# ERGODIC THEORY WITH APPLICATION TO GEOMETRY

IAN BIRINGER

## 1. Contents and a disclaimer

These are notes I wrote for a topics class I ran in Fall 2023. There's a lot of basic ergodic theory of measure preserving transformations, including discussions of recurrence, ergodicity, ergodic theorems, the ergodic decomposition theorem, unique ergodicity, and mixing. After briefly discussing the ergodic theory of more general group actions, we transition to geometry, proving ergodicity of the geodesic flow on finite volume hyperbolic manifolds, presenting some applications of mixing to lattice point counting and to counting closed geodesics. Then we finish with a very brief sketch of the Kahn-Markovic surface subgroup theorem, and its relation to the Virtual Haken Conjecture.

I'm sure there are errors in the current version of these notes. Please let me know of any you find! In the last part of the notes, some of the arguments I give are meant to convey the basic ideas rather than any precise details, but even there, if you think my description of something is inaccurate I'd love to know.

## 2. Measure preserving maps

A *measurable space* is a set $X$ equipped with a $\sigma$-algebra $\Sigma$. A *measure* on $(X, \Sigma)$ is a function $\mu : \Sigma \longrightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever the $A_i$ are pairwise disjoint. A triple $(X, \Sigma, \mu)$ that is a measurable space equipped with a measure is called a *measure space*. We'll sometimes suppress $\Sigma$ in notation, writing measure spaces as $(X, \mu)$, and referring to elements of $\Sigma$ as *measurable sets*. When $X$ is a topological space, $\Sigma$ will always be the $\sigma$-algebra of Borel[1] sets, unless otherwise specified, and then $\mu$ is called a *Borel measure* on $X$. We call $(X, \Sigma, \mu)$ a *probability space* and $\mu$ a *probability measure* if $\mu(X) = 1$.

If $(X, \Sigma), (X', \Sigma')$ are measurable spaces, a function $T : X \longrightarrow X'$ is *measurable* if $T^{-1}(\Sigma') \subset \Sigma$. The map $T$ is called a *measurable isomorphism* if it is a bijection and its inverse is also measurable. If $(X, \Sigma, \mu), (X', \Sigma', \mu')$ are measure spaces, the map $T$ is called *measure preserving* (briefly, *m.p.*) if it is measurable and $\mu(T^{-1}(A')) = \mu'(A')$ for all $A' \in \Sigma'$. A bijection $T$ is a *measure isomorphism* if $T, T^{-1}$ are both measure-preserving. From another perspective, if $T$ is measurable, then we can push forward the measure $\mu$ by setting

$$T_*\mu(B') := \mu(T^{-1}(B'),$$

and then $T$ is m.p. if and only if $T_*\mu = \mu'$.

In this course, we are mostly interested in the dynamics under iteration of measure preserving self maps $T : (X, \mu) \longrightarrow (X, \mu)$ of a measure space.

[1] The 'Borel $\sigma$-algebra' is the smallest one containing all open sets, and includes all sets you're likely to explicitly construct in a proof.

**Example 2.1.** *Let $S^1 = \mathbb{Z}\backslash\mathbb{R}$ be the circle, endowed with its Lebesgue probability measure $\mu$. Given $\alpha \in \mathbb{R}$, let $T_\alpha : S^1 \longrightarrow S^1$, $T_\alpha([x]) = [x+\alpha]$. Geometricallly, $T_\alpha$ 'rotates' the circle, translating points along it a distance of $\alpha$.*

**Example 2.2.** *Given $m \in \mathbb{N}$, let $D_m : S^1 \longrightarrow S^1$, $T_\alpha([x]) = [mx]$, so that $D_m$ wraps $S^1$ around itself $m$ times. Perhaps surprisingly, $D_m$ is measure preserving: if $A \subset S^1$, then $D_m^{-1}(A)$ is a union of $m$ sets, each of which maps bijectively onto $A$, and each of which has measure $\frac{1}{m}\mu(A)$. Note that $D_m$ is not invertible.*

One can similarly define m.p. maps $T_\alpha$ and $D_m$ on the $n$-torus $T^n := \mathbb{Z}^n\backslash\mathbb{R}^n$. Also, for any $A \in GL(n,\mathbb{Z})$, the automorphism $\mathcal{T}_A : T^n \longrightarrow T^n$, where $\mathcal{T}_A([x]) = [Ax]$ is measure preserving, since $\det A = \pm 1$.

Suppose $(X,\mu)$ is a measure space and $T : X \longrightarrow X$ is measure preserving. Then for any $p = 1, 2, \ldots$ we get a linear map

$$T^* : L^p(X,\mu) \longrightarrow L^p(X,\mu), \quad T^*(f) = f \circ T.$$

Here, $L^p(X,\mu)$ is the set of $p$-integrable functions $X \longrightarrow \mathbb{R}$, up to almost everywhere equivalence, and considered as a Banach space with the norm $|f|_p := \left(\int f^p\right)^{1/p}$. In fact, $T^*$ is an isometry, which you can verify using the definition of the Lebesgue integral as a limit of integrals of simple functions, and the fact that $T$ is measure preserving is equivalent to $T^*$ preserving the norms of characteristic functions.

**Example 2.3** (Bernoulli shifts)**.** *Let $S$ be a finite set, equipped with a probability measure $\nu$. Let $S^{\mathbb{N}} := \{f : \mathbb{N} \longrightarrow S\}$, equipped with the product topology. We can also regard elements of $S^{\mathbb{N}}$ as one-sided infinite 01-sequences $(x_i)$ of elements $x_i \in S$. The topology of $S^{\mathbb{N}}$ is generated by* cylinders

$$C[a_0, \ldots, a_n] := \{(x_i) \in S^{\mathbb{N}} \mid x_i = a_i \text{ for } i = 0, \ldots, n\},$$

*where here $a_i \in S$. Topologically, $S^{\mathbb{N}}$ is homeomorphic to a Cantor set. For instance, when $S = \{0,1\}$ we can map $(x_i)$ to the point in the middle thirds Cantor set that has the form $.y_0y_1y_2\ldots$ in ternary, where $y_i = 2x_i$. You can also verify quickly that $S^{\mathbb{N}}$ is perfect, compact, metrizable and totally disconnected, which characterizes the Cantor set.*

*There's a natural probability measure on $S^{\mathbb{N}}$, the countable product measure determined by $\nu$. This is the unique measure $\mu$ on $S^{\mathbb{N}}$ such that*

$$\mu(C[a_0, \ldots, a_{n-1}]) = \Pi_{i=0}^n \nu(a_i);$$

*to show it exists, you can show that $\mu$ thus defined is $\sigma$-additive and $\sigma$-finite on the semi-algebra of cylinders, and then appeal to Carathéodory's extension theorem. See e.g. pg 10 of Sarig's notes on ergodic theory. The shift map*

$$\sigma : S^{\mathbb{N}} \longrightarrow S^{\mathbb{N}}, \quad \sigma(a_0a_1a_2\ldots) = a_1a_2\ldots$$

*is measure preserving, since*

$$\sigma^{-1}(C[a_0, \ldots, a_{n-1}]) = C[0, a_0, \ldots, a_{n-1}] \cup C[0, a_0, \ldots, a_{n-1}]$$

*and both the original cylinder $C[a_0, \ldots, a_{n-1}]$ and the union on the right side have measure $2^{-(n-1)}$. There are also variants of this example using a probability space like $S = [0,1]$ instead of a finite set, and one can also use $\mathbb{Z}$ instead of $\mathbb{N}$, giving a two-sided shift instead of a 1-sided shift.*

Two measure spaces $(X_i, \mu_i)$ are *isomorphic mod 0* if they have measurably isomorphic subsets $X_i' \subset X_i$ with *full measure*, meaning $\mu_i(X_i \setminus X_i') = 0$. A measure space $(X, \mu)$ equipped with a m.p. transformation $T : X \longrightarrow X$ is a *m.p. transformation*, or m.p.t. Two m.p.t.'s $(X_i, \mu_i, T_i)$ are *isomorphic mod 0* if there are full measure sets $X_i' \subset X_i$ and a measure isomorphism $f : X_1' \longrightarrow X_2'$ such that $f \circ T_1 = T_2 \circ f$.

**Fact 2.4.** *The doubling map $(S^1, m, D_2)$ and the Bernoulli shift $(\{0,1\}^{\mathbb{N}}, \mu, \sigma)$ are isomorphic.*

This is the starting point of 'symbolic dynamics', where shift spaces are used to model a priori more complicated dynamical systems.

*Proof.* Take $x_0x_1 \ldots \in \{0,1\}^{\mathbb{N}}$ to the element of $[0,1]$ with that binary expansion. This is a measure isomorphism onto the complement of the dyadic rationals. □

## 3. Poincaré recurrence

Suppose that $(X, \mu)$ is a finite measure space (that is, $\mu(X) < \infty$) and $T : X \longrightarrow X$ is measure preserving.

**Theorem 3.1** (Poincaré recurrence)**.** *For any measurable set $E \subset X$ and for almost every $x \in E$, there are infinitely many $n \in \mathbb{N}$ such that $T^n(x) \in E$.*

The assumption that $\mu(X) < \infty$ is important : if $T : \mathbb{R} \longrightarrow \mathbb{R}$, $T(x) = x + 1$, then the conclusion of the theorem doesn't hold.

*Proof of Theorem 3.1.* Let $B = \{x \in E \mid T^n(x) \notin E \text{ for all } n \geq 1\}$. This is a measurable set, since it's the intersection of all the sets $T^{-n}(X \setminus E)$, $n \geq 1$.

Any two iterates $T^{-m}(B)$ and $T^{-n}(B)$ are disjoint: assuming $m < n$, if $x$ lies in both sets then $T^m(x) \in E$ and $T^{n-m}(T^m(x)) \in E$, a contradiction.

Since $T$ is measure preserving, all $T^{-n}(B)$ have the same measure, and since they're disjoint and $X$ is a finite measure space, it follows that $\mu(B) = 0$.

But the set of all $x \in E$ that do not satisfy the conclusion of the theorem is the union of all $T^{-n}(B)$, and therefore also has measure zero. □

There's a sense in which Poincaré recurrence is just a measure theoretic version of the pigeonhole principle. Indeed, the third paragraph in the proof above is essentially that: if $\mu(B)$ had positive measure, some of the sets $T^{-n}(B)$ would have to intersect.

**Corollary 3.2.** *If $X$ is a second countable topological space with $\mu$ a Borel measure on $X$, and $T : X \longrightarrow X$ is m.p., then for almost every $x \in X$, the orbit $(T^n(x))$, $n \in \mathbb{N}$, accumulates onto $x$.*

*Proof.* Let $(B_k)$ be a countable basis for the topology. Let $R_k \subset B_k$ be the set of points $x$ such that $T^n(x) \in B_k$ for some infinitely many $n$. By Poincaré recurrence,

$$R := \bigcap_k R_k \cup (X \setminus B_k)$$

is an intersection of full measure sets in $X$, so has full measure. If $x \in R$, then for any basis element $B_k$ containing $x$, we have $x \in R_k$, so there's some $n$ such that $T^n(x) \in B_k$. Hence, the orbit $(T^n(x))$, $n \in \mathbb{N}$, accumulates onto $x$. □

**Remark 3.3.** *Recurrence has the following somewhat paradoxical consequence. Imagine you have a box in which you place a piece of paper, light it, and then quickly seal the box (this is time $t = 0$). Naively, the configuration space of atoms in the box and their velocities is compact, and if you could apply classical mechanics to the movement of the atoms, you'd get a finite measure that's preserved*[2] *as $t$ increases. Then Poincaré seems to say that if you perturb all the atoms in the box slightly, then at some point in the future, the contents of the box will return to that of a just-lit piece of paper. The problem here is the simplicity of the model, and the fact that the return times promised by Poincaré's theorem are so large in this case that the model would have to be basically perfect for the conclusion to apply.*

Here's a sort of quantitative variant of the recurrence theorem.

**Proposition 3.4.** *Suppose $(X, \mu)$ is a probability space and $T : X \longrightarrow X$ is m.p., and let $E \subset X$. Then $\limsup_{n\to\infty} \mu(E \cap T^{-n}(E)) \geq \mu(E)^2$.*

In particular, if $E$ has positive measure, then $E$ and $T^{-n}(E)$ intersect (in a positive measure set) for arbitrary large $n$, as also follows from Theorem 3.1. The bound on the right is optimal: for 'mixing' $(X, \mu, T)$ that we'll study later,

$$\lim_{n\to\infty} \mu(E \cap T^{-n}(E)) = \mu(E)^2.$$

This is the case for the doubling map $D_2 : S^1 \longrightarrow S^1$, for instance. Intuitively, $E$ and $T^{-n}(E)$ are becoming 'independent' in $S^1$, so the probability that a random point lies in both is just the product $\mu(E)\mu(T^{-n}(E)) = \mu(E)^2$. The proposition above says that in general, *some* iterates $T^{-n}(E)$ are nearly independent from $E$.

*Proof.* Suppose for the moment that $T$ is invertible. For any $N \geq 1$, we have

$$\int \sum_{n=1}^{N} 1_{T^{-n}(E)} \, d\mu = N\mu(E), \implies \int (\sum_{n=1}^{N} 1_{T^{-n}(E)})^2 \, d\mu \geq N^2 \mu(E)^2,$$

by Cauchy-Schwartz, applied in $L^2$ to $\sum_{n=1}^{N} 1_{T^{-n}(E)}$ and $1_X$. Here, we're using that $\mu$ is a probability measure. But we have

$$\begin{aligned}
\int (\sum_{n=1}^{N} 1_{T^{-n}(E)})^2 \, d\mu &= \sum_{n,m=1}^{N} 1_{T^{-n}(E)\cap T^{-m}(E)} \\
&= \sum_{n,m=1}^{N} \mu(T^{-n}(E) \cap T^{-m}(E)) \\
&= \sum_{n,m=1}^{N} \mu(E \cap T^{n-m}(E)) \qquad (1) \\
&\leq (\limsup_{n\to\infty} \mu(E \cap T^{-n}(E)) + o(1))N^2,
\end{aligned}$$

where $o(1)$ indicates a function that goes to zero as $N \to \infty$. The Prop follows.

If $T$ isn't invertible, the proof is almost the same, but in (1) the terms of the sum should be allowed to be either $\mu(E \cap T^{n-m}(E))$ or $\mu(E \cap T^{m-n}(E))$, so that the exponent is always negative. Otherwise, you can't use the m.p. property of $T$ to relate (1) to the previous line. □

[2] This is called Louiville's Theorem.

Here's a much more powerful version of recurrence due to Furstenberg [8].

**Theorem 3.5** (Furstenberg's Multiple Recurrence). *Suppose that $(X, \mu)$ is a probability space and $T$ is m.p., and $E \subset X$ is measurable, with $\mu(E) > 0$. Then for every $k \in \mathbb{N}$, there's some $n$ such that*

$$\mu(E \cap T^{-n}(E) \cap \cdots \cap T^{-kn}(E)) > 0.$$

Proposition 3.4 implies that given a positive measure subset $E \subset X$, after replacing $T$ by a power, we can assume that $E$ intersects $T(E)$ in a positive measure set. Theorem 3.5 implies that we can even pass to a power of $T$ so that *the first $k$ iterates of $E$ all intersect.*

As an application, the *upper density* of a subset $E \subset \mathbb{N}$ is

$$\bar{d}(E) := \limsup_{n \to \infty} |E \cap \{0, \ldots, n\}|/(n+1).$$

For instance, $\bar{d}(5\mathbb{N}) = 5$, while $\{n^2 \mid n \in \mathbb{Z}\}$ has upper density zero.

**Theorem 3.6** (Szemerédi's Theorem). *If $E \subset \mathbb{N}$ has positive upper density, then for each $k \in \mathbb{N}$, there are $m \in \mathbb{N}, n \in \mathbb{N}_{>0}$ such that $\{m, m+n, \ldots, m+kn\} \subset E$.*

In words, subsets of the natural numbers with positive upward density contain arbitrarily long arithmetic progressions.

*Proof.* The point is to apply Furstenberg's theorem to a $\sigma$-invariant measure on $X = \{0,1\}^{\mathbb{N}}$ that has something to do with the subset $E \subset \mathbb{N}$. Here, $\sigma$ is the shift map. One way to construct a measure from $E$ is to set $e = (e_i) \in X$ to be the point $e_i = 1 \iff i \in E$, and let $\delta_e$ be the Dirac measure supported on $e$. Of course, this is not shift invariant unless $E = \mathbb{N}$. But we can try to make it shift invariant by averaging its pushforwards by iterates of $\sigma$ and taking a limit. Namely, let

$$\mu_n := \frac{1}{n+1} \sum_{i=0}^{n} \sigma_*^i(\delta_e) = \frac{1}{n+1} \sum_{i=0}^{n} \delta_{\sigma^i(e)}.$$

These are all probability measures. We now use:

**Theorem 3.7.** *If $X$ is a compact metric space, the space of probabilty measures $M(X)$ on $X$ is compact in the weak$^*$ topology.*

Here, we say that $\mu_i \to \mu$ in the weak$^*$ topology if $\int f d\mu_i \to \int f d\mu$ for all bounded continuous functions $f : X \longrightarrow \mathbb{R}$. We'll discuss this result in greater detail after finishing the proof of Szemerédi's theorem.

We now want to extract a subsequential limit of $(\mu_n)$. However, in order to exploit the condition of positive upper density, first find a subsequence $n_i$ such that

$$\lim_{i \to \infty} |E \cap \{0, \ldots, n_i\}|/(n_i + 1) > 0,$$

and then pass to a further subsequence so that $\mu_{n_i} \to \mu$ in the weak$^*$ topology. The limit measure $\mu$ is $\sigma$-invariant, since

$$\sigma_* \mu = \lim_{i \to \infty} \sigma_* \mu_{n_i} = \lim_{i \to \infty} \left( \mu_{n_i} + \frac{1}{n_i + 1} \left( \sigma_*^{n_i+1}(\delta_e) - \delta_e \right) \right) = \mu,$$

noting that technically above we should be integrating everything above against a bounded continuous function, and that standard measure theoretic arguments imply this suffices for the equality of probability measures.

Consider the cylinder $C = C[1]$, i.e. the set of sequences beginning with 1. Then

$$\mu(C) = \lim_{i\to\infty} \mu_{n_i}(C) = \lim_{i\to\infty} \frac{1}{n_i+1}|E \cap \{0,\dots,n_i\}| > 0.$$

Multiple recurrence then implies that for each $k$, there's some $n$ such that

$$\mu(I) > 0, \quad I := C \cap \sigma^{-n}(C) \cap \dots \cap \sigma^{-kn}(C).$$

Suppose that for some $m$, the iterate $\sigma^m(e) \in I$. Then

$$\{m, m+n, \dots, m+kn\} \subset E$$

as desired. So, it suffices to show that the orbit $O = \{\sigma^m(e) \mid m \in \mathbb{N}\}$ intersects $I$. But the measure $\mu$ is supported on the closure $\overline{O} \subset \{0,1\}^{\mathbb{N}}$, so since $\mu(I) > 0$ we have $I \cap \overline{O} \neq \emptyset$. And since $I$ is open, the intersection $\overline{O} \cap I$ is open in $\overline{O}$, and as it's nonempty, it must intersect the dense subset $O \subset \overline{O}$ as desired. $\square$

3.1. **Compactness of the set of probability measures.** The proof of Szemerédi's theorem above uses weak* compactness of the space of probability measures on $\{0,1\}^{\mathbb{N}}$. Let's discuss why this is true.

Let $A$ be a compact Hausdorff space. Let $C(A)$ be the Banach space of continuous functions on $A$, with the sup norm, and let $C(A)^*$ be the dual space of all continuous linear functionals $C(A) \longrightarrow \mathbb{R}$, regarded with the operator norm

$$|L| = \sup_{f\in C(A), |f|_\infty \le 1} |L(f)| < \infty.$$

**Theorem 3.8** (Riesz-Markov-Kakutani)**.** *The map*

$$\mu \in \mathcal{P}(A) \longmapsto (f \mapsto \int f\, d\mu) \in C(A)^*$$

*is injective, with image the set of positive, unit norm operators $L \in C(A)^*$.*

Here, $L \in C(A)^*$ is *positive* if we have $L(f) \ge 0$ whenever $f \ge 0$. Note that if $\mu \in \mathcal{P}(A)$, then the functional $f \mapsto \int f\, d\mu$ is positive, and has unit norm since if $|f|_\infty \le 1$ we have $\int f\, d\mu \le \int |f|\, d\mu \le \int 1\, d\mu = 1$, with equality when $f = 1$.

The dual space $C(A)^*$ has a natural *weak* topology*, where $L_i \to L$ iff $L_i(f) \to L(f)$ for all $f \in C(X)$. Requiring the map in the theorem above to be a homeomorphism onto its image, we have a corresponding *weak* topology* on $\mathcal{P}(X)$, where

$$\mu_i \to \mu \iff \int f\, d\mu_i \to \int f\, d\mu \quad \forall f \in C(X)$$

**Theorem 3.9** (Banach-Alaoglu)**.** *If $V$ is a Banach space, the unit ball in $V^*$ is compact in the weak* topology.*

As a corollary of the above theorems, we get:

**Theorem 3.10.** *$\mathcal{P}(X)$ is compact in the weak* topology.*

*Proof.* $\mathcal{P}(X)$ is identified with a subset of the compact unit ball in $C(X)^*$, so we just have to show this subset is closed. But if $L_i \to L$ weakly and $L_i$ are positive, unit norm, then for any $f \ge 0$ we have $L(f) = \lim_i L_i(f) \ge 0$, and positivity implies

$$|L| = |L(1)| = \lim_i |L_i(1)| = 1. \quad \square$$

## 4. Ergodicity

**Definition 4.1** (Ergodic)**.** Suppose that $(X, \mu)$ is a measure space and $T : X \longrightarrow X$ is measure preserving. Then $(X, \mu, T)$ is *ergodic* if whenever $A \subset X$ is a $T$-invariant measurable subset, we have $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

As a dumb example, if $T$ acts transitively on $X$, it acts ergodically. More generally, ergodicity is a sort of measure theoretic irreducibility condition; if there is a $T$-invariant subset $A \subset X$ that has positive but not full measure, then the system $(X, \mu, T)$ breaks into two pieces, $(A, \mu|_A, T|_A)$ and $(X - A, \mu|_{X-A}, T|_{X-A})$. While for ergodic systems, any such decomposition is measure theoretically trivial. Note that it is then easy to construct examples of nonergodic systems, by taking the union of two arbitrary p.m.p. systems, say.

