MALLORCA LECTURES

ANTHONY QUAS

1. Measure-preserving transformations

Throughout these notes, we will be talking about dynamical systems on *probability spaces*. That is: there will be a space X, a σ-algebra, $\mathcal{B}$ and a *probability measure* μ defined on $\mathcal{B}$.

[In case this is unfamiliar, these objects behave very intuitively: a typical example is μ is 'length measure' on $[0,1]$. μ satisfies some natural-sounding properties like $\mu(A \cup B) = \mu(A) + \mu(B)$ provided A and B are disjoint; the role of the σ-algebra is that one gets into trouble if one tries to define the measure of *every* set. Instead, a measure is defined on a large sub-collection of sets (the *measurable* sets). Unless you are a descriptive set theorist, every set you can think of will be measurable.]

Given a [measurable] map T, we say the measure μ is *invariant* under T (or "T preserves μ"; or "μ is an invariant measure for T" or "μ is T-invariant" or "T is a measure-preserving transformation of (X, μ)") if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$. On the face of it, this is hard to check because there are lots of measurable sets. However it suffices to check the invariance condition for nice sets A. Formally, one needs to check the invariance condition for all A belonging to a generating semi-algebra of $\mathcal{B}$. (A *semi-algebra* is a non-empty collection $\mathcal{S}$ of sets that is closed under finite intersections, and such that the complement of an element of $\mathcal{S}$ is a finite union of elements of $\mathcal{S}$. A semi-algebra is *generating* if the smallest σ-algebra containing $\mathcal{S}$ is $\mathcal{B}$. This condition is often easy to check in the common case where $\mathcal{B}$ is the Borel σ-algebra. An *algebra* is a semi-algebra that is also closed under taking finite unions.)

The following Lemma is very useful for checking that a measure is invariant:

Lemma 1. *Let T be a measurable map from a probability space $(X, \mathcal{B}, \mu)$ to itself. If $\mathcal{S}$ is a generating semi-algebra and $\mu(T^{-1}A) = \mu(A)$ for each A in $\mathcal{S}$, then μ is invariant under T.*

1.1. **Examples and applications.**

(1) The *doubling map* is the map T of $X = [0,1)$ equipped with Lebesgue measure Leb, defined by $T(x) = 2x \bmod 1$. Here one may take the semi-algebra to be the collection of half open intervals: $\mathcal{S} = \{[a,b)\colon 0 \le a < b \le 1\}$. It is easy to check that Leb is invariant under T. The same is true for $x \mapsto kx \bmod 1$ for any $k = 2, 3, \ldots$.

(2) (Doubling map again) Notice that $\frac{1}{5} \to \frac{2}{5} \to \frac{4}{5} \to \frac{3}{5} \to \frac{1}{5}$ is a periodic orbit under T of period 4. The measure $\mu = \frac{1}{4}(\delta_{\frac{1}{5}} + \delta_{\frac{2}{5}} + \delta_{\frac{4}{5}} + \delta_{\frac{3}{5}})$ is T-invariant. So is $\nu = \delta_0$ and so is $\frac{1}{2}(\mu + \nu)$.

The doubling map, not surprisingly, can be used to obtain information about the base 2 expansion of a point, and similarly $x \mapsto kx \bmod 1$ yields information about the base k expansion. Our next measure-preserving transformation gives information about the continued fraction expansion of a point. The continued fraction expansion of a point gives important information about Diophantine approximation: how closely can the point be approximated by rationals?

(3) (Gauss transformation): Let T be the map $[0,1) \to [0,1)$ defined by $T(x) = \mathrm{frac}(1/x)$. Then T preserves the probability measure μ given by $\mu(A) = \frac{1}{\log 2}\int_A \frac{1}{1+x}\,dx$.

(4) (Circle rotation): Let $\alpha \in [0,1)$ and let $T\colon [0,1) \to [0,1)$ be defined by $T(x) = x + \alpha \bmod 1$. Then T preserves Lebesgue measure.

If μ_{pre} is a set function defined on an algebra that is *additive* ($\mu_{\mathrm{pre}}(A \cup B) = \mu_{\mathrm{pre}}(A) + \mu_{\mathrm{pre}}(B)$ for any two disjoint elements of $\mathcal{A}$), and μ_{pre} has the additional property that $A_1 \supset A_2 \supset \ldots$ and $\bigcap A_n = \emptyset$ imply $\mu_{\mathrm{pre}}(A_n) \to 0$, then μ_{pre} may be uniquely extended to a measure μ on $\mathcal{B}$. If μ_{pre} satisfies $\mu_{\mathrm{pre}}(T^{-1}A) = \mu_{\mathrm{pre}}(A)$, then μ is T-invariant.

(5) (Coin tossing) The next measure-preserving transformation comes from probability. Imagine a (possibly biased) coin with probability of heads given by p. The space, Ω, is the set of bi-infinite sequences of 1's and 0's (with 1 representing heads and 0 representing tails). For $\omega \in \Omega$, we write $\omega = \ldots\omega_{-2}\omega_{-1}\cdot\omega_0\omega_1\omega_2\ldots$. The coordinate ω_i represents the outcome of the coin-toss at time i.

A *cylinder set* is a set of the form $[x]_a^b := \{\omega \in \Omega\colon \omega_a = x_a, \ldots, \omega_b = x_b\}$ (these sets are compact clopen sets forming a basis for the product topology). We set $\mathcal{A}$ to be the algebra of finite unions of cylinder sets. Setting $p_0 = 1 - p$ and $p_1 = p$, we define $\mu_{\mathrm{pre}}([x]_a^b) = p_{x_a}\cdots p_{x_b}$. By the above, this may be extended to a measure μ on Ω. We then define a transformation T by $T(\omega)_n = \omega_{n+1}$. That is: the sequence is moved to the left by one place. This is the *shift* transformation. The measure μ_{pre} is invariant: for each element of the generating algebra, it's easy to check that $\mu_{\mathrm{pre}}(C) = \mu_{\mathrm{pre}}(T^{-1}C)$ for each $C \in \mathcal{S}$, so that μ_{pre} extends to a T-invariant measure.

(6) (Geodesic flow) If M is a smooth compact Riemannian manifold and T_1M is its unit tangent space, then there is a natural 'time one' map on T_1M, "following the geodesic in the given direction for one unit of time". This map preserves Liouville measure on the unit tangent bundle.

(7) (Percolation) This one is a bit different. In the situation described above, the semi-group $\mathbb{N}_0$ acts on X by $n \cdot x := T^n(x)$; or if T is invertible, $\mathbb{Z}$ acts on X. This can be generalized to actions of other groups.

The space is $\Omega = \{0,1\}^{\mathbb{Z}^2}$. You should the think of this as the space of configurations of 0's and 1's on the integer lattice. Suppose that each site has a 1 with probability p independently of all of the others as in Example (5). This gives a measure $\mathbb{P}$ on Ω. The group $\mathbb{Z}^2$ naturally acts on ω by $(T^{\mathbf{k}}\omega)_{\mathbf{n}} = \omega_{\mathbf{n}+\mathbf{k}}$, that is, $T^{\mathbf{k}}$ shifts the configuration by the vector $-\mathbf{k}$. Essentially the same proof as previously shows that the group $\mathbb{Z}^2$ acts on Ω in a measure-preserving way, that is: for each $\mathbf{k} \in \mathbb{Z}^2$, $\mathbb{P}$ is $T^{\mathbf{k}}$-invariant.

(8) (Continuous map on a compact space) If X is a compact Hausdorff space and T is a continuous transformation, it is known that there is always at least one invariant measure on X. The proof uses the compactness of $\mathcal{M}_1(X)$, the set of Borel probability measures on X. Fix any $x \in X$, and define $\mu_n = \frac{1}{n}(\delta_x + \ldots + \delta_{T^{n-1}x})$. Any weak*-limit of this sequence is an invariant measure.

(9) (Szemerédi's theorem) This is a famous theorem in 'additive combinatorics': if S is a subset of $\mathbb{N}$, its (upper) density is $\bar{\rho}(S) = \limsup_{N\to\infty} \#(S \cap \{1,\ldots,N\})/N$. Szemerédi's theorem states that if $\bar{\rho}(S) > 0$, then for each $k \in \mathbb{N}$, S contains an arithmetic progression of length k. Furstenberg gave a second proof of this theorem using ergodic theory.

Defining a point x in $\Omega = \{0,1\}^{\mathbb{N}}$ by $x_n = 1$ if $n \in S$ and 0 otherwise. One can then use the shift map and the compactness argument above to obtain a shift-invariant measure on Ω

as above. Furstenberg's proof is based on establishing general properties of *all* shift-invariant measures on Ω.

(10) (Toral automorphisms) The map $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ mod 1 is a diffeomorphism from the torus to itself. Since the determinant is 1, it preserves Lebesgue measure on the torus.

(11) (Anosov diffeomorphisms) The map $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.01 \sin(2\pi(x+y)) \\ 0 \end{pmatrix}$ mod 1 also maps the torus to itself. By (8), there is at least one invariant measure. (In fact there are many such measures and it becomes important to pick a 'good one'). More on this later...

(12) (Skew product) If σ is a measure-preserving transformation of $(\Omega, \mathbb{P})$ and for each ω, T_ω is a map from a space X to itself, then the *skew product* is the map

$$\mathbf{T}(\omega, x) = (\sigma\omega, T_\omega(x)).$$

Iterating this, one can see that $\mathbf{T}^2(\omega, x) = (\sigma^2\omega, T_{\sigma\omega} T_\omega(x))$ and more generally

$$\mathbf{T}^n(\omega, x) = (\sigma^n\omega, T_{\sigma^{n-1}\omega} \circ \ldots \circ T_\omega(x)).$$

Skew products can be used to model random dynamical systems. For example suppose $\Omega = \{0,1\}^{\mathbb{Z}}$ and σ is the shift transformation and T_0 and T_1 are two maps from a space X to itself. Then define $T_\omega = T_1$ if $\omega_0 = 1$ and T_0 if $\omega_0 = 0$. Now the sequence of 0's and 1's gives the sequence of T's to apply to x.

