Oseledec's Multiplicative Ergodic Theorem

Jim Kelliher

Initial Version, Fall 2002

This document was essentially complete in February 2003, though significant changes were made in September 2003. It was re-compiled on ***March 28, 2011*** to add active links and to deal with an issue with Lemma 6.1.

These are notes for a talk in the Junior Geometry seminar at UT Austin on Oseledec's multiplicative ergodic theorem given in Fall 2002. The purpose of the notes is to insure that I know, or at least am convinced that I think I know, what I am talking about. They contain far more material than the talks themselves, constituting a complete proof of the discrete-time version of the multiplicative ergodic theorem. Perhaps sometime in the future I will work through the argument required to adapt that proof to the continuous-time version of the theorem.

To motivate the theorem, I start with a discussion of Lyapunov exponents, whose existence follows from an application of the continuous-time multiplicative ergodic theorem to the differential map on the tangent bundle of a compact Riemannian manifold. Since the intended audience for the talk was geometers, I felt this motivation was needed.

I then give a proof of the multiplicative ergodic theorem that closely follows [1], though I have filled in quite a large number of details.

Contents

1. Geometric Setting — 2
2. The Multiplicative Ergodic Theorem — 6
3. Proof of the Discrete-Time Multiplicative Ergodic Theorem — 9
4. The Fundamental Lemma — 12
5. A Bunch of Lemmas — 13
6. Grassman Manifolds — 16
7. Proof of the Fundamental Lemma — 18
8. The Hard Work — 21
9. Sources — 26

1. Geometric Setting

Our setting is a (smooth) compact Riemannian manifold M of dimension d, on which we have a probability measure, ρ—*i.e.*, $\rho(M) = 1$. This is not a talk on measure theory, so it is adequate to think, if we wish, of the measure as being the volume induced by the Riemannian metric on our manifold normalized so that the manifold has unit volume.

Throughout, we let $\varphi : \mathbb{R} \times M \to M$ be a flow on M. We will write $\varphi_t(x)$ or $\varphi(t, x)$, depending on whether we wish to view t as a parameter and x as an argument or to view φ as a function on $\mathbb{R} \times M$. Notice that we are assuming completeness of the trajectories of φ.

We will also assume that φ is measure-preserving, which we can think of as volume-preserving.

We could allow X to be time-varying.

Let us review the definition of a flow. If X is a vector field, which we can view as a map from the manifold to the tangent space at each point, then we would like to solve the differential equation,

$$\begin{aligned} &\frac{d}{dt}(\varphi(t, x)) = X(\varphi(t, x)), \\ &\varphi(0, x) = x. \end{aligned}$$

Only in special circumstances can we do this globally, but because a manifold is locally diffeomorphic to $\mathbb{R}^n$, we can always do it locally.

In any case, we will do nothing with the vector field X, and just take the flow φ as given. For each value of t, then, φ_t is a diffeomorphism. Because φ is measure-preserving, $\det d\varphi_t = 1$ for all t or $\det d\varphi_t = -1$ for all t, $\det d\varphi_t$ being the Jacobian which appears in the change of variables formula for a volume integral. But φ_0 is the identity, so $\det d\varphi_t = 1$ for all t.

A very intuitive physical model is the flow of an incompressible fluid (whether perfect or not). In fact, in a sense, this is the only physical model.

An example of a measure-preserving flow on a compact manifold is the geodesic flow on the unit tangent bundle of a compact d-dimensional manifold. The unit tangent bundle is compact because the manifold is compact and S^{d-1} is compact. (The geodesic flow is on the unit tangent bundle to the manifold, not on the manifold itself. The flow can also be viewed, though, as being on the tangent bundle, which is not compact, and so does not serve as an example.)

Now associate to a flow φ a *cocycle* T, which we define as follows:

Definition. Let $\pi : E \to M$ be a vector bundle over M where $\pi^{-1}(x) \simeq \mathbb{R}^m$ for all $x \in M$. Let $\{T^t\}$ be a family of bundle maps from E to E parameterized by time $t \in \mathbb{R}$ such that the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{T^t} & E \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow[\varphi_t]{} & M \end{array}$$

Geometric Setting

Write T_x^t for the map from the fiber over x to the fiber over $\varphi_t(x)$. If we look at local trivializations about x and about $\varphi_t(x)$, we can write, for $u \in T_xM$,

$$\begin{aligned} T_x^t &: \pi^{-1}(x) \to \pi^{-1}(\varphi_t(x)), \\ &(x,u) \mapsto (\varphi_t(x), T_x^t u). \end{aligned}$$

Here we are using T_x^t both for the map on the fiber and for the map on the vector spaces in the local trivializations. It should be clear from context which map we are referring to.

This is a convenience that has a dark side.

The collection $\{T^t\}_{t\in\mathbb{R}}$ is called a (linear skew-product) cocyle over φ if the properties we list below are satisfied. Observe, though, that the first component of T_x^t as a fiber map is $\varphi_t(x)$, which satisfies all these properties as well, so the properties apply to both our views of T_x^t. These properties are:

(1) $T_x^0 =$ identity for all $x \in M$.

(2) $T_x^{s+t} = T_{\varphi_s(x)}^t T_x^s$ for all $s, t \geq 0$.

(3) $T : \mathbb{R} \times E \to E$ is measurable.

Comment: We will need to assume that the local trivializations in our vector bundle are isometries on fibers—that is, preserve length and so the inner product. The inner product on the fibers is derived from the Riemannian metric.

The prototypical example of a cocycle, and the one we will use to illustrate the geometric meaning of the multiplicative ergodic theorem, is the differential map of the tangent bundle to itself, $T^t := d\varphi_t$, where

$$\begin{aligned} &d\varphi_t : TM \to TM, \\ &d\varphi_t(x,u) = (\varphi_t(x), (d\varphi_t)_x u). \end{aligned}$$

Then

$$T_x^t = (d\varphi_t)_x,$$

the Jacobian of the diffeomorphism φ_t. $\{T^t\}$ satisfies (1) because φ_0 is the identity and thus so is its Jacobian. It satisfies (3) because of the smoothness of φ. It satisfies the critical defining condition, (2), because of the chain rule for differentials:

$$\begin{aligned} T_{\varphi(s,x)}^t T_x^s &= (d\varphi_t)_{\varphi(s,x)} (d\varphi_s)_x \\ &= (d(\varphi_t \circ \varphi_s))_x \\ &= (d\varphi_{s+t})_x \\ &= T_x^{s+t}. \end{aligned}$$

Because the flow is measure-preserving and is the identity at time zero, $\det T_x^t = 1$ for all t and x.

Consider the discrete-time version of this example with a time step of 1. Let x and y be two nearby points in M sharing a common coordinate chart, (U, ψ), where U is an open subset of M and $\psi : U \to \mathbb{R}^m$. If $u = \psi(x) - \psi(y)$, then $(D\varphi_1)_x u$ is a vector $T_{\varphi_1(x)}M$ that measures the approximate displacement of $\varphi_1(y)$ from $\varphi_1(x)$ in a chart around $\varphi_1(x)$. $(d\varphi_n)_x u = (D\varphi_1^n)_x u$, then, approximates the displacement after time n, and $\|(d\varphi_n)_x u\|$ measures how far the two nearby points x and y have moved apart—using the inner product, on $T_x M$ in all cases, induced from the Riemannian metric—after time 1. Of course, as n increases, we will have to insure that y is closer and closer to x to insure an accurate approximation.

To try to understand how $\|(d\varphi_n)_x u\|$ might vary with time, we consider the simplest possible example in which $(D\varphi_1)_x$ is equal to the constant matrix A for all x (and, necessarily, $\det A = 1$). Then

$$\begin{aligned}(d\varphi_n)_x u &= (d\varphi_1^n)_x u \\ &= (d\varphi_1)_{\varphi_{n-1}(x)} (d\varphi_1)_{\varphi_{n-2}(x)} \cdots (d\varphi_1)_x u \\ &= AA \cdots Au = A^n u.\end{aligned}$$

Suppose u is an eigenvector of A with corresponding eigenvalue, μ. Then $(d\varphi_n)_x u = \mu^n u$, so

$$\|(d\varphi_n)_x u\| = |\mu|^n \|u\|.$$

At every time step, the vector expands (or contracts) by a factor of $|\mu|$; or, equivalently, the vector has a constant relative rate of expansion of $|\mu|$. (It is a rate because the time unit is 1.)

