TCC: DYNAMICS ON LIE GROUPS

Overview

This course is intended to serve as an introduction to dynamics on Lie groups, where throughout the course our main focus will be on the group $\mathrm{SL}_2\mathbb{R}$. We will start with a brief introduction to measurable dynamics (ergodic theory).

By the end of the second lecture, I hope we will have developed all of the necessary (ergodic-theoretic) definitions and theorems, not just in the context of iterates of a single map, but for group actions on a probability space. We will move from here to the definition of a Lie group and its corresponding Lie algebra. At this point, we should be prepared to study

(1) the action of the diagonal subgroup of $\mathrm{SL}_2\mathbb{R}$ on $\mathrm{SL}_2\mathbb{R}/\Gamma$, where Γ is a lattice in $\mathrm{SL}_2\mathbb{R}$. (The definition of a lattice will come, but you may think that $\Gamma \subset \mathrm{SL}_2\mathbb{R}$ is something like $\mathbb{Z}^2 \subset \mathbb{R}^2$. Note that $\mathbb{R}^2/\mathbb{Z}^2$ has "finite volume" for a natural choice of volume.)

(2) the action of a unipotent subgroup of $\mathrm{SL}_2\mathbb{R}$ on $\mathrm{SL}_2\mathbb{R}/\Gamma$ (think "horocycle flow").

(3) the action of $\mathrm{SL}_2\mathbb{R}$ on $\mathrm{SL}_2\mathbb{R}/\Gamma$.

(4) and eventually, the action of a lattice $\Gamma \subset \mathrm{SL}_2\mathbb{R}$ on an arbitrary probability space.

There will be a few key technical high points in the course, including (but perhaps not limited to):

(1) Hopf Argument

(2) Mautner phenomenon

(3) Howe–Moore Theorem (vanishing of matrix coefficients).

The arguments we give may hint at a well-developed theory of harmonic analysis, at least in the setting of $\mathrm{SL}_2\mathbb{R}$, but that will not be the focus of this course.

I plan to give at least one key application of the Howe–Moore Theorem, which will involve two things:

(1) "Equidistribution of circles" in $\mathbb{H}/\Gamma$.

(2) The hyperbolic lattice point counting problem (non-effective, at least at first).

Time permitting, we may talk a bit about how to make some of these results "effective" or "quantitative," by which I mean understanding error terms associated with ergodic theorems, theorems that show convergence in some way, and asymptotics.

1. Lecture 1: Recurrence and Ergodicity

We start with a definition, a few examples, and a key property that motivates the study of dynamics.

Definition 1.1. A *measure-preserving transformation* on a probability space $(X, \mathcal{B}, \mu)$ is a measurable map $T : X \to X$ such that for any $E \in \mathcal{B}$,

$$\mu(T^{-1}(E)) = \mu(E).$$

Remark 1.2.

(1) Equivalently, $T_*\mu = \mu$ (the pushforward of the measure μ by T is μ). This means, equivalently, that for any $f \in L^2(X, \mu)$,

$$\int_X f(Tx)\, d\mu = \int_X f\, d(T_*\mu) = \int_X f\, d\mu.$$

(See Einsiedler–Ward, Lemma 2.6.)

(2) We don't really need $(X, \mathcal{B}, \mu)$ to be a probability space, but since our dynamics will be happening on a measure space with finite measure, we may as well focus on probability spaces. (We can get from a finite measure to a measure μ s.t. $\mu(X) = 1$ by renormalizing.)

Examples.

Example 1.3 (Circle Rotations)**.** For any $\alpha \in \mathbb{R}$, define the *circle rotation* R_α by

$$R_\alpha : S^1 \longrightarrow S^1, \qquad e^{2\pi i\theta} \longmapsto e^{2\pi i(\theta+\alpha)},$$

where $d\theta$ is the angle measure. (Or, if you prefer,

$$R_\alpha : \mathbb{R}/\mathbb{Z} \longrightarrow \mathbb{R}/\mathbb{Z}, \qquad t \longmapsto t + \alpha \pmod 1,$$

where dm is Lebesgue measure.) These maps preserve the natural measures on the space (Lebesgue).

Exercise 1.4. Show this! Hint: Consider $\int_{\mathbb{R}/\mathbb{Z}} f \circ R_\alpha\, dm$. Expand f using a Fourier series, and deduce that $\int_{\mathbb{R}/\mathbb{Z}} f \circ R_\alpha\, dm = \int_{\mathbb{R}/\mathbb{Z}} f\, dm$, i.e. $(R_\alpha)_* m = m$.