Theorem 2 (Poincaré recurrence theorem)**.** *Let T be a measure-preserving transformation of a probability space. Then for all A with $\mu(A) > 0$, and for μ-almost every $x \in A$, x visits A infinitely often.*

Proof. Let $A_{\text{last}} = A \cap \bigcap_{n\ge 1} T^{-n}A^c$, be the set of points that are in A, but will never visit again. Now if $m < n$, we have $T^{-m}A_{\text{last}} \cap T^{-n}A_{\text{last}} = T^{-m}(A_{\text{last}} \cap T^{-(n-m)}A_{\text{last}})$, but by definition, $A_{\text{last}} \cap T^{-(n-m)}A_{\text{last}}$ is the empty set. Hence the sets $T^{-n}A_{\text{last}}$ are pairwise disjoint, and of equal measure (by the measure-preserving property). Since they sum to at most 1, they must all have measure 0. Hence A_{last} is of measure 0 and so (by countable additivity of measures) is $A_{\text{finite}} = \bigcup_{n\ge 0} T^{-n}A_{\text{last}}$, the set of points that make finitely many visits to A. $\square$

Corollary 3. *Suppose that the differential equation $\dot{x} = f(x)$ has a Lyapunov function ϕ: $\phi(x_0) = 0$ and $\phi(x) > 0$ for all $x \neq x_0$; $\frac{d}{dt}\phi(x(t)) = \langle \nabla\phi, f\rangle < 0$ whenever $x(t) \neq x_0$.*

Then the only invariant probability measure is a δ-measure supported at the attracting fixed point.

2. Ergodicity and Ergodic theorems

Notice that the Poincaré recurrence theorem says that the system will return to near where it is now.

The Boltzmann ergodic hypothesis says that the system will hit every possible state. This is fairly obviously false for discrete time dynamical systems (and is equally false even for simple continuous time systems). The Ehrenfests formulated a quasi-ergodic hypothesis, requiring that almost every orbit should be dense in the phase space. The ergodic theorem of von Neumann and subsequently Birkhoff were attempts to correctly formalize and demonstrate this.

2.1. Ergodic Theorems.

Theorem 4 (von Neumann mean ergodic theorem). *Let T be a measure-preserving transformation of a probability space (X, μ). Then for $f \in L^2(\mu)$, define $A_n f = \frac{1}{n}(f + f \circ T + \ldots + f \circ T^{n-1})$. Then $(A_n f)$ is a convergent sequence in $L^2(\mu)$.*

It is not hard to show that the limit of $A_n f$ is a T-invariant function (that is a function satisfying $g \circ T = g$).

Very shortly afterwards, Birkhoff established an extremely useful pointwise ergodic theorem:

Theorem 5 (Birkhoff pointwise ergodic theorem). *Let T be a measure-preserving transformation of a probability space (X, μ). Then for $f \in L^1(\mu)$, for μ-a.e. x, $(A_n f(x))$ is a convergent sequence.*

Again, one can show that if $\tilde{f}(x) = \lim_{n\to\infty} A_n f(x)$, then $\tilde{f}(x)$ is a T-invariant function.

2.2. Ergodicity.

Here is a critical definition: suppose T is a measurable transformation of X and μ is an invariant measure. Then μ is *ergodic* if for each T-invariant subset A of X, (i.e. $T^{-1}A = A$), $\mu(A)$ is 0 or 1.

Lemma 6. *Let T be a measure-preserving transformation of (X, μ). Then the following are equivalent:*

(a) T is ergodic;
(b) If f is an invariant measurable function (i.e. $f \circ T = f$) then f is constant on a set of full measure.
(c) If f is an invariant L^2 function, then f is constant on a set of full measure.

This lemma allows us to show ergodicity of many examples (including examples (1), (4) – provided α is irrational – and (10)). In fact, one can show that in most natural cases that invariant measures can always be described as a combination of ergodic invariant measures ("ergodic decomposition"). In these cases, when proving theorems, it is often sufficient to prove things only for ergodic measures. The measures μ and ν in Example (2) are ergodic, but $\frac{1}{2}(\mu+\nu)$ is not – see the exercises for more on this; The measure μ in Example (3) is ergodic. Under some geometric conditions on the manifold (such as negative curvature), example (6) is also ergodic.

Corollary 7. *Let T be an ergodic measure-preserving transformation of (X,μ). Then for all $f \in L^1(X)$, for μ-a.e. x,*

$$\frac{1}{n}\big(f(x)+\ldots+f(T^{n-1}x)\big) \to \int f\,d\mu.$$

Or the slogan version:

> *For ergodic systems, time averages and space averages coincide.*

Corollary 8. *Let T be an ergodic measure-preserving transformation of (X,μ). Then for all $A \in \mathcal{B}$, for μ-a.e. x,*

$$\frac{1}{n}\#\{j \in [0,n)\colon T^j(x) \in A\} \to \mu(A).$$

2.3. **Example: Normal numbers.** Recall that in Example (1), Lebesgue measure is invariant and ergodic. Notice also that if x has binary expansion $0.x_0x_1x_2\ldots$, then $T(x)$ has binary expansion $0.x_1x_2\ldots$. In particular, the nth digit of the binary expansion of x is 1 if $T^n(x) \in [\frac{1}{2},1)$ and 0 if $T^n(x) \in [0,\frac{1}{2})$.

Since T is ergodic, Corollary 8 implies that for Leb-a.e. x, the limiting proportion of 1's in the binary expansion of x is $\mathsf{Leb}([\frac{1}{2},1)) = \frac{1}{2}$ and hence the proportion of 0's is $\frac{1}{2}$ also. There is nothing special about doubling here: analogous conclusions about base k expansions (for $k > 1$) follow from studying the multiplication by k map, $x \mapsto kx \bmod 1$. If S_k denotes the subset of $[0,1)$ whose base k expansion has each possible digit occurring with equal frequency, we have $\mathsf{Leb}(S_k) = 1$ for each $k > 1$ and hence $\mathsf{Leb}(\bigcap_{k>1} S_k) = 1$. These numbers have the property that they have equal digit frequencies in all bases, and they are called *absolutely normal*. We just established that the absolutely normal numbers have Lebesgue measure 1, but *not a single example of a normal number is known!*

Example (2) gave additional examples of invariant measures for the doubling map. For example Corollary 8, we deduce that for ν-a.e.

x, the proportion of 1's in the binary expansion of x is 0. How can this be squared with for Leb-a.e. x, the proportion of 1's in the binary expansion of x is $\frac{1}{2}$? (because Leb and ν 'live on' disjoint subsets of [0,1]).

The conclusions of the ergodic theorem are only as good as the invariant measure that one applies it to.

To get useful conclusions, one would like to have ergodic invariant measures that are spread out as much as possible over [0,1] (like Leb). If Leb is not invariant, all may not be lost: we can look for absolutely continuous ergodic invariant measures (such as μ in Example (3)).

2.4. **Example: Continued fractions.** Recall that each irrational number $x \in [0,1)$ has a unique continued fraction expansion:

$$x = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}}$$

where $a_0, a_1, \ldots$ are natural numbers. The map T in Example (3) may be used to compute the a's: Notice that

$$T(x) = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}.$$

From this, we see $a_n = \lfloor 1/T^n(x) \rfloor$. In particular, the nth term of the continued fraction expansion of x is 1 if and only if $T^n x \in (\frac{1}{2}, 1)$. Since μ is ergodic, the proportion of 1's in the continued fraction expansion of a.e. x is $\frac{1}{\log 2} \int_{1/2}^{1} \frac{1}{1+x}\, dx = \log \frac{4}{3} / \log 2$. The proportion of n such that $a_n = a_{n+1} = 1$ may also be found: this happens when $T^n(x) \in (\frac{1}{2}, 1)$ and $T^{n+1}(x) \in (\frac{1}{2}, 1)$, which happens when $T^n(x) \in (\frac{1}{2}, \frac{2}{3})$. The measure of this set is $\log \frac{10}{9} / \log 2$, so that for μ-a.e. $x \in (0,1)$, the proportion of 1's in its continued fraction expansion is $\log \frac{4}{3} / \log 2$ and the proportion of two consecutive 1's in its continued fraction expansion is $\log \frac{10}{9} / \log 2$. Since the ratio between μ and Lebesgue measure is bounded above and below, the same holds for Lebesgue-a.e. x in $(0,1)$.

Continued fraction digit sequences are not independent: the probability of having two consecutive 1's is not the square of the probability of having a 1.

2.5. **Example: percolation.** Consider $X = \{0,1\}^{\mathbb{Z}^2}$ equipped with the measure μ_p where each coordinate is 1 with probability p and 0 with probability $1-p$. Here μ_p is invariant under the group of transformations $(T_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^2}$ where $T_{\mathbf{k}}$ translates a configuration x through the vector $-\mathbf{k}$ (so that the symbol at the origin of $T_{\mathbf{k}}x$ was the symbol at $\mathbf{k}$ in x). By an argument similar to exercise (8), the measure μ_p is ergodic under the group $(T_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^2}$ (that is, the only sets that are invariant under the group of transformations are of measure 0 or 1). A version of Lemma 6 still applies in this context.

If $x \in X$, let B_x be the set of coordinates with a 1. If $\mathbf{m}$ and $\mathbf{n}$ are two elements of B_x, we say they are connected if there is a path joining $\mathbf{m}$ to $\mathbf{n}$ staying in B_x moving by $\pm\mathbf{e}_1$ and $\pm\mathbf{e}_2$ at each step. The *clusters* in B_x are the connected components. We define $f(x)$ to be the number of infinite clusters in x. Clearly $f(T_{\mathbf{k}}x) = f(x)$ as translating the configuration does not alter the number of infinite clusters. Hence f takes on a constant value for μ_p-almost $x \in X$. **Important: the values might differ for different p's as the μ_p are supported on different sets**. Call the value $N(p)$, so that $f(x) = N(p)$ for μ_p-a.e. x.

A very nice argument (which applies much more generally than the 2-dimensional case here) shows that $N(p)$ can only be 0, 1, or ∞; and a really beautiful argument of Burton and Keane shows that $N(p)$ can only be 0 or 1. See the exercises for more on this.

3. Perron-Frobenius operators

In some dynamical systems, there are obvious invariant measures. In others, such as the Gauss transformation, finding an invariant measure is non-obvious. In fact, it is unclear how Gauss discovered the invariant measure in Example (3).

3.1. **Absolutely continuous invariant measures.** In this section, we will discuss a technique for finding and studying *absolutely continuous invariant measures*. This generally refers to invariant measures that are absolutely continuous with respect to Leb (possibly in higher dimensions), that is measures μ such that $\mu(A) = \int_A h(x)\,d\mathsf{Leb}(x)$. The function h is called the *density*. It's sometimes written $\frac{d\mu}{d\mathsf{Leb}}$.

Given any measure (not necessarily invariant), one may build a new measure, the *push-forward of μ under T*, $T_*\mu := \mu \circ T^{-1}$. Recall, a measure is invariant if $\mu = \mu \circ T^{-1}$, so that it is fixed under the push-forward operation.

We now find a formula for $\mu \circ T^{-1}$ in the case that μ is absolutely continuous with respect to Leb, and T is a differentiable (or piecewise monotone, piecewise differentiable) map of the interval or the circle with $T'(x) \neq 0$ for almost all x. Suppose that T has monotone intervals $I_1, \ldots, I_k$ and write $T_i = T|_{I_i}$. Finally, let $d\mu(x) = h(x)d\mathsf{Leb}(x)$.