**Example 4.2** (Bernoulli shifts are ergodic)**.** *Let $S$ be a finite set with a probability measure $\nu$, and let $\mu$ be the product measure on $S^{\mathbb{N}}$. We claim that the shift map $\sigma : (S^{\mathbb{N}}, \mu) \longrightarrow (S^{\mathbb{N}}, \mu)$ acts ergodically.*

*To see this, let $C = C[a_0, \ldots, a_n]$ be a cylinder in $S^{\mathbb{N}}$. Then for $N \geq n + 1$,*

$$\mu(C \cap \sigma^{-N}(C)) = \mu(C)^2,$$

*since $\sigma^{-N}(C)$ is the set of all sequences $(x_i)$ where $x_N = a_0, \ldots, x_{N+n} = a_n$, and for $N \geq n + 1$ these conditions are independent of those defining $C$. The same formula holds for finite unions of cylinders, for large enough $N$.*

*Take an arbitrary measurable $\sigma$-invariant subset $E \subset S^{\mathbb{N}}$ and let $\epsilon > 0$. Then there is*[3] *a finite union of cylinders $D$ such that $\mu(E \Delta D) < \epsilon$. For any $N$,*

$$\mu(E \Delta T^{-N}(D)) = \mu(T^{-N}(E \Delta D)) = \mu(E \Delta D) < \epsilon.$$

*However, for large $N$ we also have that $\mu(D \cap T^{-N}(D)) = \mu(D)^2$, so*

$$\mu(D) - \mu(D)^2 = \frac{1}{2}\mu(D \Delta T^{-N}(D)) \leq \frac{1}{2}(\mu(E \Delta T^{-N}(D)) + \mu(E \Delta D)) < \epsilon.$$

*When $\epsilon$ is small, the left side is close to $\mu(E) - \mu(E)^2$, while the right side is close to $0$. Hence, $\mu(E) - \mu(E)^2 = 0$, implying $\mu(E) = 0$ or $1$.*

**Proposition 4.3.** *Given $(X, \mu, T)$, the following are equivalent.*

1. *$(X, \mu, T)$ is ergodic,*
2. *for any measurable $B \subset X$, we have $\mu(T^{-1}(B) \Delta B) = 0$ (in which case we say $B$ is* almost $T$-invariant*) if and only if $\mu(B) = 0$ or $\mu(X - B) = 0$,*
3. *for any positive measure sets $A, B \subset X$, there's $n$ with $\mu(T^{-n}(A) \cap B) > 0$,*
4. *for any measurable function $f : X \longrightarrow \mathbb{R}$, if $f \circ T = f$ almost everywhere then $f$ is constant almost everywhere.*

*Proof.* For (1) $\iff$ (2), let's say $A, B$ are the same mod 0 if $\mu(A \Delta B) = 0$. If $B, T^{-1}(B)$ are the same mod 0, then inductively $B, T^{-n}(B)$ are the same mod 0, and hence are the same mod 0 as $B_\cup := \cup_{n \geq 0} T^{-n}(B)$. But then set

$$B_{\cap\cup} := \cap_{N \geq 1} \cup_{n \geq N} T^{-n}(B),$$

which is $T$-invariant. We have

$$\mu(B_{\cap\cup}) = \lim_{N \to \infty} \mu(T^{-N}(B_\cup)) = \mu(B).$$

[3] The point here is that cylinders generate the topology and hence the Borel $\sigma$-algebra. The set of $E$ that are 'approximable' as above is closed under intersections and unions.

Since $T$ is ergodic, $B_{\cap\cup}$ has either 0 or full measure, so the same is true of $B$.

For (1) $\iff$ (3), note that if $T$ acts ergodically and $B$ has positive measure then $B_{\cap\cup}$ is invariant and has positive measure, so has full measure and hence intersects $B$. Conversely, if (3) holds then apply it to a $T$-invariant set and its complement.

For (4) $\implies$ (2), apply (4) to the characteristic function of an almost $T$-invariant set. For the other direction, suppose that $f$ is a $T$-invariant measurable function, and that it's not constant a.e. Then there's some $x \in \mathbb{R}$ such that $f^{-1}(-\infty, x]$ and $f^{-1}(x, \infty)$ both have positive measure, and the sets are both almost $T$-invariant, so $(X, \mu, T)$ isn't ergodic. □

The following examples use some Fourier analysis. As a brief refresher, recall that the Hilbert space $L^2(S^1, \mu, \mathbb{C})$ has an orthonormal basis (meaning an orthonormal set whose linear span is dense) consisting of the functions $t \mapsto e^{2\pi int}$, where $n \in \mathbb{Z}$. One then gets that any $f \in L^2(S^1, \mu, \mathbb{C})$ can be expressed uniquely as

$$f(t) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi int}, \quad c_n = \langle f, e^{2\pi int} \rangle = \int f(t) e^{-2\pi int} dt.$$

Moreover, the map $f \mapsto (c_n)$ is an isometry $L^2(S^1, \mu, \mathbb{C}) \longrightarrow \ell^2(\mathbb{C}^2)$, so in particular

$$|f|_2^2 = \sum |c_n|^2,$$

which is called *Parseval's formula*.

**Example 4.4.** *If $S^1 = \mathbb{Z}\backslash\mathbb{R}$, the circle rotation $T_\alpha : S^1 \longrightarrow S^1$, $T_\alpha([x]) = [x + \alpha]$ is ergodic with respect to Lebesgue measure $\mu$ if and only if $\alpha$ is irrational.*

*First, assume that $\alpha$ is rational and write $\alpha = 2p/q$, with $q$ even. Then*

$$E = [0, 1/q] \cup [2/q, 3/q] \cup \cdots \cup [(q-2)/q, (q-1)/q] \subset S^1$$

*is $T_\alpha$ invariant and has measure $\frac{1}{2}$.*

*Next, suppose $\alpha$ is irrational, and that $A \subset S^1$ is measurable and $T$-invariant. We want to show that $A$ either has zero or full measure. Write*

$$1_A = \sum_{n\in\mathbb{Z}} a_n e^{2\pi int}$$

*as a Fourier expansion, where the sum converges in $L^2(S^1, \mu, \mathbb{C})$. By invariance, we have $1_A \circ T_\alpha = 1_A$, so we have*

$$\sum_{n\in\mathbb{Z}} a_n e^{2\pi in(t+\alpha)} = \sum_{n\in\mathbb{Z}} a_n e^{2\pi int},$$

*implying that $a_n = e^{2\pi in\alpha} a_n$ for all $n$. Since $\alpha$ is irrational, we only have $e^{2\pi in\alpha} = 1$ when $n = 0$, so this implies $a_n = 0$ for all $n \neq 0$. Hence, $1_A$ is equal to $a_0$ almost everywhere, and hence $A$ has either zero or full measure, depending on whether $a_0 = 0$ or $a_0 = 1$.*

**Example 4.5.** *Let $D_m : S^1 \longrightarrow S^1$ be the map $D_m([x]) = [mx]$. Again, we take a $D_m$-invariant measurable set $A$ and write*

$$1_A = \sum_{n\in\mathbb{Z}} a_n e^{2\pi int}$$

*with convergence in $L^2(S^1, \mu, \mathbb{C})$. By $D_m$-invariance, we have that*

$$\sum_{n\in\mathbb{Z}} a_n e^{2\pi int} = \sum_{n\in\mathbb{Z}} a_n e^{2\pi inmt}$$

*implying that $a_n = 0$ for all $n \neq 0$. So, $D_m$ acts ergodically as above.*

Let $A \in GL(n, \mathbb{Z})$ and let $T_A : T^k \longrightarrow T^k$ be the map $T_A([x]) = [Ax]$, where here we regard $T^k = \mathbb{Z}^k \backslash \mathbb{R}^k$. We say that $T_A$ is *hyperbolic* if $A$ has no eigenvalue on the unit circle. An example is Arnold's "cat map", where $A = (2\ 1; 1\ 1)$.

**Claim 4.6.** *If $T_A$ is hyperbolic, then $T_A$ acts ergodically on $(T^k, \mu)$, where $\mu$ is the Lebesgue measure on $T^k$.*

*Proof.* Here, we use a higher dimensional analogue of Fourier series. If $E \subset T^k$ is $T_A$-invariant, we have a Fourier expansion of the form

$$1_E = \sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle n, x \rangle} = \sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle n, Ax \rangle} = \sum_{n \in \mathbb{Z}^k} c_n e^{2\pi i \langle A^t n, x \rangle},$$

Note that the transpose $A^t$ is also in $GL(k, \mathbb{Z})$, and hence gives an automorphism of $\mathbb{Z}^k$, and the above says that $c_n = c_{A^t n}$ for all $n$. But Parseval's formula says that $\sum c_n^2 = |1_E|_2^2 < \infty$, so in particular there are only finitely many $c_n$ above any given positive value. So, if some $c_n \neq 0$, then the $A^t$-orbit of $n$ is finite. This only happens for nonzero $n$ when $A$ has an eigenvalue on the unit circle. □

The converse isn't true, and the point is that a matrix $A \in GL(k, \mathbb{Z})$ can have eigenvalues on the unit circle even while all $A$-orbits on $\mathbb{Z}^k$ are infinite, since the complex eigenspace of this eigenvector can intersect $\mathbb{Z}^k \subset \mathbb{C}^k$ trivially. An example is the following matrix in $GL(4, \mathbb{Z})$.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 8 & -6 & 8 \end{pmatrix}$$

## 5. Ergodic theorems

Suppose $(X, \mu)$ is a probability space and $T : (X, \mu) \longrightarrow (X, \mu)$ is ergodic. Intuitively, ergodicity says that $T$ acts transitively on $X$ in some measure-theoretic sense. So, if you take a random point $x \in X$ and start translating it around by $T$, you'd expect it to go basically everywhere in $X$.

One way to make this precise is as follows.

**Theorem 5.1** (Birkhoff's ergodic theorem)**.** *Suppose that $f : X \longrightarrow \mathbb{C}$ is an integrable function. Then for a.e. $x \in X$, we have*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) = \int f \, d\mu.$$

So, the orbit of $x$ distributes uniformly enough with respect to $\mu$ that taking the average value of $f$ over larger and larger subsets approximates $\mu$. Sometimes people describe this theorem as saying that the 'time average' of $f$ (on the left) limits to the 'space average' of $f$ (on the right).

Here is a number theoretic application.

**Definition 5.2** (Normal number)**.** An element $x \in \mathbb{R}$ is called *normal* if for all $b = 2, 3, \dots$ and all length $k$ strings $\omega \in \{0, 1, \dots, b-1\}^k$, if $x_0 x_1 \dots x_m . x_{m+1} \dots$ is the base $b$ expansion of $x$, then we have

$$\lim_{N \to \infty} \frac{\{i \in \{0, \dots, N-1\} \mid (x_i \dots, x_{i+k}) = \omega\}}{N} = \frac{1}{b^k}.$$

Note that rational numbers are not normal; it is also easy to see that there are unaccountably many non-normal numbers. For a specified base $b$, one can explicitly construct a number that is normal in base $b$ by concatenating together all the base $b$ expansions of natural numbers, in order. However, it's open whether this construction results in a number that is normal as defined above (in all bases), and in general there is no 'known' normal number, although there is a normal number whose digits are in principle computable, by Becher-Figuera (2002).

As an application of the ergodic theorem, we prove:

**Proposition 5.3.** *Almost every $x \in \mathbb{R}$ is normal.*

*Proof.* It suffices to work with $x \in [0,1]$, since normality only depends on the tail of the base $b$ expansion, so we can always take the decimal point to be at the front. It then suffices to show that almost every string in $X = \{0,1,\dots,b-1\}^{\mathbb{N}}$ is normal (in the obvious sense), since the map that takes such a string to the corresponding real number is measure preserving.

Consider $X$ equipped with the product measure $\mu$ of the uniform probability measure on $\{0,1,\dots,b-1\}$, and let $\sigma : X \longrightarrow X$ be the shift map. We have shown that $(X,\mu,\sigma)$ is ergodic. Given a word $\omega = (w_0,\dots,w_{k-1})$, let $C = C[\omega]$ be the corresponding cylinder. Then $\mu(C) = 1/b^k$. Applying Birkhoff's theorem to $f = \chi_C$ proves the proposition. $\square$

One can also use Birkhoff's theorem to estimate the frequency of digits in continued fraction expansions of almost every real number, of the frequency of digits in decimal expansions of powers of 2.

The point of the rest of this section is to prove a more general version of Birkhoff's theorem that applies to non-ergodic systems. We'll start with an easier theorem due to Von Neumann that asserts something similar, but where there's no specific $x$ chosen and convergence is in $L^2$.

5.1. **Mean Ergodic Theorems.** Suppose that $(X,\mu)$ is a probability space and $T : X \longrightarrow X$ is measure preserving. Given $N > 0$, consider the *averaging operator*

$$A_N : L^2(X,\mu) \longrightarrow L^2(X,\mu), \quad A_N(f) = \frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n.$$

Note that $A_N$ is linear, and has norm 1. Intuitively, as $N \to \infty$, you expect $f$ to become more and more $T$-invariant. But which $T$-invariant function should you expect to get? Let

$$L^2(X,\mu)^T := \{f \in L^2(X,\mu) \mid f \circ T = f\},$$

which is a closed subspace of $L^2(X,\mu)$. Let

$$\pi : L^2(X,\mu) \longrightarrow L^2(X,\mu)^T$$

be the orthogonal projection.

**Theorem 5.4** (Von Neumann's Mean Ergodic Theorem in $L^2$)**.** *For $f \in L^2(X,\mu)$,*

$$\lim_{N\to\infty} A_N(f) = \pi(f).$$

Here, convergence is in $L^2$. Note that the claim is trivially true if we start with a function $f \in L^2(X,\mu)^T$, since then $A_N(f) = f$ for all $N$.

Let's see what the theorem says when $(X, \mu, T)$ is ergodic. In this case,

$$L^2(X, \mu)^T = \{ \text{ constant functions } \},$$

and if $f \in L^2(X, \mu)$ then

$$\pi(f) = \int f \, d\mu,$$

since the function $f - \int f \, d\mu$ has zero integral, and hence is $L^2$-orthogonal to the set of constant functions. So, we have:

**Corollary 5.5.** *If $(X, \mu, T)$ is ergodic and $f \in L^2(X, \mu)$, then*

$$\lim_{N \to \infty} A_N(f) = \int f \, d\mu,$$

*where the right hand side is a constant function, and convergence is in $L^2$.*

*Proof of VNMET.* Let's write $I = L^2(X, \mu)^T$ for the space of invariant functions. It suffices to show that if $f \in I^\perp$, then $\lim_{N \to \infty} A_N(f) = 0$, as any element of $L^2$ can be written as a sum of an invariant function and an orthogonal function, and the theorem is linear on both sides and true for invariant functions.

Here's the trick. We claim that $I^\perp$ is the closure of the linear subspace

$$D = \{g \circ T - g \mid g \in L^2(X, \mu)\}.$$

It suffices to show that $I = D^\perp$, since in a Hilbert space $(D^\perp)^\perp = \overline{D}$. To do this, suppose $f$ is $T$-invariant and note that

$$\langle f, g \circ T - g \rangle = \langle f \circ T, g \circ T \rangle - \langle f, g \rangle = 0$$

by $T$-invariance of $f$ and the $L^2$ inner product. Conversely, suppose that $f$ is orthogonal to $D$. Then in particular we have

$$0 = \langle f \circ T - f, f \rangle = \langle f \circ T, f \rangle - \langle f, f \rangle,$$

so since $|f|_2 = |f \circ T|_2$ this implies that

$$\langle f \circ T, f \rangle = |f \circ T|_2 |f|_2,$$

implying that $f \circ T, f$ are linear combinations of each other, but then they have to be equal, since they have the same integral.

So, back to the proof. For elements of $D$, we have

$$A_N(g \circ T - g) = \frac{1}{N}(g \circ T^{N+1} - g) \to 0,$$

noting that $|g|_2 = |g \circ T^{N+1}|_2$ is fixed, so the norm of the difference is at most $2|g|_2$. For a general element $f \in \overline{D} = I^\perp$, we can take $f_i \in D$ with $f_i \to f$, and then

$$|A_N(f)|_2 = |A_N(f - f_i)|_2 + |A_N(f_i)|_2 \le |f - f_i|_2 + |A_N(f_i)|_2,$$

since $A_N$ has norm 1. But if we take $i$ large, and then $N$ much larger, both terms on the right are arbitrarily close to zero, so $\lim_{N \to \infty} A_N(f) = 0$. $\square$

In the ergodic case, the limit in the mean ergodic theorem is just the integral of $f$. So, one might expect the theorem to be true whenever the integral of $f$ exists, i.e. for $f \in L^1(X, \mu)$ rather than in $L^2$. Here, recall that when $(X, \mu)$ is a probability space, $L^2(X, \mu) \subset L^1(X, \mu)$ but the inclusion may be strict, e.g. $f(x) = 1/\sqrt{x}$ is integrable on $[0, 1]$ but not square integrable. However, one can extend the mean ergodic theorem to $L^1$ via approximation, as follows.

**Theorem 5.6** (Mean Ergodic Theorem, $L^1$ version)**.** *If $f \in L^1(X,\mu)$, then the sequence $(A_N(f))$ converges in $L^1$ and the limit is $T$-invariant. In particular, if $(X,\mu,T)$ is ergodic, then $A_N(f) \to \int f\,d\mu$ in $L^1$.*

Note that since $(X,\mu)$ is a probability space, $|g|_2 \geq |g|_1$, so convergence in $L^2$ implies convergence in $L^1$ to the same limit. Hence, if $f$ is in $L^2$, the limit functions in Theorems 5.4 and 5.6 are the same.

*Proof.* We show that $(A_N(f))$ is Cauchy. Let $\epsilon > 0$ and pick some $g \in L^2(X,\mu)$ such that $|f-g|_1 < \epsilon/4$. Then for all $N, M$, linearity of $A_N$ implies

$$\begin{aligned} |A_N(f) - A_M(f)|_1 &\leq |A_N(f-g)|_1 + |A_N(g) - A_M(g)|_1 + |A_M(f-g)|_1 \\ &\leq |A_N(g) - A_M(g)|_1 + \epsilon/2, \end{aligned}$$

where here we use that $A_N$ has norm 1. For large $N, M$, this quantity is at most $\epsilon$, showing that $(A_N(f))$ is Cauchy. It's immediate that the limit is $T$-invariant. When $(X,\mu,T)$ is ergodic, the limit is constant a.e. but its integral is $\int A_N(f) = \int f$, so the constant value is $\int f\,d\mu$. $\square$

5.2. **Birkhoff's theorem.** In this section, we upgrade $L^1$-convergence in Theorem 5.6 to convergence pointwise almost everywhere. Here, $f_n \to f$ *pointwise almost everywhere* if there is a full measure set $E \subset X$ such that $f_n(x) \to f(x)$ for all $x \in E$.

**Example 5.7.** *There are sequences of bounded functions $f_n : [0,1] \longrightarrow [0,\infty)$ that converged to zero in $L^1([0,1])$, but do not converge to anything pointwise almost everywhere. For example, consider the characteristic functions of the intervals*

$$[0,1], [0,1/2], [1/2,2/2], [0,1/3], [1/3,2/3], [2/3,3/3], [0,1/4], \ldots$$

*The integrals of these functions are $1, 1/2, 1/2, 1/3, 1/3, \ldots$, which converge to zero, but every point in $[0,1]$ is in infinitely many of these intervals, and also not in infinitely many of them, so for no $x$ does the sequence of evaluations on $x$ of these characteristic functions converge.*

However, we have the following.

**Lemma 5.8.** *Suppose $(X,\mu)$ is a measure space and $f_n : X \longrightarrow \mathbb{R}$ is a sequence of integral functions that converges in $L^1$ to some function $f$. Then there is some subsequence $(f_{n_i})$ that converges to $f$ pointwise almost everywhere.*

*Proof.* It suffices to assume that $f_n \to 0$ in $L^1$, and show that there's a subsequence that converges to zero pointwise almost everywhere. Well, pass to a subsequence such that $|f_n|_1 < 1/2^n$ and set

$$g(x) = \sum_{n=1}^{\infty} |f_n(x)|.$$

By the monotone convergence theorem, $\int g(x)\,d\mu = \sum_{n=1}^{\infty} |f_n|_1 < \infty$. So, $g$ is finite almost everywhere, which implies that $f_n(x) \to 0$ almost everywhere. $\square$

We are now ready to state the main theorem of the section.

**Theorem 5.9** (Birkhoff's ergodic theorem)**.** *Suppose that $(X,\mu)$ is a finite measure space and $T : X \longrightarrow X$ is measure preserving. If $f \in L^1(X,\mu)$, then the sequence $(A_N(f))$ converges pointwise almost everywhere.*

By Lemma 5.8, the limit function in Birkhoff's theorem is the same as the limit function in the $L^1$-version of the mean ergodic theorem. In particular, when $(X, \mu, T)$ is an ergodic probability space, the limit is constant with output $\int f\,d\mu$. We will discuss later how to interpret the limit in the nonergodic case.

Below, we give two proofs of Birkhoff's theorem. To set up the proofs, let

$$A^* = \limsup_{N\to\infty} A_N(f)(x), \quad A_*(x) = \liminf_{N\to\infty} A_N(f)(x).$$

Note that $A^*, A_*$ are both $T$-invariant, since

$$A_N(f) \circ T(x) = \frac{N+1}{N} A_{N+1}(f)(x) - \frac{1}{N} f(x),$$

so the sequences $A_N(f) \circ T(x)$ and $A_{N+1}(f)(x)$ are asymptotic, and therefore have the same limsup and liminf.