In general, we will not have a constant rate of expansion. Even in this simple example where $(d\varphi_1)_x = A$, the rate is not constant in directions other than those of the eigenvectors. And when $(d\varphi_n)_x$ is not constant over space, we would not expect any direction to show constant expansion (though it is still possible). What we are interested in knowing is the mean, long-term rate of expansion. The mean we will choose to use is the geometric mean, simply because it is more workable in this context.

The arithmetic mean is bounded below by the geometric mean.

The geometric mean rate of expansion after n time steps is given by

$$\begin{aligned}&\left(\frac{\|(d\varphi_1)_x u\|}{\|u\|} \frac{\|(d\varphi_2)_x u\|}{\|(d\varphi_1)_x u\|} \cdots \frac{\|(d\varphi_n)_x u\|}{\|(d\varphi_{n-1})_x u\|} \right)^{1/n} \\ &\quad = \left(\frac{\|(d\varphi_n)_x u\|}{\|u\|} \right)^{1/n}.\end{aligned}$$

Because $(d\varphi_n)_x$ is always nonsingular, we never get division by zero.

Geometric Setting

It is actually the logarithm of the geometric mean rate of expansion (or, equivalently, the arithmetic mean of the logarithms), in the limit as n approaches infinity, that is traditionally used:

$$\begin{aligned}\lambda_u := \lim_{n\to\infty} \log \left[\left(\frac{\|(d\varphi_n)_x u\|}{\|u\|} \right)^{1/n} \right] \\ = \lim_{n\to\infty} \frac{1}{n} \left(\log \|(d\varphi_n)_x u\| - \log \|u\| \right) \\ = \lim_{n\to\infty} \frac{1}{n} \log \|(d\varphi_n)_x u\| .\end{aligned}$$

These limits are called Lyapunov, or characteristic, exponents. Notice that they depend upon the direction of u but not upon its length. (This is true even before taking the limit since $(d\varphi_n)_x$ is linear.)

We have been taking limits over n an integer, but we would get the same result if we took limits over t a positive real number. For if $t \in \mathbb{R}$, write $t = n + s$ with $0 \leq s < 1$, and again by the chain rule,

$$(d\varphi_t)_x = (d\varphi_s)_{\varphi_n(x)} (d\varphi_n)_x u.$$

But $(d\varphi_s)_{\varphi_n(x)}$ is a linear operator with a norm uniformly bounded in the interval $s \in [0, 1]$, so

$$\lim_{n\to\infty} \frac{1}{n} \log \|(d\varphi_n)_x u\| = \lim_{t\to\infty} \frac{1}{t} \log \|(d\varphi_t)_x u\| ,$$

the first limit being over the integers, the second over the reals, so the same limit defines the average rate of growth for both discrete and continuous time.

We have no good reason to expect that these limits exist, but the multiplicative ergodic theorem will tell us that they do almost everywhere. It also gives us some information about the limits.

2. The Multiplicative Ergodic Theorem

Theorem 2.1 (Discrete-time Multiplicative Ergodic Theorem)**.** *Let T be a measurable function from M to the space of all real $m \times m$ matrices, such that*

$$\log^+ \|T(\cdot)\| \in L^1(M, \rho).$$

Let $\tau : M \to M$ be a measure-preserving map and let

$$T_x^n = T_{\tau^{n-1}(x)} \cdots T_{\tau(x)} T_x.$$

Then there is a $\Gamma \subseteq M$ with $\rho(\Gamma) = 1$ and such that $T(\Gamma) \subseteq \Gamma$, and the following holds for all $x \in \Gamma$:

(1) $\Lambda_x := \lim_{n\to\infty}((T_x^n)^* T_x^n)^{1/2n}$ *exists.*
(2) *Let* $\exp \lambda_x^{(1)} < \cdots < \exp \lambda_x^{(s)}$ *be the eigenvalues of Λ_x, where $s = s(x)$, the $\lambda_x^{(r)}$ are real, and $\lambda_x^{(1)}$ can be $-\infty$, and $U_x^{(1)}, \dots, U_x^{(s)}$ the corresponding eigenspaces. Let $m_x^{(r)} = \dim U_x^{(r)}$. The functions $x \mapsto \lambda_x^{(r)}$ and $x \mapsto m_x^{(r)}$ are τ-invariant. Let $V_x^{(0)} = \{0\}$ and $V_x^{(r)} = U_x^{(1)} \oplus \cdots \oplus U_x^{(r)}$ for $r = 1, \dots, s$. Then for $u \in V_x^{(r)} \setminus V_x^{(r-1)}$, $1 \le r \le s$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \|T_x^n u\| = \lambda_x^{(r)}.$$

Comment: The norm on matrices we use in this theorem and throughout is the operator norm in Euclidean space, which is identical in value to the spectral norm, as we will show in Lemma 5.2. A critical property of this norm is that it is a true matrix norm—that is, it is submultiplicative ($\|AB\| \le \|A\|\,\|B\|$).

Comment: By τ-invariant, we mean that a function f on M satisfies $f(\tau(x)) = f(x)$. Applying this relation repeatedly, it follows that the function is constant on the forward orbit of the point x under τ. If τ is invertible, then the function is the same on the entire orbit, forward and backward.

We prove Theorem 2.1 in the next section, but our real goal is the continuous-time version of the theorem, the proof of which requires an adaptation to flows and cocycles of the proof of the discrete-time version. This is not worked out in Ruelle, though he says it is "easily adapted," and I have not worked it out. But the statement of the theorem is as follows:

Theorem 2.2 (Continuous-time Multiplicative Ergodic Theorem)**.** *Let T be a cocyle over the measure-preserving flow φ on a compact manifold M as* *we described above, and assume that the functions*

$$g = \left(x \mapsto \sup_{0\le t\le 1} \log^+ \left\|T_x^t\right\|\right) \text{ and}$$

$$h = \left(x \mapsto \sup_{0\le t\le 1} \log^+ \left\|T_{\varphi(t,x)}^{1-t}\right\|\right)$$

Our prototypical example satisfies these conditions because of the compactness of M and of $[0, 1]$

are . in $L^1(M, \rho)$, where the norm on V_x is the operator norm.

Then there is a $\Gamma \subseteq M$ with $\rho(\Gamma) = 1$ and such that $\varphi_t\Gamma \subseteq \Gamma$ for all $t \ge 0$, and the following holds for all $x \in \Gamma$:

Ruelle never states $\rho(\Gamma) = 1$, but I believe that is just an oversight.

(1) $\Lambda_x := \lim_{t\to\infty}((T_x^t)^* T_x^t)^{1/2t}$ *exists (* is the adjoint operator).*

(2) *Let $\exp\lambda_x^{(1)} < \cdots < \exp\lambda_x^{(s)}$ be the eigenvalues of Λ_x, where $s = s(x)$, the $\lambda_x^{(r)}$ are real, and $\lambda_x^{(1)}$ can be $-\infty$, and $U_x^{(1)}, \ldots, U_x^{(s)}$ the corresponding eigenspaces. Let $m_x^{(r)} = \dim U_x^{(r)}$. The functions $x \mapsto \lambda_x^{(r)}$ and $x \mapsto m_x^{(r)}$ are φ_t-invariant (for all t). Let $V_x^{(0)} = \{0\}$ and $V_x^{(r)} = U_x^{(1)} \oplus \cdots \oplus U_x^{(r)}$ for $r = 1, \ldots, s$. Then for $u \in V_x^{(r)} \setminus V_x^{(r-1)}$, $1 \le r \le s$,*

$$\lim_{t\to\infty} \frac{1}{t} \log \left\|T_x^t u\right\| = \lambda_x^{(r)}.$$

Comment: The subspaces $\{V_x^{(r)}\}_{r=0}^s$ are nested as

$$\{0\} = V_x^{(0)} \subseteq \cdots \subseteq V_x^{(s)} = \mathbb{R}^m,$$

forming what is called a filtration. Because of Theorem 2.2 and the ordering of the eigenvalues,

$$V_x^{(r)} = \{u \in \mathbb{R}^m : \lim_{t\to\infty} \frac{1}{t} \log \left\|T_x^t u\right\| \le \lambda_x^{(r)}\},$$

for $r = 1, \ldots, s$.

Comment: As a corollary, if T_x^t is invertible for all t and x, then

$$T_x^t V_x^{(r)} = V_{\varphi(t,x)}^{(r)},$$

This is essentially Remark 1.8 page 34 of [1]

for all r, t, x. It does not follow, however, that

$$T_x^t U_x^{(r)} = U_{\varphi(t,x)}^{(r)},$$

for all r, t, x.