If J is an interval, then

$$\begin{aligned}
\mu(T^{-1}J) &= \int_{T^{-1}J} h(x)\, d\mu(x) \\
&= \sum_{i=1}^{k} \int_{T^{-1}J \cap I_i} h(x)\, dx \\
&= \sum_{i=1}^{k} \int_{T(I_i)\cap J} \frac{h(T_i^{-1}y)}{|T'(T_i^{-1}y)|}\, dy \\
&= \int_J \sum_{i=1}^{k} \mathbf{1}_{T(I_i)}(y) \frac{h(T_i^{-1}y)}{|T'(T_i^{-1}y)|}\, dy \\
&= \int_J \sum_{x \in T^{-1}y} \frac{h(x)}{|T'(x)|}\, dy,
\end{aligned}$$

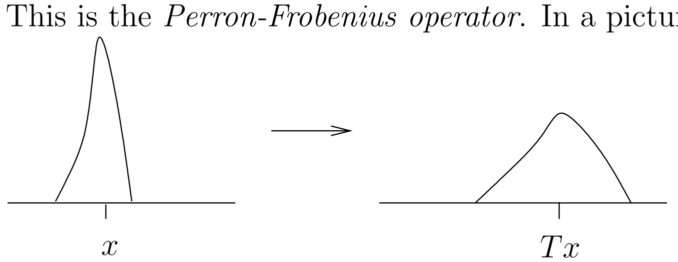
where we made the substitution $y = T(x)$ in the third line. That is, $T_*\mu$ is absolutely continuous with respect to Lebesgue measure with density given by $\mathcal{L}(h)$ where

$$\mathcal{L}(h)(y) = \sum_{x \in T^{-1}y} \frac{h(x)}{|T'(x)|}.$$

Notice that $\mathcal{L}$ maps positive functions to positive functions and $\int \mathcal{L}(h)\, d\lambda = \mu(T^{-1}I) = \mu(I) = \int h\, d\lambda$, so that $\mathcal{L}$ maps $L^1(I)$ to itself and is an operator of norm 1.

If $\mu = h \cdot \mathsf{Leb}$, then $\mu \circ T^{-1} = \mathcal{L}(h) \cdot \mathsf{Leb}$.

This is the *Perron-Frobenius operator*. In a picture, $\mathcal{L}$ acts like:



Above, we showed $\int \mathbf{1}_J(T(x))h(x)\, dx = \int \mathbf{1}_J(x)\mathcal{L}(h)(x)\, dx$. It follows that for all $f \in L^1$ and $h \in L^\infty$,

$$\int f(x) h \circ T(x)\, dx = \int \mathcal{L}(f)(x)(h)(x)\, dx.$$

Notice that $\mathcal{L}(h) = h$ if and only if $\mu = h \cdot \mathsf{Leb}$ is an invariant measure. Since $\|\mathcal{L}\|_{L^1 \to L^1} = 1$ and we are looking for an eigenfunction with eigenvalue 1, this suggests we might be able to use spectral methods.

We will use a crude argument of this type to show that Lebesgue measure on $[0, 1)$ is ergodic for the doubling map.

Theorem 9. Leb *is an ergodic invariant measure for the doubling map.*

Proof. We have already established that Leb is an invariant measure. It remains to prove the ergodicity. Notice that for the doubling map $\mathcal{L}(f)(x) = \frac{1}{2}\big(f(\frac{x}{2}) + f(\frac{x+1}{2})\big)$ We can then check $\mathcal{L}^n f(x) = \frac{1}{2^n}\sum_{i=0}^{2^n-1} f(\frac{x+i}{2^n})$.

Suppose, for a contradiction, that Leb is not ergodic. Then there is an invariant set A of Lebesgue measure between 0 and 1. It follows that $h \cdot \mathsf{Leb}$ is an invariant measure, where $h = \mathbf{1}_A/\mathsf{Leb}(A)$ (see the exercises). Since h is the density of an absolutely continuous invariant measure, $\mathcal{L}(h) = h$. Now since continuous functions on [0,1) are dense in L^1, there exists, for any $\epsilon > 0$, a continuous function f on the circle such that $\|f - h\|_1 < \epsilon$. Since f is continuous on a compact space, it is uniformly continuous: there exists a function $q(\delta)$ such that $q(\delta) \to 0$ as $\delta \to 0$ and if $|x - y| \le \delta$, $|f(x) - f(y)| \le q(\delta)$. Since $\mathcal{L}$ is of norm 1, we have $\|\mathcal{L}^n f - h\|_1 = \|\mathcal{L}^n f - \mathcal{L}^n h\| \le \epsilon$ for all n.

But on the other hand for any x, y,

$$|\mathcal{L}^n f(x) - \mathcal{L}^n f(y)| = \left|\frac{1}{2^n}\sum_{i=0}^{2^n-1}\big(f(\tfrac{x+i}{2^n}) - f(\tfrac{y+i}{2^n})\big)\right| \le q(2^{-n}),$$

so that $\mathcal{L}^n f$ is within $q(2^{-n})$ of a constant function. It follows that h may be arbitrarily closely approximated by a constant function, so that h is a constant, providing the contradiction. $\square$

3.2. **Interlude: Markov Chain refresher.** Recall: a (finite state discrete time) Markov chain is defined by a set of states S (here labeled $1, \ldots, k$), an initial distribution p, and a stochastic (i.e. with non-negative entries and each row summing to 1) transition matrix P satisfying $\mathbb{P}(X_0 = i) = p_i$ for each i and

$$\begin{aligned}&\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0)\\ =&\mathbb{P}(X_n = j | X_{n-1} = i) = P_{ij}\end{aligned}$$

The Markov chain is said to be *irreducible* if for each i and j, there exists an $n \ge 0$ such that $P^n_{ij} > 0$. It is *aperiodic* if $d := \gcd(\{n\colon P^n_{ii} > 0 \text{ for some } i\}) = 1$. (If $d > 1$, then S can be partitioned into d subsets, S_1, S_2,…, S_d such that from a state in S_i, the only possible transitions are to states in S_{i+1}). If the Markov chain is aperiodic and irreducible, then the matrix P is *primitive*: that is there exists an n_0 such that for all $n \geq n_0$, P^n has all strictly positive entries.

The matrix P governs the evolution of densities: if the probability distribution of X_n is given by p_n (that is $\mathbb{P}(X_n = i) = (p_n)_i$), then $p_{n+1} = p_n P$. A *stationary distribution* for the Markov chain is a vector π such that $\pi P = \pi$.

Theorem 10. *Consider an aperiodic irreducible Markov chain. Then there is a unique stationary distribution, π. Further, for any initial distribution p, $pP^n \to \pi$.*

This theorem has many proofs, but one nice way to see it is using the Perron-Frobenius theorem for matrices: If P is a primitive non-negative matrix, then there exists a simple positive real eigenvalue λ_1 such that all other eigenvalues have strictly smaller absolute value. Corresponding to λ_1, there is a a left eigenvector v with all positive entries. In the Markov chain case, $\lambda_1 = 1$ and one can show by expanding p with respect to the (possibly generalized) eigenvectors that $pP^n - \pi$ decays at an exponential rate of at worst polynomial times $|\lambda_2|^n$, where λ_2 is the eigenvalue with second largest absolute value.

We now suppose that the Markov chain is started in its stationary distribution (that is $p = \pi$) and consider the correlation between the state at time 0 and the state at time n.

More specifically, we will compute $\mathbb{P}(X_0 = a_0, \ldots, X_r = a_r$ and $X_n = b_0, \ldots, X_{n+s} = b_s) - \mathbb{P}(X_0 = a_0, \ldots, X_r = a_r)\mathbb{P}(X_n = b_0, \ldots, X_{n+s} = b_s)$. This is given by

$$\begin{aligned} &\pi_{a_0} P_{a_0 a_1} \cdots P_{a_{r-1} a_r} P^{n-r}_{a_r b_0} P_{b_0 b_1} \cdots P_{b_{s-1} b_s} \\ &- \pi_{a_0} P_{a_0 a_1} \cdots P_{a_{r-1} a_r} \pi_{b_0} P_{b_0 b_1} \cdots P_{b_{s-1} b_s} \\ =&\pi_{a_0} P_{a_0 a_1} \cdots P_{a_{r-1} a_r} (P^{n-r}_{a_r b_0} - \pi_{b_0}) P_{b_0 b_1} \cdots P_{b_{s-1} b_s}. \end{aligned}$$

In particular, the correlation decays at worst at an exponential rate $|\lambda_2|^n$ (possibly multiplied by a polynomial factor).

3.3. **Decay of correlation.** The heart of Theorem 9 was to find a Banach space that was a dense subspace of L^1 where everything converges nicely to the invariant density. By approximation, it follows that for every f in L^1, $\mathcal{L}^n f$ approaches a constant. *Might we hope for exponential convergence like for Markov chains?*

Answer: not in L^1 (or $C([0,1))$). To see this, notice that $h(x) = \cos(2\pi x)$ lies in the kernel of $\mathcal{L}$ and $\mathcal{L}(g \circ T) = g$ for any g. Now for any summable sequence (a_k), $f = 1 + \sum_k a_k h \circ T^k$ lies in L^1, but $\mathcal{L}^n f = 1 + \sum_{k \geq n} a_k h \circ T^{k-n}$.

In particular, by taking (a_k) to be the sequence of powers of $\frac{1}{2}$ interspersed by large blocks of 0, then $\mathcal{L}^n f$ can be made to converge to the limit, 1, arbitrarily slowly.

A beautiful idea due to Doeblin and Fortet, refined by Ionescu and Marinescu-Tulcea; then Lasota and Yorke and many others gives us this exponential convergence.

A bounded linear operator $\mathcal{L}$ on a Banach space is called *quasi-compact* if its essential spectral radius is strictly smaller than its spectral radius (its spectral radius is $\rho(\mathcal{L}) = \lim_{n\to\infty} \|\mathcal{L}^n\|^{1/n}$ and its essential spectral radius is $\rho_{\text{ess}}(\mathcal{L}) = \lim_{n\to\infty} \left(\inf_K \|\mathcal{L}^n - K\|\right)^{1/n}$, where the infimum is taken over all of the compact operators).

This implies that for any $r > \rho_{\text{ess}}(\mathcal{L})$, $\mathcal{L}$ has finitely many eigenvalues with absolute value greater than r, each with finite multiplicity.