In both proofs, we try to prove $A_* = A^*$ a.e. by showing something like

$$\int A^*\,d\mu \le \int f\,d\mu \le \int A_*\,d\mu,$$

which together with $A_* \le A^*$ implies that $A_* = A^*$ a.e. The first proof is a bit shorter and goes through an inequality called the 'maximal ergodic theorem'. The second, essentially due to Keane [16], is a bit more intuitive.

5.3. **First proof of Birkhoff's theorem.** Suppose $(X, \mu)$ is a finite measure space, $T : X \longrightarrow X$ is measure preserving. For $f \in L^1(X, \mu)$, write

$$A_n(f)(x) = \frac{1}{n}\left(f(x) + \cdots + f(T^{n-1}(x))\right).$$

The trickiest part of the argument is the following lemma.

**Lemma 5.10** (The maximal inequality)**.** *Let*

$$P = \{x \in X \mid A_n(x) > 0 \textit{ for some } n \in \mathbb{N}\}.$$

*Then* $\int_P f\,d\mu \ge 0$.

*Proof.* It's easier here to replace $A_n(f)$ with

$$S_n(x) := f(x) + \cdots + f(T^{n-1}(x)) = nA_n(f)(x).$$

For each $N \in \mathbb{N}$, let

$$P_N := \{x \in X \mid S_n(x) \ge 0 \text{ for some } n \le N\}.$$

It suffices to show that $\int_{P_N} f\,d\mu \ge 0$ for all $N$, since

$$\int_{P_N} f\,d\mu = \int_X \chi_{P_N} \cdot f\,d\mu \to \int_X \chi_P \cdot f = \int_P f\,d\mu,$$

by the dominated convergence theorem.

Set $M_N(x) := \max_{0\le n\le N} S_n(x)$, where $S_n(x) := 0$. If $x \in P_N$, then

$$M_N(T(x)) = \max_{1\le n\le N+1} S_n(x) - f(x) \ge M_N(x) - f(x), \tag{2}$$

where in the last inequality we use $x \in P_N$ to reinsert the index $n = 0$ into the max. Since $M_N$ is positive and vanishes on $X \setminus P_N$,

$$\begin{aligned}\int_{P_N} f\,d\mu &\geq \int_{P_N} M_N(x)\,d\mu - \int_{P_N} M_N(T(x))\,d\mu \\ &\geq \int_X M_N(x)\,d\mu - \int_X M_N(T(x))\,d\mu \\ &= 0. \end{aligned}$$
$\square$

So, let's now try to prove that the sequence $(A_N(f))$ converges pointwise almost everywhere. To do this, let

$$A^* = \limsup_{N\to\infty} A_N(f)(x), \quad A_*(x) = \liminf_{N\to\infty} A_N(f)(x)$$

and our goal is to prove that $A^* = A_*$ almost everywhere. For this, it suffices to show that for every rational numbers $a < b$ the set

$$E_{a,b} := \{x \in X \mid A_*(x) < a < b < A^*(x)\}$$

has measure zero, since the set where $A^* \neq A_*$ is a countable union of these sets. Since both $A_*, A^*$ are $T$-invariant, so is $E_{a,b}$.

Apply the lemma to the triple $(E_{a,b}, \mu, T)$ and the function $f - b$, noting that $A_n(f-b) = A_n(f) - b$. For every $x \in E_{a,b}$, we have $A_n(f-b)(x) = A_n(f)(x) - b$, which is positive for some $n$ by definition of the limsup. Hence, the lemma says

$$\int_{E_{a,b}} f - b\,d\mu \geq 0, \implies \int_{E_{a,b}} f\,d\mu \geq b\mu(E_{a,b}).$$

But also, we can apply the lemma to the function $a - f$. Again for $x \in E_{a,b}$ the value $A_n(a-f)(x) = a - A_n(f)(x)$ is positive for some $n$, so

$$\int_{E_{a,b}} a - f\,d\mu \geq 0 \implies \int_{E_{a,b}} f\,d\mu \leq a\mu(E_{a,b}).$$

Since $a < b$, the only way these statements can both be true is if $\mu(E_{a,b}) = 0$.

5.4. **Second proof of Birkhoff's theorem, due to Keane.** Again, our approach is to show that $A_* = A^*$ a.e. by using that $A_* \leq A^*$ and proving that

$$\int A^*\,d\mu \overset{(1)}{\leq} \int f\,d\mu \overset{(2)}{\leq} \int A_*\,d\mu,$$

Since an arbitrary $f$ is the difference of two nonnegative functions, we'll assume everywhere below that $f \geq 0$.

Let's work on (1) first. Here's the idea. Pick some moderate size $M > 0$ and some huge $N > 0$, and consider the sequence

$$f(x), f(T(x)), \ldots, f(T^{N-1}(x)). \tag{3}$$

Within the sequence, we look for stretches of at most $M$ consecutive terms where the average of those terms close to $A^*(x)$. Note that as long as $M$ is large (and $N$ is even larger), most points in the sequence above are contained in such stretches, since $A^*(x)$ is the limsup, and since $A^*$ is $T$-invariant. When computing $A_N(f)(x)$, we then compute the averages of (a disjoint collection of) such stretches first, and then say that on average over $X$, the remaining points of the sequence don't influence the result much when $M, N$ is large, so that

$$\int A_N(f)\, d\mu = \int f\, d\mu \approx \int A^*(x)\, d\mu$$

and then we take $N \to \infty$ to prove the result.

More rigorously, let $\epsilon > 0$ and for $x \in X$ set

$$A^*_\epsilon(x) = \min\{A^*(x), 1/\epsilon\} - \epsilon, \quad \tau(x) = \min\{n \mid A_n(f)(x) > A^*_\epsilon(x)\}$$

Label the points $x, T(x), \ldots, T^{N-1}(x)$ as *good*, *bad*, or *unlabeled* as follows. Starting from the left, label everything bad until we arrive at some iterate $T^k(x)$ where

$$\tau_k := \tau(T^k(x)) \leq M.$$

Label $T^k(x), \ldots, T^{k+\tau_k-1}(x)$ good, and start again with the $k + \tau_k + 1$ iterate. Continue this process until all $T^k(x)$ with $k \leq N - M$ are labeled, leaving possibly a terminal string of at most $N - M$ unlabeled points. Call the number of good and bad points $G, B$, respectively, so that $G + B \geq N - M$. Then we have

$$NA_N(f + A^*_\epsilon \cdot 1_{\tau > M})(x) \geq G \cdot A^*_\epsilon(x) + B \cdot A^*_\epsilon(x) \geq (N - M) \cdot A^*_\epsilon(x).$$

Here, we divide the terms of the sum into good stretches and bad terms, and note that both terms of $f + A^*_\epsilon \cdot 1_{\tau>M}$ are positive, so on the right side of the first inequality we can apply $f$ to the good terms and $A^*_\epsilon \cdot 1_{\tau>M}$ to the bad terms, noting that all bad iterates $T^k(x)$ lie in $\{\tau > M\}$. Integrating and using the fact that the $A_N$ operator doesn't change integrals, we get

$$\int f + A^*_\epsilon \cdot 1_{\tau>M}\, d\mu \geq \frac{N - M}{N} \int A^*_\epsilon\, d\mu.$$

Letting $N \to \infty$, and then subtracting the (finite!) integral $\int A^*_\epsilon \cdot 1_{\tau>M}\, d\mu$, we get

$$\int f\, d\mu \geq \int_{\tau \leq M} A^*_\epsilon\, d\mu,$$

and then taking first $M \to \infty$, and then $\epsilon \to 0$, proves (1).

The proof of (2) is almost the same. By Fatou's lemma, $\int A_*\, d\mu \leq \int f\, d\mu$, so $A_* < \infty$ almost everywhere. We then set

$$A^\epsilon_* = A_* + \epsilon, \quad \tau(x) = \min\{n \mid A_n(f)(x) < A^\epsilon_*\}$$

and label points $x, T(x), \ldots, T^{N-1}(x)$ as good, bad, or unlabeled as before. Set $f_M(x) = \max\{f(x), M\}$, and then compute

$$NA_N(f_M \cdot 1_{\tau\leq M})(x) \leq G \cdot A^\epsilon_*(x) + B \cdot 0 + M^2 \leq NA^\epsilon_* + (N - M) \cdot M,$$

where for the first inequality we divide into good stretches, bad points, and unlabeled points. Dividing by $N$, taking integrals, and letting $N \to \infty$, we get

$$\int f_M \cdot 1_{\tau\leq M}\, d\mu \leq A^\epsilon_*.$$

First taking $M \to \infty$ and then letting $\epsilon \to 0$ proves (2).

5.5. **Conditional expectations and the limit function.** In Theorem 5.9 the limit function is not named, except in the ergodic case. When $f \in L^2(X,\mu)$, it follows from Theorem 5.4 that the limit is $\pi(f)$, the orthogonal projection onto the $T$-invariant functions, but for $f \in L^1(X,\mu)$ there's no such projection.

**Example 5.11.** *Suppose $(X,\mu,T)$ is a m.p. system and $X = \sqcup_i X_i$ is a countable union of subset $X_i$ that are all $T$-invariant, have positive measure, and where $(X_i, \mu|_{X_i}, T|_{X_i})$ are ergodic. If $f \in L^1(X,\mu)$, then for each $i$ and a.e. $x \in X_i$,*

$$\lim_{N\to\infty} A_N(f)(x) = \frac{1}{\mu(X_i)} \int_{X_i} f \, d\mu,$$

*by Birkhoff's theorem.* Since the union is countable*, the above equation actually holds for a.e. $x \in X$, where on the right hand side $i$ is such that $x \in X_i$. So, the function that assigns to $x \in X_i$ the average of $f$ over $X_i$ is the Birkhoff limit.*

**Example 5.12.** *As another example, set $\mu$ to be Lebesgue measure on $S^1$, set $T^2 = S^1 \times S^1$, equipped with $\mu \times \mu$, and consider the map*

$$T : T^2 \longrightarrow T^2, \quad T(x,y) = x + \alpha,$$

*where $\alpha$ is irrational. Then for an absolutely integrable function $f : T^2 \longrightarrow \mathbb{R}$, if we fix $y \in S^1$ then for almost every $x \in S^1$ we have*

$$\lim_{N\to\infty} A_N(f)(x,y) = \int_{x \in S^1} f(x,y)\, d\mu,$$

*by applying Birkhoff's ergodic theorem to the action of $T$ on the circle $S^1 \times y$, equipped with Lebesgue measure $\mu$. By Fubini's theorem, the set of $(x,y)$ where the above does not hold has zero $\mu \times \mu$ measure, so the limit function in Birkhoff's theorem just averages $f$ over all the circles $S^1 \times y$.*

So, how do you formulate this more generally? The second example above indicates that not every m.p. system decomposes into countably many ergodic pieces, and the description of the limit function in that case is very particular, since $\mu \times \mu$ is a product measure. In general, one way to describe the limit function is through a framework called *conditional expectation*, which we now describe.

Say we have a measure space $(X,\mathcal{B},\mu)$, where we now emphasize the $\sigma$-algebra $\mathcal{B}$. Say $\mathcal{A} \subset \mathcal{B}$ is a sub-$\sigma$-algebra. We associate to $f \in L^1(X,\mathcal{B},\mu)$ an element

$$E(f,\mathcal{A}) \in L^1(X, \mathcal{A}\mu|_{\mathcal{A}})$$

called the *(conditional) expectation of $f$ with respect to $\mathcal{A}$*, as follows. Let $\nu$ be the measure on $(X,\mathcal{A},\mu)$ given by the formula

$$\nu(S) := \int_S f \, d\mu.$$

Then $\nu << \mu|_{\mathcal{A}}$, so the Radon-Nikodym theorem implies that there's some $\mathcal{A}$-measurable function $g : (X,\mathcal{A}) \longrightarrow \mathbb{R}$ such that

$$\nu(S) := \int_S g \, d\mu|_{\mathcal{A}}.$$

Moreover, $g$ is unique up to $\mu|_{\mathcal{A}}$-a.e. equality, and we set $E(f,\mathcal{A}) := g$. Note that $E(f,\mathcal{A}) \in L^1$, since $\int E(f,\mathcal{A})\, d\mu|_{\mathcal{A}} = \int f d\mu < \infty$.

**Example 5.13.** *Suppose $\mathcal{A} = \{\emptyset, X\}$. Then $E(f, \mathcal{A})$ is the constant function $\int f\,d\mu$. If $\mathcal{A}$ is generated by a partition $X = \sqcup X_i$ into set of positive measure, then $E(f, \mathcal{A})$ takes on the value $\int_{X_i} f\,d\mu$ on each $X_i$, and if $T^2 = S^1 \times S^1$, equipped with the product measure $\mu \times \mu$, and $\mathcal{A}$ is the $\sigma$-algebra that's the pullback of the Borel $\sigma$-algebra of $S^1$ under the projection onto the second factor, then*

$$E(f, \mathcal{A})(x, y) = \int_{x \in S^1} f(x, y)\,d\mu.$$

Now let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \longrightarrow X$ be measure preserving. Let $\mathcal{B}^T \subset \mathcal{B}$ be the sub-$\sigma$-algebra consisting of all $A \in \mathcal{B}$ that are *almost $T$-invariant*, meaning that $\mu(A \Delta T^{-1}(A)) = 0$.

**Claim 5.14.** *A $\mathcal{B}$-measurable function $f : X \longrightarrow \mathbb{R}$ is $\mathcal{B}^T$-measurable if and only if $f \circ T = f$ almost everywhere.*

In particular, the conditional expectation $E(f, \mathcal{B}^T)$ is always almost $T$-invariant. Note that when $T$ acts ergodically, $\mathcal{B}^T$ is the $\sigma$-algebra consisting of all sets that have either zero or full measure, and $f : X \longrightarrow \mathbb{R}$ is $\mathcal{B}^T$-measurable if and only if it's constant a.e.

*Proof.* Suppose we have $f \circ T = T$ almost everywhere. If $B \subset \mathbb{R}$ is Borel, then $f^{-1}(B)$ is almost $T$-invariant, so we're done.

Conversely, suppose $f$ is $\mathcal{B}^T$-measurable (so the same is true for $f \circ T$), but we don't have $f \circ T = T$ almost everywhere. Then for some $\epsilon > 0$ the set

$$E = \{x \in X \mid |f \circ T(x) - f(x)| \geq 2\epsilon\}$$

has positive measure; note that $E \in \mathcal{B}^T$. There is then also some $\epsilon$-ball $U \subset \mathbb{R}$ such that $f^{-1}(U) \cap E \in \mathcal{B}^T$ has positive measure. But

$$T^{-1}(f^{-1}(U) \cap E) \approx_0 T^{-1}(f^{-1}(U)) \cap E = (f \circ T)^{-1}(U) \cap E,$$

which is disjoint from $f^{-1}(U) \cap E$ since a point in $E$ can't map into $U$ by both $f$ and $f \circ T$. This contradicts that $f^{-1}(U) \cap E \in \mathcal{B}^T$. $\square$

We can now reformulate Birkhoff's theorem as follows.

**Theorem 5.15** (Birkhoff's ergodic theorem, limit specified)**.** *If $f : X \longrightarrow \mathbb{R}$ is absolutely integrable, then $\lim_{N\to\infty} A_N(f) = E(f, \mathcal{B}^T)$ pointwise a.e. and in $L^1$.*

*Proof.* By our earlier version of Birkhoff's theorem, we know $(A_N(f))$ limits pointwise a.e. and in $L^1$ to some function $A_\infty(f)$ that is $T$-invariant, hence $\mathcal{B}^T$-measurable by the Claim above. And if $B \in \mathcal{B}^T$, we have

$$\int_B A_\infty(f)\,d\mu = \lim_{N\to\infty} \int_B A_N(f)\,d\mu = \int_B f\,d\mu,$$

where the first equality uses $L^1$-convergence and the second uses that $B$ is almost $T$-invariant. So, we have $A_\infty(f) = E(f, \mathcal{B}^T)$ by definition. $\square$

## 6. Ergodic Decomposition

In this section we fix a measurable map $T : X \longrightarrow X$ of a measurable space, and show that any $T$-invariant probability measure on $X$ can be written as a sort of convex combination of $T$-ergodic measures. For the most part we'll take the perspective of infinite dimensional convex geometry, working with continuous actions on compact metric spaces, but we'll mention at the end another approach to such an ergodic decomposition.

6.1. **The set of invariant measures.** Suppose $(X, \mathcal{B})$ is a measurable space and $T : X \longrightarrow X$ is measurable. Let $\mathcal{P}(X)$ be the set of all probability measures on $(X, \mathcal{B})$, and let $\mathcal{P}(X)^T \subset \mathcal{P}(X)$ be the subset of $T$-invariant measures. Let $\mathcal{E}(X) \subset \mathcal{P}(X)^T$ be the $T$-ergodic measures.

**Example 6.1.** *Suppose $X = \{0,1\}^{\mathbb{Z}}$ and $\sigma$ is the shift map. Then the product measure is shift invariant, but there are also many other shift invariant measures, e.g. the uniform measure on the (finite) orbit of any periodic sequence.*

**Remark 6.2.** *It's possible that $\mathcal{P}(X)^T$ is empty! Take $T(x) = x + 1$ on $X = \mathbb{R}$. Then there's no $T$-invariant probability measure on $\mathbb{R}$: some interval $[a, b]$ has positive measure, but it's disjoint from infinitely many of its translates under $T$.*

If $V$ is a vector space, a subset $A \subset V$ is *convex* if whenever $x, y \in A$, we have $tx + (1 - t)y \in A$ for all $t \in [0, 1]$. In other words, $A$ is convex if the line segment between two points of $A$ is always contained in $A$.

We can regard $\mathcal{P}(X)$ as a convex subset of the vector space

$$\mathbb{R}^{\mathcal{B}} := \{f : \mathcal{B} \longrightarrow \mathbb{R}\}.$$

It's convex since if $\mu, \nu \in \mathcal{P}(X)$ then $t\mu + (1-t)\nu \in \mathcal{P}(X)$ as well for any $t \in [0, 1]$; the measure axioms are easily verified, and in particular the constraint $t \in [0, 1]$ ensures that the resulting function takes positive values. Moreover, if $\mu, \nu$ are $T$-invariant, so is any convex combination of them, so $\mathcal{P}(X)^T$ is also convex.

So, how do ergodic measures appear in this picture? If $C$ is a convex set, an *extreme point* of $C$ is a point $p \in C$ such that whenever we have $p = tx + (1 - t)y$, with $x, y \in C$, we have either $x = p$ or $y = p$.

**Example 6.3.** *The measures $\delta_x$, $x \in X$, are the extreme points in $\mathcal{P}(X)$. For any measure $\mu$ not of this form has a subset $A \subset X$ with $\mu(A), \mu(X \setminus A) > 0$, and then restricting to those sets gives a nontrivial convex combination giving $\mu$.*

**Theorem 6.4.** *A measure $\mu \in \mathcal{P}(X)^T$ is ergodic if and only if it's an extreme point of $\mathcal{P}(X)^T$.*

*Proof.* If $\mu \in \mathcal{P}(X)^T$ isn't ergodic, then $\mu = \mu|_A + \mu|_{X \setminus A}$, where $A$ is any $T$-invariant set with $\mu(A), \mu(X \setminus A) > 0$, and hence $\mu$ isn't an extreme point.

Now assume $\mu$ is ergodic and $\mu = t\alpha + (1 - t)\beta$, where $\alpha, \beta \in \mathcal{P}(X)^T$. We can assume $t \neq 0$, say, since the $t \neq 1$ case is similar. Since $\alpha \leq \frac{1}{t}\mu$, we have $\alpha << \mu$, so by Radon-Nikodym, we have a measurable functions $f$ on $X$ such that $\alpha(S) = \int_S f \, d\mu$ for all measurable $S$. Since $\alpha, \mu$ are $T$-invariant, we have

$$\int_S f \circ T \, d\mu = \int_{T^{-1}(S)} f \, d\mu = \alpha(T^{-1}(S)) = \alpha(S) = \int_S f \, d\mu$$

for all measurable $S$, so $f = f \circ T$ $\mu$-a.e. By ergodicity, $f$ is constant $\mu$-a.e., so $\mu = \alpha$ since both are probability measures. $\square$

We now review some more subtle facts about convex sets in vector spaces, and afterwards we'll apply them to our study of $\mathcal{P}(X)^T$.

6.2. **Convex geometry.** Recall that if $V$ is a vector space, a *convex combination* of $p_1, \ldots, p_k \in V$ is a linear combination of the form

$$v = \sum_{i=1}^{k} t_i p_i, \text{ where } \sum_i t_i = 1.$$

The *convex hull* of a subset $E \subset V$ is the smallest convex set $CH(E)$ that contains $E$, and it's easy to verify that $CH(E)$ is exactly the set of all convex combinations of finite subsets of $E$. Indeed, the set of all such combinations is convex, and is contained in any convex subset containing $E$.

In the finite dimensional setting, a famous theorem of Minkowski states:

**Theorem 6.5.** *If $A \subset \mathbb{R}^n$ is compact and convex, then $A$ is the convex hull of its set $\mathcal{E}(A)$ of extreme points.*

Note that compactness is necessary here: e.g. $\mathbb{R} \subset \mathbb{R}$ has no extreme points. Towards the proof, given $A \subset \mathbb{R}^n$ and $p \in \partial A$, a *support plane* for $A$ through $p$ is an $(n-1)$-dimensional hyperplane $P \subset \mathbb{R}^n$ such that $p \in P$ and $A$ is contained in the closure of some component of $\mathbb{R}^n \setminus P$.