Comment: If φ is ergodic (so the only measurable subsets of M that are mapped into themselves by $\varphi_t(x)$ for almost all t and x are M and the empty set) then the functions $s(x)$ and $\lambda_x^{(r)}$ are constant almost everywhere.

Comment: When $m = 1$ the matrices are real numbers and

$$\Lambda_x = \lim_{t\to\infty}((T_x^t)^* T_x^t)^{1/2t} = \lim_{t\to\infty}(\left|T_x^t\right|)^{1/t},$$

so

$$\log \Lambda_x = \lim_{t\to\infty} \frac{1}{t} \log T_x^t,$$

which exists by the "ordinary" (Birkhoff's) ergodic theorem. Thus, the multiplicative ergodic theorem is a generalization of the ordinary ergodic theorem.

Comment: The measure that we use to define "almost everywhere" needn't be Lebesgue measure—we could apply the theorem to a flow that preserves some other measure while not preserving volume. Thus, we could in principle start with a non-volume preserving flow, find a measure that preserves it, and then say something about the long-term behavior of the flow. Unfortunately, it is usually hard to determine a nontrivial measure that is preserved by a flow, and the measure will often be singular with respect to the volume (Lebesgue) measure, and thus possibly of less physical or geometric interest. This kind of thing is done, however, in dynamical systems.

Comment: When applied to our prototypical example, the $\{\lambda_x^{(r)}\}$ correspond to the limits we defined earlier as λ_u and are called *Lyapunov* (variants, *Liapunov*, *Ljapunov*) or *characteristic exponents*.

Comment: Ruelle states that any norm on the tangent spaces in part (2) of the theorem will produce the same result. I haven't chased down the details, but this makes some intuitive sense since a change in norm should (more-or-less) only introduce a multiplicative constant, which will become an additive constant after taking the log and thus average to zero. This also means that the conclusion of the theorem is independent of our choice of Riemannian metric. This I need to think about.

Comment: If we want the measure ρ to be volume on a Riemannian manifold then the manifold must be compact to be able to normalize the volume to 1 as required. As we observed above, compactness also insures that the conditions on the functions g and h in the statement of Theorem 2.2 hold, though compactness is sufficient but not necessary.

If the measure is a probability measure other than volume, I do not believe that the manifold needs to be compact. In [4], compactness is not assumed in this theorem, and the proof follows along the same lines.

3. Proof of the Discrete-Time Multiplicative Ergodic Theorem

Our proof of Theorem 2.1 uses the following extension of the classical (Birkhoff's) ergodic theorem made by Kingman in 1968.

Theorem 3.1 (Kingman's Subadditive Ergodic Theorem, 1968)**.** *Let $\tau : M \to M$ be a measurable map preserving ρ-measure, and let $\{f_n\}_{n>0}$ be a sequence of measurable functions, $f_n : M \to \mathbb{R} \cup \{-\infty\}$, satisfying the conditions:*

1. *integrability: $f_1^+ \in L^1(M, \rho)$;*
2. *subadditivity: $f_{k+n} \le f_k + f_n \circ \tau^k$ almost everywhere.*

Then there exists a τ-invariant measurable function $f : M \to \mathbb{R} \cup \{-\infty\}$ such that

(a) $f^+ \in L^1(M, \rho)$,

(b) $\lim_{n\to\infty} \frac{1}{n} f_n = f$ *a.e., and*

(c) $\lim_{n\to\infty} \frac{1}{n} \int f_n(x)\, d\rho = \inf_n \frac{1}{n} \int f_n(x)\, d\rho = \int f(x)\, d\rho.$

Comment: Birkhoff's ergodic theorem can be seen as a special case of Kingman's theorem by applying Kingman's theorem to $f_n(x) = g(x) + g(\tau(x)) + \cdots + g(\tau^{n-1}(x))$, where g is a measurable function in $L^1(M, \rho)$. The main conclusion (which we use below) is that if $g_k(x) = g(\tau^k(x))$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g_k(x)$$

exists and is finite for almost all x.

We will use the following corollary of Theorem 3.1, which is a version of the Furstenberg-Keston theorem (which historically came before Kingman's subadditive theorem).

Corollary 3.2 (Furstenberg-Keston, 1960)**.** *Let $\tau : M \to M$ be a measurable map preserving ρ-measure, and let T be a measurable function from M to the space of all $m \times m$ real matrices such that $\log^+ \|T(\cdot)\| \in L^1(M, \rho)$. Let*

$$T_x^n = T_{\tau^{n-1}(x)} \cdots T_{\tau(x)} T_x,$$

as in Theorem 2.1.

Then there exists a τ-invariant measurable function $\chi : M \to \mathbb{R} \cup \{-\infty\}$ such that

(a) $\chi^+ \in L^1(M, \rho)$,

(b) $\lim_{n\to\infty} \frac{1}{n} \log \|T_x^n\| = \chi(x)$ *for almost all* x, *and*

(c) $\lim_{n\to\infty} \frac{1}{n} \int \log \|T_x^n\|\, d\rho = \inf_n \frac{1}{n} \int \log \|T_x^n\|\, d\rho = \int \chi(x)\, d\rho.$

Proof: Let $f_n(x) = \log \|T_x^n\|$. Then

$$\begin{aligned} f_{k+n}(x) &= \log \|T_x^{k+n}\| \\ &= \log \|T_{\tau^{k+n-1}(x)} \cdots T_{\tau^k(x)} T_{\tau^{k-1}(x)} \cdots T_{\tau(x)} T_x\| \\ &\leq \log \Big(\|T_{\tau^{k+n-1}(x)} \cdots T_{\tau^k(x)}\| \|T_{\tau^{k-1}(x)} \cdots T_{\tau(x)} T_x\| \Big) \\ &= \log \|T_{\tau^{k+n-1}(x)} \cdots T_{\tau^k(x)}\| + \log \|T_{\tau^{k-1}(x)} \cdots T_{\tau(x)} T_x\| \\ &= \log \|T^n_{\tau^k(x)}\| + \log \|T_x^k\| \\ &= f_k + f_n \circ \tau^k, \end{aligned}$$

where we used the fact that the spectral (operator) norm is a matrix norm—$\|AB\| \leq \|A\|\|B\|$ for all A, B. The corollary then follows immediately from applying Theorem 3.1. □

Definition (Exterior power of a matrix (or operator))**.** Let A be an $m \times m$ real matrix, let $1 \leq q \leq m$, and let $\{v_1, \dots, v_m\}$ be a basis for $\mathbb{R}^m$ endowed with the Euclidean metric. Then

$$\{v_{i_1} \wedge \cdots \wedge v_{i_q} : i_1 < \cdots < i_q\}$$

is a basis for $\wedge^q \mathbb{R}^m$. We define $A^{\wedge q}$ to be a real $\binom{m}{q} \times \binom{m}{q}$ matrix whose value is defined by its action on each basis vector of $\wedge^q \mathbb{R}^m$ as follows:

This action defines $A^{\wedge q}$ as a linear map from $\wedge^q \mathbb{R}^m$ to $\wedge^q \mathbb{R}^m$.

$$A^{\wedge q}(v_{i_1} \wedge \cdots \wedge v_{i_q}) = Av_{i_1} \wedge \cdots \wedge Av_{i_q}.$$

By linearity it follows that

$$A^{\wedge q}(x_1 \wedge \cdots \wedge x_q) = Ax_1 \wedge \cdots \wedge Ax_q$$

for any vectors $x_1, \dots, x_q \in \mathbb{R}^m$.

$\wedge^q \mathbb{R}^m$ is called the q-fold exterior power of $\mathbb{R}^m$ and $A^{\wedge q}$ the q-fold exterior power of A.