Theorem 11. *(Lasota–Yorke) Let T be a piecewise map from $[0,1]$ to itself, with finitely many branches and each branch being monotonic, expanding and C^2. Then the restriction of $\mathcal{L}_T$ to $BV[0,1]$ is quasi-compact. Further if the branches of T map onto $[0,1]$, then 1 is a simple eigenvalue; all other eigenvalues lie strictly inside the unit circle.*

Although this theorem does not directly apply to example **??** (as that example has infinitely many branches), there exist extensions of that theorem that do cover this example.

Corollary 12 (Exponential decay of correlations)**.** *Let T be as in the statement of Theorem 11 with onto branches and suppose that in addition* Leb *is T-invariant. Then there exists $r < 1$ such that for all $f \in BV$ and $g \in L^1$,*

$$\int f \cdot g \circ T^n \, d\mathsf{Leb} - \int f \, d\mathsf{Leb} \int g \, d\mathsf{Leb} = O(poly \times \lambda_2^n),$$

4. Sub-additive ergodic theorems

Kingman (in 1968, following earlier work of Hammersley and Welsh) proved the Sub-Additive Ergodic Theorem.

If $\sigma\colon \Sigma \to \Sigma$ is a measure-preserving transformation (of a probability space), a sequence of functions $(f_n)_{n\geq 1}$ is *sub-additive* (with respect to σ) if

$$f_{n+m}(\omega) \leq f_n(\omega) + f_m(\sigma^n \omega).$$

Examples of sequences of functions satisfying this condition?

(1) (Fekete's lemma) Let (a_n) be a sequence of real numbers such that $a_{n+m} \leq a_n + a_m$. (Here the functions (f_n) are constant functions(!)). Then $a_{mk+r} \leq m a_k + a_r$, so that $\limsup a_n/n \leq$

a_k/k for each k. Now $\limsup a_n/n \le \inf_k a_k/k \le \liminf a_n/n \le \limsup a_n/n$. Hence a_n/n converges to a value $a = \inf a_k/k \in [-\infty, \infty)$.

(2) (Birkhoff averages) If f is an L^1 function on Ω, then $f_n(\omega) = f(\omega) + \ldots + f(\sigma^{n-1}\omega)$ is an *additive* sequence:

$$f_{n+m}(\omega) = f_n(\omega) + f_m(\sigma^n\omega).$$

(3) (First passage percolation) Let $\mathcal{E}$ denote the collection of edges in the $\mathbb{Z}^2$ lattice. Let ν be a probability measure on $(0, \infty)$ (with $\int x\, d\nu(x) < \infty$).

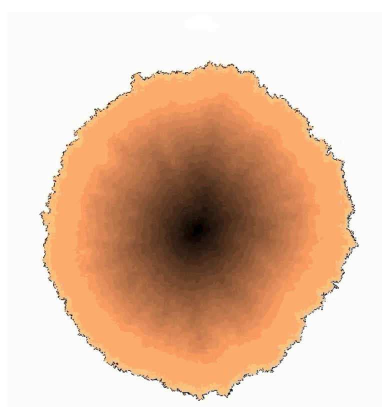
Let $\Omega = (0, \infty)^{\mathcal{E}}$ be the collection of all weightings of $\mathcal{E}$ and equip X with the probability measure $\nu^{\mathcal{E}}$ (so that in a realization $\omega \in \Omega$, each edge is assigned a weight from the distribution ν independently of all other edges). Define a $\mathbb{Z}^2$ action, $\tau_{\mathbf{v}}$ on Ω that translates the pattern of edge weightings through $-\mathbf{v}$.

Now for $\mathbf{v} \in \mathbb{Z}^2$, define $F_{\mathbf{v}}(\omega)$ to be the length of the shortest path from $\mathbf{0}$ to $\mathbf{v}$ (where the length of a path is the sum of the weights of the edges). Then

$$F_{\mathbf{u}+\mathbf{v}}(\omega) \le F_{\mathbf{u}}(\omega) + F_{\mathbf{v}}(\tau_{\mathbf{u}}(\omega)).$$

In particular, if $\mathbf{v}$ is any non-zero integer vector, then defining $\sigma = \tau_{\mathbf{v}}$ and $f_n(\omega) = F_{n\mathbf{v}}(\omega)$, (f_n) is a sub-additive sequence for the ergodic dynamical system $\sigma\colon (0, \infty)^{\mathcal{E}} \to (0, \infty)^{\mathcal{E}}$.

Hammersley and Welsh interpreted this as a 'wetting time': a rock is modelled by $\mathbb{Z}^2$. 'Water' is in contact with the rock at $\mathbf{0}$. The edge label determines the time it takes water to pass from one vertex to its neighbour. They were interested in the geometry of $\{\mathbf{v}\colon F_{\mathbf{v}}(\omega) < T\}$.



(image due to Jéremie Bettinelli)

(4) (Matrix products) If $\sigma\colon (\Omega,\mathbb{P})\to(\Omega,\mathbb{P})$ is a measure-preserving transformation, and $A\colon (\Omega,\mathbb{P})\to M_{d\times d}(\mathbb{R})$ is measurable, then one can form the *matrix cocycle*:

$$A^{(n)}(\omega)=A(\sigma^{n-1}\omega)\cdots A(\omega) \text{ for } n\in\mathbb{N} \text{ and } \omega\in\Omega.$$

Notice that

$$A^{(n+m)}(\omega)=A^{(m)}(\sigma^n\omega)A^{(n)}(\omega) \text{ (the cocycle relation)}.$$

defining $f_n(\omega)=\log\|A^{(n)}(\omega)\|$, you obtain

$$f_{n+m}(\omega)\le f_n(\omega)+f_m(\sigma^n\omega)$$

(providing $\|\cdot\|$ satisfies $\|AB\|\le\|A\|\|B\|$ (e.g. operator norm))

Theorem 13 (Kingman Sub-additive ergodic theorem, 1968)**.** *Let $\sigma\colon(\Omega,\mathbb{P})\to(\Omega,\mathbb{P})$ be an ergodic measure-preserving transformation. Let $(f_n)_{n\ge1}$ be a sub-additive sequence of integrable functions. Then*

(1) $\lim_{n\to\infty}\frac1n\int f_n(\omega)\,\mathbb{P}(\omega)$ converges to a constant $c\in[-\infty,\infty)$.
(2) For $\mathbb{P}$-a.e. $\omega\in\Omega$, $\frac1n f_n(\omega)\to c$.

Theorem 14 (Furstenberg, Kesten, 1960)**.** *Let $\sigma\colon(\Omega,\mathbb{P})\to(\Omega,\mathbb{P})$ be an ergodic measure-preserving transformation. Let $A\colon\Omega\to M_{d\times d}(\mathbb{R})$ be a matrix-valued function with $\int\log\|A(\omega)\|\,d\mathbb{P}(\omega)<\infty$. Then*

$$\tfrac1n\log\|A(\sigma^{n-1}\omega)\cdots A(\omega)\|\to E \text{ for a.e. } \omega,$$

where $E=\lim_{n\to\infty}\frac1n\int\log\|A(\sigma^{n-1}\omega)\cdots A(\omega)\|\,d\mathbb{P}(\omega)$.

The Birkhoff and Furstenberg–Kesten theorems are immediate corollaries of Kingman's theorem (but were both proved earlier).

5. Lyapunov Exponents and the Multiplicative ergodic theorem

If $T\colon I\to I$ is a differentiable self-map of the interval, then the chain rule gives $(T^n)'(x)=T'(T^{n-1}x)\cdot T'(T^{n-2}x)\cdots T'(x)$. The nth root of $(T^n)'(x)$ is a 'geometric average stretching rate' per step.

The *Lyapunov exponent for the one-dimensional map, T at x* is the logarithm of the limit of these rates: $\lambda(T,x)=\lim_{n\to\infty}\frac1n\log|(T^n)'(x)|$ (if it exists). That is: the derivative of T^n should be (logarithmically) close to $e^{n\lambda}$ (where 'logarithmically close' means between $e^{n(\lambda-\epsilon)}$ and $e^{n(\lambda+\epsilon)}$ for large n.)

We have

$$\lambda(T,x)=\lim_{n\to\infty}\frac1n\sum_{i=0}^{n-1}\log|T'(T^ix)|,$$

the Birkhoff sum of $\log|T'|$ along the orbit.

In the particular case, $I = [0,1]$ and $T(x) = 4x(1-x)$, $|T'|(x) = 4|1-2x|$. T has an ergodic absolutely continuous invariant measure, μ with density $1/(\pi\sqrt{x(1-x)})$.

Now we can apply Birkhoff's theorem (writing $\phi(x) = \log|T'(x)| = \log 4 + \log|1-2x|$) to get

$$\begin{aligned}\lambda(T,x) &= \lim_{n\to\infty}\frac{1}{n}\sum_0^{n-1}\phi(T^i x)\\ &= \int_0^1 \phi(t)\,d\mu(t)\\ &= \log 4 + \int_0^1 \frac{\log|1-2t|}{\pi\sqrt{t(1-t)}}\,dt\\ &= \log 2\end{aligned}$$

for μ-a.e. $x \in [0,1]$.

We therefore expect $|T^{50}(0.3+10^{-30}) - T^{50}(0.3)|$ to be of the order of $e^{50\lambda}\cdot 10^{-30} = 2^{50}10^{-30} \approx 1.13\times 10^{-15}$. In fact, $|T^{50}(0.3+10^{-30}) - T^{50}(0.3)| \approx 3.44\times 10^{-16}$ (so the prediction was correct to one order of magnitude).

A positive Lyapunov exponent is one of the (many and inequivalent) definitions of 'chaos'.

Now we'll consider the case of a differentiable map, T, from a subset of $\mathbb{R}^d$ to itself (or a differentiable map from a manifold to itself). Writing $DT(x)$ for the Jacobian matrix of T at x, the Chain rule gives

$$DT^n(x) = DT(T^{n-1}x)\cdots DT(T(x))\cdot DT(x).$$

We'd like to make sense of how fast these matrices grow. We can apply the Furstenberg-Kesten theorem as soon as we have an invariant measure for T.

On the other hand, if A is a single matrix $\|A^n v\|$ grows at different rates depending on the eigenvectors that make up v. This suggests we might expect $DT^n(x)v$ to grow at different rates for different subspaces of $\mathbb{R}^d$.

In the case of Example (10), $DT(x) = A := \left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)$ for every x. The eigenvalues of A are $\lambda_1 := \phi^2$ and $\lambda_2 := \phi^{-2}$ where ϕ is the golden mean, with eigenvectors $v_1 := \left(\begin{smallmatrix}\phi\\1\end{smallmatrix}\right)$ and $v_2 := \left(\begin{smallmatrix}1\\-\phi\end{smallmatrix}\right)$ respectively. Vectors in the space over x spanned by v_1 are mapped to vectors in the space over Tx also spanned by v_1, scaled by λ_1. Similarly vectors in the subspace spanned by v_2 are scaled by λ_2 and mapped to vectors spanned by v_2. The collection of subspaces in either direction (sub-vector bundles of the tangent space of $\mathbb{T}^2$) are called *equivariant families* of subspaces (or the whole collection is an *invariant sub-bundle*). Note that any vector that is *not* in the v_2 direction has asymptotic expansion at rate λ_1.