**Lemma 6.6.** *If $A \subset \mathbb{R}^n$ is closed and convex, and $p \in \partial A$, there's a support plane for $A$ through $p$.*

*Proof.* Take $p_i \to p$, $p_i \in \mathbb{R}^n \setminus A$. Let $\pi(p_i) \in \partial A$ be a closest point to $p_i$ within $A$, and let $P_i$ be the hyperplane perpendicularly bisecting the line segment $[p_i, \pi(p_i)]$ at its midpoint $m_i$. Taking a subsequence, $P_i \to P$, a plane through $P$. Each $P_i$ is disjoint from $A$, since if $x \in P_i \cap A$ then on the right triangle $m_i, \pi(p_i), x$ there's a point a little closer to $m_i$ than $\pi(p_i)$ on the opposite edge, which lies in $A$, contradicting the definition of $\pi(p_i)$. Hence $P$ is a support plane. $\square$

*Proof of Theorem 6.5.* This is proved via induction on $n$. For the inductive case, we can assume that $A \subset \mathbb{R}^n$ has nonempty interior (since otherwise it's contained in a hyperplane). Given $x \in A$, we divide into two cases:

(1) if $x \in \partial A$, pick a support plane $P \subset \mathbb{R}^n$ through $x$, i.e. a hyperplane through $x$ where $A$ lies in the closure of one component of $\mathbb{R}^n \setminus P$. Then $P \cap A$ is compact and convex, so is the convex hull of its extreme points by induction, and any extreme point in $P \cap A$ is extreme in $A$, so we're done.
(2) if $x \in int(A)$, write $x = ty + (1-t)z$ for $y, z \in \partial A$, and apply (1) to $y, z$, to get a convex combination of extreme points that equals $x$. $\square$

So, is this true in infinite dimensions? No!

**Example 6.7.** *Let $B = [-1,1]^{\mathbb{N}}$, regarded as a subset of the vector space $\mathbb{R}^{\mathbb{N}}$, equipped with the product topology. Then $B$ is compact and convex.*

*We claim that the extreme points of $B$ are exactly those $x \in B$ with $x_i = \pm 1$ for all $i$. Indeed, any such $x$ is extreme, since if $x = ty + (1-t)z$ then for each $i$, we have $\pm 1 = ty_i + (1-t)z_i$, where $y_i, z_i \in [-1,1]$, implying that $y_i = z_i = x_i$. And if we have $x$ with $x_k \neq \pm 1$ for some $k$, then $x$ is the average of of two sequences $y, z$ defined by $y_i = z_i = x_i$ for $i \neq k$, and $y_k = x_k + \epsilon, z_k = x_k - \epsilon$, where $\epsilon$ is small enough so both $y, z \in B$.*

*Now, the convex hull of the extreme points is not equal to $B$, since any convex combination of finitely many extreme points is a sequence that takes on only finitely* *many values. However, you can check that $B$ is the* closure *of the convex hull of its set of extreme points.*

A *topological vector space* (TVS) is a vector space $V$ with a topology such that the vector space operations are continuous. We say that $V$ is *locally convex* if there is a neighborhood basis of 0 (equivalently, of any point) that consists of convex open sets.

**Example 6.8.** *All normed vector spaces are locally convex, where balls around the origin are the convex open sets in question. More generally, if $V$ is a Banach space, then $V^*$ is locally convex in the weak* topology: here, $L_i \to L$ if $L_i(v) \to L(v)$ for all $v \in V$. The reason is that a neighborhood basis for $0 \in V^*$ is given by convex open sets obtained by fixing $\epsilon > 0$ and a finite set $S \subset V$, and defining*

$$O(S, \epsilon) := \{L \in V^* \mid |L(v)| < \epsilon, \ \forall v \in S\}.$$

*A non-example is the toplogical vector space*

$$\ell_p := \left\{ x = (x_i, i \in \mathbb{N}) \;\Big|\; \sum_i |x_i|^p < \infty \right\}, \quad p \in (0,1),$$

*regarded as a topological vector space induced by the distance function*

$$d(x,y) = |x-y|_p^p, \quad |x|_p^p := \sum_i |x_i|^p.$$

*Note that $|x|_p^p$ isn't a norm, since $|rx|_p^p = |r|^p |x|_p^p$, but it does satisfy the triangle inequality, so the resulting $d$ is a metric. This isn't locally convex: in fact, suppose we have some convex set $U$ with $B_\delta(0) \subset U \subset B_1(0)$. Then if $(e^i)$ is the standard basis, we have $\delta^{1/p} e^i \in B_\delta(0) \subset U$, so therefore*

$$v_N := \sum_{i=1}^N \frac{\delta^{1/p}}{N} e^i \in U \subset B_1(0),$$

*but $|v_N|_p^p := N^{1-p}\delta \to \infty$ as $N$ increases.*

Here's why we care about locally convex TVSs.

**Theorem 6.9** (Hahn Banach)**.** *Suppose $V$ is a locally convex, Hausdorff TVS and $A \subset X$ is compact and convex, while $v \in V \setminus A$. Then there's a continuous linear functional $L : V \longrightarrow \mathbb{R}$ such that $\sup_{x \in A} L(x) < L(v)$.*

Geometrically, if we pick $s$ with $\sup_{x \in A} L(x) < s < L(v)$, then the closed hyperplane $L^{-1}(s)$ separates $v$ from $A$. This can be used, for instance, to prove that points in $\partial A$ have supporting hyperplanes, just like in the lemma above. The theorem is true without the locally convex assumption *as long as $A$ has nonempty interior*, but that's not the case in our intended application.

So, is it true that in any locally convex TVS, a compact, convex subset is the convex hull of its extreme points?

Indeed, this is a general phenomenon.

**Theorem 6.10** (Krein-Millman)**.** *Any compact, convex subset $A$ of a Hausdorff, locally convex topological vector space $V$ is the closure of the convex hull of its set $\mathcal{E}(A)$ of extreme points; in symbols, $A = \overline{CH(\mathcal{E}(A))}$.*

*Proof.* Say $A$ is a compact, convex subset of a Banach space $V$. A *face* of $A$ is a nonempty, compact, convex subset $F \subset A$ such that whenever $x, y \in A$ and some convex combination $tx + (1-t)y = z \in F$, where $t \in [0, 1]$, then we actually have $x, y \in F$. Here are some examples:

- If $x \in A$ is an extreme point, then $\{x\}$ is a face of $A$.
- When $L : V \longrightarrow \mathbb{R}$ is a continuous linear functional, then the set
$$A_L := \{x \in A \mid L(x) \text{ is maximal}\}$$
is a face, since it is closed in $A$, is convex, and $L$ is maximized on the endpoints of any segment.
- A face of a face of $A$ is a face of $A$.

Note that any *minimal* face of $A$ is a singleton set, and then the point it contains is exteme, by definition. Indeed, if $F \subset A$ is a face and has more than one point, Hahn-Banach implies there's $L \in V^*$ that's nonconstant on $F$, and then $F_L \subset F$ is a strictly smaller face of $A$.

Assuming $A$ is nonempty, we first show that it *has* an extreme point, by showing that it has a minimal face. This is a Zorn's lemma argument. Ordering faces by reverse inclusion, any chain of faces
$$F_1 \supset F_2 \supset \cdots$$
has nonempty intersection (as all these sets are compact), and the intersection is itself a face. So, the assumption in Zorn's lemma is satisfies, implying there's a minimal face $F$, which is an extreme point as noted above.

Finally, let $C \subset A$ be the closure of the convex hull of the extreme points of $A$. Suppose there's some $x \in A \setminus C$. Then Hahn-Banach implies that there's some $L \in V^*$ such that $L(x) > L(C)$, where here we use that $C$ is *closed*, hence compact. The face $A_L$ has an extreme point by the argument above, which lies outside of $C$, and hence we have a contradiction. $\square$

Krein-Millman says that any point in a compact, convex subset $A$ is a limit of a sequence of convex combinations of finitely many extreme points of $A$.

**Definition 6.11.** A point $x \in A$ is *represented by* a Borel probability measure $\mu$ on $A$ if for every $f \in V^*$, we have
$$\int f \, d\mu = f(x).$$

As an example, say that $x = \sum_i t_i e_i$ is a finite convex combination of elements of $A$. Then $x$ is represented by the finitely supported measure $\mu = \sum_i t_i \delta_{e_i}$ since
$$\int f \, d\mu = \sum_i t_i f(e_i) = f(\sum_i t_i e_i) = f(x).$$
The following is equivalent to Krein-Millman.

**Corollary 6.12.** *Every point $x \in A$ is represented by a Borel probability measure $\mu$ on $A$ that is concentrated on the closure $\overline{\mathcal{E}(A)}$.*

Here, $\mu$ is *concentrated on* $E$ if $\mu(E) = 1$.

*Proof.* Since $A$ is compact and Hausdorff, the space $\mathcal{P}(A)$ of probability measures on $A$ is compact in the weak topology, by Riesz-Markov-Kakutani (below) and Banach Alaoglu. (See below and the next section for stuff about this, or just believe it.) So, given $x \in A$, use Krein-Millman to get a sequence $x_i = \sum_i t_i e_i$ of convex combintions of extreme points with $x_i \to x$. After passing to a subsequence, we can assume that the measures $\mu_i := \sum_i t_i \delta_{e_i}$ converge weakly to some $\mu$. Since all the $\mu_i$ are supported on $\mathcal{E}(A)$, the limit measure $\mu$ is supported on $\overline{\mathcal{E}(A)}$. And if $f \in V^*$,

$$\int f\, d\mu = \lim_{i\to\infty} \int f\, d\mu_i = \lim_{i\to\infty} f(x_i) = f(x). \quad \square$$

However, there's a more subtle theorem of Choquet that represents any point of $A$ by a measure just on the set of extreme points, rather than its closure.

**Theorem 6.13** (Choquet)**.** *If $V$ is a locally convex TVS and $A \subset V$ is compact, convex and metrizable, then for every $x \in A$ there's a Borel probability measure $\mu_x$ on $A$ that represents $x$ and is concentrated on $\mathcal{E}(A)$.*

**Exercise 6.14.** *Take a point $x = (x_1, x_2, \dots) \in [0,1]^{\mathbb{N}}$ in the Hilbert cube, and write down a probability measure supported on the vertices of the cube that represents $x$.*

*Proof Sketch of Choquet's theorem.* As a stupid warmup, look at

$$C(A) \longrightarrow \mathbb{R}, \quad f \mapsto f(x),$$

which is a positive, unit norm linear functional on $A$. RMK says this is represented by a measure on $A$. Of course, here the measure is just the Dirac measure $\delta_x$, so it's usually not concentrated on the extreme points of $A$. Damn.

Here's a more intelligent approach. We want to change our liner functional $f \mapsto f(x)$ by adding on second term, in such a way that it pushes the support of the resulting measure out to $\mathcal{E}(A)$. To do this, pick a strictly convex function

$$c : A \longrightarrow \mathbb{R}.$$

One way to produce $c$ is to take a countable dense subset $h_n$ of the set of affine functions on $A$ with sup norm 1, and then set $c = \sum_n 2^{-n} h_n^2$. Although we'll skip the details here, this is where we're using metrizability of $A$! Namely, metrizability implies that $C(A)$ is separable, implying separability of the subspace of affine functions with norm 1. In fact, metrizability of $A$ is really equivalent to the existence of a strictly convex function on $A$, so one can't get round this.

Let $\mathrm{Aff}(A) \subset C(X)$ be the subspace of affine functions. For $f \in C(X)$, set

$$\overline{f} = \inf\{h \in \mathrm{Aff}(A) \mid h \geq f\}.$$

We call $\overline{f}$ the *upper envelope* of $f$. Note that $\overline{f}$ is concave, since it's the inf of a bunch of (nonstrictly) concave functions. Define a linear functional

$$m_x : \mathrm{Aff}(A) + \mathbb{R}c \longrightarrow \mathbb{R}, \quad m(h + tc) = h(x) + t\overline{c}(x).$$

So, this $m_x$ (currently only defined on a subspace of $C(A)$) is like the functional $f \mapsto f(x)$ in the affine term, but then we add on $t\overline{c}(x)$ in the second. How are extreme points related to the second term? The key is that the set

$$E = \{x \in A \mid \overline{c}(x) = c(x)\}$$

is contained in $\mathcal{E}(A)$: indeed, if $x = \frac{1}{2}(e + f)$, where $e, f \in A$, then we have

$$c(a) < \frac{1}{2}(c(e) + c(f)) \leq \frac{1}{2}(\overline{c}(e) + \overline{c}(f)) \leq \overline{c}(a).$$

Continuing our proof, note tht $m_x(h + tc) \leq \overline{(h + tc)}(x)$, and the RHS is the restriction of the subadditive linear functional $g \mapsto \overline{g}$ on $C(X)$, so by $m_x$ extends by Hahn Banach (see its Wikipedia page) to a continuous linear functional on $C(X)$, which we'll also call $m_x$, such that $m_x(g) \leq \overline{g}(x)$ for all $g \in C(X)$. This $m_x$ is positive, since if $g \leq 0$ then $m_x(g) \leq \overline{g}(x) \leq 0$, and it's unit norm since $\overline{(h+tc)}(x) \leq 1$ if $h + tc \leq 1$, and $m_x(1) = 1(x) = 1$.

Let $\mu_x \in \mathcal{P}(A)$ be a probability measure such that $m_x(f) = \int f \, d\mu$ for al $f \in C(X)$, as given by RMK. Then $\mu_x$ represents $x$, since if $f \in V^*$, then

$$\int f \, d\mu = m_x(f) = f(x),$$

just by definition of $m_x$ since $f \in \mathrm{Aff}(A)$. Moreover, $\mu_x$ is concentrated on the set $E \subset \mathcal{E}(A)$ of points where $c = \overline{c}$, which was discussed above. Indeed, $c \leq \overline{c}$, but

$$\int \overline{c} \, d\mu = m(\overline{c}) \leq \inf_{\substack{h \in \mathrm{Aff}(A) \\ h \geq c}} m(h) = \inf_{\substack{h \in \mathrm{Aff}(A) \\ h \geq c}} h(x) = \overline{c}(x) = m(c) = \int c \, d\mu, \tag{4}$$

so the set of points where $c < \overline{c}$ must have $\mu$-measure zero, as desired. □

6.3. **An ergodic decomposition on compact spaces.** All the theorems in the previous section are about compact subsets of (locally convex) topological vector spaces, while in the first section we looked at probability measures on $(X, \mathcal{B})$ as a subset of $\mathbb{R}^{\mathcal{B}}$, which doesn't even come with a reasonable topology.

However, let now $X$ be a compact metric space, and $T : X \longrightarrow X$ be continuous. Let $\mathcal{P}(X), \mathcal{P}(X)^T$ be the sets of all Borel probability measures on $X$, and all $T$-invariant Borel probability measures, respectively. In Theorem 3.10 we discussed the fact that $\mathcal{P}(X)$ is compact in the weak* topology.

**Proposition 6.15.** *$\mathcal{P}(X)^T \subset \mathcal{P}(X)$ is closed in the weak topology, hence is also compact.*

*Proof.* If $\mu_i$ are $T$-invariant and $\mu_i \to \mu$, then for every continuous $f : X \longrightarrow \mathbb{R}$,

$$\int f \, d(T_*\mu) = \int f \circ T \, d\mu = \lim_i \int f \circ T \, d\mu_i = \lim_i \int f \, d\mu_i = \int f \, d\mu,$$

and two measures are determined by their integrals on bounded continuous function, by Riesz-Markov-Kakutani. □

The space $C(X)^*$, equipped with the weak* topology, is a locally convex TVS. Its unit ball, in the weak topology, is metrizable, for instance by the function

$$d(L, L') = \sum_{i=1}^{\infty} \frac{1}{2^i} |L(f_i) - L'(f_i)|,$$

where $(f_i)$ is a countable dense subset of the unit ball in $C(X)$. (Note that since the unit ball in $C(X)^*$ is compact in the weak* topology, it's sufficient to show that the metric $d$ is a continuous with respect to the weak topology, since then the fact that it actually *induces* the weak topology follows from the fact that any continuous bijective function from a compact space to a Hausdorff space is a homeomorphism.)

So, Choquet's Theorem applies, giving:

**Theorem 6.16** (Existence of ergodic decomposition, compact case)**.** *If $X$ is compact, $T : X \longrightarrow X$ continuous and $\mu \in \mathcal{P}(X)^T$, then there's a probability measure $\nu$ on the subset $\mathcal{E}(X,T) \subset \mathcal{P}(X)^T$ of ergodic measures such that for any $f \in C(X)$,*

$$\int f \, d\mu = \int_{\eta \in \mathcal{E}(X)} \left( \int f \, d\eta \right) d\nu,$$

In other words, $\mu$ is represented by the probability measure $\nu$ on $\mathcal{E}(X)$. Here, every $f \in C(X)$ gives a continuous linear functional $L \longmapsto L(f)$ on $C(X)^*$, which we can use when talking about representing point by measures in Choquet's theorem.

**Example 6.17.** *Say we have $T : T^2 \longrightarrow T^2$, where $T^2 = S^1 \times S^1$ and $T(x, y) = (x + \alpha, y)$, and $\alpha$ is irrational. Let $\mu$ be Lebesgue measure on $S^1$, and for each $y \in S^1$, let's denote by $\mu_y$ the Lebesgue measure on $S^1 \times y \subset T^2$. Then each $\mu_y$ is $T$-ergodic, and we let $\nu$ be the probability measure on $\mathcal{E}(X, T)$ that's the pushforward of $\mu$ under the map $y \mapsto \mu_y$. Then Fubini says*

$$\begin{aligned}\int f\, d(\mu \times \mu) &= \int_{y \in S^1} \int_{x \in S^1} f(x, y)\, d\mu_y\, d\mu \\ &= \int_{\eta \in \mathcal{E}(X)} (\int f\, d\eta)\, d\nu.\end{aligned}$$

In fact, the measure $\nu$ is uniquely determined by $\mu$! To see this, let's say that a compact, convex, metrizable subset $A \subset V$ is a *Choquet simplex* if every point in $A$ is represented by a *unique* measure on the extreme points of $A$. For example, an affine simplex in $\mathbb{R}^n$ is a Choquet simplex, but an $n$-cube is not. We want to say $\mathcal{P}(X)^T$ is a Choquet simplex, so what's special about $\mathcal{P}(X)^T$ as a convex set? Think about $\mathcal{P}(X)^T$ as a 'base' for the cone $\mathcal{M}(X)^T \subset C(X)^*$ of all $T$-invariant finite measures on $X$. Given $\mu_1, \mu_2 \in \mathcal{M}(X)^T$, one can always construct

$$\mu_{max} \in \mathcal{M}(X)^T, \quad \mu_{max}(B) = \max\{\mu_1(B), \mu_2(B)\}.$$

The existence of this operation is what makes the base $\mathcal{P}(X)^T$ of the cone into a Choquet simplex. Indeed, if $A \subset V$ is the base of a cone $C$, then you can define a partial order on $V$, where $x \le y$ if there's some $z \in C$ such that $x + z = y$. It turns out that $A$ is a Choquet simplex if and only if for every pair of points $x, y \in A$, there's a least upper bound for $\{x, y\}$ with respect to this partial order. And for $\mathcal{M}(X)^T$, it's easy to verify that $\mu_{max}$ above is the desired least upper bound.

**Exercise 6.18.** *Convince yourself that the simplex in $\mathbb{R}^n$ spanned by $0$ and the standard basis vectors has the least upper bound property above, and find an example of a cone over a compact convex set in $\mathbb{R}^n$ that doesn't have that property.*

6.4. **Equidistribution and generic points.** Say that $X$ is a compact metric space and $\mu$ is a Borel probability measure on $X$. A sequence of points $(x_n)$ in $X$ is *equidistributed* with respect to $\mu$ if for every $f \in C(X)$,

$$\frac{1}{N} \sum_{n=0}^{N-1} f(x_n) \to \int f\, d\mu.$$

In other words, the atomic measures $\frac{1}{N} \sum_{n=0}^{N-1} \delta_{x_n} \to \mu$ in the weak topology.

The following is a consequence of Birkhoff's theorem.

**Theorem 6.19.** *Suppose $(X, \mu, T)$ is ergodic. Then for $\mu$-almost every $x \in X$, the orbit $(T^n(x))$ is equidistributed with respect to $\mu$.*

Such points $x \in X$ are therefore called *generic points* for $\mu$.

*Proof.* Given $f \in C(X) \subset L^1(X, \mu)$, Birkhoff's theorem says that

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x) \to \int f \, d\mu \tag{5}$$

for $\mu$-a.e. $x \in X$. However here $f$ is fixed, while we want this convergence to be true for all $f$ simultaneously. We can fix this using the fact that $C(X)$ is separable, where the topology on $C(X)$ is given by the sup norm.

Pick a countable dense subset $D \subset C(X)$. For each $f \in D$, there's some full measure set $E_f \subset X$ such that (5) hold for all $x \in E_f$. Set $E = \cap_{f \in D} E_f$, which still has full measure. Given $x \in E$, $g \in C(X)$, and $\epsilon > 0$, choose some $f \in D$ with $|g - f| < \epsilon$, and then for large $N$ we have

$$\frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n(x) < \left( \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x) \right) + \epsilon < \int f \, d\mu + 2\epsilon < \int g \, d\mu + 3\epsilon,$$

and an inequality in the other direction follows similarly. □

**Example 6.20.** *Under an irrational rotation $T_\alpha : S^1 \longrightarrow S^1$, $\underline{\text{every}}$ point $x \in S^1$ is generic for Lebesgue measure $\mu$. Why? We know some point $x$ is generic. And if $\beta \in \mathbb{R}$, then $T_\beta(x)$ is also generic for $\mu$, since as $T_\alpha, T_\beta$ commute,*

$$\frac{1}{N} \sum_{n=0}^{N-1} \delta_{T_\alpha^n \circ T_\beta(x)} = (T_\beta)_* \frac{1}{N} \sum_{n=0}^{N-1} \delta_{T_\alpha^n(x)} \to (T_\beta)^* \mu = \mu.$$

We say that a measurable map $T : X \longrightarrow X$ is *uniquely ergodic* if there's a unique Borel probability measure on $X$ such that $(X, \mu, T)$ is ergodic. Note that by Krein-Millman, this happens if and only if there's a unique $T$-invariant probability measure on $X$.

**Fact 6.21.** *Suppose $X$ is a compact metric space, $\mu$ is a Borel probability measure on $X$, and $T : X \longrightarrow X$ is measure preserving. Then $\mu$ is the unique $T$-ergodic probability measure on $X \iff$ every point of $x$ is $\mu$-generic.*

*Proof.* For the forwards direction, fix $x \in X$ and consider the sequence

$$\mu_N := \frac{1}{N} \sum_{n=0}^{N-1} \delta_{T^n(x)}.$$

Compactness of the space of $T$-invariant probability measures on $X$ says that any subsequence of $\mu_N$ has a subsequence that converges to some $T$-invariant probability measure on $X$, and hence to $\mu$. So, $\mu_N \to \mu$, implying $x$ is $\mu$-generic.