Example 1.5. Observe that $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} \in \mathbb{Z}^2$ for $\begin{pmatrix} m \\ n \end{pmatrix} \in \mathbb{Z}^2$. So $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ preserves the integer lattice, and it descends to a map on $\mathbb{R}^2/\mathbb{Z}^2 \cong \mathbb{T}^2$. Since $\det \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = 1$, this map preserves the Lebesgue measure on $\mathbb{T}^2$.

Example 1.6 (2-times map)**.** The 2-*times map* on $\mathbb{R}/\mathbb{Z} \cong \mathbb{T}^1$:

$$T_2 : \mathbb{T}^1 \longrightarrow \mathbb{T}^1, \qquad x \longmapsto 2x \pmod{\mathbb{Z}}.$$

Note: Surjective, not injective, but still preserves the Lebesgue measure! (We are interested in pre-images. For example,

$$T_2^{-1}\big([0, \tfrac{1}{2}]\big) = \big[0, \tfrac{1}{4}\big] \sqcup \big[\tfrac{1}{2}, \tfrac{3}{4}\big]\,.)$$

Exercise 1.7. Show this measure is preserved!

Example 1.8 (3-IET)**.** Divide $[0, 1]$ into three pieces a, b, c of lengths ℓ_1, ℓ_2, ℓ_3 where $\ell_1 + \ell_2 + \ell_3 = 1$. The map T_3 permutes the pieces in a specified way, sending $(a, b, c) \to (b, c, a)$. We can write T_3 as a piecewise function, where each piece is just a translation. This preserves the Lebesgue measure on $[0, 1]$.

Exercise 1.9. Using an analogous definition, show that a 2-IET is a circle rotation. (Lebesgue is invariant under translations.)

Example 1.10 (Bernoulli Shifts)**.** See, for example, Einsiedler–Ward, Examples 2.8, 2.9.

Motivation. A key question in dynamics revolves around understanding the orbits of points under a transformation. Measurable dynamics seeks to address these questions through measures on the space, whereas topological dynamics seeks to answer these questions through topological properties (e.g., dense orbits?). Our focus will be primarily developing the measurable perspective (ergodic theory). We start with a foundational result in measurable dynamics, which is akin to a "pigeon-hole principle" for measure-preserving transformations on a probability space.

Theorem 1.11 (Poincaré Recurrence)**.** *Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ and let $E \in \mathcal{B}$ be any measurable set. Then a.e. $x \in E$ returns to E infinitely often: there exists $F^{\text{meas}} \subseteq E$ with $\mu(F) = \mu(E)$ such that for every $x \in F$, there exist integers $0 < n_1 < n_2 < \cdots$ such that $T^{n_i}(x) \in E$ for all $i \geq 1$.*

In other words, the "finite measure" space is "small" enough that almost all points in measurable sets recur (to the starting set) infinitely often under the transformation T.

Proof. Let $B = \{x \in E \mid T^n(x) \notin E \; \forall \, n \geq 1\}$, the set of points in E that *never* return (the "bad set"). We can write B as

$$B = E \cap T^{-1}(X \setminus E) \cap T^{-2}(X \setminus E) \cap \cdots$$

Thus, since B is the intersection of measurable sets, B is measurable.

Consider the following sets for $n \geq 1$:

$$T^{-n}(B) = T^{-n}(E) \cap T^{-(n-1)}(X \setminus E) \cap \cdots$$

(points in E after n iterates, but never return to E again; i.e., $T^n(x)$ is the last time we land in E.)

Notice:

(1) $\mu(T^{-n}(B)) = \mu(B)$ since T is measure-preserving.

(2) $\mu(B)$, $\mu(T^{-1}(B))$, $\ldots$, $\mu(T^{-n}(B))$, $\ldots$ are all disjoint. Indeed, if $x \in T^{-n}(B) \cap T^{-k}(B)$, then for all $n, k \geq 0$ with $k < n$, $x \in T^{-k}(B) \Rightarrow x \notin T^{-n}(B)$ ($T^k(x)$ is the last time in E, and $T^n(x) \notin E$).

