Lecture #2

We begin by restating the Mean Ergodic Theorem using the Koopman operator.

<u>THM1</u> (Mean Ergodic Theorem, von Neumann)
Assume $T: X \longrightarrow X$ is an ergodic measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Then for any $f \in L^2(X,\mu)$,

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \xrightarrow[L^2]{} \int_X f d\mu \cdot \mathbb{1},$$

where U_T is the Koopman operator corresponding to the transformation T.

When T is unitary (and $U_T^* = (U_T)^{-1} = U_{T^{-1}}$), the mean ergodic theorem follows from the following general theorem.

<u>THM2</u> (Abstract Ergodic Theorem)
Let $\mathcal{H}$ be a Hilbert space, let $U: \mathcal{H} \longrightarrow \mathcal{H}$ be a unitary operator on $\mathcal{H}$. Let

$$\mathcal{H}_0 = \{ f \in \mathcal{H} \mid Uf = f \}$$

be the subspace of elements invariant under U, and let $P: \mathcal{H} \longrightarrow \mathcal{H}_0$ be the orthogonal projection onto $\mathcal{H}_0$.

Then, for any $f \in \mathcal{H}$,

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n f \xrightarrow[L^2]{} Pf.$$

First, we will show that THM 2 implies a special case of THM 1 – when T is invertible. Indeed, T invertible implies U_T is unitary, which is the setting of THM 2. More concretely,

pf (THM 2 $\Rightarrow$ $\overbrace{\text{special case of THM 1}}^{\text{Assume } T \text{ invertible}}$):

Let $\mathcal{H} := L^2(X,\mu)$ and let $U := U_T$. Then

$$\mathcal{H}_0 = \{ f \in L^2(X,\mu) : U_T(f) = f \}$$

$$= \mathbb{C}\cdot\mathbb{1} \quad \text{(constant functions since } T \text{ ergodic)}$$

Moreover, $P : \mathcal{H} \longrightarrow \mathcal{H}_0$ is given by

$$Pf = \left(\int_X f\,d\mu\right)\cdot\mathbb{1}.$$

Indeed, for $Pf := \left(\int_X f\,d\mu\right)\cdot\mathbb{1}$, we have

① $\left(\int_X f\,d\mu\right)\cdot\mathbb{1} \in \mathbb{C}\cdot\mathbb{1}$ since $L^2 \subset L^1$ in a probability space (i.e. $\int_X f\,d\mu$ is convergent.)

and

② $\langle f - Pf, g\rangle = 0 \;\; \forall\, g \in \mathbb{C}\cdot\mathbb{1}$:

Assume $g \in \mathbb{C} \cdot \mathbb{1}$. Then $g \equiv c \cdot \mathbb{1}$ for some $c \in \mathbb{C}$.

Then,

$$
\begin{aligned}
\langle f - Pf, g \rangle &= \langle f - Pf, c\mathbb{1} \rangle \\
&= \int_X (f - Pf) \cdot \overline{c\mathbb{1}} \, d\mu \\
&= \bar{c} \int_X (f - Pf) \, d\mu \\
&= \bar{c} \int_X \left(f - \left(\int_X f \, d\mu \right) \cdot \mathbb{1} \right) d\mu \\
&= \bar{c} \left[\int_X f \, d\mu - \int_X \left(\int_X f \, d\mu \right) d\mu \right] \\
&= \bar{c} \left[\int_X f \, d\mu - \int_X f \, d\mu \cdot \underbrace{\int_X d\mu}_{=1} \right] \\
&= 0.
\end{aligned}
$$

Thus, THM 1 when T is invertible follows from THM 2.

$\square$

Now, we turn our attention to proving THM 2.

<u>pf (THM 2)</u>:

The key observation is the following: the Hilbert space $\mathcal{H}$ naturally decomposes with respect to the unitary operator into two subspaces: One subspace is the subspace of functions that are invariant under the unitary operator. The other is a subspace of functions whose averages decay to zero under successive applications of the unitary operator. The key insight is how to represent this second subspace.