The backwards direction follows from Theorem 6.19, since any other ergodic measure has to have a generic point, but all points are generic for $\mu$. □

So, an irrational rotation $T_\alpha : S^1 \longrightarrow S^1$ is uniquely ergodic. For a nonexample, the shift action on a Bernoulli space $\{0, 1\}^{\mathbb{N}}$ is not uniquely ergodic, since one can produce different measures by varying the weights on 0 and 1, or by taking a measure supported on a finite orbit.

6.5. **A general ergodic decomposition theorem.** Sometimes one wants an ergodic decomposition theorem in a more general setting that for continuous maps of compact metric spaces. Here we describe a bit of that theory.

A measurable space $(X, \Sigma)$ is called *standard Borel* if it is measurably isomorphic to a Borel subset of a complete separable metric space, equipped with its Borel $\sigma$-algebra. By a theorem of Kuratowski, (see e.g. Srivastava [31], Theorem 3.3.13) any such measure space is actually measurably isomorphic to either $\mathbb{Z}, \mathbb{R}$ or a finite set. Moreover, one can show that whenever $\mu$ is a $\sigma$-finite measure on such an $(X, \Sigma)$, the space $(X, \Sigma, \mu)$ is isomorphic to the union of a (possibly empty, possibly unbounded) interval in $\mathbb{R}$ (equipped with Lebesgue measure) with an at most countable set of atoms. We'll call such measure spaces *standard*[4].

Basically every measure space that you'll encounter in nature is standard; the assumption rules out pathological examples like measure spaces with cardinality bigger than that of the continuum, or measurable spaces like $\mathbb{R}/\mathbb{Q}$, equipped with the quotient $\sigma$-algebra. As another non-standard example, consider the measure space $\{0,1\}^I$, where $I$ is uncountable, equipped with the product topology and the Borel $\sigma$-algebra $\mathcal{B}$, and a product measure guaranteed by Kolmogorov's extension theorem. This measure space is not even isomorphic mod 0 to a standard measure space. Indeed, any isomorphism mod 0 of measure spaces induces an isometry of $L^2$-spaces, and $L^2(X)$ is separable whenever $X$ is standard (e.g. when $X = [0,1]$ the $\mathbb{Q}$-span of the characteristic functions of intervals with rational endpoints is dense), but $L^2(\{0,1\}^I)$ is not (e.g. for $i \in I$, the $i^{th}$-coordinate functions $\phi_i$ are all 1-apart in $L^2$).

**Theorem 6.22** (Ergodic decomposition, standard borel case)**.** *Suppose $(X, \mathcal{B})$ is standard Borel and $\mathcal{P}(X)^T$ is nonempty. Then there's a map $X \longrightarrow \mathcal{E}(X,T), x \mapsto \eta_x$, that has the following properties:*

*(1) $\eta_{T(x)} = \eta_x$ for all $x \in X$,*
*(2) if $A \subset X$ is measurable, then $x \mapsto \eta_x(A)$ is measurable,*
*(3) for every $\mu \in \mathcal{P}(X)^T$ and every measurable $A \subset X$,*

$$\mu(A) = \int_X \eta_x(A)\, d\mu.$$

One thing that's different in this formulation is that, essentially, we're defining the measure on $\mathcal{E}(X)$ is defined to be the pushforward of $\mu$ under some map. So, how does one prove such a theorem? At least if $X$ is a compact metric space, it turns out that one can take the measures $\eta_x$ to be the weak limits

$$\eta_x = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \delta_{T^n(x)},$$

at least whenever the RHS is defined, $T$-invariant, and ergodic. In general, Varadarajan's compact model theorem says that if $(X, \mathcal{B})$ is a standard Borel space and $T : X \longrightarrow X$ is measurable, then there's a compact metric space $X'$, a continuous map $T' : X' \longrightarrow X'$, and a $T$-invariant Borel subset $Y \subset X'$ such that $(X,T)$ is isomorphic to $(Y, T')$ as measurable dynamical systems. One can use this theorem,

[4]There's a slightly different notion of a 'standard probability space' in the literature which is defined using axioms developed by Rokhlin in 1940, see e.g. the Wikipedia page or Section 9.4 in [5], but its function is similar.

then, to port the result from the setting of compact metric spaces to the setting of standard Borel spaces, as presented above.

As an example, consider $T^2 = S^1 \times S^1$ equipped with $\mu \times \mu$ and the action

$$T : T^2 \longrightarrow T^2, \quad T(x, y) = (x + \alpha, y),$$

where $\alpha$ is irrational. Then for every $(x, y) \in T^2$, the measure $\eta_{(x,y)}$ above is just the Lebesgue measure $\mu_y$ on $S^1 \times y$, by the fact that irrational circle rotations are uniquely ergodic for Lebesgue measure. So, the theorem above predicts that

$$\mu \times \mu(A) = \int_{(x,y)\in T^2} \mu_y(A)\, d(\mu \times \mu) = \int_{y\in S^1} \int_{x\in S^1} \mu_y(A)\, d\mu\, d\mu = \int_{y\in S^1} \mu_y(A)\, d\mu,$$

which is just true by Fubini's theorem.

## 7. Mixing systems

Suppose $(X, \mu)$ is a probability space and $T : X \longrightarrow X$ is measure preserving.

**Fact 7.1.** $(X, \mu, T)$ *is ergodic if and only if for all measurable* $A, B \subset X$*,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

*Proof.* If $(X, \mu, T)$ isn't ergodic and $X = A \sqcup B$, where both sets are $T$-invariant and positive measure, then the left side of the equality is zero for all $N$, but the right side is nonzero, so the statement above can't hold.

To prove that ergodicity implies the limit statement above, let's see how to write the limit in terms of the indicator functions $1_A, 1_B$. For $f \in L^2(X, \mu)$, recall that

$$A_N(f) := \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n.$$

We can then write

$$\langle A_N(1_A), 1_B\rangle_2 = \int \left( \frac{1}{N} \sum_{n=0}^{N-1} 1_A \circ T^n \right) \cdot 1_B\, d\mu = \int \left( \frac{1}{N} \sum_{n=0}^{N-1} 1_{T^{-n}(A)\cap B} \right) d\mu,$$

and after distributing the integral over the sum, this becomes the left hand side of the equality in the statement of the fact.

By the mean ergodic theorem,

$$\lim_{N\to\infty} A_N(1_A) \to \int 1_A\, d\mu = \mu(A),$$

where the limit $\mu(A)$ is interpreted as a constant function, and convergence is in $L^2$. Then we're done, since

$$\langle A_N(1_A), 1_B\rangle_2 \to \langle \mu(A), 1_B\rangle_2 = \mu(A)\mu(B). \quad \square$$

A measure preserving dynamical system $(X, \mu, T)$ is called *mixing* if for all measurable $A, B \subset X$, we have

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

In other words, $(X, \mu, T)$ is mixing if for all $A, B$, the event that $x \in T^{-n}(A)$ is nearly independent from the event that $x \in B$, for large $n$. Note that

$$\text{mixing} \implies \text{ergodic},$$

either because of the above proposition, or just directly, since if $X = A \sqcup B$ where $A, B$ are $T$-invariant and positive measure, then we have $\mu(T^{-n}(A) \cap B) = 0$ for all $n$, while $\mu(A)\mu(B) > 0$.

**Remark 7.2.** *Although we won't study it much, for culture let's mention that* $(X, \mu, T)$ *is* weakly mixing *if for all measurable* $A, B \subset X$*, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| = 0.$$

*So, mixing* $\implies$ *weak mixing* $\implies$ *ergodic.*

**Example 7.3** (Irrational circle rotations aren't mixing)**.** *Say we look at a rotation* $T : S^1 \longrightarrow S^1$, $T(x) = x + \alpha$*, where* $\alpha$ *is irrational, and let* $\mu$ *be Lebesgue measure on* $S^1$*. Then* $(S^1, \mu, T)$ *is ergodic, but it's not mixing. To see this, just take a small interval* $[0, \epsilon] \subset S^1$*. Then the numbers* $\mu(T^{-n}(I) \cap I)$ *are usually zero, but there are infinitely many* $n$ *where* $\alpha n \in [0, \epsilon/2]$*, and for those* $n$ *we have*

$$\mu(T^{-n}(I) \cap I) \geq \epsilon/2.$$

*With more work, you can even show they aren't weakly mixing.*

To produce examples of mixing systems, we need a lemma. Suppose that $(X, \mathcal{B}, \mu)$ is a probability space. A *semialgebra* $\mathcal{B}' \subset \mathcal{B}$ is a subset that's closed under finite unions and intersections. We say $\mathcal{B}'$ *generates* $\mathcal{B}$ if there is no sub-$\sigma$-algebra of $\mathcal{B}$ that contains $\mathcal{B}'$.

**Lemma 7.4.** *Suppose that* $T : X \longrightarrow X$ *is measure preserving. Then in the definitions of mixing and weak mixing, it suffices to take* $A, B$ *within a semialgebra* $\mathcal{B}'$ *that generates* $\mathcal{B}$*.*

*Proof.* Briefly, it's a quick exercise to show that given $A \in \mathcal{B}$ and $\epsilon > 0$, there's some $A' \in \mathcal{B}'$ such that $\mu(A \Delta A') < \epsilon$. Indeed, one just shows that the set of elements of $\mathcal{B}$ that are approximable like this is closed under countable unions, intersections, and complements. Given this, we want to show that for $A, B \in \mathcal{B}$ and $\delta > 0$,

$$\limsup_{n\to\infty} \mu(T^{-n}(A) \cap B) < \mu(A)\mu(B) + \delta,$$

and a similar statement with the liminf.

For the limsup statement, fix $A, B \in \mathcal{B}$, $\epsilon > 0$, and find $A', B' \in \mathcal{B}'$ such that

$$\mu(A \Delta A'), \mu(B \Delta B') < \epsilon.$$

Since $T$ is measure preserving, $T^{-n}(A) \Delta T^{-n}(A') < \epsilon$, which implies that

$$|\mu(T^{-n}(A') \cap B') - \mu(T^{-n}(A) \cap B)| < 2\epsilon.$$

We then get that

$$\limsup_{n\to\infty} \mu(T^{-n}(A) \cap B) \leq \limsup_{n\to\infty} \mu(T^{-n}(A') \cap B') + 2\epsilon = \mu(A')\mu(B') + 2\epsilon,$$

and if $\epsilon$ is small relative to $\delta$, this is at most $\mu(A)\mu(B) + \delta$ as desired. $\square$

We can now show:

**Theorem 7.5.** *Is* $S$ *is a finite set endowed with a probability measure* $\nu$*, and* $(S^{\mathbb{N}}, \mu, \sigma)$ *is the associated Bernoulli shift, with product measure* $\mu$ *and shift map* $\sigma$*, then* $(S^{\mathbb{N}}, \mu, \sigma)$ *is mixing.*

*Proof.* It suffies to check the definition of mixing on the semialgebra consisting of finite unions of cylinders. If $C = C[a_0, \ldots, a_{n-1}]$ is a cylinder, let's call $n$ the *length* of the cylinder. If $A = \cup_i C_i$ is a finite unions of cylinders, let

$$\mathrm{length}(A) = \max_i \mathrm{length}(C_i).$$

The point then is that $A$ consists of all sequences $(x_i)$ subject to some constraints on $x_0, \ldots, x_{\mathrm{length}(A)-1}$, and for $n > 0$, the preimage $\sigma^{-n}(A)$ consists of all sequences $(x_i)$ subject to some constraints on $x_n, \ldots, x_{n+\mathrm{length}(A)-1}$.

Fix now $A, B$ that are both finite unions of cylinders and take $n > \mathrm{length}(B)$. Then $\sigma^{-n}(A)$ and $B$ are sets of sequences determined by constraints on *disjoint* sets of terms, so membership in the two sets are independent events, implying

$$\mu(\sigma^{-n}(A) \cap B) = \mu(\sigma^{-n}(A))\mu(B) = \mu(A)\mu(B).$$

That is, $(S^{\mathbb{N}}, \mu, \sigma)$ is mixing, where the sequence in the definition of mixing is eventually constant. □

As a corollary, the doubling map on the circle is also mixing. One cool fact about mixing systems is that the diagonal action on the product is also mixing.

**Fact 7.6.** *Suppose that $(X, \mu, T)$ is mixing, then the system*

$$(X \times X, \mu \times \mu, T \times T), \quad T \times T(x, x) = (T(x), T(x))$$

*is also mixing.*

*Proof.* It suffices to check the statement on the subalgebra of finite unions of sets of the form $A \times A'$, where $A, A' \subset X$ are measurable. For simplicity, say you just have sets of the form $A \times A'$ and $B \times B'$. Then

$$\begin{aligned}
\lim_n \mu \times \mu((T \times T)^{-n}&(A \times A') \cap (B \times B')) \\
&= \lim_n \mu \times \mu((T^{-n}(A) \cap B) \times (T^{-n}(A') \cap B')) \\
&= \lim_n \mu(T^{-n}(A) \cap B)\mu(T^{-n}(A') \cap B') \\
&= \mu(A)\mu(B)\mu(A')\mu(B') \\
&= \mu(A \times A')\mu(B \times B').
\end{aligned}$$

The reader can extend this to finite unions. □

Note that in contrast, the diagonal action on a product of ergodic systems need not be ergodic. For instance, the system $(S^1 \times S^1, \mu \times \mu, T_\alpha \times T_\alpha)$, where $T_\alpha(x) = x + \alpha$, is not ergodic, since it preserves the foliation of $T^2 = S^1 \times S^1$ by circles of slope 1, so we can divide $T^2$ into two positive measure invariant sets by dividing the set of such circles in half.

To conlude the section, let's briefly survey some deeper examples.

7.1. **Random walks on finite graphs.** Say that $S$ is a finite set of $n$ 'vertices', and that $P$ is a 'stochastic matrix' on $S$, meaning that

$$P = (P_{ab}, \ a, b \in S),$$

where each $P_{ab} \in [0, 1]$ and for all $b$, we have $\sum_{a \in S} P_{ab} = 1$. Graphically, we draw $S$ as the set of vertices of a directed graph $G$, where there's an edge labeled $P_{ab}$ from $a$ to $b$ when $P_{ab} > 0$. Regard $P_{ab}$ as the probability that if we're currently at vertex $a$, we next 'walk' to state $b$. Iterating, we regard the data $S, P$ as encoding a $P$-*random walk* on the finite directed graph $G$.

**Fact 7.7** (Informal)**.** *Given $k \in \mathbb{N}$, the entry $(P^k)_{ab}$ represents the probability that a $P$-random walk starting at $a$ ends at $b$ after $k$ steps.*

*Proof.* It's just matrix multiplication and induction. The fact is true by definition for $k = 0, 1$, and assuming it's true for $k$, we have

$$(P^{k+1})_{ab} = \sum_{c \in S} P^k_{ac} P_{cb}.$$

Here, $P^k_{ac}$ is the probability you walk from $a$ to $c$ in $k$ steps, and then $P_{cb}$ is the probability that afterwards we walk from $c$ to $b$. Summing over all possible $c$ gives the probability that we walk from $a$ to $b$ in $k+1$ steps. □

A vector $v = (v_a, a \in S)$ is a *probability vector* if its entries are in $[0, 1]$ and $\sum_a v_a = 1$. You can regard each entry $v_a$ as indicating the probability that a walker starts at vertex $a$. We say $v$ is $P$-*stationary* if $vP = v$. Here, $(vP)_a = \sum_b v_b P_{ba}$ is the probability that after starting at a $v$-random vertex, we then walk to state $b$, so the equality says that the distribution of our unknown object is the same under one step of the walk.

As an example, say that $G$ is a finite $d$-regular (undirected) graph, and label each orientation of each edge of $G$ with $1/d$. Then the associated random walk is called the *simple random walk* on $G$, and the uniform probability vector $v = (1, 1, \ldots, 1)$ is $P$-stationary, as you can easily check.

**Fact 7.8.** *For any stochastic matrix $P$, there's a $P$-stationary probability vector $v$.*

*Proof.* Set $D = \{$ probability vectors $v$ $\}$. Then topologically, $D$ is a disk (it's an $(n-1)$-simplex in $\mathbb{R}^n$ obtained by intersecting the plane given by the equation $\sum_a v_a = 1$ with the positive cone) and the right action of $P$ is a continuous action on $D$, so there's a fixed point by Brouwer's theorem. □

So, suppose $v$ is a $P$-stationary probability vector, and look at $S^{\mathbb{N}}$, the set of all sequences $(a_0, a_1, \ldots)$, where $a_i \in S$. We define a probability measure $\mu$ on $S^{\mathbb{N}}$ that indicates the probability that a sequence $(a_0, a_1, \ldots)$ occurs as the itinerary under iterated transitions, after we pick a state randomly according to $v$ and transition randomly according to $P$. Rigorously, $\mu$ is defined on cylinders, and

$$\mu(C[a_0, \ldots, a_n]) := v_{a_0} P_{a_0 a_1} P_{a_1 a_2} \cdots P_{a_{n-1} a_n}.$$

You can check this definition is finitely additive, so Carathéodory's extension theorem implies there's a unique Borel probability measure $\mu$ on $S^{\mathbb{N}}$ taking these values on cylinders.

**Fact 7.9.** *$\mu$ is invariant under the shift map $\sigma : S^{\mathbb{N}} \longrightarrow S^{\mathbb{N}}$.*

*Proof.* We just have to check $\sigma_* \mu = \mu$ on cylinders. But

$$\begin{aligned}
\sigma_* \mu(C[a_1, \ldots, a_n]) &= \sum_{a_0 \in A} \mu(C[a_0, \ldots, a_n]) \\
&= \sum_{a_0 \in A} v_{a_0} P_{a_0 a_1} \cdots P_{a_{n-1} a_n} \\
&= P_{a_1 a_2} \cdots P_{a_{n-1} a_n} \\
&= \mu(C[a_1, \ldots, a_n]),
\end{aligned}$$

where the second to last equality uses that $v$ is $P$-stationary. $\square$

So, when is $(S^{\mathbb{N}}, \mu, \sigma)$ ergodic, or mixing? We say $P$ is *irreducible* if for all $a, b \in S$, we have $(P^k)_{ab} > 0$ for some $k$. Intuitively, if we start at $a$, there's a positive probability we'll eventually walk to $b$. We say $P$ is *aperiodic* if it's NOT the case that there's some $a \in S$ and some $m \in \mathbb{N}$ such that $(P^k)_{aa} > 0$ only when $m|k$. Here, the latter condition means that some $a$ can only come back to itself at times $m, 2m, 3m, \ldots$.

**Theorem 7.10.** *If $v$ is strictly positive, the system $(S^{\mathbb{N}}, \mu, \sigma)$ is ergodic if and only if $P$ is irreducible, and is mixing if and only if $P$ is irreducible and aperiodic.*

The assumption $v > 0$ is necessary for the 'only if' part. E.g. if $S = \{a, b\}$, $P = Id$ and $v_a = 1$, $v_b = 0$, then our system starts at $a$ with full probability and with full probability walks back to $a$, so our measure $\mu$ is atomic on $(a, a, \ldots)$, and is ergodic, even though $P$ isn't irreducible.

Note that if $v > 0$ and $P$ isn't irreducible, say $P^k_{ab} = 0$ for all $k$, then the set

$$E = \{(a_0, a_1, \ldots) \mid a_i \neq b \text{ for large } i\}$$

is shift invariant and has measure at least $v_a > 0$, since any sequence that starts with $a$ almost surely has no $b$'s in it. Also, if $NB_n((a_i)) = 1$ if $a_n \neq b$ and is zero otherwise, then we have

$$\mu(E) = \int \liminf_{n\to\infty} NB_n \, d\mu \leq \liminf_{n\to\infty} \int NB_n \, d\mu = 1 - v_b,$$

where the inequality is Fatou's Lemma, and the last equality is because the probability that $a_0 = b$ is $1 - v_b$, and the measure is shift invariant. So $(S^{\mathbb{N}}, \mu, \sigma)$ isn't ergodic.

Also, suppose that $P$ is irreducible but isn't aperiodic, so there's some $a \in S$ and some $m \in \mathbb{N}$ such that $(P^k)_{aa} > 0$ only when $m|k$. Then if $C = C[a]$, we have $\mu(\sigma^{-k}(C) \cap C) = 0$ whenever $m \not| k$, so these numbers can't converge to $\mu(C)^2 > 0$.

7.2. **Interval exchange transformations, or IETs.** Fix a permutation $\pi$ of $\{1, \ldots, n\}$ and a vector $\lambda \in \mathbb{R}^n_+$ with $\sum_{i=1}^n \lambda_i = 1$. Using the data $(\pi, \lambda)$, create two partitions of the interval $[0, 1)$ into $n$ subintervals, written from left to right as

$$[0, 1) = I_1 \cup I_2 \cup \cdots \cup I_n, \quad [0, 1) = J_1 \cup J_2 \cup \cdots \cup J_n,$$

where for each $i$, the interval $I_i$ has length $\lambda_i$ and the interval $J_i$ has length $\lambda_{\pi(i)}$. By convention, let's consider all intervals as half open, closed on the left and open on the right. Let

$$f : [0, 1) \longrightarrow [0, 1)$$

be the map that takes each $I_i$ to $J_{\pi(i)}$ isometrically and orientation preservingly. This $f$ is called the *interval exchange transformation*, or *IET*, associated to the data $(\pi, \lambda)$. Informally, we're just taking a partition of $I$ into $n$ subintervals of varying lengths, and then $f$ is the map obtained by reordering the intervals according to the permutation $\pi$. Note that since $f$ is piecewise isometric, it preserves the Lebesgue probability measure on $[0, 1)$.