Lemma 3.3. *For any A, B real $m \times m$ matrices and $c \in \mathbb{R}$,*

$$\begin{aligned} (AB)^{\wedge q} &= (A^{\wedge q})(B^{\wedge q}), \\ (A^{\wedge q})^{-1} &= (A^{-1})^{\wedge q}, \\ (cA)^{\wedge q} &= c^q A^{\wedge q}. \end{aligned}$$

Proof of the Discrete-Time Multiplicative Ergodic Theorem

Proof: The first and third properties are almost immediate from the definition of $A^{\wedge q}$. The second property follows from showing that $(A^{-1})^{\wedge q}$ acts on a basic element of $\wedge^q \mathbb{R}^m$ the same as $(A^{\wedge q})^{-1}$ does. $\square$

We can also define an inner product on $\wedge^q \mathbb{R}^m$ by

$$\langle u_1 \wedge \cdots \wedge u_q, w_1 \wedge \cdots \wedge w_q \rangle = \det(\langle u_i, w_j \rangle)_{i,j=1}^q.$$

Properties such as orthogonality of matrices are preserved under the operator $\wedge^q : \mathbb{R}^m \to \wedge^q \mathbb{R}^m$. This operator is discussed at greater length in [4] p. 118-120.

$A^{\wedge q}$ is written $\wedge^q A$ in [4].

We are now in a position to give the proof of the discrete-time multiplicative ergodic theorem, although we refer to a key result, which we call the "Fundamental Lemma," that will take all of the remaining sections to prove.

Proof of Theorem 2.1: Let $f_n = \log^+ \|T(\tau^{n-1}x)\|$, $n = 1, 2, \ldots$. By assumption, $f_n \in L^1(M, \rho)$, so by Birkhoff's ergodic theorem (see the comment following Theorem 3.1), there is a measurable function f on M such that

$$\frac{1}{n} \sum_{k=1}^{n} f_k(x) \to f(x)$$

for x in some $\Gamma_1 \subseteq M$ such that $\tau\Gamma_1 \subseteq \Gamma_1$ and $\rho(\Gamma_1) = 1$. But,

$$\frac{1}{n} \sum_{k=1}^{n} f_k(x) = \left(\frac{n-1}{n}\right) \left(\frac{1}{n-1}\right) \sum_{k=1}^{n-1} f_k(x) + \frac{1}{n} f_n(x).$$

Taking the limit as $n \to \infty$ of both sides, we conclude that

$$\frac{1}{n} f_n(x) = \frac{1}{n} \log^+ \|T(\tau^{n-1}x)\| \to 0$$

for all $x \in \Gamma_1$.

By Lemma 3.3,

$$\begin{aligned}(T^{\wedge q})_x^n &= T^{\wedge q}_{\tau^{n-1}(x)} \cdots T^{\wedge q}_{\tau(x)} T^{\wedge q}_x = (T_{\tau^{n-1}(x)} \cdots T_{\tau(x)} T_x)^{\wedge q} \\ &= (T_x^n)^{\wedge q},\end{aligned}$$

so we can apply Corollary 3.2 to $T^{\wedge q}$ to conclude that there is also a $\Gamma_2 \subseteq M$ such that $\tau\Gamma_2 \subseteq \Gamma_2$ and $\rho(\Gamma_2) = 1$, and, for $q = 1, \ldots, m$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|(T_x^n)^{\wedge q}\| = \lim_{n\to\infty} \frac{1}{n} \log \|(T^{\wedge q})_x^n\|$$

exists and is a τ-invariant function of x.

Let $\Gamma = \Gamma_1 \cap \Gamma_2$. Then Theorem 2.1 follows from the Fundamental Lemma—Lemma 4.1—applied to $T_n = T(\tau^{n-1}x)$. $\square$

4. The Fundamental Lemma

The following lemma is the key to the proof of the multiplicative ergodic theorem. We state the lemma now, and in the remaining sections prove it.

This is Proposition 1.3 of [1].

Lemma 4.1 (Fundamental Lemma)**.** *Let* $\{T_n\}_{n>0}$ *be a sequence of real* $m \times m$ *matrices such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_n\| \le 0.$$

We write

$$T^n = T_n \cdots T_2 T_1,$$

and assume that the limits,

$(T^n)^{\wedge q}$ is defined in the previous section.

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|(T^n)^{\wedge q}\right\|,$$

exist for $q = 1, \ldots, m$*. Then:*

(1) $\Lambda := \lim_{n\to\infty}((T^n)^* T^n)^{1/2n}$ *exists (* is matrix transposition).*
(2) *Let* $\exp\lambda^{(1)} < \cdots < \exp\lambda^{(s)}$ *be the eigenvalues of* Λ*, where the* $\lambda^{(r)}$ *are real and* $\lambda^{(1)}$ *can be* $-\infty$*, and* $U^{(1)}, \ldots, U^{(s)}$ *the corresponding eigenspaces. Let* $V_x^{(0)} = \{0\}$ *and* $V_x^{(r)} = U_x^{(1)} \oplus \cdots \oplus U_x^{(r)}$ *for* $r = 1, \ldots, s$*. Then for* $u \in V_x^{(r)} \setminus V_x^{(r-1)}$*,* $1 \le r \le s$*,*

$$\lim_{n\to\infty} \frac{1}{n} \log \|T^n u\| = \lambda^{(r)}.$$

The proof of Lemma 4.1 is long and hard, and makes use of a series of lemmas, which we present in the next section. In the section following, we define a metric on Grassman manifolds, which we use in the proof of Lemma 4.1. Last comes the "hard work" (to quote [4]), which appears in a lemma that we state only after the proof of Lemma 4.1, since the lemma only makes sense in the context of that proof.

5. A Bunch of Lemmas

Most of the following lemmas are standard results with which the reader may already be familiar. We include them because the author was only vaguely aware of them before he discerned the need for some such results in trying to figure out what was going on in [1].

Lemma 5.1. *Let A be an $m \times m$ real matrix, n a positive integer. Then $(A^*A)^{1/k}$ exists for any positive integer k and is positive-definite and self-adjoint.*

Proof: A^*A is self-adjoint since $(A^*A)^* = A^*A$. It is also positive-definite, because if v is an eigenvector of A^*A with eigenvalue μ, then

$$\mu = \langle A^*Av, v\rangle = \langle Av, Av\rangle \geq 0.$$

By the spectral theorem, which states that any self-adjoint matrix has an orthonormal basis (consisting of eigenvectors of the matrix), there is a unitary matrix U (unitary means $U^* = U^{-1}$) and diagonal matrix D such that $A^*A = UDU^{-1}$. Since the diagonal of D contains the eigenvalues of A^*A, which are nonnegative, $\sqrt{D}$ exists and is a diagonal matrix whose diagonal entries are the non-negative square roots of those of A^*A. Then

Since A^*A is real, U can be assumed to be orthogonal.

$$(U\sqrt{D}U^{-1})^2 = U\sqrt{D}U^{-1}U\sqrt{D}U^{-1} = UDU^{-1} = A^*A,$$

so $U\sqrt{D}U^{-1}$ provides the square root of A^*A. This argument extends to any k-th root, so $UD^{1/k}U^{-1} = (A^*A)^{1/k}$.

Also, $(A^*A)^{1/k}$ is self-adjoint, since

$$\begin{aligned}((A^*A)^{1/k})^* &= (UD^{1/k}U^{-1})^* = (U^{-1})^*(D^{1/k})^*U^* \\ &= (U^*)^*(D^{1/k})U^{-1} = UD^{1/k}U^{-1} = (A^*A)^{1/k},\end{aligned}$$

where we used the fact that $D^{1/k}$ is diagonal and real and so self-adjoint. $D^{1/k}$ also contains only nonnegative values along the diagonal and so all the eigenvalues of $(A^*A)^{1/k}$ are non-negative. That is, $(A^*A)^{1/k}$ is self-adjoint and positive-definite. □

Lemma 5.2. *Let A be an $m \times m$ real matrix, which we view as a linear map from $\mathbb{R}^m$ to itself endowed with the Euclidean metric. Then $\|A\|$ is equal to the largest eigenvalue of $\sqrt{A^*A}$ (the so-called spectral norm).*

Comment: Remember that we have defined the norm to be the operator norm, $\sup_{\|v\|=1} \|Av\|$, where each norm in the supremum is assumed to be the Euclidean norm.