Let's start with the first, the subspace of invariant functions:

$$\mathcal{H}_0 = \{ f \in \mathcal{H} \mid Uf = f \}.$$

For any vector $f \in \mathcal{H}_0$,

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n f = \frac{1}{N} \sum_{n=0}^{N-1} f = f.$$

Moreover, $f = Pf$ since $f \in \mathcal{H}_0$. Thus,

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n f = Pf$$

and clearly $\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n f = Pf$. It is then somewhat straightforward that the theorem holds for any $f \in \mathcal{H}_0$.

Let's turn our attention to the orthogonal subspace. We claim

$$\mathcal{H}_0 = \{Uf - f \mid f \in \mathcal{H}\}^{\perp},$$

meaning, in particular, that $\mathcal{H}_0^{\perp} = \overline{\{Uf - f \mid f \in \mathcal{H}\}}$.

To see this, observe the following. First, for any $v \in \mathcal{H}_0$, $U^{-1}v = v$. Indeed, $Uv = v$ so $U^{-1} \circ U v = U^{-1} v$ and so $U^{-1} v = v$. Second, or rather, using this first fact, we have that for any $v \in \mathcal{H}_0$, and any $f \in \mathcal{H}$,

$$\begin{aligned}
\langle Uf - f, v \rangle &= \langle Uf, v \rangle - \langle f, v \rangle \\
&= \langle f, U^{-1} v \rangle - \langle f, v \rangle \quad (U \text{ is unitary}) \\
&= \langle f, U^{-1} v - v \rangle \\
&= \langle f, \underbrace{v - v}_{0} \rangle \\
&= 0.
\end{aligned}$$

This shows that $\mathcal{H}_0 \subset \{Uf - f \mid f \in \mathcal{H}\}^{\perp}$. To see the reverse inclusion, let $v \in \mathcal{H}$ such that for all $f \in \mathcal{H}$,

$$\langle Uf - f, v \rangle = 0.$$

Then, $0 = \langle Uf - f, v \rangle = \langle f, U^{-1} v - v \rangle$ (as above), meaning $U^{-1} v = v$, and we must have $v \in \mathcal{H}_0$.

As desired, we have the decomposition

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp, \quad \text{where}$$

$$\mathcal{H}_0 = \{Uf - f \mid f \in \mathcal{H}\}^\perp$$

and, consequently,

$$\mathcal{H}_0^\perp = \overline{\{Uf - f \mid f \in \mathcal{H}\}}.$$

Now, consider $h \in \{Uf - f \mid f \in \mathcal{H}\}$. We have

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n h = \frac{1}{N}\sum_{n=0}^{N-1} U^n(Uf - f)$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} U^{n+1}f - U^n f$$

$$= \frac{1}{N}\left((Uf - f) + (U^2 f - Uf) + \cdots + (U^N f - U^{N-1} f)\right)$$

telescoping

$$= \frac{1}{N}(U^N f - f).$$

Consider the norm of the sequence of averages.

$$\left\| \frac{1}{N}\sum_{n=0}^{N-1} U^n h \right\| \leq \frac{1}{N}\left(\|U^n f\| + \|f\|\right)$$

U unitary

$$= \frac{1}{N}\left(\|f\| + \|f\|\right)$$

$$= \frac{2}{N} \cdot \|f\|.$$

Thus, $\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=0}^{N-1} U^n h \right\| \leq \lim_{N\to\infty} \frac{2}{N}\cdot\|f\| = 0.$

Exercise Show that this holds for any $h \in \overline{\{Uf - f \mid f \in \mathcal{H}\}}$.

Assuming the exercise, we are essentially done. For any $f \in \mathcal{H}$, write $f = f_1 + f_2$ for $\underbrace{f_1}_{\boxed{f_1 = Pf}^{(*)}} \in \mathcal{H}_0$ and $f_2 \in \mathcal{H}_0^\perp$.

Then $\frac{1}{N}\sum_{n=0}^{N-1} U^n f = \frac{1}{N}\sum_{n=0}^{N-1} U^n f_1 + \frac{1}{N}\sum_{n=0}^{N-1} U^n f_2$.

Thus, $\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=0}^{N-1} U^n f \right\| = \|f_1\| = \|Pf\|$, as desired.