**Example 7.11.** *If $n = 2$, then $\pi$ is either the identity (in which case $f$ is the identity), or the transposition of $\{1, 2\}$. In the latter case, $f$ just translates to the right by $\lambda_1$, mod $L$, so is conjugate to a rotation of the circle, and we understand its dynamics pretty well. You can also understand IETs with $n = 3$ well using circle* *rotations. E.g. if $\pi$ is the transposition $(12)$ of $\{1,2,3\}$ then the corresponding IET fixes the third interval and acts in a way conjugate to a circle rotation on the union of the first two, while a 3-cycle acts in a way conjugate to a circle rotation on the whole interval.*

When $n \geq 4$ the dynamics of IETs can be quite complicated! For instance, let's say that a continuous dynamical system is *minimal* if every orbit is dense. Minimal circle rotations (i.e. irrational ones) are always uniquely ergodic, but Keynes and Newton [17] constructed a minimal IET with $n = 5$ that has two distinct invariant probability measures. More generally, in the 80s Masur [22] and Veech [36] independently proved that for a given 'irreducible' $\pi$, almost every $\lambda$ gives a uniquely ergodic IET. Here, $\pi$ is *irreducible* if there's no $k < n$ such that

$$\pi(\{1,\dots,k\}) = \{1,\dots,k\}.$$

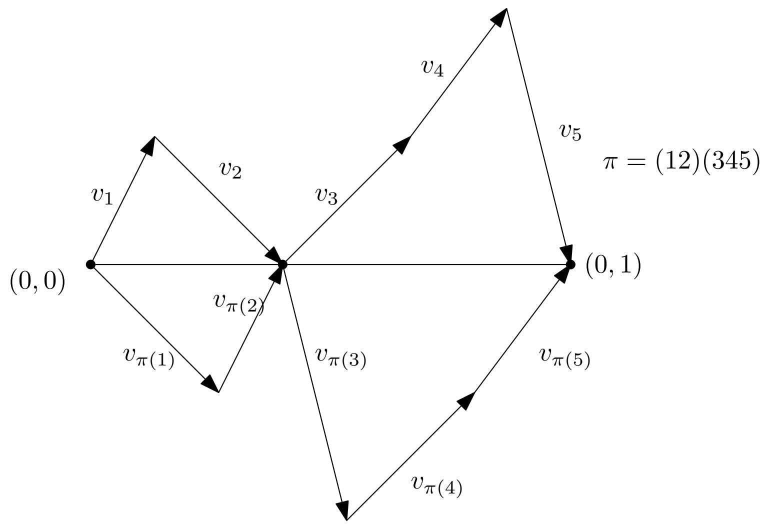
Note that if $\pi$ is reducible, then the corresponding IET $f$ preserves the union of the first $k$ intervals, and its complement, so we can scale the Lebesgue measures on the union and its complement independently to get a one parameter family of probability measures, each of which is preserved by $f$. However, in 1980 Katok [14] showed that like circle rotations, IETs are never mixing. But in 2007, Avila-Forni they're a.e. weakly mixing.

There's a strong connection between IETs and the 'vertical flows' on '(singular) translation surfaces'. Here, a *(singular) translation surface* is a surface $S$ obtained by isometrically gluing sides of a Euclidean polygon $P$ via translations. Any such surface $S$ inherits a well-defined *vertical flow*. Informally, this flow moves a point in the interior of $P$ straight up at unit speed, and because the sides of $P$ are identified by translations, the flow is well defined on interiors of edges too. However, it's not defined at the vertices of $P$, so usually we just restrict to the (full measure) subset of $S$ consisting of points that do not flow into vertices.

Any IET can be realized as a 'first return map' for the vertical flow of some translation surface. From $(\pi, \lambda)$, create a polygon $P$ as follows. Create vectors

$$v_i := (\lambda_i, \tau_i), \quad \tau_i := \pi(i) - i$$

and then create a (possibly degenerate) polygon $P$ as in the following picture. You can check that because of the definition of $\tau_i$, the top sides of the polygon are all above height 0, while the bottom ones are all below height 0, so there are no intersecting edges. For each $i$, glue the $v_i$ side to the other side represented by the same vector, giving a translation surface. Then vertical flow from a point $x \in [0,1]$ on the horizontal line first returns to $[0,1]$ at the point $f(x)$, where $f$ is the IET corresponding to $(\pi, \lambda)$. This process is called 'suspending' the IET to a translation surface. Note that there's some flexibility in how we define the $\tau_i$; they could be perturbed and the picture would still work. Conversely, if $S$ is a translation surface where the vertical flow is 'minimal' (all vertical leaves are dense in $S$) then if $\gamma \subset S$ is an open horizontal segment, any $x \in \gamma$ is guaranteed to eventually flow back to $\gamma$, and the first return map to $\gamma$ is an IET. It's a theorem of Veech that these two operations are inverses in a sense: for any translation surface with minimal vertical flow, there's a horizontal segment where the first return IET suspends to $S$.




## 8. Measure preserving group actions

Previously, we considered only dynamical systems given by iterating a single m.p. map $T : X \longrightarrow X$. What changes if we replace this by a group action?

**Definition 8.1.** Let $G$ be a locally compact second countable group, and $(X, \mathcal{B})$ a measurable space. Endow $G$ with the Borel $\sigma$-algebra and $G \times X$ with the product $\sigma$-algebra, generated by products of measurable sets in the factors $G, X$.

An action $G \curvearrowright X$ is *measurable* if the action map $G \times X \longrightarrow X$ is measurable. If $\mu$ is a measure on $(X, \mathcal{B})$, we say that the measurable action $G \curvearrowright (X, \mu)$ is *measure preserving* if for each $g \in G$, the action of $g$ is measure preserving.

For example, $X$ will usually be a topological space, with the $G$-action

$$G \times X \longrightarrow X$$

continuous. Possible $G$'s that will arise are countable groups with the discrete topology (e.g. $\mathbb{Z}$, or $\mathbb{Z}^n$) or Lie groups, in particular $\mathbb{R}$. A continuous $\mathbb{R}$-action on a topological space $X$ is usually called a *flow*. Often, if $t \in \mathbb{R}$ instead of writing $t(x)$ for the action of $t$, as we might for a general group action, we'll write $\phi_t(x)$.

**Example 8.2** (General Bernoulli actions)**.** *Say that $\Gamma$ is a countable group, and $S$ is a finite set equipped with a probability measure $\nu$. Then*

$$\{0,1\}^\Gamma := \{f : \Gamma \longrightarrow \{0,1\}\}$$

*comes equipped with a natural $\Gamma$-action, namely $f \mapsto f \circ \gamma$, where $\gamma$ acts on $\Gamma$ on the left. The product measure $\mu$ associated to $\nu$ is $\Gamma$-invariant.*

**Example 8.3** (Translation flow on a torus)**.** *Let $T^n$ be the $n$-torus, $T^n = \mathbb{Z}^n \backslash \mathbb{R}^n$, and define a flow on $T^n$ by picking a vector $v \in \mathbb{R}^n$ and setting $t(x) = x + tv$. This flow preserves Lebesgue measure. More generally, if $S$ is a (singular) translation surface as considered in the last section and $v \in \mathbb{R}^n$, then there's a similar flow defined on the complement of the set of points that flow into or out of vertices, and this flow also preserves Lebesgue measure.*

**Definition 8.4.** Suppose $G \curvearrowright (X, \mu)$ is measure preserving. We say that

(1) the action is *ergodic* if any $G$-invariant subset $A \subset X$ has either zero or full measure,
(2) the action is *mixing* if whenever $g_n \to \infty$ in $G$, we have for all measurable $A, B \subset X$ that $\mu(g_n(A) \cap B) \to \mu(A)\mu(B)$.

Here, $g_n \to \infty$ means that it exits every compact subset of $G$: that is, for every compact $K \subset G$, we have $g_n \in G \setminus K$ for all large $n$. As before, mixing easily implies ergodic. The ergodic decomposition theorem still holds, so any $G$-invariant probability measure can be written as an integral of ergodic ones. For flows, there's even an ergodic theorem that follows from Birkhoff's theorem.

**Theorem 8.5.** *Suppose we have a measure preserving flow $(\phi_t)$ on a probability space $(X, \mu)$. Given $f \in L^1(X, \mu)$, define*

$$A_N(f)(x) := \frac{1}{N} \int_0^N f \circ \phi_t(x)\, dt, \quad N \in \mathbb{R}_+$$

*Then as $N \to \infty$, the functions $A_N(f)$ converge to some function $f^*$, both in $L^1$ and pointwise a.e. This $f^*$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of almost $(\phi_t)$-invariant subsets of $X$, and if $(X, \mu, \phi_t)$ is ergodic, then we can take $f^*(x) = \int f\, d\mu, \forall x$.*

Alternatively, one could average over $[-N, N]$ if desired.

*Proof Sketch.* Set $g(x) = \int_0^1 f \circ \phi_t(x)\, dx$, which using Fubini you can show is finite a.e. and in $L^1$. Then we have

$$A_N(f) = \frac{1}{N} \sum_{n=0}^{N-1} g \circ \phi_n(x).$$

Since $\phi_n = (\phi_1)^n$, the right side is just one of the averages in Birkhoff's theorem, so the averages converge for $N = 1, 2, \ldots$ to some $f^*$ pointwise a.e. and in $L^1$. You can then show that actually, $A_N(f)$ is almost $\phi_t$ invariant for any fixed $t$, from which it follows that it's the conditional expectation referenced above, and you can use this to conclude that $A_N(f)$ converges to this conditional expectation not just for integer $N$, but for $N \in \mathbb{R}$ as $N \to \infty$. □

Are there ergodic theorems for more general group actions? Like, say $\Gamma = \langle S \rangle$ is a group generated by a finite set $S$, and you have a m.p. action $\Gamma \curvearrowright (X, \mu)$ on a probability space. For $\gamma \in \Gamma$, set $|\gamma|$ to be the minimal length of a word in $S$ that represents $\gamma$, and consider the ball

$$B_N := \{\gamma \in \Gamma \mid |\gamma| \leq N\},$$

Given $f \in L^1(X, \mu)$, we could then define

$$A_N(f)(x) := \frac{1}{|B_N|} \sum_{\gamma \in B_N} f \circ \gamma(x),$$

and we could hope that the functions $A_N(f)$ converge to something as $N \to \infty$.

This doesn't always work, unfortunately. For instance, take the action of the rank 2 free group

$$F = \langle a, b \rangle, \quad F \curvearrowright X := \{-1, 1\},$$

where both $a, b$ act as the nontrivial transposition. If $f : X \longrightarrow \{-1, 1\} \subset \mathbb{R}$ is the identity function, then we get

$$A_N(f)(1) := \frac{1}{|B_N|} \sum_{w \in B_N} w(1) = \frac{1}{|B_N|} \sum_{w \in B_N} (-1)^{|w|} = \frac{1}{|B_N|} \sum_{n=0}^{N} (-1)^n |S_n|,$$

where $S_n := \{w \in F \mid |w| = n\}$. But for $n \geq 1$, we have $|S_n| = 4 \cdot 3^{n-1}$, while $|S_0| = 1$, and we have $|B_N| = \sum_{n=0}^{N} |S_N|$, so you can just calculate this, giving

$$\begin{aligned} \cdots &= \frac{1}{1 + 2(3^N - 1)} \left( 1 + \sum_{n=1}^{N} (-1)^n 4 \cdot 3^{n-1} \right) \\ &\approx \frac{1}{2 \cdot 3^N} \cdot 4 \cdot \sum_{n=0}^{N-1} (-3)^n \\ &= \frac{1}{2 \cdot 3^N} \cdot 4 \cdot \frac{(-3)^N - 1}{(-3) - 1} \\ &\approx \frac{1}{2} (-1)^N, \end{aligned}$$

so the sequence $A_N(f)(1)$ oscillates back and forth between approximately $\frac{1}{2}$ and approximately $-\frac{1}{2}$, so doesn't converge. Similarly, $A_N(f)(-1) = -A_N(f)(1)$ doesn't converge. One way to fix this particular example is by averaging (say) over the sphere $S_{2n}$ and letting $n \to \infty$, and sort of amazingly, it turns out that if you do this sort of averaging you do get an ergodic theorem for free group actions on probability spaces, at least for functions in $L^p$, where $p > 1$, see Nevo-Stein [23], if not for $L^1$, see Tao [32]. Any action of a finitely generated group $\Gamma$ induces an action of some free group, defined so that it factors through a chosen surjection $F_k \longrightarrow \Gamma$, so in some sense Nevo-Stein does give ergodic theorems for arbitrary group actions, but the averaging process isn't particularly natural to $\Gamma$.

The real issue in the rank 2 free group counterexample above is that the sphere $S_N$ takes up a definition proportion of $B_N$: indeed, we have

$$|S_N|/|B_N| \approx 4 \cdot 3^{N-1}/(2 \cdot 3^N) \to 2/3, \quad \text{as } N \to \infty.$$

Another way to eliminate this counterexample is to restrict to 'amenable groups'. In the finitely generated setting, say, a group $\Gamma = \langle S \rangle$, where $|S| < \infty$ and $S = S^{-1}$, is called *amenable* if it contains a sequence of finite subsets

$$F_1 \subset F_2 \subset \cdots, \quad \cup_N F_N = \Gamma, \text{ where } \lim_{N \to \infty} \frac{|F_N \Delta S F_N|}{|F_N|} = 0.$$

Note that if we take the balls $B_N$ in the rank 2 free group $F = \langle S \rangle$, where $S = \{a, b, a^{-1}, b^{-1}\}$, then $B_N \Delta S B_N = S_{N+1}$, so the limit is positive. A chain $(F_N)$ as above is called an (increasing) *Følner sequence* for $\Gamma$. Any group $\Gamma = \langle S \rangle$ as above where $N \mapsto |B_N|$ is subexponential is amenable, and you can just take $(B_N)$ as your Følner sequence. This covers for instance all abelian, or nilpotent groups. More general, all solvable groups are amenable, but as there are solvable groups of exponential growth, you can't always take your Følner sequence to be balls.

For amenable groups, we have the following:

**Theorem 8.6** (Mean ergodic theorem for amenable groups)**.** *Suppose $\Gamma = \langle S \rangle$ is amenable with Følner sequence $(F_N)$, and that $\Gamma \curvearrowright (X, \mu)$ is a measure preserving* *action on a probability space. If $[f] \in L^1(X,\mu)$, and we define*

$$A_N(f) := \frac{1}{N} \sum_{\gamma \in F_N} f \circ \gamma,$$

*then $A_N(f)$ converges in $L^1$ to the conditional expectation of $f$ with respect to the $\sigma$-algebra of almost $\Gamma$-invariant subsets of $X$. In particular, if the action is ergodic, then $A_N(f) \to \int f \, d\mu$.*

There are also pointwise ergodic theorems like Birkhoff's that work for amenable groups, but they typically only work for special Følner sequences. For instance, let's say that $(F_N)$ is *tempered* if there's some $C > 0$ such that

$$|\bigcup_{n<N} F_n^{-1} F_N| \leq C|F_N|.$$

One can show that any Følner sequence has a tempered subsequence.

**Theorem 8.7** (Lindenstrauss [20]). *In the setting of the theorem above, if $(F_N)$ is tempered then $A_N(f)$ converges pointwise a.e.*

There are also specific examples of nontempered Følner sequences where pointwise convergence holds, e.g. for $\mathbb{Z}$ and $F_N = \{-N, \dots, N\}$, we get pointwise convergence by Birkhoff, even though $(F_N)$ isn't tempered, and Tempelman [33] proved a similar theorem for $\mathbb{Z}^d$ and $F_N = \{-N, \dots, N\}^d$, see also Sarig's notes on ergodic theory for a proof in English.

## 9. Geodesic flow

Let $M$ be a Riemannian $n$-manifold. Associated to the Riemannian metric on $M$, there's an operation called *covariant derivative* as follows. If $\gamma : I \longrightarrow M$ is a path and $X : I \longrightarrow TM$ is a vector field over $\gamma$, so $\pi \circ X = \gamma$ where $\pi$ is the natural projection from $TM$ to $M$, then for every $t \in I$, there's a vector

$$D_t X \in TM_{\gamma(t)}$$

that records how the vector field $X$ is changing along $\gamma$. The covariant derivative $D_t$ has some nice properties: for instance, it's linear and we have

(1) $D_t(fX) = f D_t X + f' X$, if $f : I \longrightarrow \mathbb{R}$ is smooth,
(2) $\frac{d}{dt}\langle X, Y\rangle = \langle X, D_t Y\rangle + \langle D_t X, Y\rangle$, if $X, Y$ are two vector fields over $\gamma$.

You can write down a definition of $D_t$ in coordinates, using the coordinate description of the Riemannian metric on $M$, and then verify these properties, see e.g. Lee [18]. We won't go into it here, but for example, if $M = \mathbb{R}^n$ with the Euclidean metric, then $D_t X = \frac{d}{dt} X$, where we write $X$ as a function $\mathbb{R}^n \longrightarrow \mathbb{R}^n$, and if $M \subset \mathbb{R}^n$ is an embedded submanifold with the Riemannian metric inherited from $\mathbb{R}^n$, then $D_t X$ is the orthogonal projection of $\frac{d}{dt} X$ onto the subspace $TM_{\gamma(t)} \subset T\mathbb{R}^n_{\gamma(t)}$.

A vector field $X$ over a path $\gamma$ is called *parallel* if $D_t X = 0$ for all $t$. Intuitively, $X$ is not really changing along $\gamma$, it's just being dragged along $\gamma$ as efficiently as possible, with respect to the given metric. A *geodesic* in $M$ is a smooth path $\gamma : I \longrightarrow M$ where $I \subset \mathbb{R}$ is some interval, such that $D_t \gamma'(t) = 0$ for all $t$; so, in other words, the velocity vector field is parallel. Intuitively, since $D_t$ is a metric-defined derivative, geodesics are paths that have 'zero acceleration'. As such, they're 'straight' paths in $M$. For example, if $M = \mathbb{R}^n$, then the geodesic equation is just

$\gamma''(t) = 0$, which describes constant speed parametrizations $t \mapsto p + tv$ of lines in $\mathbb{R}^n$. In general, geodesics always have constant speed, since by (2) above

$$\frac{d}{dt}\langle \gamma'(t), \gamma'(t)\rangle = 2\langle D_t\gamma'(t), \gamma'(t)\rangle = 0.$$

There's also an explicit metric characterization of geodesics: they are exactly the constant speed paths $\gamma$ that are *locally distance minimizing*, meaning that for every $t \in I$, we have that for $s \approx t$, we have that

$$\text{length}(\gamma|_{[s,t]}) = d_M(\gamma(s), \gamma(t)).$$

The equation $D_t\gamma'(t) = 0$ is a second order ODE, so locally, solutions to it exist and are unique given first order data, which in this case are the initial point $\gamma(0) = 0$ and the initial vector $\gamma'(0) = v$. In other words, given $p \in M$, $v \in TM_p$, there's a unique geodesic starting at $p$ with initial velocity $v$, at least up to a choice of domain $I \subset \mathbb{R}$. When $M$ is a complete Riemannian manifold, the Hopf-Rinow theorem says that actually, geodesics always exist for all time, so for every $(p, v) \in TM$, there's a unique geodesic $\gamma_{p,v} : \mathbb{R} \longrightarrow M$ with $\gamma_{p,v}(0) = p$ and $\gamma'_{p,v}(0) = v$. The *geodesic flow* on $TM$ is then defined to be the flow $(\phi_t)$, where

$$\phi_t : TM \longrightarrow TM, \quad \phi_t(p, v) = \gamma'_{p,v}(t).$$

This is a continuous flow, just by ODE theory since the solutions to an ODE depend continuously on the inputs.

As the geodesic flow is a flow on the tangent bundle of $M$, it's helpful to understand $T(TM)$. The covariant derivative on $M$ determines a *canonical splitting*

$$T(TM)_{(p,v)} = H_{(p,v)} \oplus V_{(p,v)} \tag{6}$$

into 'horizontal' and 'vertical' subspaces, where here $H_{(p,v)} \subset T(TM)_{(p,v)}$ is the set of tangent vectors of parallel vector fields $X : I \longrightarrow TM$ over paths, and where $V_{(p,v)} = T(TM_p)_{(p,v)}$, the tangent space to the vector subspace $TM_p$, regarded a $n$-submanifold of the $2n$-manifold $TM$. Here, if $\pi : TM \longrightarrow M$ is the projection, then $d\pi$ restricts to an isomorphism $H_{(p,v)} \longrightarrow TM_p$. The subspace $V_{(p,v)}$ also comes with a canonical isomorphism to $TM_p$, since the tangent space to a vector space is the vector space. Note that the subspace $V_{(p,v)}$ is well defined independent of the metric, but that $H_{(p,v)}$ isn't.

With respect to the canonical splitting, the geodesic flow is generated by the vector field $\sigma$ on $TM$, where $\sigma(p, v) = (v, 0) \in H_{(p,v)} \oplus V_{(p,v)}$. Indeed, the path $t \mapsto \phi_t(p, v)$ is a geodesic, so the velocity field $t \mapsto \frac{d}{dt}\phi_t(p, v) \in TM$ is parallel, and hence its initial tangent vector lies in $H_{(p,v)}$, and projects to $v$. The canonical splitting also allows us to define a Riemannian metric on $TM$ called the *Sasaki metric*, defined by endowing $H_{(p,v)}$ with the pullback under $d\pi$ of the inner product on $TM_p$, and endowing $V_{(p,v)} \cong TM_p$ with the same inner product, and defining the two factors to be orthogonal.

Any Riemannian manifold $M$ has a natural *Riemannian measure* $\mu$, where in local coordinates $(x^i)$, if the metric is given at each point $p$ by the matrix

$$(g_{ij})_p, \quad g_{ij} = \langle \frac{d}{dx^i}, \frac{d}{dx^j}\rangle_p,$$

then the Riemannian measure $\mu$ is obtained by scaling Lebesgue measure in the coordinate chart by the function $p \mapsto |det(g_{ij})_p|$. There's also an induced *Liouville* *measure* on $TM$, which is just the Riemannian measure of the Sasaki metric. Perhaps more intuitively, every $n$-dimensional inner product space $V$ has a canonical *Lebesgue measure*, given by pushing forward the usual Lebesgue measure under an inner-product preserving isomorphism $\mathbb{R}^n \to V$. Liouville measure is then the fiberwise product of the Riemannian measure on $M$ with the Lebesgue measures on all the $TM_p$.