Proof: Let v be any vector in $\mathbb{R}^m$. Then

$$\begin{aligned}\|Av\|^2 = \langle Av, Av\rangle = \langle A^*Av, v\rangle &= \left\langle \sqrt{A^*A}\sqrt{A^*A}v, v\right\rangle \\ &= \left\langle \sqrt{A^*A}v, \sqrt{A^*A}v\right\rangle = \|\sqrt{A^*A}v\|^2,\end{aligned}$$

so $\|Av\| = \|\sqrt{A*A}v\|$, where we used Lemma 5.1 to conclude that $\sqrt{A^*A}$ exists and is self-adjoint. Therefore,

$$\|Av\| = \sup_{\|v\|=1} \|Av\| = \sup_{\|v\|=1} \|\sqrt{A^*A}v\|.$$

Since $\sqrt{A^*A}$ is self-adjoint, $\sqrt{A^*A} = UDU^{-1}$ for some orthogonal matrix, U, where D is diagonal and contains the eigenvalues of $\sqrt{A^*A}$ along its diagonal. Since an orthogonal matrix preserves norms, $\|\sqrt{A^*A}v\| = \|Dv\|$ for all $v \in \mathbb{R}^m$. But D is diagonal, so it is easy to see that $\sup_{\|v\|=1}\|Dv\|$ occurs for the eigenvector corresponding to the largest eigenvalue of $\sqrt{A^*A}$, which is also the value of the supremum. (Because $\sqrt{A^*A}$ is positive-definite, this eigenvalue is positive.) □

We will use the following corollary of Lemma 5.2 :

Corollary 5.3. *Let A be an $m \times m$ real matrix. Then*

$$\|A\| = \|\sqrt{A^*A}\|.$$

Proof: By Lemma 5.2, $\|A\|$ equals the largest eigenvalue of $\sqrt{A^*A}$. By the same lemma, $\|\sqrt{A^*A}\|$ equals the largest eigenvalue of $\sqrt{(\sqrt{A^*A})^*\sqrt{A^*A}}$. But by Lemma 5.1, $\sqrt{A^*A}$ is self-adjoint, so

$$\sqrt{(\sqrt{A^*A})^*\sqrt{A^*A}} = \sqrt{\sqrt{A^*A}\sqrt{A^*A}} = \sqrt{A^*A},$$

so, in fact, these two norms are equal. □

Lemma 5.4. *Let $\{x_1, \ldots, x_m\}$ be a complete set of eigenvectors of A with corresponding, not necessarily distinct, eigenvalues, $\{\mu_1, \ldots, \mu_m\}$. Then $\left\{x_{i_1} \wedge \cdots \wedge x_{i_q} : i_1 < \cdots < i_q\right\}$ is a complete set of eigenvectors for $A^{\wedge q}$ with eigenvalues*

$$\left\{\mu_{i_1} \cdots \mu_{i_q} : i_1 < \cdots < i_q\right\}.$$

Proof: This follows from

$$\begin{aligned}A^{\wedge q}(x_{i_1} \wedge \cdots \wedge x_{i_q}) &= Ax_1 \wedge \cdots \wedge Ax_q \\ &= \mu_{i_1}x_{i_1} \wedge \cdots \wedge \mu_{i_q}x_{i_q} \\ &= \mu_{i_1} \cdots \mu_{i_q}(x_{i_1} \wedge \cdots \wedge x_{i_q}).\end{aligned}$$

□

A Bunch of Lemmas

Lemma 5.5. *Let A be an $m \times m$ real matrix and $q \in [1, m]$ be an integer. Then*

$$\left\| A^{\wedge q} \right\| = \sigma_m \cdots \sigma_{m-q+1},$$

where $\sigma_m, \ldots, \sigma_1$ are the eigenvalues of $\sqrt{A^*A}$ arranged so that $\sigma_1 \leq \cdots \sigma_m$.

Proof: Let $x_1, \ldots, x_q$ be any vectors in $\mathbb{R}^m$. Then

$$\begin{aligned}(\sqrt{A^*A})^{\wedge q}(\sqrt{A^*A})^{\wedge q} & [x_1 \wedge \cdots \wedge x_q] \\ &= (\sqrt{A^*A})^{\wedge q} \left[\sqrt{A^*A}x_1 \wedge \cdots \wedge \sqrt{A^*A}x_q \right] \\ &= A^*Ax_1 \wedge \cdots \wedge A^*Ax_q \\ &= A^*A \left[x_1 \wedge \cdots \wedge x_q \right].\end{aligned}$$

This means that $(\sqrt{A^*A})^{\wedge q}$ acts the same on a basic element of $\wedge^k \mathbb{R}^m$ as $\sqrt{(A^*A)^{\wedge q}}$ does, so the two are equal. But $\sqrt{(A^*A)^{\wedge q}} = \sqrt{(A^*)^{\wedge q} A^{\wedge q}}$ by Lemma 3.3. Therefore, $\|A^{\wedge q}\|$, which equals the maximum eigenvalue of $\sqrt{(A^*)^{\wedge q} A^{\wedge q}}$, also equals the maximum eigenvalue of $(\sqrt{A^*A})^{\wedge q}$. By Lemma 5.4, the eigenvalues of $\sqrt{(A^*A)^{\wedge q}}$ form the set

$$\{\sigma_{i_1} \wedge \cdots \wedge \sigma_{i_q} : i_1 < \cdots < i_q\}.$$

The maximum of these eigenvalues is $\sigma_m \cdots \sigma_{m-q+1}$. □

Lemma 5.6. *Let $u \in \mathbb{R}^m$ and V, W be subspaces of $\mathbb{R}^m$. Then*

$$\| \operatorname{Proj}(u, W) \| \leq \| \operatorname{Proj}(u, V) \| + \| \operatorname{Proj}(\operatorname{Proj}(u, V^\perp), W) \|,$$

where $\operatorname{Proj}(u, W)$ is the projection of u onto the subspace W.

Proof: $u = \operatorname{Proj}(u, V) + \operatorname{Proj}(u, V^\perp)$, so

$$\begin{aligned}\| \operatorname{Proj}(u, W) \| &= \| \operatorname{Proj}(\operatorname{Proj}(u, V) + \operatorname{Proj}(u, V^\perp), W) \| \\ &\leq \| \operatorname{Proj}(\operatorname{Proj}(u, V), W) \| + \| \operatorname{Proj}(\operatorname{Proj}(u, V^\perp), W) \| \\ &\leq \| \operatorname{Proj}(u, V) \| + \| \operatorname{Proj}(\operatorname{Proj}(u, V^\perp), W) \|.\end{aligned}$$

□

6. Grassman Manifolds

We will cast the proof of the Fundamental Lemma that appears in [1] as a statement about the convergence of a sequence of subspaces in a Grassman manifold, something that Ruelle is doing in [1] without ever stating it. To do this, we need to define a metric on Grassman manifolds that is compatible with Ruelle's proof. I don't have a decent reference on Grassman manifolds so I am making this up as I go.

Define the Grassman manifold, Gr_n^m, to be the set of all linear subspaces of dimension n, $0 \le n \le m$, of the Euclidean space $\mathbb{R}^m$. Gr_n^m is diffeomorphic to Gr_n^{n-m} once a suitable differentiable structure is put on the spaces. We, however, need only deal with a metric structure for the spaces.

$v^{\perp}$ is just a mnemonic device—the $\perp$ is not some kind of operator.

Define

$$d : \mathrm{Gr}_n^m \times \mathrm{Gr}_n^m \to \mathbb{R}^{\ge 0},$$
$$d(U,V) = \max\{|\langle u, v^{\perp}\rangle| : u \in U, v^{\perp} \in V^{\perp}, \|u\| = \|v^{\perp}\| = 1\}.$$

We use d for the function regardless of the value of n, since there is no real possibility of confusion about the domain of the function.

Lemma 6.1. *For all subspaces $U, V \in Gr_n^m$,*

$$d(U,V) = d(U^{\perp}, V^{\perp}).$$

Proof: My original proof for this is nonsense, so until I can come up with a clean proof, here is a geometric argument. The geometric interpretation of the metric is that $d(U,V) = |\sin\theta|$, where θ is the angle formed between the two subspaces, U and V. This is clearly the same as the angle formed between their orthogonal complements in $\mathbb{R}^m$. $\square$

Lemma 6.2. *The function, d, defined above is a metric on Gr_n^m.*

Proof: If $U = V$, then clearly $d(U,V) = 0$. If $d(U,V) = 0$, then for all $u \in U$ and $v^{\perp} \in V^{\perp}$, $\langle u, v^{\perp}\rangle = 0$, so $V^{\perp} \supseteq U^{\perp}$. But $\dim U = \dim V$ so $V^{\perp} = U^{\perp}$ and hence $U = V$. That is, $d(U,V) = 0 \Leftrightarrow U = V$.