$\boxtimes$

A few additional remarks are in order. First, the expression $\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n(x) = \frac{1}{N}\sum_{n=0}^{N-1} U_T^n f(x)$ is often called an ergodic average. Second, as hinted at in the proof, the subspace $\{Uf - f \mid f \in \mathcal{H}\}$ is a natural subspace to study. Functions of this form are called cocycles, and the point of the proof is that cocycles "decay." In fact, the rate at which the cocycles decay is tied to a quantitative form of the theorem (eg. a "quantitative" or "effective" mean ergodic theorem for iterates of a particular system, a theorem that would capture an "error rate" associated to the convergence to the ergodic average). Some well-known study precisely this, for instance, Forni's work "On the deviation of ergodic averages..."

Let's return to the main story. We have proven a special case of the Mean Ergodic Theorem (THM 1), which is all we will need in this course (all of our ergodic transformations will be invertible).

<u>Exercise</u> Attempt a proof of the general case of the theorem (THM 1) where U_T is not necessarily unitary, but <u>is</u> an isometry. (Alternatively, see Einsiedler-Ward ______ for the necessary details.)

Now that we've seen a proof of the Mean Ergodic theorem, it makes sense to see the full statement of the pointwise ergodic theorem (all bells + whistles included).

<u>THM</u> (Pointwise Ergodic Theorem, Birkhoff)
Let $T: X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. If $f \in L^1(X, \mu)$, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\circ T^n(x) = f^*(x)$$

converges almost everywhere, and in $L^1(X,\mu)$, to a T-invariant function $f^* \in L^1(X,\mu)$, where

$$\int_X f^* d\mu = \int_X f\, d\mu .$$

Moreover, if T is ergodic, then

$$f^*(x) = \int_X f\, d\mu \qquad \xrightarrow{(*)} \qquad \left(\text{and } \int_X f^* d\mu = \int_X \left(\int_X f\,d\mu\right) d\mu = \int_X f\,d\mu \cdot \int_X d\mu = \int_X f\,d\mu. \right)$$

almost everywhere.

For a proof of this theorem, see Einsiedler-Ward §2.6 or Bekka-Mayer THM 2.5.

We move on to slightly finer variants of ergodic systems: Mixing, k-fold Mixing, mixing of all orders, and weak-mixing. Each of these provides a more refined description of how $T^{-n}(E)$ moves (or "spreads") around the space.

DEF A measure-preserving transformation $T: X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is mixing if for any $A, B \in \mathcal{B}$,

$$\mu(A \cap T^{-n}B) \longrightarrow \mu(A) \cdot \mu(B)$$

as $n \to \infty$.

Recall the notion of independence in probability: two sets $A, B \in \mathcal{B}$ are independent if

$$\mu(A \cap B) = \mu(A) \cdot \mu(B).$$

In the above definition of mixing, we can see how one might say that the sets $T^{-n}B$ and A become "asymptotically independent." In fact, this perspective is helpful when studying dynamical systems through the lens of probability.

As with ergodicity, mixing has the following "functional interpretation."

Prop $T: X \to X$ is mixing on a probability space $(X, \mathcal{B}, \mu)$ if and only if

$$\lim_{n \to \infty} \overbrace{\langle U_T^n f, g \rangle}^{\int_X (f \circ T^n) \cdot g \, d\mu} \longrightarrow \int_X f \, d\mu \cdot \int_X g \, d\mu$$

for all f, g lying in a dense subset of $L^2(X, \mu)$.

NOTE Also holds for $f = g$ above! i.e. $\langle U_T^n f, f \rangle \xrightarrow{n \to \infty} \left(\int_X f \, d\mu \right)^2$ is equivalent to mixing!

Exercise Prove the prior proposition.

DEF A measure-preserving transformation $T: X \longrightarrow X$ on a probability space $(X, \mathcal{B}, \mu)$ is called mixing of order k if for any $A_0, A_1, \ldots, A_k \in \mathcal{B}$,

$$\mu(A_0 \cap T^{-n_1} A_1 \cap \cdots \cap T^{-n_k} A_k) \longrightarrow \mu(A_0) \cdot \cdots \cdot \mu(A_k)$$

as

$$n_1,\ n_2 - n_1,\ n_3 - n_2,\ n_4 - n_3,\ \ldots,\ n_k - n_1 \longrightarrow \infty.$$

Note that mixing of order 1 is just mixing. We say T is mixing of all orders if T is mixing of order k for all $k \in \mathbb{N}$.