In general, the matrix $DT^n(x)$ depends on x, so we might expect the subspaces to depend on the point x.

Theorem 15 (Oseledets – non-invertible, 1965)**.** *Let σ be an ergodic measure-preserving transformation of $(\Omega, \mathbb{P})$. Let $A\colon \Omega \to M_{d\times d}(\mathbb{R})$ be a matrix-valued function with $\int \log \|A(\omega)\|\, d\mathbb{P}(\omega) < \infty$. Then there exist* Lyapunov exponents $\infty > \lambda_1 > \ldots > \lambda_k \geq -\infty$*;* multiplicities $m_1, \ldots, m_k \in \mathbb{N}$ *satisfying $m_1 + \ldots + m_k = d$ and a measurable family of subspaces $F_1(\omega)$, $F_2(\omega)$, $\ldots F_k(\omega)$ such that*

(1) filtration*:* $\mathbb{R}^d = F_1(\omega) \supset F_2(\omega) \supset \ldots \supset F_k(\omega) \supset F_{k+1}(\omega) = \{0\}$*;*
(2) dimension*:* $\dim F_i(\omega) = m_i + \ldots + m_k$*; for a.e.* ω
(3) equivariance*:* $A(\omega)F_i(\omega) \subset F_i(\sigma(\omega))$ *for a.e.* ω
(4) growth*: If $v \in F_i(\omega) \setminus F_{i+1}(\omega)$ then $\frac{1}{n}\log \|A^{(n)}_\omega v\| \to \lambda_i$ for a.e. ω, where $A^{(n)}_\omega = A(\sigma^{n-1}\omega)\cdots A(\omega)$.*

The quantities λ_i are called *Lyapunov exponents* and the subspaces $F_i(\omega)$ are the collection of vectors expanding **at rate** λ_i **or less**.

The sequence of subspaces $F_1(\omega) \supset F_2(\omega) \supset \ldots \supset F_k(\omega)$ is called a *flag*.

Theorem 16 (Oseledets – invertible, 1965)**.** *Let σ be an* invertible *ergodic measure-preserving transformation of $(\Omega, \mathbb{P})$. Let $A\colon \Omega \to GL(d, R)$ be a matrix-valued function with $\int \log \|A(\omega)\|\, d\mathbb{P}(\omega) < \infty$ and $\int \|(A(\omega))^{-1}\|\, d\mathbb{P}(\omega) <$*

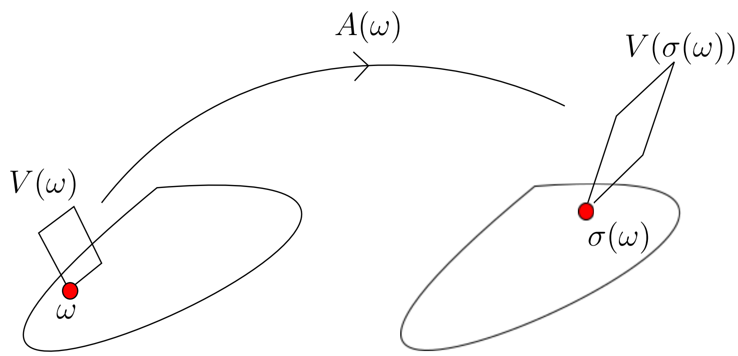


FIGURE 1. The Multiplicative Ergodic Theorem gives "A dynamical Jordan normal form decomposition."

∞. *Then there exist* Lyapunov exponents $\infty > \lambda_1 > \ldots > \lambda_k > -\infty$*;* multiplicities $m_1, \ldots, m_k \in \mathbb{N}$ *satisfying* $m_1 + \ldots + m_k = d$ *and measurable families of subspaces* $V_1(\omega)$, $V_2(\omega)$, $\ldots$, $V_k(\omega)$ *such that*

(1) decomposition*:* $\mathbb{R}^d = V_1(\omega) \oplus V_2(\omega) \oplus \cdots \oplus V_k(\omega)$*;*
(2) dimension*:* $\dim V_i(\omega) = m_i$ *for a.e.* ω*;*
(3) equivariance*:* $A(\omega)V_i(\omega) = V_i(\sigma(\omega))$ *for a.e.* ω
(4) growth*: If* $v \in V_i(\omega) \setminus \{0\}$ *then*

$$\frac{1}{n}\log \|A_\omega^{(n)} v\| \to \lambda_i \text{ and } \frac{1}{n}\log \|A_\omega^{(-n)} v\| \to -\lambda_i \text{ as } n \to \infty \text{ for a.e. } \omega,$$

where

$$A_\omega^{(n)} = A(\sigma^{n-1}\omega) \cdots A(\omega) \text{ for } n \geq 0; \text{ and}$$
$$A_\omega^{(-n)} = A(\sigma^{-n}\omega)^{-1} \cdots A(\sigma^{-1}\omega)^{-1} \text{ for } n > 0.$$

The $V_i(\omega)$ are the vectors expanding **at rate** λ_i. These are the *Oseledets subspaces.*

6. Deducing Oseledets from Kingman: Preliminaries

In the next section, we'll sketch an argument of Raghunathan, giving a proof of the non-invertible form of Oseledets' theorem from the subadditive ergodic theorem.

As a warm-up, we need some reminders about the *singular value decomposition* of a matrix; and definition of the Grassmannian of a vector space; and the exterior algebra of a vector space.

6.1. Singular Value Decomposition.

Theorem 17 (Singular Value Decomposition)**.** *Let* $A \in M_{d\times d}(\mathbb{R})$*. Then there exist orthogonal matrices* O_1 *and* O_2 *and a diagonal matrix* D *with non-negative entries such that* $A = O_1 D O_2$*.*

Proof. The matrix A^TA is symmetric, and so there is an orthonormal basis of $\mathbb{R}^d$ consisting of eigenvectors. If $A^TAv = cv$, then $c = \langle A^TAv, v\rangle = \langle Av, Av\rangle \geq 0$, so that all eigenvalues are non-negative. Let the eigenvalues be $\lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_d^2$ with corresponding orthonormal eigenvectors $v_1, \ldots, v_d$. Let O_2 be the matrix with rows consisting of $v_1, \ldots, v_d$; D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_d$. Let k be the largest index such that $\lambda_k > 0$. For $i \leq k$, let $u_i = Av_i/\lambda_i$. If $k < d$, let $u_{k+1}, \ldots u_d$ be an orthonormal basis for $A(\mathbb{R}^d)^\perp$. Let O_1 be the matrix whose columns are $u_1, \ldots, u_d$.

Since the rows of O_2 are orthonormal, we see that $(O_2O_2^T)_{ij} = \langle v_i, v_j\rangle = \delta_{ij}$, so that O_2 is orthogonal. We have $O_1DO_2v_i = O_1De_i = O_1\lambda_ie_i = \lambda_iu_i = Av_i$, so that $O_1DO_2 = A$. Finally, notice that for $i < j \leq k$, $\lambda_i\lambda_j\langle u_i, u_j\rangle = \langle Av_i, Av_j\rangle = \langle v_i, A^Av_j\rangle = \lambda_j^2\langle v_i, v_j\rangle = 0$ for $i \neq j$, so that the first k columns of O_1 are orthonormal (and so are the rest by construction), so $O_1^TO_1 = I$ as required. □

The *singular values* of A, $\sigma_1(A) \geq \ldots \geq \sigma_d(A)$, are the entries of D. The *singular vectors* are the rows of O_2 and their images are multiples of the columns of O_1, so that $Av_i = \sigma_i(A)u_i$.

Remark. Singular value decomposition (SVD) also makes sense for non-square matrices.

Lemma 18. *Let the singular values of A be $\sigma_1 \geq \ldots \geq \sigma_d$. Then for $1 \leq k \leq d$,*

$$\sigma_k = \max_{\dim V = k}\left(\min_{x \in V : \|x\|=1} \|Ax\|\right); \text{ and}$$
$$\sigma_k = \min_{\operatorname{codim} V = k-1}\left(\max_{x \in V :\ \|x\|=1} \|Ax\|\right).$$

Proof. (Exercise) □

In particular, from this characterization, you can see that $\sigma_1(A)$ is $\max_{\|x\|=1} \|Ax\|$, the norm of A and the first singular vector is a vector that is expanded most by A. By continuity, any vector close to v_1 is also expanded a lot by A, but v_2 is a vector in $\operatorname{lin}(v_1)^\perp$ that is expanded the most by A. etc.:

Lemma 19. *v_k is a vector in $\operatorname{lin}(v_1, \ldots v_{k-1})^\perp$ that is expanded the most by A.*

Proof. (Exercise) □

6.2. **Grassmannian of a vector space.** The k-dimensional Grassmannian of $\mathbb{R}^d$, $\mathrm{Gr}(d,k)$ is the collection of all k-dimensional subspaces of $\mathbb{R}^d$. This is a very nice space: a compact metric space, a smooth manifold etc.

To define a metric, we'll go with the most intuitive one: $d_{\mathrm{Gr}}(V,V') = d_H(V\cap S, V'\cap S)$, where S is the unit ball and d_H is the Hausdorff distance: for two non-empty compact sets, their Hausdorff distance is defined by $d_H(K,K') = \max\big(\max_{x\in K}\min_{y\in K'} d(x,y), \max_{y\in K'}\min_{x\in K} d(x,y)\big)$.

6.3. **Exterior power of a vector space.** A very useful construction in multiplicative ergodic theory is that of an exterior power of a vector space. For the formal construction of the kth exterior power, if V is a vector space, you form the free vector space F with basis consisting of all elements of the form $e_{v_1,\dots,v_k}$ for $(v_1,\dots,v_k)\in V^k$ (so that a typical element is $17e_{v_1,\dots,v_k} + 2e_{24v_1,v_2,\dots,v_k} + 12e_{0,v_2,\dots,v_k}$). We then let Z be a subspace of elements of F that we want to identify with 0: Z is the subspace of F spanned by elements of the form

$$\begin{aligned} e_{v_1,\dots,cv+c'v',\dots,v_k} - ce_{v_1,\dots,v,\dots,v_k} - c'e_{v_1,\dots,v',\dots,v_k} &\text{ (multilinearity)}\\ e_{v_1,\dots,v_i,\dots,v_j,\dots,v_k} + e_{v_1,\dots,v_j,\dots,v_i,\dots,v_k} &\text{ (antisymmetry)} \end{aligned}$$

The kth exterior power of V, $\bigwedge^k V$ is then F/Z. We write $v_1\wedge\cdots\wedge v_k$ for $e_{v_1,\dots,v_k}+Z$.

