasonable definition of entropy of the transformation T with respect to the partition $\{P_i\}$, and under some conditions, it's in fact the entropy of the partition, which is what the Shannon-McMillan-Breiman theorem states.

Theorem 3 (Shannon-McMillan-Breiman)**.** *If* T *is a measure preserving ergodic transformation on* X*, and* $\mathcal{P}$ *is a finite partition, then for almost all* $x \in X$*, the following identity holds.*

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\mu(\mathcal{A}^n(x))\right) = h_\mu(T, \mathcal{P})$$

Hochman's notes [Hoc] have a couple of proofs of this theorem; the combinatorial proof is rather elementary, and not too hard to follow, which is what we did in class.

3 Topological Entropy

Fill this in later, from the notes. Also write about the Local entropy formula and the variational principle.

4 Volume Entropy

From all the definitions of entropy we've seen so far, it's clear that it measures the exponential growth rate of some prescribed quantity. The name volume entropy suggests that it is the exponential growth rate of the volume of some set, as a parameter like the diameter varies. It is in fact precisely that, stated in the context of Riemannian manifolds.

Definition 4.1 (Volume Entropy)**.** Let (M, g) be a compact Riemannian manifold, and let $\widetilde{M}$ be its universal cover, and pick a point $\widetilde{x}_0 \in \widetilde{M}$. Then the volume entropy of (M, g) is defined in the following manner.

$$h_V((M, g)) = \lim_{R \to \infty} \frac{1}{R} \log\left(\mathrm{vol}(B_R(\widetilde{x}_0))\right)$$

In this formula, B_R is the ball of radius R around $\widetilde{x}_0$ in $\widetilde{M}$, and the volume is measured with respect to the Riemannian volume form.

It turns out that the volume entropy is actually independent of the choice of basepoint $\widetilde{x}_0$. Suppose we pick another base point $\widetilde{x}_1$, and the distance between the two basepoints is d. By the triangle inequality, we get the following two inequalities, which when plugged into the volume entropy formula show they're equal.

$$\mathrm{vol}(B_R(\widetilde{x}_0)) \leqslant \mathrm{vol}(B_{R+d}(x_1))$$
$$\mathrm{vol}(B_R(\widetilde{x}_1)) \leqslant \mathrm{vol}(B_{R+d}(x_0))$$

We must also show that the limit in the definition actually exists. For this, we'll outline Manning's proof [Man79] of the fact. The first thing to do when showing such results is get a subadditivity inequality, like the following.

$$\log \mathrm{vol}(B_{r+s}) \leqslant \log \mathrm{vol}(B_r) + \log \mathrm{vol}(B_{s+A})$$

Here, S is some constant depending on the size of the fundamental domain, and the curvature. Now that we have subadditivity, we can invoke Fekete's Subadditive Lemma to conclude that the limit exists.

Let's compute the volume entropy of some manifolds.

Example 1 (Volume entropy of $\mathbb{R}^n$)**.** The volume of a ball of radius r in $\mathbb{R}^n$ is a polynomial in r of degree n. That means its exponential growth rate is 0, and hence its volume entropy is 0.

Example 2 (Volume entropy of $\mathbb{H}^2$ with curvature -1)**.** In this case, one looks at the disc model of the space, and can compute the volume of a radius r ball centred at 0, and doing that gives the following relation.

$$\mathrm{vol}(B_r) \sim e^r$$

This means the volume entropy of this space is 1. A similar computation for the hyperbolic n-space tells us that the volume entropy of $\mathbb{H}^n$ is $n-1$.

4.1 Geodesic Flow and Volume Entropy

The main reason why we care about volume entropy is that it is intricately connected with geodesic flow on a compact manifold. We can give a lower bound to the topological entropy of the geodesic flow in terms of the volume entropy, and in certain cases, the lower bound is actually an equality.