Sometimes, it's more useful to work with the *unit tangent bundle* $T^1M$ instead of $TM$. The geodesic flow preserves $T^1M$, so we get a restricted geodesic flow defined just on $T^1M$. There's a natural *Sasaki metric* on $T^1M$ obtained by restricting the one on $TM$, and there's a natural *Liouville measure* on $T^1M$, which is the Riemannian measure of the Sasaki metric. Alternatively, the Liouville measure is the fiberwise product of Riemannian measure on $M$, and the natural Riemannian measures on the spheres $T^1M_p$, which are Riemannian submanifolds of the inner product space $TM_p$.

**Theorem 9.1** (Liouville's Theorem)**.** *The geodesic flow $(\phi_t)$ preserves the Liouville measures on $TM$ and $T^1M$.*

Briefly, the point is that if a flow $(\phi_t)$ on a Riemanian manifold is generated by a vector field $X$, then $(\phi_t)$ is volume preserving if and only if its 'divergence' is zero. In Euclidean space $\mathbb{R}^n$, the divergence of a vector field $X = (X^i)$ is just

$$\operatorname{div}(X) := \sum_{i=1}^{n} \frac{\partial X^i}{\partial x_i} \in \mathbb{R},$$

and the fact that vector fields with zero divergence are volume preserving is often included in a multivariable calculus course. On a Riemannian manifold, $\operatorname{div}(X)_{(p,v)}$ is the trace of the map $TM_p \longrightarrow TM_p$, defined by taking $v \in TM_p$ to the derivative $D_t X|_{\gamma_v}$ at $t = 0$, where $\gamma_v$ is a path starting at $p$ with initial velocity $v$. We saw above that the geodesic flow is generated by the vector field $\sigma(p,v) = (v,0) \in T(TM)_{(p,v)}$, in the coordinates given by the canonical splitting above. Since $v$ doesn't depend on $p$, and $0$ doesn't depend on $v$, when you calculate the appropriate trace (using the covariant derivative associated to the Sasaki metric), you'll get zero.

In some examples, it's easy to see that geodesic flow is volume preserving explicitly, without using the argument sketched above.

**Example 9.2** (Geodesic flow on $\mathbb{R}^n$)**.** *Here, the Riemannian measure on $\mathbb{R}^n$ is just Lebesgue measure, and the Liouville measure on $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$ is just $2n$-dimensional Lebesgue measure. In the natural coordinates above, geodesic flow is the linear map $\phi_t(p,v) = (p+tv, v)$, which we can represent as a block matrix*

$$\phi_t \begin{pmatrix} p \\ v \end{pmatrix} = \begin{pmatrix} I & tI \\ 0 & I \end{pmatrix} \begin{pmatrix} p \\ v \end{pmatrix},$$

*Since the matrix has determinant $1$, the map $\phi_t$ preserves the Liouville (i.e. Lebesgue) measure on $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$.*

*For the unit tangent bundle, one can argue as follows. Note that*

$$T^1\mathbb{R}^n \cong \mathbb{R}^n \times S^{n-1},$$

*with Liouville measure the product of Lebesgue measure and the Riemannian measure on the unit sphere. Here, the Riemannian measure on $S^{n-1}$ is obtained by* *integrating the volume form of $S^{n-1}$, which is the restriction of the $(n-1)$-form on $\mathbb{R}^n$, regarded with coordinates $(v_1, \ldots, v_n)$, given by the formula*

$$\omega = \sum_i (-1)^{i-1} v_i dv_1 \wedge \cdots \widehat{dv_i} \cdots \wedge dv_n.$$

*So, Liouville measure on $T^1M \subset \mathbb{R}^n \times \mathbb{R}^n$ is given by integrating the form*

$$\omega' = dx_1 \wedge \cdots \wedge dx_n \wedge \omega.$$

*And using our formula for the linear map $\phi_t$ above, we can see that*

$$\phi_t^* dx_i = dx_i + t dv_i, \quad \text{and} \quad \phi_t^* dv_i = dv_i, \implies \phi^* \omega = \omega$$

*and then when we go to compute $\phi_t^* \omega'$, we can FOIL out the sums $dx_i + tdv_i$ in the first iterated wedge, and all terms with repeated $v_i$'s die, so we're left with*

$$\phi_t^* \omega' = \omega' + t \sum_i \pm v_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n \wedge dv_1 \wedge \cdots \wedge dv_n.$$

*But the term $dv_1 \wedge \cdots \wedge dv_n$ on the right vanishes on $T^1M$, since it's impossible to select $n$ tangent vectors to $T^1M$ that are linearly independent in the $v$-factor, so we get that $\phi_t^* \omega' = \omega'$ on $T^1M$, implying that $\phi_t$ preserves Liouville measure.*[5]

The advantage of looking at $T^1M$ instead of the full tangent bundle is that if $M$ has *finite volume*, meaning that $\mu(M) < \infty$, where $\mu$ is the Riemannian measure, then $T^1M$ also has finite volume with respect to the Liouville measure. Note that if $M$ is compact, it has finite volume. So, when $M$ has finite volume, we have a continuous flow $(\phi_t)$ on a finite measure space $T^1M$, which is the setting for ergodic theory that we've been discussing.

**Example 9.3** (Geodesic flow on the unit sphere)**.** *Let $S^n$ be the unit sphere in $\mathbb{R}^{n+1}$. Then for every geodesic $\gamma$ on $S^n$, the image of $\gamma$ is the intersection $P \cap S^n$, where $P$ is a hyperplane through the origin. So, the geodesic flow on $T^1M$ is periodic with period $2\pi$.*

**Example 9.4** (Geodesic flow on a torus)**.** *Let's look at the geodesic flow on the torus $T^n := \mathbb{Z}^n \backslash \mathbb{R}^n$. Here, $\mathbb{Z}^n$ acts isometrically on $\mathbb{R}^n$, so the Riemannian metric on $\mathbb{R}^n$ descends to a 'flat' metric on $T^n$. (Here, a Riemannian metric is* flat *if it's locally isometric to the Euclidean metric on $\mathbb{R}^n$.) Since $\mathbb{Z}^n$ acts on $T^1\mathbb{R}^n \cong \mathbb{R}^n \times S^{n-1}$ by maps that are the identity in the second coordinate, we have global coordinates*

$$T^1T^n \cong T^n \times S^{n-1}$$

*in which geodesic flow is given by $\phi_t(p, v) = (p + tv, v)$. This flow is more interesting than the geodesic flow on $S^n$. For instance, if $v \in S^{n-1} \cap \mathbb{Q}^n$, then we have $mv \in \mathbb{Z}^n$ for some minimal $m$, and then the flow line $t \mapsto \phi_t(p, v)$ have period $m$. So there are flow lines with arbitrarily large periods, and indeed if $v \in S^{n-1}$ does not have coordinates that are all linearly dependent over the rationals, then the flow line of every $(p, v)$ projects to a dense subset of $T^n$, although it's not dense in the second factor.*

[5]Feels like there should be a way to say that the theorem for the unit tangent bundle follows formally from the theorem for the full tangent bundle, or just that there should be an easier proof of this... Let me know if so.

9.1. **Hyperbolic geometry.** Hyperbolic $n$-space $\mathbb{H}^n$ is the unique simply connected, complete Riemannian $n$-manifold with sectional curvatures equal to $-1$, where unique means up to isometry. Two standard models for $\mathbb{H}^n$ are:

(1) the upper half space $H^n := \{x = (x_i) \in \mathbb{R}^n \mid x_n > 0\}. \subset \mathbb{R}^n$, endowed with the Riemannian metric that has norm
$$|\cdot|_{\mathbb{H}^n} = \frac{1}{x_n}|\cdot|_{\mathbb{R}^n},$$
(2) the open unit disc $D^n := \{x \in \mathbb{R}^n \mid |x| < 1\}$, endowed with the metric
$$|\cdot|_{\mathbb{H}^n} = \frac{2}{1-|x|^2}|\cdot|_{\mathbb{R}^n}.$$

In the two models, the *boundary* $\partial\mathbb{H}^n$ of hyperbolic $n$-space can be seen as the circle
$$\partial H^n := \{x \in \mathbb{R}^n \mid x_n = 0\} \cup \infty, \quad \partial D^n := \{x \in \mathbb{R}^n \mid |x| = 1\}.$$

The images of geodesics in $\mathbb{H}^n$ appear in both models as line segments and arcs of circles that are orthogonal to $\partial\mathbb{H}^n$. So for instance, in the upper half space model, geodesics are either vertical lines or semicircles perpendicular to the boundary, while in the disc model, geodesics are either line segments through 0, or the intersections with $D^n$ of a circle orthogonal to $\partial D^n$. Note that given two points $x, y$ in $\mathbb{H}^n \cup \partial\mathbb{H}^n$, there is a unique geodesic with endpoints at $x, y$.

The isometry group $\mathrm{Isom}(\mathbb{H}^n)$ acts transitively on $\mathbb{H}^n$, with stabilizers isomorphic to $O(n)$. Each isometry $f : \mathbb{H}^n \longrightarrow \mathbb{H}^n$ extends continuously to the closure $\mathbb{H}^n \cup \partial\mathbb{H}^n$, which is homeomorphic to a closed ball. Brouwer's fixed point theorem then says that $f$ has a fixed point in this ball, and one can classify isometries into three types, depending on the number and location of the fixed points.

(1) $f$ is *elliptic* if there's a fixed point $p$ in $\mathbb{H}^n$. Here, $f$ preserves the foliation of $\mathbb{H}^n$ by hyperbolic spheres centered at $p$. Example: any element of $O(n)$ acting linearly in the disc model.
(2) $f$ is *parabolic* if it has a single fixed point $\xi$ in $\partial\mathbb{H}^n$. Here, $f$ preserves the foliation of $\mathbb{H}^n$ by *horospheres* centered at $\xi$, which in the two models are planes or spheres in $\mathbb{H}^n$ that are tangent to $\partial\mathbb{H}^n$ at $\xi$. Example: take any fixed-point-free isometry $g$ of $\mathbb{R}^{n-1}$, and act on the upper half space model $H^n$ by $f = g \times id$, preserving the last coordinate.
(3) $f$ is *hyperbolic type* if it has no fixed point in $\mathbb{H}^n$ and two fixed points in $\partial\mathbb{H}^n$. Here, $f$ translates along the geodesic $\alpha$ connecting the two fixed points, and preserves the $r$-equidistant sets $E_r := \{x \in \mathbb{H}^n \mid d(x, \alpha) = r\}$. Example: a dilation $f(x) = \lambda x$, $\lambda > 0$, in the half space model, where the axis is the $x_n$-axis, and the equidistant sets are vertical cones.

9.2. **Geodesic flow on $\mathbb{H}^2$.** Let's consider the upper half space (really, plane) model for $\mathbb{H}^2$, in which the isometry group has a particularly nice representation. Given a matrix $A \in PSL(2, \mathbb{R})$, consider the *fractional linear transformation*
$$f_A : H^2 \longrightarrow H^2, \text{ where if } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ then } f_A(z) = \frac{az+b}{cz+d}$$
using complex coordinates. You can check that $f_A$ is an isometry, and it's orientation preserving since they're holomorphic, so the map $A \mapsto f_A$ defines an embedding of $PSL(2, \mathbb{R})$ in $\mathrm{Isom}^+(\mathbb{H}^2)$. Moreover, since $PSL(2, \mathbb{R})$ acts transitively on $\mathbb{H}^2$ and the stabilizer of $i \in H^2 \subset \mathbb{C}$ is $SO(2)$ (check this), which is the entire set of o.p.

orthogonal isomorphisms of $TH_i^2$, one can show that any o.p. isometry of $\mathbb{H}^2$ can be written as a fractional linear transformation, so the map

$$PSL(2,\mathbb{R}) \longrightarrow \mathrm{Isom}^+(\mathbb{H}^2), \quad A \mapsto f_A$$

is an isomorphism.

This perspective allows for a convenient parametrization of the unit tangent bundle $T^1\mathbb{H}^2$. Namely, $PSL(2,\mathbb{R}) \curvearrowright T^1\mathbb{H}^2$ simply transitively, so if we fix a base vector $v_0 \in T^1\mathbb{H}^2$, then the orbit map

$$O : PSL(2,\mathbb{R}) \longrightarrow T^1\mathbb{H}^2, \quad A \mapsto df_A(v_0)$$

is a homeomorphism.

**Fact 9.5** (Algebraic representation of geodesic flow)**.** *Suppose we take as our base vector the vertical vector $v_0 = i \in TH_i^2$, and let $O$ be the orbit map above. Let $(\phi_t)$ be the geodesic flow on $T^1\mathbb{H}^2$. Then we have*

$$\phi_t(v) = O\left(O^{-1}(v) \cdot A_t\right), \text{ where } A_t := \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

*In other words, $O$ conjugates the flow on $PSL(2,\mathbb{R})$ given by right multiplication by $A_t$ to the geodesic flow on $T^1\mathbb{H}^2$.*

*Proof.* First, note that $df_{A_t}(v_0) = \phi_t(v_0)$, since the geodesic in the direction of $v_0$ is the vertical line, $v_0$ has unit length, and $f_{A_t}(i) = e^t i$ lies along that geodesic at a distance of $t$ from $i$. In general, note that for $A \in PSL(2,\mathbb{R})$, the orbit map $O$ conjugates the action of $A \curvearrowright PSL(2,\mathbb{R})$ by *left* multiplication to the action of $f_A$ on $T^1\mathbb{H}^2$. Left multiplication commutes with right mulltiplication, and geodesic flow commutes with isometries, so if we write $v = df_A(v_0) = O(A)$, we have

$$\phi_t(v) = \phi_t(df_A(v_0)) = df_A \circ \phi_t(v_0) = df_A \circ df_{A_t}(v_0) = df_{A\cdot A_t}(v_0) = O(A \cdot A_t). \quad \square$$

One can use the algebraic perspective to give a concrete proof that geodesic flow preserves Liouville measure on $T^1\mathbb{H}^2$. Namely, the orbit map $O$ conjugates *left multiplication* by $A$ on $PSL(2,\mathbb{R})$ to the action of $df_A$ on $T^1\mathbb{H}^2$, and since $f_A$ is an isometry, the map $df_A$ preserves Liouville measure. So, Liouville measure pulls back under $O$ to a Radon measure on $PSL(2,\mathbb{R})$ that is invariant under left translation. However, $PSL(2,\mathbb{R})$ is a *unimodular Lie group*, meaning that every left invariant Radon measure on $G$ is also right invariant, so in particular this measure is invariant under right multiplication by the matrix $A_t$ in the Fact above, and hence the Liouville measure on $T^1\mathbb{H}^2$ is invariant under geodesic flow. One way to see that $PSL(2,\mathbb{R})$ is unimodular is to show:

- $PSL(2,\mathbb{R})$ is simple, i.e. has no nontrivial, proper normal subgroups. You can do this explicitly using matrices or using hyperbolic geometry. E.g. if $N \leq PSL(2,\mathbb{R})$ is normal, and it contains some hyperbolic type isometry with translation distance $\tau$, it contains all hyperbolic type isometries with translation distance $\tau$, and you can then start composing them to get parabolic isometries and hyperbolic isometries with other translation distances, etc..., eventually proving that $N = PSL(2,\mathbb{R})$.
- Simple Lie groups $G$ are unimodular. For this, note that if $\mu$ is a left invariant Radon measure on $G$, then for each $g \in G$, the pushfoward $(R_g)^*\mu$

by the right multiplication map $R_g : G \longrightarrow G,\ h \mapsto hg$ is also left invariant, and therefore a scale of $\mu$, i.e.

$$(R_g)^*\mu = \lambda_g \mu, \quad \lambda_g \in \mathbb{R}_+.$$

The map $G \longrightarrow \mathbb{R}_+,\ g \mapsto \lambda_g$ is a homomorphism, so simplicity says its kernel must be trivial (it's not if $G \neq 1$, since simple nontrivial $G$ aren't abelian) or everything, which means $\mu$ is right invariant.

9.3. **Hyperbolic manifolds.** A *hyperbolic n-manifold* is a Riemannian manifold $M$ such that each point in $M$ as a neighborhood isometric to an open set in $\mathbb{H}^n$. Equivalently, $M$ has an atlas of charts in $\mathbb{H}^n$ where all transition maps are local isometries. If $\Gamma$ acts properly disontinuously and freely by isometries on $\mathbb{H}^n$,

$$\pi : \mathbb{H}^n \longrightarrow M := \Gamma \backslash \mathbb{H}^n$$

is a covering map, and the Riemannian metric on $\mathbb{H}^n$ pushes down to a hyperbolic metric on $\mathbb{H}^n$. Also, the quotient $M$ is complete as a metric space. Conversely, it's a standard fact that every complete hyperbolic $n$-manifold is isometric to such a quotient, see e.g. [35, Ch 3]. In low dimensions, hyperbolic manifolds can often be constructed via gluings of hyperbolic polyhedra. For instance, there is a regular hyperbolic octagon $P \subset \mathbb{H}^2$ with all interior angles equal to $\pi/4$. One can glue up opposite sides of $P$ to give a genus 2 surface $S$, and then construct charts into $\mathbb{H}^2$ with isometric transition maps by taking the identity chart around interior points of $P$, piecing together two half-charts around points on the interiors of edges, and gluing together neighborhoods of the 8 vertices of $P$ to give a chart around the identified vertex of $S$.

The *volume* of $M$ is the total mass of its Riemannian measure, so in particular $M$ has *finite volume* if its Riemannian measure is a finite measure. Any compact hyperbolic manifold has finite volume, but there are also noncompact manifolds with finite volume. For instance, suppose that $T$ is an *ideal triangle* in $\mathbb{H}^2$, i.e. a region bounded by three bi-infinite geodesics that limit to three distinct points on $\partial\mathbb{H}^2$. All ideal triangles are *congruent*, i.e. they all differ by isometries of $\mathbb{H}^2$, since one can show that $Isom(\mathbb{H}^2)$ acts transitively on triples of points in $\partial\mathbb{H}^2$. So taking the vertices to be $-1, 1, \infty$ in the upper half plane model, a quick computation shows that the area of $T$ is $\pi$. One can then produce noncompact finite volume hyperbolic surfaces by gluing finitely many ideal triangles together along their boundary components. Such gluings may not always be complete, but you can check at least that doubling an ideal triangle gives a complete hyperbolic surface homeomorphic to a sphere with three punctures. Note that if $M$ has finite volume, then the Liouville measure on $T^1M$ is also a finite measure.

**Theorem 9.6.** *The geodesic flow on the unit tangent bundle of any finite volume hyperbolic manifold $M$ is ergodic.*

In contrast, note that the geodesic flow on a round sphere $S^n$ and a torus $T^n$ are not ergodic. On the sphere, take a point $p \in S^n$, a vector $v \in S^n_p$, and note that if $U \ni (p, v)$ is a small neighborhood, then $\cup_t \phi_t(U)$ is invariant, and has positive but not full measure. We leave the torus as an exercise.

Here's the general strategy. If $(\phi_t)$ is a flow on a metric space $V$, the *stable set* $S_+(v)$ and *unstable set* $S_-(v)$ of a point $v \in V$ are the subsets

$$S_\pm(v) = \{w \in V \mid \lim_{t\to\pm\infty} d(\phi_t(v), \phi_t(w)) = 0\}.$$

For example, take the flow $\phi_t(x, y) = (x + t, e^{-t}y)$ on $\mathbb{R}^2$. Then we have that the stable set $S_+(x, y) = \{(x, y') \mid y' \in \mathbb{R}\}$, while $S_-(x, y) = \{(x, y)\}$.

**Lemma 9.7.** *Suppose $V$ is locally compact, $\mu$ is a finite Borel measure and $f \in L^2(V)$ is $(\phi_t)$-invariant. Then there is a measure zero subset $N \subset V$ such that whenever $v, w \in V \setminus N$, we have*

$$w \in S_\pm(v) \implies f(w) = f(v).$$

So, a flow invariant function is invariant mod 0 on stable and unstable sets.

*Proof.* The given assumptions on $V, \mu$ imply that the set of continuous functions on $V$ with compact support are dense in $L^1(V)$. So given $m \in \mathbb{N}$, pick a continuous function $h_m$ on $V$ with $|f - h_m|_1 < \frac{1}{m}$. By the Birkhoff theorem,

$$h_m^+(v) = \lim_{T\to\infty} \frac{1}{T} \int_0^T h_m(\phi_t(v))\, dt$$

exists for all $v$ outside a measure zero subset $N_m \subset V$. Note that since $h$ is uniformly continuous, if $v \in V \setminus N_m$ and $w \in S_+(v)$, then $h_m^+(v) = h_m^+(w)$.