Then, consider that $\bigsqcup_{n \geq 0} T^{-n}(B) \subseteq X$. We have

$$1 = \mu(X) \geq \mu\left(\bigsqcup_{n \geq 0} T^{-n}(B)\right) \overset{(2)}{=} \sum_{n=0}^{\infty} \mu(T^{-n}(B)) \overset{(1)}{=} \sum_{n=0}^{\infty} \mu(B),$$

and we can conclude $\mu(B) = 0$, i.e., the set of points that *never* return has measure 0. Consequently, the measure of points that return at least once has full measure in E. Call this set F_1, where $\mu(F_1) = \mu(E)$.

Apply the same argument to the measure-preserving transformations $T^2, T^3, \ldots$ and get a full measure set for each map. Call these $F_2, F_3, \ldots \subseteq E$, where

$$\mu(F_2) = \mu(F_3) = \cdots = \mu(E).$$

Consider

$$F := \bigcap_{n \geq 1} F_n \subseteq E.$$

For any $x \in F$, there exists a sequence of positive integers $k_1, k_2, \ldots$ such that

$$x \in F \iff \begin{cases} x \in E, \\ T^{k_1}(x) \in E, \\ T^{2k_2}(x) \in E, \\ \quad\vdots \\ T^{nk_n}(x) \in E, \\ \quad\vdots \end{cases}$$

The claim is that the set $\{nk_n\}_{n\in\mathbb{N}}$ contains infinitely many distinct integers. Assume otherwise. Then there exists $N \in \mathbb{N}$ such that $nk_n \leq N$ for any $n \in \mathbb{N}$. Pick $n > N$. Since k_n is a positive integer, $nk_n > Nk_n \geq N$, a contradiction.

So points in F land back in E infinitely often. Lastly,

$$\mu(F) = \mu\left(\bigcap_{n\in\mathbb{N}} F_n\right) = \mu(E),$$

since each F_n has full measure in E. To see this, consider

$$\mu\left(\bigcup_{n\geq 1}(E \setminus F_n)\right) \leq \sum_{n\geq 1}\mu(E \setminus F_n) = 0,$$

and apply De Morgan's.) □

Poincaré recurrence tells us that in this sufficiently constrained setting, orbits of points are interesting. Already, with few assumptions on the map, we know that a.e. point in a measurable set comes back to that set infinitely often. Naturally, we can ask many more questions. For example:

(1) If a.e. $x \in E$ recurs under T infinitely often, does any $x \in X$ hit the set E infinitely often? What is special about constraining ourselves to E?

(2) If $x \in E$ recurs under T infinitely often, can we say precisely how often? What is the "return time" to E for each point in the full measure set F? Is it uniformly bounded? Above? Below? Are there certain points in the set F which take unbearably long to return to E? (Interpret "unbearable" however you wish.)

In some sense, it should be obvious that the answers to such questions depend on the map T. However, for a large class of maps, we can clarify (1). Consider the following examples which highlight a few ways (1) can fail:

- $R_{1/2} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, $x \mapsto x + \frac{1}{2} \pmod{\mathbb{Z}}$. Each orbit has 2 points! In fact, each point has a period of 2, meaning $T^2(x) = x$ for all $x \in \mathbb{R}/\mathbb{Z}$. So while every $x \in E^{\text{meas}} \subseteq \mathbb{R}/\mathbb{Z}$ recurs to E (to itself!) infinitely often, we can have points $y \notin E$ which never hit E!
- $R_\pi \times R_{2\pi} : \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$, $(x, y) \mapsto (x + \pi,\, y + 2\pi) \pmod{\mathbb{Z}^2}$. It seems like points in each component move around a lot, but pairs of points seem trapped in a smaller set.

We need a definition which amounts to saying when a measure-preserving transformation is "indecomposable," i.e., points in measurable sets aren't stuck in these sets. (Otherwise, we may as well decompose the dynamical system and study the behavior of orbits in smaller sets!)

Definition 1.12. A measure-preserving transformation $T : X \to X$ of a probability space $(X, \mathcal{B}, \mu)$ is *ergodic* if for any $B \in \mathcal{B}$,

$$T^{-1}(B) = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1.$$

Proposition 1.13. *The following are equivalent:*

(1) T is ergodic.

(2) For all $B \in \mathcal{B}$, $\mu(T^{-1}(B)\triangle B) = 0 \implies \mu(B) = 0$ or $\mu(B) = 1$.