Theorem 4 (Manning, 1979 [Man79]). *Let M be a compact Riemannian manifold, and let f be the geodesic flow on SM. Then the topological entropy of the transformation f_1 is bounded below by the volume entropy of M.*

$$h_{top}(f_1) \geqslant h_V(M)$$

Furthermore, if M has all sectional curvatures less than or equal to 0, then we have an equality.

$$h_{top}(f_1) = h_V(M)$$

Sketch of proof. To show that $h_{top}(f_1)$ is greater than or equal to $h_V(M)$, we need to find a maximal δ separated set in SM, and show that it's cardinality is asymptotically greater than $c\exp(h_V(M) - \epsilon)r$ as r goes to infinity. The way we do that is look at an annulus of radii $\left(r, r + \frac{\delta}{2}\right)$ centred at x_0, and find the cardinality of the maximal 2δ separated set. We get a lower bound on this cardinality from the volume entropy, and now we look at the geodesics from x_0 to each point in the maximal set. We show that these geodesics are δ separated in SM, and their cardinality is lower bounded by a function of $h_V(M)$, which gives us the result we want, modulo performing some technical calculations.

The equality in the case of non-positive curvature comes from the fact that if in a space of non-positive curvature we have two geodesics σ_1 and σ_2, the distance between $\sigma_1(t)$ and $\sigma_2(t)$ is bounded above by $d(\sigma_1(s), \sigma_2(s)) + d(\sigma_1(u), \sigma_2(u))$, where $s < t < u$. Using this inequality lets us prove the entropy inequality in the opposite direction. □

5 The Anosov Closing Lemma

5.1 Anosov diffeomorphisms

A special class of diffeomorphisms on a compact smooth Riemannian manifold are Anosov diffeomorphisms.

Definition 5.1 (Anosov diffeomorphism). A diffeomorphism $K : M \to M$ is called an Anosov diffeomorphism if the tangent bundle TM splits as two sub-bundles T^sM (called the stable distribution), and T^uM (called the unstable distribution) such that for some $\lambda \in (0, 1)$ (λ is the hyperbolicity constant), and some positive real number C, the following conditions are satisfied for all $x \in M$.

(i) The map DK sends T^sM to T^sM, and T^uM to T^uM.

(ii) $\|DK^n_x(v^s)\| \leqslant C\lambda^n \|v^s\|$ for $v^s \in T^s_xM$ and $n \geqslant 0$.

(iii) $\|DK^{-n}_x(v^u)\| \leqslant C\lambda^n \|v^u\|$ for $v^u \in T^u_xM$ and $n \geqslant 0$.

The idea behind the definition is to model dynamical systems where there is exponential decay in one direction as time goes forward, and exponential decay in another direction as time goes backward.

A major result of smooth dynamical systems theory is the *Stable Manifold Theorem*, which states that the stable and unstable distributions give rise to foliations, which are called the stable and unstable manifold respectively. Anosov diffeomorphisms also satisfy another nice technical property, which is called the *local product structure*. What that means is that there are small enough constants ϵ and δ such that for all $x, y \in M$, with $d(x, y) < \delta$, $W^s_\epsilon(x) \cap W^u_\epsilon(y)$ consists of exactly one point, and the intersection is transversal, where $W^s_\epsilon(x)$ is the ϵ ball around x in its stable manifold, and $W^u_\epsilon(y)$ is the ϵ ball around y in its unstable manifold. This property of Anosov diffeomorphisms has several consequences, one of which is the Anosov closing lemma. Some of the other consequences, and generalizations of this notion are outlined in [BS02], chapter 5.

5.2 Statement and proof of the Anosov Closing Lemma

The Anosov closing lemma is a remarkable result that shows that any orbit under an Anosov diffeomorphism that approximately closes can be perturbed by a small amount such that the perturbed orbit exactly closes up. Here's the more formal statement.