In fact, if $e_1,\dots,e_d$ is a basis for V then $\{e_{i_1}\wedge\cdots\wedge e_{i_k} : i_1<\ldots<i_k\}$ forms a basis for $\bigwedge^k V$, but proving this goes through a universal algebraic property of $\bigwedge^k V$.

Some elements of $\bigwedge^k V$ may be expressed in the form $v_1\wedge\cdots\wedge v_k$. Others can only be expressed as a sum of elements of this form (cf matrices expressed as sums of rank 1 matrices). A 'pure' vector $v_1\wedge\cdots\wedge v_k$ can be roughly thought of as defining an element of $\mathrm{Gr}(d,k)$ (i.e. $\mathrm{lin}(v_1,\dots,v_k)$) and a magnitude.

If A is a linear self-map of V, then $\bigwedge^k A$ is a self-map of $\bigwedge^k V$ satisfying $\left(\bigwedge^k A\right)(v_1\wedge\cdots\wedge v_k) = (Av_1)\wedge\cdots(Av_k)$ for each $v_1\wedge\cdots\wedge v_k$.

The space $\bigwedge^k\mathbb{R}^d$ can be turned into a Euclidean space by letting $\{e_{i_1}\wedge\cdots\wedge e_{i_k} : i_1<\ldots<i_k\}$ be an orthonormal basis, where $e_1,\dots,e_d$ is the standard basis.

It's completely non-obvious that if $f_1,\dots,f_d$ is any orthonormal basis, then $\{f_{i_1}\wedge\cdots\wedge f_{i_k} : i_1<i_2<\ldots<i_k\}$ is orthonormal with respect to this inner product. But it's true!

6.4. **SVD of Exterior powers.** Singular value decomposition and exterior powers play *extremely* nicely together. Let A be a $d\times d$ matrix with singular values $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_d$ and singular vectors $v_1, \ldots, v_d$. Recall that these are orthonormal. By what we just said, $\{v_{i_1} \wedge \cdots \wedge v_{i_k} : i_1 < \ldots < i_k\}$ forms an orthonormal basis of $\bigwedge^k \mathbb{R}^d$.

Recall also that $Av_i = \sigma_i u_i$, where the u_i's are also orthonormal. This means that

$$\bigwedge^k A(v_{i_1} \wedge \cdots \wedge v_{i_k}) = (\sigma_{i_1} \cdots \sigma_{i_k}) u_{i_1} \wedge \cdots \wedge u_{i_k}.$$

The $\{u_{i_1} \wedge \cdots \wedge u_{i_k}\}$ are orthonormal also, so that we obtain

Lemma 20. *The singular values of $\bigwedge^k A$ are $\{\sigma_{i_1} \cdots \sigma_{i_k} : i_1 < \ldots < i_k\}$ and the singular vectors are $\{v_{i_1} \wedge \cdots \wedge v_{i_k} : i_1 < \ldots < i_k\}$.*

In particular,

$$\left\| \bigwedge^k A \right\| = \sigma_1 \cdots \sigma_k. \tag{1}$$

7. Deducing non-invertible Oseledets from Kingman

7.1. **The Raghunathan trick.** Recall the notation $A^{(n)}_\omega = A(\sigma^{n-1}\omega) \cdots A(\omega)$. For each $1 \le k \le d$, define $f^{\wedge k}_n(\omega) = \log \| \bigwedge^k A^{(n)}_\omega \|$. Since $A^{(n+m)}_\omega = A^{(m)}_{\sigma^n\omega} A^{(n)}_\omega$ and $\bigwedge^k (AB) = \bigwedge^k A \bigwedge^k B$, we see that

$$f^{\wedge k}_{n+m}(\omega) \le f^{\wedge k}_m(\sigma^n \omega) + f^{\wedge k}_n(\omega).$$

Hence the Kingman sub-additive ergodic theorem (or Furstenberg-Kesten theorem) applies. There exist $L_1, \ldots, L_d$ such that $f^{\wedge k}_n(\omega)/n \to L_k$ for each k and a.e. ω.

Notice also that by (1), $f^{\wedge k}_n(\omega) = \log \| \bigwedge^k A^{(n)}_\omega \| = \sum_{i=1}^k \log \sigma_i(A^{(n)}_\omega)$. Hence $f^{\wedge k}_n(\omega) - f^{\wedge (k-1)}_n(\omega) = \log \sigma_k(A^{(n)}_\omega)$. Dividing by n and taking the limit, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \sigma_k(A^{(n)}_\omega) = L_k - L_{k-1}.$$

Define $\mu_k = L_k - L_{k-1}$. By the above, we have

$$\infty > \int \| \log A(\omega) \| \, d\mathbb{P}(\omega) > \mu_1 \ge \mu_2 \ge \ldots \ge \mu_d \ge -\infty.$$

These are the *Lyapunov exponents*. It is useful to group them by multiplicity:

$$\begin{aligned} &\{\lambda_1, \ldots, \lambda_k\} = \{\mu_1, \ldots, \mu_d\} \\ &\infty > \lambda_1 > \lambda_2 > \ldots > \lambda_k \ge -\infty \\ &\mu_{m_1+\ldots+m_{i-1}+j} = \lambda_i \text{ for } 1 \le j \le m_i. \end{aligned}$$

7.2. **Equivariant subspaces.** That was the easy part! Now we need the subspaces... We're trying to find equivariant spaces $F_j(\omega)$ of dimension $m_j + \ldots + m_k$ consisting of vectors expanding at rate λ_j or lower, *"the jth slow space"*. We'll get at these using the slow singular vectors of $A_\omega^{(n)}$.

Let $M_{j-1} = m_1 + \ldots + m_{j-1}$ for $1 \le j \le k$. This is the dimension of the "$(j-1)$*st fast space*", the number of exponents larger than λ_j. The jth slow space should be spanned by singular vectors with exponents λ_j and below: by the $(M_{j-1}+1)$st to dth singular vectors. Let $O_j = m_j + \ldots + m_k$.

The idea is to define $F_j^{(n)}(\omega)$ to be the space spanned by the $(M_{j-1} + 1)$st to dth singular vectors of $A_\omega^{(n)}$, and prove:

(1) these subspaces converge to a limit, $F_j(\omega)$, as $n \to \infty$;
(2) $F_j(\omega)$ is equivariant: $A(\omega)F_j(\omega) \subset F_j(\sigma(\omega))$;
(3) if $v \notin F_j(\omega)$, then $\liminf_{n\to\infty} \frac{1}{n} \log \|A_\omega^{(n)} v\| \ge \lambda_{j-1}$;
(4) if $v \in F_j(\omega)$, then $\limsup_{n\to\infty} \frac{1}{n} \log \|A_\omega^{(n)} v\| \le \lambda_j$.

Of these, (1), (2) and (3) are relatively straightforward, while (4) is the trickiest.

7.3. **A sketch of** (1)**.** Remember that $\mathrm{Gr}(d, O_j)$ is compact metric (hence complete). The idea is to show that the distance from $F_j^{(n)}(\omega)$ to $F_j^{(n+1)}(\omega)$ is $O(e^{-n(\lambda_{j-1}-\lambda_j-\epsilon)})$. Then the subspaces form a 'fast Cauchy sequence'.

How to do this? Take a unit vector, v, in $F_j^{(n)}(\omega)$ and write it as an orthogonal sum $u + w$ of a part u in $F_j^{(n+1)}(\omega)$ (the span of the slow singular vectors for $n+1$ step evolution) and w in $F_j^{(n+1)}(\omega)^\perp$ (the fast singular vectors). Since we know that $\|A^{(n)}v\| \lesssim e^{n\lambda_j}$, [1] it follows that $\|A^{(n+1)}v\| \lesssim e^{n\lambda_j}$. But $\|A_\omega^{(n+1)}v\|$ is the sum of the orthogonal vectors $A_\omega^{(n+1)}u$ and $A_\omega^{(n+1)}w$. Hence $\|A_\omega^{(n+1)}w\| \lesssim e^{n\lambda_j}$. Since w is in the fast space (so grows at rate λ_{j-1} or faster), this implies $\|w\| \lesssim e^{-n(\lambda_{j-1}-\lambda_j)}$. That is: every unit vector in $F_j^{(n)}(\omega)$ is $e^{-n(\lambda_{j-1}-\lambda_j-\epsilon)}$-close to something in $F_j^{(n+1)}(\omega)$.

7.4. **A sketch of** (2)**.** We show that elements of $A(\omega)(F_j^{(n+1)}(\omega))$ are exponentially close to $F_j^{(n)}(\sigma(\omega))$ and take the limit as $n \to \infty$ using claim (1).

[1] I'll write $x_n \lesssim e^{an}$ to mean for any ϵ, $x_n \le e^{(a+\epsilon)n}$ for large n. That is, the exponential growth rate is at most a.

Take $v = A(\omega)z$ in the unit ball of $A(\omega)(F_j^{(n+1)}(\omega))$; express it as $u + w$ with $u \in F_j^{(n)}(\sigma(\omega))$ and $w \in F_j^{(n)}(\sigma(\omega))^\perp$. The rest of the argument is like the previous step.

7.5. **A sketch of** (3)**.** If v is a unit vector not in $F_j(\omega)$, it is some positive distance, δ, from $F_j(\omega)$. By the triangle inequality, it is at least $\frac{\delta}{2}$ from $F_j^{(n)}(\omega)$ for all large n. That means that if v is decomposed into components in the slow space, $F_j^{(n)}(\omega)$ and the fast space, $F_j^{(n)}(\omega)^\perp$, there is a vector of length at least $\frac{\delta}{2}$ in the fast space. When you apply $A_\omega^{(n)}$, you get a vector of length $\frac{\delta}{2}e^{n(\lambda_{j-1}-\epsilon)}$, as required.

7.6. **Sketch of a sketch of** (4)**.** Write $V_i^{(n)}(\omega)$ for the space spanned by the $(M_{i-1}+1)$st to M_ith singular vectors of $A_\omega^{(n)}$. The idea is to

> Show that if v is a unit vector in $F_j(\omega)$, then the component of v in $V_i^{(n)}$ is of size at most $e^{(\lambda_j-\lambda_i+\epsilon)n}$ for each $i < j$.

Now when you apply $A_\omega^{(n)}$ to v, the vector obtained is of size at most $e^{(\lambda_j+\epsilon)n}$ (as seen working component by component and using the triangle inequality).