(3) For $A \in \mathcal{B}$, $\mu(A) > 0 \implies \mu(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$, i.e., you can "move things around enough to cover" X.

(4) For $A, B \in \mathcal{B}$, $\mu(A), \mu(B) > 0 \implies \exists\, n \geq 1$ such that $\mu(T^{-n}(A) \cap B) > 0$, i.e., for two positive measure sets, iterates of one must eventually "overlap" the other in a non-trivial (measure-theoretic) way.

(5) For $f : X \to \mathbb{C}$ such that $f \circ T = f$ a.e., f must be equal to a constant a.e.

Proof. See Einsiedler–Ward, §2.3.

Note: (1) $\Leftrightarrow$ (5) is a bootstrapping exercise. Ergodicity is a statement about invariant sets, and we can write

$$T^{-1}(B) = B \implies \int_X \chi_B \, d\mu = 0 \text{ or } \int_X \chi_B \, d\mu = 1.$$

$\square$

Proposition 1.14. *$R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is ergodic w.r.t. the Lebesgue measure m if and only if α is irrational.*

Proof. First, let $\alpha \in \mathbb{Q}$, $\alpha = p/q$ in lowest terms. Then $R_\alpha^q = \mathrm{Id}$. Consider the set $A = [0, \frac{1}{2q})$. Let $B = A \cup R_\alpha(A) \cup R_\alpha^2(A) \cup \cdots \cup R_\alpha^{q-1}(A)$. Then B is measurable, B is invariant under R_α, and $m(B) = \frac{1}{2}$. Since $m(B) \neq 0$ or 1, R_α is not ergodic.

Now assume $\alpha \notin \mathbb{Q}$. Let $f \in L^2(\mathbb{R}/\mathbb{Z}, m)$ be a function invariant under R_α. By Prop (5), it suffices to show f is constant a.e.

Assume $f \circ R_\alpha = f$. Let $f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k x}$ in $L^2(\mathbb{R}/\mathbb{Z}, m)$. Then,

$$f \circ R_\alpha(x) = f(x + \alpha) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k x} \cdot e^{2\pi i k \alpha}.$$

By assumption,

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

Thus, since the Fourier transform is an (bijective) isometry on L^2, we need

$$c_k = c_k e^{2\pi i k \alpha} \quad \forall\, k \in \mathbb{Z},$$

i.e., we need $e^{2\pi i k \alpha} = 1$. But since $\alpha \notin \mathbb{Q}$, $e^{2\pi i k \alpha} = 1$ iff $k = 0$. So to have equality, $c_k = 0$ for all $k \neq 0$, and we see that $f(x) = c_0$ in L^2. Thus $f = c_0$ a.e. $\square$

Exercise 1.15. Using a Fourier series, show that when $\alpha \in \mathbb{Q}$, R_α is not ergodic.

As it turns out, ergodicity gives us much more than suggested by the Proposition. In some sense, the definition of ergodic coupled with the Proposition tells us that "an ergodic map moves points around substantially." But how substantially? The following two theorems give us a bit of insight.

Theorem 1.16 (Mean Ergodic Theorem, von Neumann)**.** *Assume $T : X \to X$ is ergodic, where $(X, \mathcal{B}, \mu)$ is a probability space. Then, for any $f \in L^2(X, \mu)$,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \xrightarrow[L^2]{N\to\infty} \left(\int_X f \, d\mu \right) \cdot \mathbf{1}.$$

Theorem 1.17 (Pointwise Ergodic Theorem, Birkhoff)**.** *Assume $T : X \to X$ is ergodic, where $(X, \mathcal{B}, \mu)$ is a probability space. Then, for any $f \in L^1(X, \mu)$,*

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n(x) \xrightarrow{N\to\infty} \int_X f\, d\mu$$

for a.e. $x \in X$.

The first theorem is an L^2-version of the second theorem, and while the second theorem is much harder to prove, it turns out that in most applications the first theorem is sufficient. Indeed, we will see as the course progresses that it is often sufficient to work in the L^2 setting.

Regardless, the theorems tell us the following: for an ergodic transformation $T : X \to X$, the time averages of a function $f \in L^2(X, \mu)$ (or even $L^1(X, \mu)$) converge in some precise sense to the space average of the function:

- "time average" of f : $\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n(x)$,
- "space average" of f : $\int_X f\, d\mu$,

i.e., T moves points around enough such that the average $\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n$ eventually approximates the average of the function over the entire space, $\int_X f\, d\mu$.