By Lemma 6.1,

$$\begin{aligned} d(U,V) &= d(U^{\perp}, V^{\perp}) \\ &= \max\{|\langle u^{\perp}, v\rangle| : u^{\perp} \in U^{\perp}, v \in (V^{\perp})^{\perp}, \|u^{\perp}\| = \|v\| = 1\} \\ &= \max\{|\langle v, u^{\perp}\rangle| : v \in V, u^{\perp} \in U^{\perp}, \|v\| = \|u^{\perp}\| = 1\} \\ &= d(V,U). \end{aligned}$$

If A and B are subspaces of $\mathbb{R}^m$ and a is a vector in A, then

$$\begin{aligned} \|\operatorname{Proj}(a, B^{\perp})\| &= \max\{|\langle a, b^{\perp}\rangle| : b^{\perp} \in B^{\perp}, \|b^{\perp}\| = 1\} \\ &\le \|a\|\, d(A,B) \end{aligned}$$

and equality holds for some unit vector a in A.

Let U, V, and W be subspaces in Gr_n^m. Choose a unit vector u in U such that $\|\operatorname{Proj}(u, W^\perp)\| = d(U, W)$. Applying Lemma 5.6 with $W^\perp$ in place of W and $V^\perp$ in place of V gives

$$\begin{aligned} d(U,W) &= \|\operatorname{Proj}(u, W^\perp)\| \\ &\le \|\operatorname{Proj}(u, V^\perp)\| + \|\operatorname{Proj}(\operatorname{Proj}(u, (V^\perp)^\perp), W^\perp)\| \\ &= \|\operatorname{Proj}(u, V^\perp)\| + \|\operatorname{Proj}(\operatorname{Proj}(u, V), W^\perp)\| \\ &\le d(U,V) + \|\operatorname{Proj}(u, V)\| d(V,W) \\ &\le d(U,V) + d(V,W), \end{aligned}$$

where we used the fact that $\operatorname{Proj}(u, V)$ is a vector in V of norm less-than-or-equal to 1.

Thus, the triangle inequality also holds, so d is a metric. □

7. Proof of the Fundamental Lemma

Although it will not be clear until we examine the proof of the Fundamental Lemma below, the subadditive ergodic theorem, essentially a classical result, gave us the conditions needed to assure convergence of the eigenvalues of $((T_x^n)^* T_x^n)^{1/2n}$ in Theorem 2.1. The "easy work" of the proof of Lemma 4.1 consists of showing that these conditions do, indeed, imply convergence of the eigenvalues. The "hard work" in the proof of Lemma 4.1 consists of showing that they also imply that the matrix itself converges.

Proof of Lemma 4.1 (The Fundamental Lemma): By Lemma 5.1, the matrix $((T^n)^* T^n)^{1/2}$ exists and has nonnegative eigenvalues, which we label $t_n^{(1)} \le \cdots \le t_n^{(m)}$, where some eigenvalues may be repeated. By Lemma 5.5,

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| (T^n)^{\wedge q} \right\| = \lim_{n\to\infty} \frac{1}{n} \log \left(t_n^{(m)} \cdots t_n^{(m-q+1)} \right),$$

where the left-hand side exists and is finite by assumption. From the above equality for $q = 1$ and $q = 2$, we see that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log t_n^{(m)} \quad \text{and} \\ &\lim_{n\to\infty} \frac{1}{n} \log \left(t_n^{(m)} t_n^{(m-1)} \right) = \lim_{n\to\infty} \left(\frac{1}{n} \log t_n^{(m)} + \frac{1}{n} \log t_n^{(m-1)} \right) \end{aligned}$$

exist and are finite, and therefore also that

$$\lim_{n\to\infty} \frac{1}{n} \log t_n^{(m-1)}$$

exists and is finite. Continuing this argument for the remaining values of q, we conclude that each of the limits,

$$\chi^{(p)} := \lim_{n\to\infty} \frac{1}{n} \log t_n^{(p)},$$

for $p = 1, \ldots, m$ exists and is finite. Let $\lambda^{(1)} < \cdots < \lambda^{(s)}$ be the distinct values of the $\chi^{(p)}$, let

$$L^{(r)} = \left\{ p \in \{1, \ldots, m\} : \frac{1}{n} \log t_n^{(p)} \to \lambda^{(r)} \right\},$$

and let $U_n^{(r)}$ be the space spanned by the eigenvectors corresponding to the eigenvalues $\{t_n^{(p)} : p \in L^{(r)}\}$. Note that $\dim U_n^{(r)} = \dim U_{n'}^{(r)}$ for all n, n'; let m_r be their common dimension.

At this point we know that the eigenvalues, $(t_n^{(p)})^{1/n}$, of $((T^n)^* T^n)^{1/2n}$ converge to a limit, but we do not know that $((T^n)^* T^n)^{1/2n}$ itself converges. The hard work of showing this is done in Lemma 8.1, the fruits of our labors being Corollary 8.2, which states that the sequence of subspaces $(U_n^{(r)})_{n=1}^{\infty}$ is Cauchy in $\mathrm{Gr}_{m_r}^m$. Since Grassman manifolds are complete, the subspaces approach a limit, $U^{(r)}$. Knowing both the eigenspaces and the eigenvalues is enough to uniquely determine a matrix, since it tells us where each vector in a complete set of linearly independent (eigen)vectors is mapped to, which is enough to determine a linear map and its corresponding matrix (in a given basis). Therefore, the limiting matrix, Λ, exists.

Or observe that the limiting matrix is $\lambda^{(r)} I^{m_r \times m_r}$ on $U^{(r)}$.

This establishes (1) of the theorem and all of (2) except for showing that

$$\lim_{n\to\infty} \frac{1}{n} \log \|T^n u\| = \lambda^{(r)},$$

which we now show.

Let $u \in V^{(r)} \setminus V^{(r-1)}$. Then

$$u = c_1 u_1 + \cdots + c_r u_r$$

for some constants $c_1, \ldots, c_r$ with $u_i \in U^{(i)}$ and with $c_r \neq 0$. By virtue of Corollary 8.2, each u_i differs little from its projection into $U_n^{(i)}$. Also, $t_n^{(i)} \sim e^{n\lambda^{(i)}}$, so

This could all be made precise with an epsilon-delta argument, but I, for one, saw enough of that in the proof of Lemma 8.1.

$$((T^n)^* T^n)^{1/2} u \sim c_1 e^{n\lambda^{(1)}} u_1 + \cdots + c_r e^{n\lambda^{(r)}} u_r.$$

Then

$$\begin{aligned}
\frac{1}{n} \log \|T^n u\| &= \frac{1}{n} \log \|((T^n)^* T^n)^{1/2} u\| \\
&= \frac{1}{n} \log \|c_1 e^{n\lambda^{(1)}} u_1 + \cdots + c_r e^{n\lambda^{(r)}} u_r\| \\
&= \frac{1}{n} \log(c_1^2 e^{2n\lambda^{(1)}} + \cdots + c_r^2 e^{2n\lambda^{(r)}})^{1/2} \\
&= \frac{1}{2n} \log(c_1^2 e^{2n\lambda^{(1)}} + \cdots + c_r^2 e^{2n\lambda^{(r)}}) \\
&= \frac{1}{2} \frac{2c_1^2 \lambda^{(1)} e^{2n\lambda^{(1)}} + \cdots + 2c_r^2 \lambda^{(r)} e^{2n\lambda^{(r)}}}{c_1^2 e^{2n\lambda^{(1)}} + \cdots + c_r^2 e^{2n\lambda^{(r)}}} \\
&= \frac{c_1^2 \lambda^{(1)} e^{-2n(\lambda^{(r)} - \lambda^{(1)})} + \cdots + c_r^2 \lambda^{(r)}}{c_1^2 e^{-2n(\lambda^{(r)} - \lambda^{(1)})} + \cdots + c_r^2} \\
&\to \lambda^{(r)},
\end{aligned}$$

where we used L'Hospital's rule, the fact that $\lambda^{(r)}$ is the largest eigenvalue, and the fact that the $U^{(r)}$'s are mutually orthogonal, being the limits of mutually orthogonal eigenspaces. □

This limit gives us the existence of the Lyapunov exponents.

Notice how the subadditive ergodic theorem gave us the condition necessary to insure the existence in the limit of the eigenvalues of $((T^n)^* T^n)^{1/2n}$. (This condition being that $\lim_{n\to\infty} \frac{1}{n} \log \|(T^n)^{\wedge q}\|$ exists.) The existence of these eigenvalues was, in turn, needed to prove the convergence of $((T^n)^* T^n)^{1/2n}$ (as can be seen by examining the proof of Lemma 8.1). And finally, the existence of a limiting matrix for $((T^n)^* T^n)^{1/2n}$ was required to prove that the eigenvalues correspond to the possible values of the Lyapunov exponents.