Theorem 5 (Anosov Closing Lemma). *Let K be an Anosov diffeomorphism (with hyperbolicity constant λ) acting on a manifold M. Then there exists a M such that for any τ smaller than a fixed δ, and for any x such that $d(K^n x, x) < \tau$, there's a y within $M\tau$ distance of x such that $d(K^i x, y) < M\tau$ for $0 \leqslant i \leqslant n$, and $K^n y = y$.*

Proof. Since K is an Anosov diffeomorphism, (M, T) has a local product structure, i.e. there exists a δ such that any δ-ball has a stable and unstable foliations passing through each point that intersect exactly once. Now consider any point x such that for some n, $d(K^n x, x) < \tau < \delta$. Now look at the $W^s_\epsilon(x)$ (the ϵ we get from the definition of local product structure). Since this is the stable foliation, applying K^n on this foliation will give us a ball $W^s_{\lambda^n \epsilon}(K^n x)$ around $K^n x$. The unstable foliation passing through each point of $W^s_{\lambda^n \epsilon}(K^n x)$ intersects $W^s_\epsilon(x)$ at exactly one point. This gives us a continuous function from $W^s_\epsilon(x)$ to itself which is Lipschitz with Lipschitz factor λ^n. This means it has a fixed point x'. Now look at $W^u_\epsilon(K^n x')$. Applying K^{-n} to this foliation gives a subset of $W^u_{\lambda^n \epsilon}(x')$, which can be seen as a subset of $W^u_\epsilon(K^n x')$ since x' was the fixed point of the earlier map. Since this new map is also a contraction, it also has a fixed point x''. This will be our y. It's easy to see because of the way we constructed this map that $K^n y = y$. Furthermore, because we can reach $K^i y$ from $K^i x$ by going along a stable foliation followed by an unstable foliation, $d(K^i x, K^i y) < (\lambda^i + \lambda^{-i}) d(x, y) < 2\lambda^{-n}\delta$. □

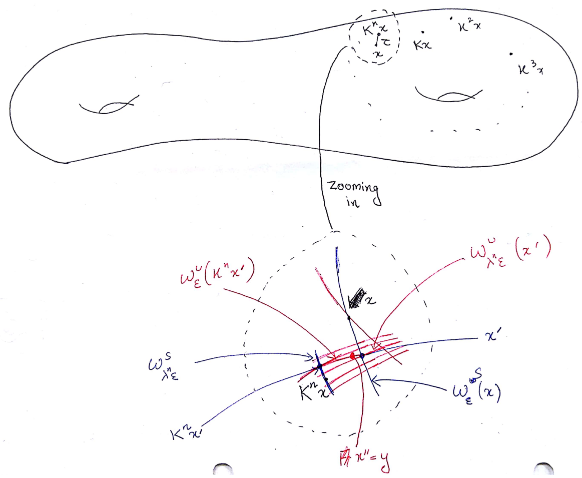

Figure 1: A pictorial representation of what's happening in the proof

In fact, the technique used in proof can be generalized to give a significantly stronger result, due to Bowen, which loosely states that in topologically transitive Anosov systems, any collection of orbit segments can ϵ-approximated by an actual orbit. A more precise formulation can be found in [HP].

6 Lyapunov exponents

Consider the dynamical system consisting of a torus ($\mathbb{R}^n/\mathbb{Z}^n$) and a map $\Gamma \in SL_n(\mathbb{Z})$. This dynamical system is in some sense completely determined by the eigenvalues of the map Γ. Could we try to do this for a more general smooth dynamical system, i.e. find a collection that dictate the behaviour of the system. To be more precise, what was happening in the example of the torus is that the eigenvalues of the matrix Γ dictate how the tangent vectors grow. We can do something similar on a compact Riemannian manifold M with a smooth transformation ϕ acting on it: the thing we get is what is called the Lyapunov exponent of ϕ. It's defined for every vector $v \in T_p M$ for every point $p \in M$.