Using the pointwise ergodic theorem, we can show how the assumption that T is ergodic allows us to answer question (1) (following Poincaré Recurrence) positively. Indeed, fix any measurable set $E \in \mathcal{B}$. Then

$$\frac{1}{N}\sum_{n=0}^{N-1} \chi_E \circ T^n(x) \xrightarrow{\text{a.e.}} \int_X \chi_E\, d\mu = \mu(E).$$

Notice that $\chi_E \circ T^n(x) = \chi_{T^{-n}(E)}(x)$, so $\frac{1}{N}\sum_{n=0}^{N-1} \chi_{T^{-n}(E)}(x)$ is an average of the number of times that x lands in E (vs. misses). In other words, the pointwise ergodic theorem tells us that a.e. $x \in X$ lands in E infinitely often, and we have some measure of "how often": x lands in E "$\mu(E)$" of the time. The proportion of iterates that land in E amounts to the measure of the set E as we let the number of iterates tend to infinity.

We will not give a proof of the pointwise ergodic theorem, but you can find one in Einsiedler–Ward (§2.6) or Bekka–Mayer (Thm 2.5). However, we will give a proof of the Mean Ergodic Theorem in the next lecture, or at least a version of it.

To prepare for this proof, the remainder of this lecture will be devoted to explaining how we interpret a measure-preserving transformation as an operator on functions. This will enable us to more readily work in function spaces later in the course.

Fix a measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$. Define an operator $U_T : L^2(X, \mu) \to L^2(X, \mu)$ by

$$U_T(f) = f \circ T \qquad \forall\, f \in L^2(X, \mu).$$

We call this the *Koopman operator* or *associated operator.*

Proposition 1.18. *Because T is measure-preserving, U_T is an isometry.*

Proof. Indeed, for any $f_1, f_2 \in L^2(X, \mu)$,

$$\begin{aligned}\langle U_T f_1,\, U_T f_2\rangle &= \int_X f_1(Tx)\, \overline{f_2(Tx)}\, d\mu(x) \\ &= \int_X f_1(y)\, \overline{f_2(y)}\, d\mu(y) \qquad (\text{let } y = T(x),\ d\mu(y) = d\mu(T(x)) = d\mu(x)) \\ &= \langle f_1,\, f_2\rangle .\end{aligned}$$

Note that for $f = f_1 = f_2$, we have $\|U_T f\|_2 = \|f\|_2$. ∎

Definition 1.19. $T : X \to X$ is an *invertible measure-preserving transformation* if T is a measure-preserving transformation, T^{-1} exists a.e., and T^{-1} is a.e. equivalent to a measurable function. (Extend T^{-1} to points where it is not defined however you wish, and this extension must be measurable.)

Proposition 1.20. *If T is an invertible measure-preserving transformation, then U_T is a unitary operator.*

Recall, an operator U is unitary if $U^* = U^{-1}$, where U^* is the adjoint, and the adjoint is defined by

$$\langle Uf_1,\, f_2\rangle = \langle f_1,\, U^* f_2\rangle \qquad \forall\, f_1, f_2 \in L^2(X,\mu).$$

Proof. For T invertible, fix any $f_1, f_2 \in L^2(X,\mu)$. Then

$$\begin{aligned}\langle U_T f_1,\, f_2\rangle &= \int_X f_1(Tx)\,\overline{f_2(x)}\,d\mu(x)\\ &= \int_X f_1(y)\,\overline{f_2(T^{-1}(y))}\,d\mu(y) \qquad (y = T(x) \Leftrightarrow x = T^{-1}(y))\\ &= \langle f_1,\, U_{T^{-1}} f_2\rangle\,.\end{aligned}$$

Thus $U_T^* = U_{T^{-1}}$. Now observe that for all $f \in L^2(X,\mu)$:

(1) $U_{T^{-1}} \circ U_T(f) = f$,

(2) $U_T \circ U_{T^{-1}}(f) = f$.

This means that $U_{T^{-1}} = (U_T)^{-1}$, and we have that

$$U_T^* = (U_T)^{-1},$$

so U_T is unitary. □

In the case where T is invertible, we call U_T the *associated unitary operator* (or just the *Koopman operator*).

It is natural to ask about the spectral properties of the operator U_T: how does the behavior of T, or rather, properties of T, influence the properties of the operator U_T? Here is a sample lemma that gives a partial answer to this question.