So we might say that proving the existence of the limiting eigenvalues was easy, as was proving their connection to Lyapunov exponents once the hard work of establishing the convergence of $((T^n)^*T^n)^{1/2n}$ was done. But, there was no way (it seems) to bypass the hard work even if all we were after were the Lyapunov exponents.

8. The Hard Work

Lemma 8.1. *Given $\delta > 0$ there is a $K > 0$ such that, for all $k > 0$,*

$$\max\{|\langle u, u'\rangle| : u \in U_n^{(r)}, u' \in U_{n+k}^{(r')}, \|u\| = \|u'\| = 1\}$$
$$\leq K \exp(-n(|\lambda^{(r')} - \lambda^{(r)}| - \delta)).$$

Comment: Since $U_n^{(r)}$ and $U_{n+k}^{(r')}$ needn't have the same dimension when $r \neq r'$, we cannot state this lemma in terms of our metric on Grassman manifolds that we defined in Lemma 6.2. But we will be able to adopt this point of view in the last lemma of this section, where we do have $r = r'$.

Proof: (We inherit all the internal definitions in the proof of Lemma 4.1 in section 7.) Let

$$\alpha = \max\{|\langle u, u'\rangle| : u \in U_n^{(r)}, u' \in U_{n+k}^{(r')}, \|u\| = \|u'\| = 1\}.$$

We can interpret α as the maximum norm of the projection of any unit vector in $U_n^{(r)}$ into $U_{n+k}^{(r')}$.

Let $u \in \oplus_{t\leq r} U_n^{(t)}$. If $v_{rr'}^k = \mathrm{Proj}(u, \oplus_{t\geq r'} U_{n+k}^{(t)})$, the orthogonal projection of u into $\oplus_{t\geq r'} U_{n+k}^{(t)}$, then

$$\|v_{rr'}^k\| = \max\{|\langle u, u'\rangle| : u' \in \oplus_{t\geq r'} U_n^{(t)}, \|u'\| = 1\}.$$

If we can show that

$$\|v_{rr'}^k\| \leq K\,\|u\| \exp(-n(|\lambda^{(r')} - \lambda^{(r)}| - \delta)), \tag{1}$$

then the theorem will follow by our interpretation of α, since $v_{rr'}^k$ is the projection of a vector from a larger space than $U_n^{(r)}$ into a larger space than $U_{n+k}^{(r')}$ (and where we include the factor of $\|u\|$ since u is now not assumed to be of unit length).

Without loss of generality, assume that $\delta < |\lambda^{(r')} - \lambda^{(r)}|$ for all $r \neq r'$. Also, let $\delta^* = \delta/s$. Since

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_n\| \leq 0$$

by assumption, it follows that there is a $C > 0$ such that

$$\log \|T_{n+1}\| \leq C + n\frac{\delta^*}{4}, \tag{2}$$

for all n.

For n sufficiently large, $t_n^{(p)} < \exp\left(n(\lambda^{(r)} + \frac{\delta^*}{4})\right)$ for each p in $\bigcup_{t\leq r} L^{(t)}$, since $\lambda^{(r)}$ is the largest eigenvalue reached in the limit by the eigenvalues corresponding to the space $\oplus_{t\leq r} U_n^{(t)}$. The eigenspaces are orthogonal and the eigenvalues measure the stretching by $((T^n)^* T^n)^{1/2}$ of the component of a vector lying in the eigenspace. Hence, a vector of a given length can be stretched by no more than a factor equal to the largest eigenvalue—the factor that would occur if the vector lied in the eigenspace of the largest eigenvalue. Hence,

$$\|T^n u\| = \|((T^n)^* T^n)^{1/2} u\| \leq \exp\left(n(\lambda^{(r)} + \frac{\delta^*}{4})\right) \|u\|,$$

where we used Lemma 5.2.

Similarly, for n sufficiently large, each $t^{(p)}_{n+1}$ in $\bigcup_{t \geq r'} L^{(t)}$ will be greater than $\exp\left((n+1)(\lambda^{(r')} - \frac{\delta^*}{4})\right)$. Also, a vector is stretched by at least as much as would occur if the vector lied in the eigenspace of the smallest eigenvalue. But we know that the component in the eigenspace of the smallest eigenvalue is no larger than $\|v^1_{rr'}\|$, so

$$\|T^{n+1} u\| \geq \|T^{n+1} v^1_{rr'}\| \geq \exp\left((n+1)(\lambda^{(r')} - \frac{\delta^*}{4})\right) \|v^1_{rr'}\|.$$

Combining these two inequalities with Equation (2), we have

$$\begin{aligned} &\exp\left((n+1)(\lambda^{(r')} - \frac{\delta^*}{4})\right) \|v^1_{rr'}\| \\ &\quad \leq \|T^{n+1} u\| \leq \|T_{n+1}\| \|T^n u\| \\ &\quad \leq \exp\left(C + n\frac{\delta^*}{4}\right) \exp\left(n(\lambda^{(r)} + \frac{\delta^*}{4})\right) \|u\|, \end{aligned}$$

so,

$$\|v^1_{rr'}\| \leq \exp\left(C - \lambda^{(r')} + \frac{\delta^*}{4} - n\lambda^{(r')} + n\lambda^{(r)} + \frac{3n\delta^*}{4}\right) \|u\|.$$

Assume now that n is large enough that $C - \lambda^{(r')} + \frac{\delta^*}{4} \leq \frac{n\delta^*}{4}$. Then for $u \in \oplus_{t \leq r} U^{(t)}_n$,

$$\begin{aligned} \|\operatorname{Proj}(u, \oplus_{t \geq r'} U^{(t)}_{n+k})\| &= \|v^1_{rr'}\| \\ &\leq \|u\| \exp\left(-n(\lambda^{(r')} - \lambda^{(r)} - \delta^*)\right). \end{aligned} \tag{3}$$

We can use this inequality to bound $\|v^2_{r,r+1}\|$ by applying Lemma 5.6 with $V = \oplus_{t \geq r+1} U^{(t)}_{n+1}$ and $W = \oplus_{t \geq r+1} U^{(t)}_{n+2}$. Then $V^\perp = \oplus_{t \leq r} U^{(t)}_{n+1}$ (this is why we needed consecutive eigenvalues) and so Lemma 5.6 gives

$$\begin{aligned} \|v^2_{r,r+1}\| \leq{}& \|u\| \exp\left(-n(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right) \\ &+ \|u'\| \exp\left(-(n+1)(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right), \end{aligned}$$

where u' is the projection of u into W. But $\|u'\| \leq \|u\|$ and, extending the above result inductively, we have

$$\begin{aligned}\|v^k_{r,r+1}\| &\leq \sum_{j=0}^{k-1} \|u\| \exp\left(-(n+j)(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right)\\ &\leq \sum_{j=0}^{\infty} \|u\| \exp\left(-(n+j)(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right)\\ &= K_1\|u\| \exp\left(-(n+j)(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right),\end{aligned}$$

where

$$K_1 = \left(1 - \exp(-(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*))\right)^{-1}.$$

Now let $V = \oplus_{t \geq r+2} U^{(t)}_{n+1}$ and $W = \oplus_{t \geq r+2} U^{(t)}_{n+2}$. Then $V^\perp = \oplus_{t \leq r+1} U^{(t)}_{n+1}$ and applying Lemma 5.6 we conclude that

$$\begin{aligned}\|v^2_{r,r+2}\| &= \mathrm{Proj}(u, W)\\ &\leq \|\mathrm{Proj}(u, V)\| + \|\mathrm{Proj}(\mathrm{Proj}(u, V^\perp), W)\|\\ &= \|\mathrm{Proj}(u, \oplus_{t \geq r+2} U^{(t)}_{n+1})\| +\\ &\qquad \|\mathrm{Proj}(\mathrm{Proj}(u, \oplus_{t \leq r+1} U^{(t)}_{n+1}), \oplus_{t \geq r+2} U^{(t)}_{n+2})\|\\ &= \|v^1_{r,r+2}\| + \|\mathrm{Proj}(\mathrm{Proj}(u, \oplus_{t \leq r+1} U^{(t)}_{n+1}), \oplus_{t \geq r+2} U^{(t)}_{n+2})\|.\end{aligned}$$