NOTE One can show that mixing of order k implies mixing of order $k-1$ (take $A_k = X$, or a full measure set invariant under T). However, it is still an open problem in ergodic theory as to whether or not mixing implies mixing of all orders!

Exercise Take a gander at the "functional" version of mixing of order k.

Lastly, we have weak-mixing.

<u>DEF</u> A measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is called <u>weak-mixing</u> if for any $A, B \in \mathcal{B}$,

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \mu(A \cap T^{-n}B) - \mu(A) \cdot \mu(B) \right| \longrightarrow 0$$

as $N \to \infty$.

i.e., instead of mixing, we have a sort of "mixing on average" property, as in, we don't require $\left|\mu(A \cap T^{-n}B) - \mu(A) \cdot \mu(B)\right| \to 0$, we only require that the averages of this sequence converge to 0. Clearly, this is a weaker condition than mixing since

$\lim_{n \to \infty} a_n \to 0$ implies $\lim_{n \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n| = 0$, but

$\lim_{n \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n| = 0$ does not imply $\lim_{n \to \infty} a_n = 0$. (Note the extra $\frac{1}{N}$ which causes this!)

There are several equivalent notions of weak-mixing, see, for example, Einsiedler-Ward THM 2.36. We will mention three of them here:

Prop Let $T: X \longrightarrow X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. TFAE

1. T is weak-mixing.
2. $T \times T$ is ergodic on the product space $X \times X$ with measure $\mu \times \mu$.
3. $T \times T$ is weak-mixing on the product space $X \times X$ with measure $\mu \times \mu$.
4. The associated operator U_T has no non-constant measurable eigenfunctions (i.e., T has continuous spectrum).

We also have the following "functional" interpretation for weak-mixing (in addition to ④ above).

Prop Let $T: X \longrightarrow X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Then T is weak-mixing if and only if for any $f, g \in L^2(X, \mu)$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \underbrace{\langle U_T^n f, g \rangle}_{\int_X (f \circ T^n) \cdot g \, d\mu} - \int_X f \, d\mu \cdot \int_X g \, d\mu \Big| = 0.$$

EXERCISE Prove the prior proposition.

We will see examples of weak-mixing in the group setting, soon. For now, we will focus on an example which is mixing: hyperbolic toral automorphisms.

THM Let $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ and let A be an $n \times n$ matrix with integer coefficients, determinant 1, and no eigenvalues on the unit circle in $\mathbb{C}$. The linear map $A: \mathbb{R}^n \to \mathbb{R}^n$ preserves the integer lattice and descends to a map $A: \mathbb{T}^n \longrightarrow \mathbb{T}^n$. The map on the quotient

(1) preserves the Lebesgue measure

and

(2) is mixing.

pf (using Fourier Analysis):

First, we show that the Lebesgue measure is preserved under iterates of A.

Let $f \in L^2(\mathbb{T}^n, \overset{(\text{Leb})}{m})$ and recall that $\{e^{2\pi i \langle k,x\rangle}\}_{k\in\mathbb{Z}^n}$ ($\langle \cdot, \cdot \rangle :=$ Euclidean inner product) forms a basis for $L^2(\mathbb{T}^n, dm)$. Let

$$f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i \langle k,x\rangle},$$

where c_k is the k^{th} Fourier coefficient.

Recall from the first lecture that A is measure-preserving if and only if for any $f \in L^2(\mathbb{T}^n, m)$,

$$\int_{\mathbb{T}^n} f(Ax)\, dm = \int_{\mathbb{T}^n} f\, dm. \qquad (\text{i.e. } A_* m = m)$$

Consider the following.

$$\int_{\mathbb{T}^n} f(Ax)\, dm = \sum_{k \in \mathbb{Z}^n} c_k \int_{\mathbb{T}^n} e^{2\pi i \langle k, Ax \rangle}\, dm$$

A^t is the transpose of A.

$$= \sum_{k \in \mathbb{Z}^n} c_k \int_{\mathbb{T}^n} e^{2\pi i \langle A^t k, x \rangle}\, dm$$

$$= c_0,$$

where the last line is justified as follows:

1. $A^t k \in \mathbb{Z}^n$ for any $k \in \mathbb{Z}^n$. (If A has integer entries, so does its transpose!)