Lemma 1.21. *Assume $T : X \to X$ is a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. T is ergodic if and only if 1 is a simple eigenvalue of the associated operator U_T.*

Proof. We have $U_T(f) = f \overset{\text{def}}{\iff} f \circ T = f$. Since T is ergodic and by Prop (5), this $\iff$ f is constant a.e. Constants are a 1-dimensional $\mathbb{C}$-subspace in $L^2(X,\mu)$, namely $\mathbb{C} \cdot \mathbf{1}$. □

Looking Ahead. In the next lecture, we will start with a proof of the Mean Ergodic Theorem, and then explore the notions of mixing and weak-mixing transformations, which are ergodic transformations with stronger dynamical properties. During our discussion of mixing, we'll encounter the Hopf argument. Time permitting, we will begin upgrading all of our definitions to the setting of group actions.

2. Lecture 2: Ergodicity, Mixing, Weak-mixing

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

Theorem 2.1 (Mean Ergodic Theorem, von Neumann)**.** *Assume $T : X \to X$ is an ergodic measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Then for any $f \in L^2(X, \mu)$,*

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f \xrightarrow[L^2]{N\to\infty} \left(\int_X f\,d\mu\right)\cdot \mathbf{1},$$

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.

Theorem 2.2 (Abstract Ergodic Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space, let $U : \mathcal{H} \to \mathcal{H}$ 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 $P : \mathcal{H} \to \mathcal{H}_0$ 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 Theorem 2.2 implies a special case of Theorem 2.1 — when T is invertible. Indeed, T invertible implies U_T is unitary, which is the setting of Theorem 2.2. More concretely,

Proof (Theorem 2.2 $\Rightarrow$ special case of Theorem 2.1, assuming T invertible). Let $\mathcal{H} = L^2(X, \mu)$ and let $U := U_T$. Then

$$\mathcal{H}_0 = \{f \in L^2(X,\mu) : U_T(f) = f\} = \mathbb{C}\cdot\mathbf{1} \qquad \text{(constant functions, since } T \text{ ergodic).}$$

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

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

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

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

(2) $\langle f - Pf, g\rangle = 0$ for all $g \in \mathbb{C}\cdot\mathbf{1}$:

Assume $g \in \mathbb{C}\cdot\mathbf{1}$, so $g = c\cdot\mathbf{1}$ for some $c \in \mathbb{C}$. Then

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

Thus Theorem 2.1 when T is invertible follows from Theorem 2.2. □

Now we turn our attention to proving Theorem 2.2.

Proof of Theorem 2.2. 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

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

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 $U^* \circ Uv = U^* v$ 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 $f \in \mathcal{H}$,

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

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

As desired, we have the decomposition

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp, \quad \text{where} \quad \mathcal{H}_0^\perp = \overline{\{Uf - f \mid f \in \mathcal{H}\}}.$$

Now, consider $h \in \{Uf - f \mid f \in \mathcal{H}\}$, i.e. $h = Uf - f$ for some $f \in \mathcal{H}$. We have

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1} U^n h &= \frac{1}{N}\sum_{n=0}^{N-1} U^n (Uf - f) \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \left(U^{n+1} f - U^n f\right) \\
&= \frac{1}{N}\left((Uf - f) + (U^2 f - Uf) + \cdots + (U^N f - U^{N-1} f)\right) \\
&= \frac{1}{N}(U^N f - f). \quad \text{(telescoping)}
\end{aligned}$$

Consider the norm of the sequence of averages:

$$\left\| \frac{1}{N}\sum_{n=0}^{N-1} U^n h \right\| \leq \frac{1}{N}\left(\left\|U^N f\right\| + \|f\|\right) = \frac{1}{N}(\|f\| + \|f\|) = \frac{2}{N}\|f\|. \quad (U \text{ unitary})$$

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

Exercise 2.3. 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 $f_1 \in \mathcal{H}_0$ (where $f_1 = Pf$) 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,

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

□

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 (e.g., 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 works 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 (Theorem 2.1), which is all we will need in this course (all of our ergodic transformations will be invertible).

Exercise 2.4. Attempt a proof of the general case of the theorem (Theorem 2.1) where U_T is not necessarily unitary, but is 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 and whistles included).

Theorem 2.5 (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 $f^*(x) = \int_X f\, d\mu$ almost everywhere.