$$\lambda^+(v) = \limsup_{n\to\infty} \log \frac{1}{n} \frac{\|d\phi^n(v)\|}{\|v\|}$$

This is a generalization of the role eigenvalues played in the case of the linear toral automorphism. For an eigenvector v of Γ corresponding to eigenvalue α, the Lyapunov exponent of v will be $\log|\alpha|$. And extending the analogy between eigenvalues and Lyapunov exponents, for a given tangent space T_pM, there are at most $\dim(M)$ many numbers which can appear as Lyapunov exponents of vectors in the tangent space. That's because if a vector v is in the span of the k vectors with Lyapunov exponents $\lambda_1^+ \geqslant \lambda_2^+ \geqslant \cdots \lambda_k^+$, then the Lyapunov exponent of v will be the highest Lyapunov exponent, i.e. λ_1^+. That means for a given tangent space, we can associate a finite collection of numbers that appear as Lyapunov exponents of vectors in that space, and call that the Lyapunov exponents of that tangent space.

That means we get a function from M to $\mathbb{R}^n$ which sends each point to the ordered collection of the Lyapunov exponents at the tangent space to that point. This function is actually measurable, since it's the limit of a sequence of measurable functions. Furthermore, this function is invariant under the action ϕ (that just follows from the definition of Lyapunov exponent). If the action ϕ is ergodic with respect to a measure μ, that means that this function is constant μ-a.e. In particular, that means that we have a finite collection of Lyapunov exponents that are the same for a measure 1 subset of M. These are what we'll call the Lyapunov exponents of ϕ.

The Lyapunov exponents we have been dealing with so far are the forward time Lyapunov exponents, since we require $n \to \infty$. We can analogously define λ^-, where $n \to -\infty$. The same arguments as before show that for an ergodic ϕ, there's a finite collection of backward-time Lyapunov exponents. The question is, are they the same as forward-time Lyapunov exponents? For the linear toral automorphism we looked at, they are indeed the same. It also holds true in the more general setting, but to show that, we'll need a multiplicative version of the ergodic theorem, to show that the forward-time and backward-time averages are indeed the same.

7 The Multiplicative Ergodic Theorem

To talk about the multiplicative ergodic theorem, we'll need to define the notion of a cocycle.

Definition 7.1 (Cocycles)**.** Given a group G acting on a space X, and another group H, a cocycle is a function $\alpha : G \times X \to H$ which satisfies the following identity for all points in the domain.

$$\alpha(g_1 g_2, x) = \alpha(g_1, g_2 x) \cdot \alpha(g_2, x)$$

Although this may seem very abstract, it can be thought of as a generalization of the chain rule in multivariable analysis.

Example 3 (Derivative cocycle)**.** Let $X = \mathbb{R}^n$, G the group of diffeomorphisms from $\mathbb{R}^n$ to itself, and H be the group $GL(n, \mathbb{R})$. Let α be the map that takes $g \in G$ and $x \in X$ to the map $Dg : T_xX \to T_{g(x)}X$, which can be canonically identified with $GL(n, \mathbb{R})$. Then it follows from the chain rule that the map α is a cocycle.

The above example can be extended to any smooth manifold provided we fix a global frame. Then the derivative cocycle we get is dependent on the choice of frame, and if we change our frame, the new cocycle is related to the old cocycle by multiplication with a section of the $GL(n, \mathbb{R})$-bundle over X.

Here's another example, called the orbit equivalence cocycle, whose importance will become clear later (I hope).

Example 4 (Orbit equivalence cocycle)**.** Let G be a group that acts on spaces X and Y freely. Let ψ be a map from X to Y such that ψ sends G-orbits to G-orbits. Then for any $g \in G$, and $x \in X$, there's a unique $g' \in G$ such that $\psi(g, x) = g' \cdot \psi(x)$. Let α be the map that sends (g, x) to the corresponding g'. Then the map α is a cocycle from $G \times X$ to G.