The vector $\mathrm{Proj}(u, \oplus_{t \leq r+1} U^{(t)}_{n+1}) = v^1_{r,r+1}$ lies in $\oplus_{t \leq r+1} U^{(t)}_{n+1}$. Thus,

$$\begin{aligned}&\|\mathrm{Proj}(\mathrm{Proj}(u, \oplus_{t \leq r+1} U^{(t)}_{n+1}), \oplus_{t \geq r+2} U^{(t)}_{n+2})\|\\ &= \|\mathrm{Proj}(v^1_{r,r+1}, \oplus_{t \geq r+2} U^{(t)}_{n+2})\|.\end{aligned}$$

By Equation (3),

$$\|\mathrm{Proj}(v^1_{r,r+1}, \oplus_{t \geq r+2} U^{(t)}_{n+2})\| \leq \|v^1_{r,r+1}\| \exp\left(-n(\lambda^{(r+2)} - \lambda^{(r+1)} - \delta^*)\right)$$

and

$$\|v^1_{r,r+1}\| \leq \|u\| \exp\left(-n(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right).$$

Therefore,

$$\begin{aligned}\|v^2_{r,r+2}\| \leq \|u\| &\exp\left(-n(\lambda^{(r+2)} - \lambda^{(r)} - \delta^*)\right)\\ &+ \|u\| \exp\left(-n(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right)\\ &\qquad \times \exp\left(-n(\lambda^{(r+2)} - \lambda^{(r+1)} - \delta^*)\right),\end{aligned}$$

where we used Equation (3) once more.

Extending this result inductively gives,

$$\|v^k_{r,r+2}\| \leq \|u\| \sum_{j=0}^{k-1} \exp\left(-(n+j)(\lambda^{(r+2)} - \lambda^{(r)} - \delta^*)\right)$$
$$+ \|u\| K_1 \sum_{j=0}^{k-1} \exp\left(-n(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right)$$
$$\times \exp\left(-(n+j)(\lambda^{(r+2)} - \lambda^{(r+1)} - \delta^*)\right).$$

But,

$$\exp\left(-n(\lambda^{(r+1)} - \lambda^{(r)} - \delta^*)\right) \exp\left(-(n+j)(\lambda^{(r+2)} - \lambda^{(r+1)} - \delta^*)\right)$$
$$= \exp\left(-j(\lambda^{(r+2)} - \lambda^{(r+1)} - \delta^*)\right) \exp\left(-n(\lambda^{(r+2)} - \lambda^{(r+1)} - 2\delta^*)\right),$$

so we have a sum of a geometric series and a common factor of $\exp(-n(\lambda^{(r+2)} - \lambda^{(r+1)} - 2\delta^*))$, from which we conclude that

$$\|v^k_{r,r+2}\| \leq K_2 \, \|u\| \exp\left(-n(\lambda^{(r+2)} - \lambda^{(r)} - 2\delta^*)\right),$$

where K_2 is a constant.

Continuing the above argument inductively, we conclude that

$$\|v^k_{r,r'}\| \leq K_{r'-r} \, \|u\| \exp\left(-n(\lambda^{(r')} - \lambda^{(r)} - (r'-r)\delta^*)\right).$$

With the symmetric result for $r' < r$, and taking K to be the largest of all the constants K_i (and with the trivial result for $r = r'$, where $K_0 = 1$), we have

$$\|v^k_{r,r'}\| \leq K \, \|u\| \exp\left(-n(|\lambda^{(r')} - \lambda^{(r)}| - |r'-r|\delta^*)\right).$$

Since $\delta^* = \delta/s$, $|r'-r|\delta^* < \delta$ and Equation (1) follows, completing the proof of the lemma. □

Corollary 8.2. *The sequence of subspaces* $\{U_n^{(r)}\}_{n=1}^{\infty}$ *is a Cauchy sequence in the metric* d *on the Grassman manifold* Gr_l^m *defined in Lemma 6.2, where* l *is the common dimension of each* $U_n^{(r)}$.

Proof: Let $\delta < |\lambda^{(r')} - \lambda^{(r)}|$ for all $r \neq r'$. Then by Lemma 8.1,

$$\max\{|\langle u, u' \rangle| : u \in U_n^{(r)}, u' \in U_{n+k}^{(r')}, \|u\| = \|u'\| = 1\}$$
$$\leq K \exp(-nC),$$

where

$$C = \min\{|\lambda^{(r')} - \lambda^{(r)}| - \delta : r \neq r'\} > 0.$$

The Hard Work

Since

$$U_n^{(r)\perp} = \bigcup_{r' \neq r} U_n^{(r')},$$

it follows that for all $k > 0$,

$$\begin{aligned} & d(U_n^{(r)}, U_{n+k}^{(r)}) \\ & \quad = \max\{|\langle u, u' \rangle : u \in U_n^{(r)}, u' \in (U_{n+k}^{(r)})^{\perp}, \|u\| = \|u'\| = 1|\} \\ & \quad \leq K \exp(-nC), \end{aligned}$$

showing that the sequence $\{U_n^{(r)}\}_{n=1}^{\infty}$ is Cauchy. $\square$

9. Sources

My proof of Oseledec's multiplicative ergodic theorem is from Ruelle's paper,

[**1**] David Ruelle, *Ergodic Theory on Differentiable Dynamical Systems*, IHES Publicationes Mathematiques, 50:275-320 (1979).

There are actually two sets of page numbers in [1]: my references are to the page numbers with the lower of the two values.

The example I use to illustrate the theorem is from Ruelle's text,

[**2**] David Ruelle, *Chaotic Evolution of Strange Attractors*, Cambridge University Press, 1989.

This example was discussed briefly in a talk by Roman Shvydkoy earlier this semester in the fluid mechanics seminar and was first mentioned to me by Dr. Vishik at the beginning of this semester.

What is now called the multiplicative ergodic theorem or Oseledec's ergodic theorem was first proved by Oseledec in

[**3**] V. I. Oseledec, *A multiplicative ergodic theorem, Ljapunov characteristic numbers for dynamical systems*, Trudy Moskov, Mat. Obšč, 19 (1968).

Ruelle states that the proof of this theorem due to Oseledec is "not appropriate for our discussion," and gives instead a proof he attributes to M. S. Raghunathan, with a reference to a paper entitled *A proof of Oseledec' multiplicative ergodic theorem* to appear in Israel. J. Math. I have not looked up either Oseledec's or Raghunathan's papers.

Late in the process—the day before my first talk—I started looking at the following text:

[**4**] Ludwig Arnold, *Random Dynamical Systems*, Springer-Verlag, 1998.

Arnold gives a very detailed proof of various versions of the multiplicative ergodic theorem. He takes the same approach as that of [1] (which he calls the "established approach"), but states that he is "basically following Goldsheid and Margulis in doing the hard work." This is a reference to

[**5**] I. Y. Goldsheid and G. A. Margulis, *Lyapunov Indices of a product of random matrices*, Russian Mathematical Surveys, 44:11-71, 1989.

Arnold also gives the reference to the published version of Raghunathan's proof:

[**6**] M. S. Raghunathan, *A Proof of Oseldec's Multiplicative Ergodic Theorem*, Israel J. Math, 32:356-362, 1979.

Ruelle gives the following reference to a paper of Kingman's in which he states and proves the sub-additive ergodic theorem that bears his name:

Sources

[7] J. F. C. Kingman, *The Ergodic Theory of Subadditive Stochastic Processes*, J. Royal Statist. Soc., B 30, 499-510, 1968.

Ruelle gives the following reference to a paper of Furstenberg and Kesten, in which they prove our Corollary 3.2:

[8] H. Furstenberg and H. Kesten, *Products of Random Matrices*, Ann. Math. Statist., 31 p. 457-469, 1960.

Ruelle is extremely sketchy on the details in [1] and I spent more time than I would want to admit filling in the details. But the flow of the proof in Ruelle is easier to see than in, say, [4], precisely because so few details are given. The most difficult part of the proof—the proof of the Fundamental Lemma, which contains what Arnold refers to in [4] as the "hard work"—I almost despaired of trying to figure out since in [1] it is proved in one page containing a series of inequalities without any explanation of where they came from (Ruelle is working in a Grassman manifold, for example, but never even mentions that fact). And in [4] it is proved in roughly 12 pages, so I opted in the end to just figure out what Ruelle had written, which was a lot more fun anyway than reading 12 pages of dense calculation.