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

2.1. **Mixing.** 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.

Definition 2.6. A measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is *mixing* if for any $A, B \in \mathcal{B}$,

$$\mu(A \cap T^{-n}B) \longrightarrow \mu(A)\cdot\mu(B) \quad \text{as } n \to \infty.$$

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."

Proposition 2.7. *$T : X \to X$ is mixing on a probability space $(X, \mathcal{B}, \mu)$ if and only if*

$$\lim_{n\to\infty} \langle U_T^n f,\, g\rangle = \int_X f\, d\mu \cdot \int_X g\, d\mu$$

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 2.8. Prove Proposition 2.7.

Definition 2.9. A measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is called *mixing of order* k if for any $A_0, A_1, \ldots, A_k \in \mathcal{B}$,

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

as $n_1,\ n_2 - n_1,\ n_3 - n_2,\ n_4 - n_3,\ \ldots,\ n_k - n_{k-1} \to \infty$.

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 $k \in \mathbb{N}$.

Remark 2.10. 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 2.11. Take a gander at the "functional" version of mixing of order k.

2.2. **Weak-mixing.**

Definition 2.12. A measure-preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is called *weak-mixing* if for any $A, B \in \mathcal{B}$,

$$\frac{1}{N} \sum_{n=0}^{N-1} \left|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\right| \longrightarrow 0 \quad \text{as } N \to \infty.$$

That is, instead of mixing, we have a sort of "mixing on average" property: we don't require $|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \to 0$; 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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |a_n| = 0 \quad \text{does not imply} \quad \lim_{n\to\infty} a_n = 0.$$

(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.

Proposition 2.13. *Let $T : X \to X$ 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).

Proposition 2.14. *Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Then T is weak-mixing if and only if for any $f, g \in L^2(X, \mu)$,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \langle U_T^n f,\, g \rangle - \int_X f \, d\mu \cdot \int_X g \, d\mu \right| = 0.$$

Exercise 2.15. Prove Proposition 2.14.

2.3. **Hyperbolic Toral Automorphisms.** 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.*

Theorem 2.16. *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 $A : \mathbb{T}^n \to \mathbb{T}^n$. The map on the quotient:*

(1) *preserves the Lebesgue measure, and*

(2) *is mixing.*

Proof (using Fourier Analysis). First, we show that the Lebesgue measure is preserved under iterates of A.

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

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

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:

$$\begin{aligned} \int_{\mathbb{T}^n} f(Ax) \, dm &= \sum_{k \in \mathbb{Z}^n} c_k \int_{\mathbb{T}^n} e^{2\pi i \langle k, Ax \rangle} \, dm \\ &= \sum_{k \in \mathbb{Z}^n} c_k \int_{\mathbb{T}^n} e^{2\pi i \langle A^t k, x \rangle} \, dm, \qquad (A^t \text{ is the transpose of } A) \end{aligned}$$

where this equals c_0, where the last equality 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 $A : \mathbb{T}^n \to \mathbb{T}^n$ 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

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

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

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

(absorbing the negative into ℓ coming from the conjugate). Note

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

This is non-zero if and only if $(A^t)^n k + \ell = 0$. 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 $n_1 < n_2 < \cdots$ such that

$$\begin{cases} (A^t)^{n_1} k = -\ell \quad \Rightarrow \quad k = (A^t)^{-n_1}(-\ell) \\ (A^t)^{n_2} k = -\ell \\ \vdots \end{cases}$$

Meaning, by plugging the first equation into the second after solving for k,

$$(A^t)^{n_2}\left((A^t)^{-n_1}(-\ell)\right) = -\ell,$$

which reduces to

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

This means $(A^t)^{n_2-n_1}$ has 1 as an eigenvalue unless $\ell = 0$, and so A^t must have a root of unity as an eigenvalue, unless $\ell = 0$.

Linear Algebra fact: 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 $\ell = 0$. By our hypothesis, we see that $\ell = 0$. However, this means that $(A^t)^{n_1} k = -\ell = 0$, 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 $k = \ell = 0$:

$$\int_{\mathbb{T}^n} c_0\, e^{2\pi i\langle 0, x\rangle} \cdot c_0\, e^{2\pi i\langle 0, x\rangle}\, dx = c_0 \cdot c_0 = \left(\int_X f\, d\mu\right) \cdot \left(\int_X f\, d\mu\right).$$

□