With the word *cocycle* in our vocabulary, we can state the multiplicative ergodic theorem, also known as Oseledet's theorem. Consider a probability space (X, μ) and let ϕ be a measure preserving ergodic transformation on it. Let α be a $GL(n, \mathbb{R})$ valued cocycle, i.e. α is a measurable map from $\mathbb{Z} \times X$ to $GL(n, \mathbb{R})$ such that the following property holds.

$$\alpha(n + m, x) = \alpha(n, \phi^m x) \cdot \alpha(m, x)$$

We shall also need an integrability condition on α.

$$\int_X \|\alpha(1,x)\| \, d\mu(x) < \infty$$

In this context, we can define the forward and backward Lyapunov exponents of the cocycle α for every vector $v \in \mathbb{R}^n$

$$\lambda^+(x,v) := \lim_{n\to\infty} \frac{1}{n} \log\left(\|\alpha(n,x)(v)\|\right)$$
$$\lambda^-(x,v) := \lim_{n\to-\infty} \frac{1}{n} \log\left(\|\alpha(n,x)(v)\|\right)$$

The multiplicative ergodic theorem states that the forward and backward Lyapunov exponents exist for almost all $x \in X$, and are equal.

Theorem 6 (Multiplicative Ergodic Theorem)**.** *For almost every $x \in X$, there exists subspaces $E_x^{\lambda_i} \subset \mathbb{R}^n$ such that the following holds.*

(i) $\oplus E_x^{\lambda_i} = \mathbb{R}^n$.

(ii) $\alpha(n,x)E_x^{\lambda_i} = E_{\phi^n x}^{\lambda_i}$.

(iii) For all $v \in E_x^{\lambda_i}$, $\lambda^+(x,v) = \lambda_i = \lambda^-(x,v)$.

An important point to note here is that when $n = 1$, the multiplicative ergodic theorem really is just a weaker version of Birkhoff's ergodic theorem. This fact is used in the proof of the multiplicative ergodic theorem. The proof in these notes is from Appendix A of [Mar91].

Proof of Theorem 6. Recall the Iwasawa decomposition of $GL(n,\mathbb{R})$, which states that any $g \in GL(n,\mathbb{R})$ can be written as kan, where $k \in O(n,\mathbb{R})$, a is a diagonal matrix with positive entries, and n is an upper triangular matrix with all the diagonal entries equal to 1. In other words, $GL(n,\mathbb{R}) = KAN$, where K, A, and N are the respective matrix groups.

We shall first prove the result if the cocycle only takes values in A. In that case, ϕ can be written as $\bigoplus_{i=1}^n \phi_i$, where each of the ϕ_i correspond to the value of ϕ in the i^{th} diagonal entry. Each ϕ_i is thus a 1-dimensional cocycle, and we can invoke Birkhoff's ergodic theorem to conclude that the forward and backward Lyapunov exponents for each of them is equal, and the Lyapunov exponents for the cocycle ϕ are the collection of Lyapunov exponents of each of the ϕ_i, repeated with multiplicity. This proves the theorem when the cocycle takes values in A.

Now suppose the cocycle takes values in AN, i.e. its values are upper triangular matrices with positive diagonal entries. We can now consider a quotient cocycle $\overline{\alpha}$ which takes values in $AN/N \cong A$, for which the multiplicative ergodic theorem holds. We claim now that α and $\overline{\alpha}$ have the same Lyapunov exponents. To see that, consider the following expression.

$$\lambda_\alpha^+(x,v) = \lim_{m\to\infty} \frac{1}{m} \log\left(\|\alpha(m,x)(v)\|\right) \tag{1}$$
$$= \lim_{m\to\infty} \frac{1}{m} \log\left(\left\|\alpha(1,\phi^{m-1}x)\cdot\alpha(1,\phi^{m-2}x)\cdots\alpha(1,x)\right\|\right) \tag{2}$$