2. $\int_{\mathbb{T}^n} e^{2\pi i \langle A^t k, x \rangle}\, dm = 0$ unless $A^t k = 0$.

   Since $\det(A) = 1$, $\det(A^t) = 1$, and $k = 0$ is the only solution.

3. $\int_{\mathbb{T}^n} e^{2\pi i \langle 0, x \rangle}\, dm = 1$.

Now, recall what the c_k-Fourier coefficients are:

$$c_k = \int_{\mathbb{T}^n} f(x)\, e^{-2\pi i \langle k, x\rangle}\, dm .$$

Thus, $c_0 = \int_{\mathbb{T}^n} f(x)\, dm$, and we have

$$\int_{\mathbb{T}^n} f(Ax)\, dm = c_0 = \int_{\mathbb{T}^n} f(x)\, dm ,$$

completing the proof that A preserves the Lebesgue measure.

Next, we show that $A : \mathbb{T}^n \longrightarrow \mathbb{T}^n$ is mixing. Recall that A mixing is equivalent to

$$\langle f \circ A^n, f \rangle \xrightarrow{n \to \infty} \left(\int_X f\, d\mu\right) \cdot \left(\int_X f\, d\mu\right)$$

for any $f \in L^2(\mathbb{T}^n, dm)$. Consider

$$\langle f \circ A^n, f \rangle = \int_{\mathbb{T}^n} \left(\sum_{k \in \mathbb{Z}} c_k\, e^{2\pi i \langle k, A^n x\rangle}\right) \overline{\left(\sum_{\ell \in \mathbb{Z}} c_\ell\, e^{2\pi i \langle \ell, x\rangle}\right)} dm .$$

This breaks into a sum of many integrals. We need to understand integrals of the form

absorb the negative into ℓ (coming from the conjugate!)

$$\int_{\mathbb{T}^n} e^{2\pi i \langle k, A^n x\rangle}\, e^{2\pi i \langle \ell, x\rangle}\, dm \qquad \text{for } k, \ell \in \mathbb{Z}^n .$$

Note

$$\int_{\mathbb{T}^n} e^{2\pi i \langle k, A^n x\rangle} e^{2\pi i \langle l, x\rangle} \, dm = \int_{\mathbb{T}^n} e^{2\pi i (\langle (A^n)^t k, x\rangle + \langle l, x\rangle)} \, dm$$

$$= \int_{\mathbb{T}^n} e^{2\pi i \langle (A^t)^n k + l, x\rangle} \, dm.$$

This is non-zero if and only if $(A^t)^n k + l = 0$. Our interest is when $n \to \infty$. For the limit of these integrals to be non-zero we would need there to be distinct integers $n_1 < n_2 < \dots$ such that

$$(A^t)^{n_1} k = -l \quad \rightsquigarrow \quad ①\ k = (A^t)^{-n_1} \cdot (-l)$$

$$②\ (A^t)^{n_2} k = -l$$

$$\vdots$$

meaning, by plugging ① into ②,

$$(A^t)^{n_2} \left((A^t)^{-n_1} \cdot (-l) \right) = -l$$

which reduces to

$$(A^t)^{n_2 - n_1} \, l = l.$$

This means $(A^t)^{n_2-n_1}$ has 1 as an eigenvalue unless $l = 0$, and so

A^t must have a root of unity as an eigenvalue, unless $l=0$.

<u>Linear Algebra fact:</u> A and A^t have the same characteristic polynomial, and, hence, the same eigenvalues.

This means A must have a root of unity as an eigenvalue unless $l=0$. By our hypothesis, we see that $l=0$.

However, this means that $(A^t)^{n_1} k = -l = 0$, so $k \in \ker((A^t)^{n_1})$. Since A is invertible, so is A^t, and $(A^t)^{n_1}$.

Consequently, $k=0$ also.

As $n \to \infty$, the only non-zero term is when $k=l=0$:

$$\int c_0 e^{2\pi i \langle 0, x\rangle} \cdot c_0 e^{2\pi i \langle 0, x\rangle} dx$$

$$= c_0 \cdot c_0$$

$$= \left(\int_X f \, d\mu\right) \cdot \left(\int_X f \, d\mu\right).$$

$\boxtimes$