Raghunathan shows the above by clever estimates on the inverse of a matrix.

As an alternative, step (1) already gives the desired estimate in the case of $F_2(\omega)$. This is enough to show that elements of $F_2(\omega)$ grow at rate λ_2 or less. Now, one can look at the restriction of $A(\omega)$ to $F_2(\omega)$ and deduce that $F_3(\omega)$ grows at rate λ_3 or less and obtain the result inductively. (This argument is carried out in a Banach space setting in papers of Alex Blumenthal, and of Cecilia González-Tokman and myself).

8. Deducing invertible Oseledets from non-invertible Oseledets

For this section, we're assuming that the base dynamics, σ, is invertible, and also that the matrices $A(\omega)$ are invertible (and $\|(A(\omega))^{-1}\|$ is log-integrable). It turns out that the first condition is crucial, whereas the second condition is not.

Recall the definition of the stable and unstable manifolds of a fixed point of an invertible map T

$$W_s(p) = \{x \colon d(T^n x, p) \to 0 \text{ as } n \to \infty\};$$
$$W_u(p) = \{x \colon d(T^n x, p) \to 0 \text{ as } n \to -\infty\}.$$

At first sight, the definition of the unstable manifold may be surprising:

Exercise. Why is the unstable manifold defined this way? (Think about the map $T(x) = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} x$ for a concrete example).

Like this, we will obtain fast spaces as the slow spaces of the inverse system.

8.1. **Non-invertible result implies invertible result using the Inverse system.** The map σ^{-1} is another ergodic measure-preserving transformation of $(\Omega, \mathbb{P})$. Define the matrix $B(\omega) = A(\sigma^{-1}\omega)$ and build the matrix cocycle $B_\omega^{(n)} = B(\sigma^{-(n-1)}\omega) \cdots B(\omega)$. Notice that $B_\omega^{(n)} = (A_{\sigma^{-n}\omega}^{(n)})^{-1}$.

Applying the one-sided Oseledets theorem to the inverse system, we obtain a family of subspaces $\mathbb{R}^d = E_k(\omega) \supset E_{k-1}(\omega) \supset \cdots E_1(\omega)$ such that:

- (dimension): $\dim E_j(\omega) = m_1 + \ldots + m_j$;
- (equivariance): $B(\omega)E_j(\omega) \subset E_j(\sigma^{-1}\omega)$;
- (growth): $v \in E_j(\omega) \setminus E_{j+1}(\omega)$ implies $\frac{1}{n} \log \|B_\omega^{(n)} v\| \to -\lambda_j$;

Since $B(\omega) = A(\sigma^{-1}(\omega))^{-1}$, the equivariance condition can be rephrased as $E_j(\omega) \subset A(\sigma^{-1}(\omega))E_j(\sigma^{-1}\omega)$, or $E_j(\sigma(\omega)) \subset A(\omega)E_j(\omega)$. Since $A(\omega)$ is invertible, and $\dim E_j(\omega) = m_1 + \ldots + m_j$ for a.e. ω, we deduce $E_j(\omega)$ *is an equivariant family.*

If $v \in E_j(\omega)$, then $\|B_\omega^{(n)} v\| \lesssim e^{-(\lambda_j - \epsilon)n}\|v\|$. Since $B_\omega^{(n)} = (A_{\sigma^{-n}\omega}^{(n)})^{-1}$, it follows that for $w \in E_j(\sigma^{-n}\omega)$ (writing w as $(A_{\sigma^{-n}\omega}^{(n)})^{-1}v$), $\|w\| \lesssim e^{-(\lambda_j - \epsilon)n}\|A_{\sigma^{-n}\omega}^{(n)} w\|$ or $\|A_{\sigma^{-n}\omega}^{(n)} w\| \gtrsim e^{(\lambda_j - \epsilon)n}\|w\|$.

This (plus a little more work) shows that $E_j(\omega)$ is the jth fast space: the vectors expanding at rate λ_j or faster.

Now: $V_j(\omega) = E_j(\omega) \cap F_j(\omega)$ is an equivariant space consisting of vectors expanding at exactly rate λ_j. The last thing to check is that it has the correct dimension, m_j. Since $\dim E_j(\omega) = m_1 + \ldots + m_j$ and $\dim F_j(\omega) = m_j + \ldots + m_k = (d - \dim E_j(\omega)) + m_j$, we see from the formula $\dim(U \cap V) = \dim U + \dim V - \dim(U + V)$ that $\dim V_j(\omega) \geq m_j$. It is not hard to see that the $(V_j(\omega))$ are mutually linearly independent: Suppose that that $v_1 + \ldots + v_k = 0$, where $v_i \in V_i(\omega)$. Suppose for a contradiction that the v_i are not all 0. Then let ℓ be the smallest index such that $v_\ell \neq 0$. Now $A_\omega^{(n)} v_\ell$ grows at rate λ_ℓ, while $A_\omega^{(n)}(v_{\ell+1} + \ldots + v_k)$ grows at rate at most $\lambda_{\ell+1}$, so that they cannot cancel for large n, contradicting the assumption that $v_1 + \ldots + v_k = 0$ (hence $A_\omega^{(n)}(v_1 + \ldots + v_k) = 0$).

A key observation: *The $E_j(\omega)$ were the slow spaces for the inverse system – that is these are determined by $(A(\sigma^n\omega))_{n<0}$, while the $F_j(\omega)$ are governed by $(A(\sigma^n\omega))_{n\geq 0}$.*

8.2. **Non-invertible result implies invertible result using duality.** In this sub-section, we'll prove the same result using duality. It is still important that σ is invertible, but we never take inverses of the matrices.

Define $C(\omega) = A(\sigma^{-1}\omega)^*$ and build a cocycle over σ^{-1}: $C_\omega^{(n)} = C(\sigma^{-(n-1)}\omega)\cdots C(\omega) = (A^{(n)}_{\sigma^{-n}\omega})^*$.

Theorem 21. *Let σ be an ergodic invertible measure-preserving transformation. Let $A\colon \Omega \to M_{d\times d}(\mathbb{R})$ be such that $\|A(\cdot)\|$ is log-integrable.*

Let $C_\omega^{(n)}$ be the dual cocycle over σ^{-1} as above. Then the Lyapunov exponents of the dual cocycle are the same as those of $A_\omega^{(n)}$.

Let the slow spaces for the dual cocycle be $G_1(\omega),\ldots G_k(\omega)$. Then:

(1) $A(\omega)G_j(\omega)^\perp = G_j(\sigma(\omega))^\perp$ for a.e. ω;
(2) $G_j(\omega)^\perp \cap F_{j-1}(\omega) = V_{j-1}(\omega)$ for a.e. ω.

Proof. To prove (1), let $v \in G_j(\omega)^\perp$ and $y \in G_j(\sigma(\omega))$. Then we have

$$\langle A(\omega)v, y\rangle = \langle v, A(\omega)^*y\rangle = \langle v, C(\sigma(\omega))y\rangle.$$

Since $C(\sigma(\omega))G_j(\sigma(\omega)) \subset G_j(\omega)$, we have $C(\sigma(\omega))y \in G_j(\omega)$, so that $\langle A(\omega)v, y\rangle = 0$ and $A(\omega)v \in G_j(\sigma(\omega))^\perp$, as required.

We'll just sketch the proof of (2). The main idea is to show that $G_j(\omega)^\perp$ has a trivial intersection with $F_j(\omega)$. Assuming this for now, since they have complementary dimensions ($F_j(\omega)$ and $G_j(\omega)$ have the same dimension as the A and C cocycles have the same Lyapunov exponents), it will then follow that $\mathbb{R}^d = F_j(\omega) \oplus G_j(\omega)^\perp$. From here (and (3) of section 7.2), it follows that everything in $G_j(\omega)^\perp$ expands at rate λ_{j-1} or faster. Now we have $G_j(\omega)^\perp \cap F_{j-1}$ is an equivariant subspace consisting of vectors expanding at rate λ_{j-1}. From the formula $\dim(U\cap V) = \dim U + \dim V - \dim(U+V)$, we see that $G_j(\omega)^\perp \cap F_{j-1}$ is of dimension m_{j-1}.

To prove the trivial intersection, let $Z = F_j(\omega)^\perp$. By section 7.2 (3), we have $\|A_\omega^{(n)}z\| \gtrsim e^{\lambda_{j-1}n}$ for all $z \in Z \cap S$. On the other hand, we have $d(A_\omega^{(n)}z, G_j(\sigma^n\omega)^\perp) = \max_{y\in G_j(\sigma^n\omega)\cap S}\langle A_\omega^{(n)}z, y\rangle$. For any $y \in G_j(\sigma^n\omega)\cap S$, we have $\langle A_\omega^{(n)}z, y\rangle = \langle z, B^{(n)}_{\sigma^n\omega}y\rangle \lesssim e^{\lambda_j n}$. Hence for any $z \in Z \cap S$, the component of $A_\omega^{(n)}z$ in the direction perpendicular to $G^\perp(\sigma^n\omega)$ is $\lesssim e^{\lambda_j n}$. We deduce $\angle(A_\omega^{(n)}Z, G^\perp(\sigma^n\omega)) \lesssim e^{-n(\lambda_{j-1}-\lambda_j)}$.

To finish, we show that elements of $A_\omega^{(n)}Z$ are forced to lie far from $F_j(\sigma^n\omega)$. One can show (with a little determinant magic) $\|\bigwedge^k A_\omega^{(n)}\| \approx$

$\| \bigwedge^k A_\omega^{(n)}|_{\bigwedge^k Z}\|$ for all large n and so $\| \bigwedge^k A_\omega^{(n)}|_{\bigwedge^k Z}\| \approx e^{n(m_1\lambda_1+\ldots+m_{j-1}\lambda_{j-1})}$. If an element z of Z had the property that $A_\omega^{(n)}z$ was e^{-an} close (in angle) to $F_j(\sigma^n\omega)$, then the above growth condition is contradicted. In particular, we deduce $A_\omega^{(n)}Z$ is far (at an exponential scale) in every direction from $F_j(\sigma^n\omega)$; but $A_\omega^{(n)}Z$ is close to $G_j(\sigma^n\omega)$. Hence $F_j(\sigma^n\omega)\cap G_j(\sigma^n\omega)=\{0\}$ for large n. Hence $F_j(\omega)\cap G_j(\omega)=\{0\}$ a.e. by the Poincaré recurrence theorem.