Each $\alpha(1,\phi^m x)$ can be written as $a_m(1+\eta_m)$, where a_m is a diagonal matrix with positive eigenvalues and η_m is a nilpotent matrix. We can write out the same expression for the quotient cocycle $\overline{\alpha}$, which gives us the following expression.

$$\lambda_{\overline{\alpha}}^+ = \lim_{m\to\infty} \frac{1}{m} \log\left(\left\|\overline{\alpha}(1,\phi^{m-1}x)\cdot\overline{\alpha}(1,\phi^{m-2}x)\cdots\overline{\alpha}(1,x)\right\|\right) \tag{3}$$

Here, each $\overline{\alpha}1, \phi^m x$ turns out to be just the matrix a_m because we quotient away the N subgroup. What we need to show now is that the right hand side of equation (2) and (3) have the same exponential growth rate. We can expand out equation (2) in terms of a_m and η_m.

$$\prod_{i=1}^{m} a_m(1+\eta_m)$$

It turns out any term having more than n η_i's will become 0, which means there are only $\binom{m}{n}$ terms in the above product, and all of their exponential growth is as fast as that of $\overline{\alpha}$, which means α and $\overline{\alpha}$ have the same exponential growth rate (since the polynomial $\binom{m}{n}$ grows only polynomially in m). This proves the theorem for cocycles which take values in AN.

For cycles which take values in $GL^+(n, \mathbb{R})$, the idea is to write this group as $\mathbb{R}_+ \times SL(n, \mathbb{R})$. On the first component, we know the result holds for the first component, so we only need to show it for the second component, which is $G = SL(n, \mathbb{R})$. To do this, we consider a new space $X \times G/AN$, along with the transformation $\widehat{T}$ which does the following to points of $X \times G/AN$.

$$\widehat{T}(x, z) = (Tx, \alpha(1, x)z)$$

We can also define a cocycle $\widehat{\alpha}$ over $X \times G/AN$ in the following manner.

$$\widehat{\alpha}(n, (x, z)) = \alpha(n, x)$$

It turns out that $\widehat{\alpha}$ is cohomologous to a cocycle β taking values in AN. To see this consider a bounded section $\sigma : G/AN \to G$ and set $c(x, d) = \sigma(d)$. Then c is the conjugation function which constructs β from $\widehat{\alpha}$, and the values lie in AN. The fact that the section is bounded ensures β is also integrable. All we need to do now is pick a measure on the skew product $X \times G/AN$ such that it projects down to μ and $\widehat{T}$ is ergodic. Such a measure would be an extreme point in the set of all measures that project down to μ. □

References

[BS02] M. Brin and G. Stuck. *Introduction to Dynamical Systems.* Cambridge University Press, 2002.

[Hoc] Michael Hochman. Lectures on dynamical systems and entropy. URL: `http://math.huji.ac.il/~mhochman/courses/dynamics2014/notes.5.pdf`. Last visited on 2018/09/27.

[HP] B. Hasselblatt and Y. Pesin. Hyperbolic dynamics: Specification and entropy. URL: `http://www.scholarpedia.org/article/Hyperbolic_dynamics#Specification_and_entropy`. Last visited on 2018/10/23.

[Kat07] Anatole Katok. Fifty years of entropy in dynamics: 1958–2007. *J. Mod. Dyn*, 1(4):545–596, 2007.

[Man79] Anthony Manning. Topological entropy for geodesic flows. *Annals of Mathematics*, 110(3):567–573, 1979.

[Mar91] Gregori A Margulis. *Discrete subgroups of semisimple Lie groups*, volume 17. Springer Science & Business Media, 1991.

[Mer] Will J. Merry. Dynamical systems lecture notes: Measure theoretic entrop. URL: `https://www.merry.io/dynamical-systems-lecture-notes/2016/12/18/26-measure-theoretic-entropy`. Last visited on 2018/10/03.