□

Corollary 22 (Oseledets theorem: semi-invertible case)**.** *Let σ be an* invertible *ergodic measure-preserving transformation of* $(\Omega,\mathbb{P})$. *Let $A\colon \Omega\to M_{d\times d}$ be a matrix-valued function with $\int \log\|A(\omega)\|\,d\mathbb{P}(\omega)<\infty$. Then there exist $\infty>\lambda_1>\ldots>\lambda_k\ge-\infty$; $m_1,\ldots,m_k\in\mathbb{N}$ satisfying $m_1+\ldots+m_k=d$ and measurable families of subspaces $V_1(\omega)$, $V_2(\omega)$, …, $V_k(\omega)$ such that*

(1) decomposition*: $\mathbb{R}^d=V_1(\omega)\oplus V_2(\omega)\oplus\cdots\oplus V_k(\omega)$;*
(2) dimension*: $\dim V_i(\omega)=m_i$ for a.e. ω;*
(3) equivariance*: $A(\omega)V_i(\omega)=V_i(\sigma(\omega))$ for a.e. ω*
(4) growth*: If $v\in V_i(\omega)\setminus\{0\}$ then $\frac{1}{n}\log\|A_\omega^{(n)}v\|\to\lambda_i$ as $n\to\infty$ for a.e. ω, where $A_\omega^{(n)}=A(\sigma^{n-1}\omega)\cdots A(\omega)$.*

The hypotheses are a hybrid of the two original Oseledets theorems: the underlying system must be invertible; there is no invertibility requirement for the matrices. The good news: we can still get a decomposition: $\mathbb{R}^d=V_1\oplus\ldots\oplus V_k$ rather than a filtration ($\mathbb{R}^d=F_1\supset\ldots\supset F_k$). We did lose something though: we have no backwards growth bounds on $\|A_\omega^{(-n)}v\|$ – the inverse matrices needn't even exist.

Theorem 23 (Oseledets theorem: Banach space version)**.** *Let σ be an invertible ergodic measure-preserving transformation of* $(\Omega,\mathbb{P})$. *Let B be a separable Banach space. Let $\mathcal{L}\colon\Omega\to L(B,B)$ be an operator-valued function with $\int\log\|\mathcal{L}_\omega\|\,d\mathbb{P}(\omega)<\infty$.*

Suppose that $\frac{1}{n}\int\log\|\mathcal{L}_\omega^{(n)}\|\,d\mathbb{P}(\omega)\to\lambda$ and $\frac{1}{n}\int\log\kappa(\mathcal{L}_\omega^{(n)})\,d\mathbb{P}(\omega)\to\alpha<\lambda$, where $\kappa(\mathcal{L})=\inf\{r\colon \mathcal{L}(B)$ can be covered by balls of radius $r\}$ and B is the unit ball.

Then there exist $1\le k\le\infty$, $\lambda_1>\lambda_2>\ldots>\lambda_k$ and equivariant subspaces $V_1(\omega),\ldots,V_k(\omega)$ and $R(\omega)$ such that $B=V_1(\omega)\oplus\ldots\oplus V_k(\omega)\oplus R(\omega)$ and the growth conditions of the matrix Oseledets theorem hold.

The proofs are based on defining suitable notions of singular values (or volume growth) for maps of linear maps on Banach spaces. There are *many* possibilities – all giving the same growth rates.

9. Exercises

(1) Show that Lebesgue measure is invariant under the doubling map. Also the measures μ, ν and $\frac{1}{2}(\mu+\nu)$ from Example (2)

(2) Show that for the Gauss map from Example (3), $T^{-1}[0,a) = \bigcup_{n\ge 1}(\frac{1}{n+a},\frac{1}{n}]$. Writing the μ-measure of the set on the right as $\frac{1}{\log 2}\sum_{n\ge 1}[(\log(n+a)-\log(n))-(\log(n+a+1)-\log(n+1))]$, show that μ is invariant.

(3) Consider the circle rotation, example (4) with $\alpha=\frac{1}{3}$. How many invariant probability measures can you find? Can you give a description of all invariant probability measures? What if $\alpha = 1/\sqrt{2}$?

(4) Let T be a measurable map from a space X to itself. Can you give a complete description of all atomic T-invariant probability measures? (A probability measure is *atomic* if there exists a countable set S such that $\mu(S)=1$ and $\mu(S^c)=0$).

(5) Suppose that T is a measure-preserving transformation of a probability space (X,μ) and f is a measurable function such that $f(Tx)\le f(x)$ for all x. Prove that for almost every x, $f(T^nx)=f(x)$ for all n.

[Hint: is it possible for the set $\{x\colon f(Tx)<f(x)-\epsilon\}\cap\{x\colon f(x)>a\}$ to have positive measure for some $\epsilon>0$ and some $a\in\mathbb{R}$? Is it possible for $\{x\colon f(Tx)<f(x)-\epsilon\}$ to have positive measure for some $\epsilon>0$?]

Hence, taking T to be the time one map of the differential equation, prove Corollary 3.

(6) Suppose that T is a continuous map from a compact metric space X to itself and that μ is a T-invariant probability measure. Show that for all $\epsilon>0$, and for μ-almost every x, there exists $n>0$ such that $d(x,T^nx)<\epsilon$.

[Hint: it may be helpful to consider the *support* of the measure, that is $\operatorname{supp}(\mu) := \{x\colon \mu(B_r(x))>0 \text{ for all } r>0\}$. It follows from the definition that this is closed, and from second countability of X together with countable additivity of μ that $\mu(\operatorname{supp}(\mu))=1$.]

(7) Suppose T is a measurable transformation of X preserving a probability measure μ, and that A is a measurable subset of X satisfying $T^{-1}A=A$ and $\mu(A)>0$. Show that the measure ν defined by $\nu(B)=\mu(A\cap B)/\mu(A)$ is also T-invariant.

(8) Let $\Omega=\{0,1\}^{\mathbb{Z}}$ and let μ_p be the coin-tossing measure with probability of 1's given by p as in example (5).

There is a standard theorem as follows:

Theorem. *If $\mathcal{A}$ is an algebra that generates a σ-algebra $\mathcal{B}$, then for each element B of $\mathcal{B}$ and each $\epsilon > 0$, there exists an $A \in \mathcal{A}$ such that $\mu\big((A \setminus B) \cup (B \setminus A)\big) < \epsilon$.*

(a) Use this to show that any T-invariant set is of μ_p-measure 0 or 1. (Hint: $B = T^{-n}B$ for any invariant set);
(b) Show that for μ_p-almost every x, the frequency of 1's is p;
(c) Deduce that there exists an uncountable collection of disjoint measurable sets B_p (one for each $p \in (0,1)$) such that $\mu_p(B_p) = 1$ and $\mu_p(\bigcup_{q \neq p} B_q) = 0$ for each p.

(9) Consider site percolation on $\mathbb{Z}^2$ (for example) where the probability that a site is occupied is p (and the measure is μ_p). By ergodicity, μ_p-a.e. configuration has exactly $N(p)$ infinite clusters. This exercise is about showing that $N(p)$ can only take values 0, 1 or ∞. Suppose for a contradiction that $N(p) = \ell$ with $1 < \ell < \infty$.
(a) Use continuity of measure (if $A_1 \subset A_2 \subset \ldots$, then $\mu_p(\bigcup A_n) = \lim \mu_p(A_n)$) to prove that there exists an k such that with positive probability at least two infinite clusters enter the $(2k+1) \times (2k+1)$ square centred at the origin.

Let Φ_k be the map that modifies a configuration by replacing all coordinates in the $(2k+1) \times (2k+1)$ square around the origin by 1's:

$$\Phi_k(x)_{\mathbf{n}} = \begin{cases} 1 & \text{if } \mathbf{n} \in [-k,k]^2 \\ x_{\mathbf{n}} & \text{otherwise.} \end{cases}$$

(b) Show that if a set S has positive measure, then $\Phi_k(S)$ has positive measure. [Hint: it may be useful to partition S into pieces according to the configuration seen on $[-k,k]^2$.]
(c) Deduce that $N(p)$ cannot take any finite value bigger than 1.

(10) Prove that $\mathrm{Gr}(d,2)$ is sequentially compact.

(11) For a fixed k-codimensional subspace, W, of $\mathbb{R}^d$, let $\mathcal{U}$ denote those elements of $\mathrm{Gr}(d,k)$ that have a trivial intersection with W.

Prove that $\mathcal{U}$ is an open subset of $\mathrm{Gr}(d,k)$.

Fix an element $V_0 \in \mathcal{U}$, a basis $e_1, \ldots, e_k$ for V_0 and a basis $f_1, \ldots, f_{d-k}$ for W. For an element $V \in \mathcal{U}$, since $V \oplus W = \mathbb{R}^d$, each e_i may be expressed uniquely as $v_i + w_i$ with $v_i \in V$ and $w_i \in W$. Form a matrix, $\Phi(V)$ whose ith column is the coefficients of w_i in the (f_j) basis.

Prove that Φ is a bijection from $\mathcal{U}$ to $\mathbb{R}^{(d-k)\times k}$.

Prove that the collection of $(\mathcal{U}, \Phi)$ (as W varies over $\mathrm{Gr}(d, d-k)$, V_0 varies over $\mathcal{U}$ and the bases vary over bases for W and V_0) forms a smooth manifold structure on $\mathrm{Gr}(d, k)$.

(12) Verify that:
- $v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k = -v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k$; and
- $v_1 \wedge \cdots \wedge (cv + c'v') \wedge \cdots \wedge v_k = v_1 \wedge \cdots \wedge v \wedge \cdots \wedge v_k + c'v_1 \wedge \cdots \wedge v' \wedge \cdots \wedge v_k$
- If $e_1, \ldots, e_d$ is a basis for V, then $\{e_{i_1} \wedge \cdots \wedge e_{i_k} : i_1 < \ldots < i_k\}$ spans $\bigwedge^k V$.

(13) Let V be a k-dimensional subspace of $\mathbb{R}^d$. Let two bases for V be $e_1, \ldots, e_k$ and $f_1, \ldots, f_k$.

Prove that $f_1 \wedge \cdots \wedge f_k = c\, e_1 \wedge \cdots \wedge e_k$, where c is the determinant of the matrix of coefficients of the f vectors in terms of the e vectors.

(14) Prove that there is a linear map, $\bigwedge^k A$ satisfying $\left(\bigwedge^k A\right)(v_1 \wedge \cdots \wedge v_k) = (Av_1) \wedge \cdots (Av_k)$ for each $v_1 \wedge \cdots \wedge v_k$..

(15) Show that if the invertible matrix A has singular values $\sigma_1 > \ldots > \sigma_d$, then A^{-1} has singular values $\sigma_d^{-1} > \ldots > \sigma_1^{-1}$.

(16) Check from the previous exercise and the Kingman sub-additive ergodic theorem that the Lyapunov exponents for this cocycle are $-\lambda_k, \ldots, -\lambda_1$, with multiplicities $m_k, \ldots, m_1$.