Since $f$ and $\mu$ are flow invariant, we have

$$\frac{1}{m} > |f - h_m|_1 = |f \circ \phi_t - h_m \circ \phi_t|_1 = |f - h_m \circ \phi_t|_1.$$

It then follows that

$$\left|f(\cdot) - \frac{1}{T} \int_0^T h_m(\phi_t(\cdot))\, dt\right|_1 < \frac{1}{m}$$

as well, since balls in $L^1$ are convex, and taking the limit, we have

$$|f - h_m^+|_1 < \frac{1}{m}.$$

Passing to a subsequence, we have by Lemma 5.8 that $h_m^+ \to f$ pointwise outside some measure zero subset $Z \subset V$. Setting $N = Z \cup \big(\cup_m N_m\big)$, which has measure zero, each $h_m^+$ is constant on stable sets outside of $N$, so the same is true for $f$. $\square$

For the geodesic flow $(\phi_t)$ on $T^1\mathbb{H}^n$, the stable and unstable sets of $v \in T^1\mathbb{H}^n$ are defined as follows. Let $\xi_\pm \in \partial\mathbb{H}^2$ be the points such that $v$ *points toward* $\xi_+$ and *away from* $\xi_-$, in the sense that

$$\lim_{t\to\infty} \gamma_v(t) = \xi_+ \in \partial\mathbb{H}^2, \quad \lim_{t\to-\infty} \gamma_v(t) = \xi_- \in \partial\mathbb{H}^2,$$

where $\gamma_v : \mathbb{R} \longrightarrow \mathbb{H}^n$ is the geodesic with $\gamma'(0) = v$. Let $C_\pm(v)$ be the horospheres centered at $\xi_\pm$, i.e. in the disk or half plane model, $C_\pm(v)$ are the Euclidean circles or lines that are tangent to $\partial\mathbb{H}^n$ at $\xi_\pm$, and let

$$S_\pm(v) \subset T^1\mathbb{H}^n|_{C_\pm(v)}$$

be the set of vectors based on $C_\pm(v)$ that point toward $\xi_+$ or away from $\xi_-$, respectively. One can verify that these are indeed the stable and unstable sets by taking $v \in T^1\mathbb{H}^n$ to be vertical in the half space model, noting by direct computation that $S_+(v)$ is as described, so the set of all vertical vectors based at points on the horizontal plane through $v$, and noting that horospheres and geodesic flow are invariant under isometries of $\mathbb{H}^n$. Since here $S_\pm(v)$ are submanifolds of $T^1\mathbb{H}^n$, we'll often call them the *stable and unstable manifolds.*

**Fact 9.8.** *Let $v \in T^1\mathbb{H}^n$, and write $G(v) = \{\phi_t(v) \mid t \in \mathbb{R}\}$. Then we have*

$$TG_v \oplus TS_-(v)_v \oplus TS_+(v)_v = T(T^1\mathbb{H}^n)_v.$$

*Proof.* The dimensions of the spaces above are $1 + (n-1) + (n-1) = 2n-1$, so that checks out at least. Applying an isometry, we can work in the upper half space model with $v$ a vector pointing straight up, and based at $p \in \mathbb{H}^n$, say. Just using Euclidean coordinates, and disregarding the hyperbolic metric, we have

$$T(T\mathbb{H}^n)_v \cong T(\mathbb{H}^n)_p \oplus T(T\mathbb{H}^n_p)_v \cong T(\mathbb{H}^n)_p \oplus T(\mathbb{H}^n)_p.$$

If $P \subset T(\mathbb{H}^n)_p$ is the horizontal $(n-1)$-dimensional subspace, then in these coordinates $TG(v)_v$ is $P^\perp \oplus 0$, $TS_+(v)$ is $P \oplus 0$, and $TS_-(v)$ is $\{(w, 2w) \mid w \in P\}$. So, the map $TG_v \oplus TS_-(v)_v \oplus TS_+(v)_v \longrightarrow T(T^1\mathbb{H}^n)_v \subset T(T\mathbb{H}^n)_v$ is injective, and we're done by the dimension count. $\square$

So, suppose now that $M = \Gamma\backslash\mathbb{H}^n$ is a compact hyperbolic $n$-manifold. If $\pi : \mathbb{H}^n \longrightarrow M$ is the covering map, then

$$d\pi : T^1\mathbb{H}^n \longrightarrow T^1M$$

is also a (regular) covering map, with deck group $\Gamma$, acting on $T^1\mathbb{H}^n$ via the derivative of its action on $\mathbb{H}^n$. The geodesic flow $(\phi_t)$ on $T^1M$ and the geodesic flow $(\tilde{\phi}_t)$ on $\mathbb{H}^n$ then satisfy $\phi_t \circ d\pi(\tilde{v}) = d\pi \circ \tilde{\phi}_t(\tilde{v})$, for every $\tilde{v} \in T^1\mathbb{H}^n$. And if $\tilde{v} \in T^1\mathbb{H}^n$ projects to $v = d\pi(\tilde{v})$, the sets $S_\pm(\tilde{v}) \subset T^1\mathbb{H}^n$ project to the stable and unstable manifolds $S_\pm(v)$, although these are only immersed submanifolds of $T^1M$, and $G(\tilde{v})$ projects to $G(v)$, the flow line through $v$. Note that we still have

$$TG_v \oplus TS_-(v)_v \oplus TS_+(v)_v = T(T^1M)_v. \tag{7}$$

To show that geodesic flow $(\phi_t)$ is ergodic on $T^1M$, it suffices to show that any $(\phi_t)$-invariant $L^1$ function $f : T^1M \longrightarrow \mathbb{R}$ is constant almost everywhere. By definition, $f$ is constant on the flow lines $G(v)$, and Lemma 9.7 tells us that $f$ is constant a.e. on the stable and unstable submanifolds $S_\pm(v)$. But (7) tells us that locally near each $v \in T^1M$, the three foliations by flow lines, stable submanifolds and unstable manifolds are transverse, and since everything in sight is smooth, there's a smooth chart around $v$ wherein these foliations are coordinate foliations in $\mathbb{R}^n$. And one can show (see below) that if outside a measure zero set, a function on $\mathbb{R}^n$ is constant in the directions of a set of the coordinate foliations, it's actually just constant outside of a measure zero set. It follows that $f$ is constant a.e.

Here's the missing statement that says that a function on $\mathbb{R}^n$ that's constant a.e. 'in the direction of the coordinate foliations' is constant a.e. We state it just in $\mathbb{R}^2$ for simplicity, but the proof is the same in general.

**Fact 9.9.** *Suppose $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is measurable and that for all points $(x, y), (x', y')$ outside some measure zero subset $N \subset \mathbb{R}^2$, we have*

$$x = x' \text{ or } y = y' \implies f(x, y) = f(x', y').$$

*Then $f$ is constant outside some measure zero set $N'$ in $\mathbb{R}^2$.*

*Proof.* Let $H_y$ and $V_x$ be the horizontal and vertical lines with fixed coordinates $y$ and $x$, respectively. By Fubini, for almost every $x$, the intersection $N \cap V_x$ has measure zero in $V_x$, and similarly for $H_y$. So, fix some $y = b$ such that $N \cap H_b$ has measure zero in $H_b$, and let

$$X = \{(x, y) \mid (x, b) \in H_b \setminus N,\ (x, y) \in V_x \setminus N\}.$$

Then $X$ has full measure in $\mathbb{R}^2$, since we're selecting a full measure set of $x$'s, and then for each $x$ we take a full measure set of $y$'s. And $f$ is constant on $X$, since if $(x, y), (x', y') \in X$, we have $f(x, y) = f(x, b) = f(x', b) = f(x', y')$. $\square$

**Remark 9.10.** *The same proof outline shows that more generally, the geodesic flow on the unit tangent bundle of any compact (say) Riemannian manifold with negative sectional curvatures is ergodic, see e.g. the appendix to* [3]*. Namely, the stable and unstable sets in the unit tangent bundle of the universal cover are still a pair of transverse foliations as above, whose tangent spaces span together with that of the flow lines. However, there's a lot of subtlety at the end getting the Fubini argument to work, because while the leaves of these foliations are $C^1$, they aren't smooth foliations, so you don't get smooth charts in which they're coordinate foliations as above. The point then becomes to show that the given foliations are well behaved enough that you can at least make such charts that send null sets to null sets. Namely, you need the foliations to be 'absolutely continuous with bounded Jacobians', as described in* [3]*.*

*From another persepective, a differentiable flow $(\phi_t)$ on a compact Riemannian manifold $M$ is called* Anosov *if each flow line is immersed, and there are constants $C > 0$ and $\lambda \in (0, 1)$ such that for each $p \in M$, we have*

$$TM_p = S_p^+ \oplus S_p^- \oplus L,$$

*where $L$ is the tangent space to the flow line $\phi_t$, and $S_p, U_p$ are continuously varying plane fields on $M$ such that*

$$|d\phi_t(v)| \leq C\lambda^t |v| \ \forall v \in S_p^+, \quad |d\phi_{-t}(v)| \leq C\lambda^t |v| \quad \forall v \in S_p^-.$$

*Geodesic flow on the unit tangent bundle of a hyperbolic manifold is Anosov, where if $v \in T^1M$ then $S_v^\pm$ are just the tangent spaces to the stable and unstable submanifolds $S_\pm(v)$. One can show that every Anosov flow that preserves the Riemannian volume is ergodic, in much the same way as in the proof sketch above.*

On a hyperbolic surface $S$, the stable and unstable submanifolds for geodesic flow on $T^1S$ are 1-dimensional, and are the flow lines of the *stable and unstable horocycle flows* on $T^1S$, denoted by $h_t^\pm$. These flows are the projections of the associated flows on $T^1\mathbb{H}^2$, which are defined as follows. Given $v \in T^1\mathbb{H}^2$, we let $\xi_\pm$ be the endpoints in $\partial\mathbb{H}^2$ of the geodesic through $v$, let $C_\pm$ be the horocycle through the basepoint of $v$ centered at $\xi_\pm$, and let $h_t^\pm(v)$ be the unit normal vector to $C_\pm$ whose basepoint lies at a length of $t$ to the right from the basepoint of $v$, along $C_\pm$. In terms of the identification $PSL(2, \mathbb{R}) \longrightarrow T^1\mathbb{H}^2$ discussed in the previous section, the horocycle flows $h_t^\pm$ are given by right multiplication as follows:

$$h_t^\pm(B) = BU_t^\pm, \quad U_t^+ = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad U_t^- = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

Another way to prove that geodesic flow on a finite volume hyperbolic surface is ergodic is via the following steps:

(1) Show that $PSL(2,\mathbb{R})$ is generated by all matrices of the form

$$U_t^{\pm}, \quad A_t := \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \text{ where } t \in \mathbb{R}$$

One can do this algebraically, or one can show the equivalent geometric statement that you can get from any $v \in T^1\mathbb{H}^2$ back to the base vector $v_0 = i \in T\mathbb{H}^2_i$ by first doing unstable horocycle flow until $v$ becomes vertical, then doing geodesic flow until it's height 1, then doing stable horocycle flow until it's based at $i$.

(2) Show that $\phi_t \circ h_s^+ \circ \phi_{-t} = h^+_{e^{-t}s}$ and $\phi_{-t} \circ h_s^- \circ \phi_t = h^-_{e^{-t}s}$, either just by matrix multiplication, or by noting that this is what happens when you conjugate horocycle flow by geodesic flow and then apply it to $v_0$: namely, for the first equality, if you move $v_0$ up vertically a hyperbolic distance of $t$, so up to height $e^t$, then move horizontally a *hyperbolic* length of $s$, so a Euclidean length of $se^{-t}$, then move back down to height 1, you'll end up a horizontal hyperbolic length of $e^{-t}s$ away from where you started.

(3) Using (2), show that if $f$ is an $L^2$ function on $T^1S$ that's $(\phi_t)$ invariant, then it's also invariant under $h^+_{\pm}$. (Prove it first for continuous functions with compact support, which are dense in $L^2$.)

(4) Any $f \in L^2$ that's invariant under all three flows is invariant under the whole right action of $PSL(2,\mathbb{R})$ by (1). But $PSL(2,\mathbb{R})$ acts transitively, and you can show this means $f$ is constant almost everywhere. (If not, pick points $u, v$ of concentration for the preimages $f^{-1}(U)$ and $f^{-1}(V)$, where $U, V \subset \mathbb{R}$ are disjoint, and find an element of $PSL(2,\mathbb{R})$ taking $u$ to $v$. You then get a contradiction, since $f$ is $PSL(2,\mathbb{R})$ invariant and the action is measure preserving.

Here are some additional results.

**Theorem 9.11.** *The geodesic flow on the unit tangent bundle of a finite volume hyperbolic manifold is mixing.*

Recall that a measure preserving flow $(\phi_t)$ on $X$ is *mixing* if for all $A, B \subset X$ we have $\lim_{t\to\pm\infty} \mu(\phi_{t_n}(A) \cap B) = \mu(A)\mu(B)$. Mixing of the geodesic flow was first shown in 1939 work of Hedlund [10], and is a consequence of a more general theorem of Howe-Moore (see [4]) about actions of semisimple Lie groups. In fact, more is true: one can show that the geodesic flow is 'exponentially mixing'. Such results were first proved by Ratner and Moore (see [29], and also Pollicott [27]). One formulation is that for $C^1$ functions $f, g$ on $T^1M$,

$$|\int (f \circ \phi_t) \cdot g \, d\mu - \int f \, d\mu \int g \, d\mu| \le Ce^{-ct}|f|_{C^1}|g|_{C^1},$$

see for instance [19], where a more general statement is proven.

**Theorem 9.12.** *The horocycle flows on the unit tangent bundle of a closed hyperbolic surface are mixing, and uniquely ergodic.*

Mixing was shown in the same paper of Hedlund mentioned above [10]. Unique ergodicity is due to Furstenberg '96, see [9]. Consequently, every horocycle is dense in $T^1S$. When $S$ has finite volume but is noncompact, this isn't true anymore, since there are closed orbits of the horocycle flow around the cusps, but there's a similar theorem in that setting.

## 10. Counting lattice points and closed geodesics

In this section we apply the mixing of the geodesic flow to counting problems in hyperbolic geometry. Here's some motivation.

**Fact 10.1.** *The number $N(R)$ of integer points in the ball $B(0,R)$ of radius $R$ around the origin in $\mathbb{R}^2$, satisfies $N(R) \sim \pi R^2$.*

Here, $f \sim g$ if $f/g \to 1$. For a proof, just note that

$$\pi R^2 \sim Area(B(0, R - \sqrt{2})) \leq N(R) \leq Area(B(0, R + \sqrt{2})) \sim \pi R^2,$$

where the two inequalities follow since the squares with lower left corner at an integer point $p \in B(0,R)$ cover $B(0,R-\sqrt{2})$, and are all contained in $B(0,R+\sqrt{2})$.

**Remark 10.2.** *The Gauss circle problem asks for better estimates on the error term $E(R) = N(R) - \pi R^2$. Gauss showed that $|E(R)| \leq 2\sqrt{2}\pi R$. Conjecturally,*

$$|E(R)| = O(R^{1/2+\epsilon}) \quad \forall \epsilon,$$

*is an optimal bound. Currently it is known that $|E(R)| = O(R^\delta)$ for $\delta = .6298\ldots$ by work of Huxley* [12]*, but not for $\delta = \frac{1}{2}$ by work of Hardy and Landau in 1915.*

For a hyperbolic version of the question above, suppose that $\Gamma$ acts properly discontinuously and freely on $\mathbb{H}^n$, that the quotient $M = \Gamma \backslash \mathbb{H}^n$ has finite volume, and fix some $p, q \in \mathbb{H}^n$. Let $N(q,R)$ be the number of points of the orbit $\Gamma q$ that lie in the ball $B(p,R) \subset \mathbb{H}^n$. How do we estimate $N(q,R)$?

Here, $vol(B(p,R)) \sim Ce^{(n-1)R}$ for some $C = C(n)$, e.g. in two dimensions $vol(B(p,R)) = 2\pi(\cosh(R) - 1) \sim \pi e^R$. If $M$ is compact, you can try running the argument above using copies of a compact fundamental domain for $\Gamma$ rather than squares. The fundamental domain will have diameter at most $D$, so we have

$$\frac{\operatorname{vol} B(p, R-D)}{\operatorname{vol}(M)} \leq N(q,R) \leq \frac{\operatorname{vol} B(p,R+D)}{\operatorname{vol}(M)}, \tag{8}$$

just like in the Euclidean case. Using the asymptotic formulas for ball volumes,

$$e^{-(n-1)D} \cdot \frac{\operatorname{vol} B(p,R)}{\operatorname{vol}(M)} \preceq N(q,R) \preceq e^{(n-1)D} \cdot \frac{\operatorname{vol} B(p,R)}{\operatorname{vol}(M)}, \tag{9}$$

However, it turns out we can do better:

**Theorem 10.3.** $N(q,R) \sim vol(B(p,R))/vol(M)$.

The proof is a special case of an argument from Margulis's thesis. It's also written up nicely in [4] and (more briefly, but intuitively) in [7]. Since the balls $B(p,R)$ grow exponentially in volume, the contribution of orbit points near $\partial B(p,R)$ is not negligible, and the point is to use mixing of the geodesic flow to prove that points in $\Gamma q$ occur more or less randomly near the boudnary $\partial B(p,R)$.

Starting on the proof, write $S_t = S(p,t)$ and let $O_t \subset T^1\mathbb{H}^n$ be the set of unit outward normals to $S_t$. Let $\tilde{\lambda}_t$ be the probability measure supported on $O_t$ that is the pushforward of the Riemannian probability measure on $T^1\mathbb{H}^n_p$ under the map $\phi_t$, where $(\phi_t)$ is geodesic flow and $\pi : T^1\mathbb{H}^n \longrightarrow \mathbb{H}^n$ is the projection.

Let $\rho : \mathbb{H}^n \longrightarrow M$ be the covering projection, so $d\rho : T^1\mathbb{H}^n \longrightarrow T^1M$ is also a covering map. Write 'vol' for the Riemannian measures on manifolds and the Liouville measures on their unit tangent bundles, and let $\lambda_t := (d\rho)_* \tilde{\lambda}_t$.

**Lemma 10.4** (Equidistribution of spheres). *Let $f : T^1M \longrightarrow \mathbb{R}$ be a continuous function with compact support. Then*

$$\int f\,d\lambda_t \to \frac{1}{\mathrm{vol}(T^1M)} \int f\,d\mathrm{vol}, \quad \text{as } t \to \infty.$$

*In other words, when conditioned against continuous functions with compact support, the measures $\lambda_t$ weakly converge to the (normalized) Liouville probability measure on $T^1M$, namely* $\mathrm{vol}/\mathrm{vol}(T^1M)$.

In other words, the projections $d\rho(O_t)$ are equidistributed in $T^1M$ as $t \to \infty$. As an aside, a similar result is also true for large radius circles in $\mathbb{R}^2$, say, when projected into the torus $T^2 = \mathbb{Z}^2\backslash\mathbb{R}^2$, see e.g. [28]. The proof is very different, though. Our proof here uses mixing of the geodesic flow, but that's not true on the flat torus.

*Proof Sketch.* Lift $f$ to a continuous $\Gamma$-invariant map $\tilde{f} : T^1\mathbb{H}^n \longrightarrow \mathbb{R}$, and let $\epsilon > 0$. Since $f$ is uniformly continuous, so is $\tilde{f}$, and there's some $\delta > 0$ with

$$d(x,y) < \delta \implies |\tilde{f}(x) - \tilde{f}(y)| < \epsilon.$$

Let $V \subset T^1\mathbb{H}^n$ be a small open neighborhood of $T^1\mathbb{H}^n_p$. We claim that

$$O_t \subset \phi_t(V) \subset \mathcal{N}_\delta(O_t),$$

if $V$ is sufficiently small. The first inclusion is immediate since $T^1\mathbb{H}^n_p \subset V$. For the second, note that any vector $v$ close enough to $T^1\mathbb{H}^n_p$ is obtained from some vector in $w \in T^1\mathbb{H}^n_p$ by first moving within $S_+(v)$ a length less than $\delta/2$, and then flowing geodesically for time less than $\delta/2$, so in particular we can assume this is true for all $v \in V$, and then $d(\phi_t(v), \phi_t(w)) < \delta$, but $\phi_t(w) \in O_t$. Moreover, if we take $V$ to be a 'symmetric neighborhood' invariant under all isometries of $\mathbb{H}^n$ fixing $p$, then $\phi_t(V)$ is also invariant under all such isometries, which implies

$$\left| \int_{O_t} \tilde{f}\,d\tilde{\lambda}_t - \frac{1}{\mathrm{vol}(\phi_t(V))} \int_{\phi_t(V))} \tilde{f}\,d\mathrm{vol} \right| < \epsilon.$$

For small $\delta$, the set $V$ projects injectively into $T^1M$ under $d\rho$, in which case

$$\begin{aligned} \int_{\phi_t(V)} \tilde{f}\,d\mathrm{vol} &= \int_V \tilde{f} \circ \phi_t\,d\mathrm{vol} \\ &= \int_{d\rho(V)} f \circ \phi_t\,d\mathrm{vol}, \\ &\to \mathrm{vol}(d\rho(V)) \cdot \frac{1}{\mathrm{vol}(T^1M)} \cdot \int_{T^1M} f\,d\mathrm{vol}. \end{aligned}$$

Since $\phi_t$ is volume preserving, we have $\mathrm{vol}(d\rho(V)) = \mathrm{vol}(V) = \mathrm{vol}(\phi_t(V))$, so

$$\int f d\lambda_t = \int \tilde{f}\,d\tilde{\lambda}_t \to \frac{1}{\mathrm{vol}(M)} \int f\,d\mathrm{vol}. \quad \square$$

As a direct consequence, we have a similar equidistribution result for spheres within $M$, rather than in the unit tangent bundle.

**Corollary 10.5.** *If $\pi : T^1M \longrightarrow M$ is the projection, then the measures $\nu_t := \pi_*\lambda_t$ weakly converge to the normalized Riemannian probability measure on $M$.*

*Proof.* $\pi$ is continuous and proper, and $\pi$ pushes forward normalized Liouville measure on $T^1M$ to the normalized Riemannian measure on $M$. $\square$

We now prove the theorem. Pick some $\tilde{q} \in \mathbb{H}^n$ and let $q = \rho(\tilde{q}) \in M$ be the projection, and abusing notation, let $N(q, R) := N(\tilde{q}, R)$, noting that this only depends on $q$, not on $\tilde{q}$. Fix $\epsilon > 0$ such that the projection

$$\rho : B(\tilde{q}, \epsilon) \longrightarrow B(q, \epsilon)$$

is an isometry, and let $\alpha : M \longrightarrow \mathbb{R}$ be a nonnegative function supported in $B(q, \epsilon)$ that has integral 1. For $x \in B(q, \epsilon)$, we have

$$N(q, R - \epsilon) \leq N(x, R) \leq N(q, R + \epsilon),$$

so integrating, we get

$$N(q, R - \epsilon) \leq \int \alpha(x) N(x, R)\, d\mathrm{vol} \leq N(q, R + \epsilon).$$

But if $\tilde{\alpha} : \mathbb{H}^n \longrightarrow \mathbb{R}$ is the lift $\tilde{\alpha} = \alpha \circ \rho$, then we have

$$\begin{aligned} \int \alpha(x) N(x, R)\, d\mathrm{vol} &= \int_{B(\tilde{q},\epsilon)} \tilde{\alpha}(x) N(x, R)\, d\mathrm{vol} \\ &= \int_{B(\tilde{q},\epsilon)} \tilde{\alpha}(x) \sum_{\gamma \in \Gamma} 1_{B(p,R)}(\gamma(x))\, d\mathrm{vol} \\ &= \int_{B(p,R)} \tilde{\alpha}(x)\, d\mathrm{vol} \\ &= \int_0^R \mathrm{vol}_{n-1}(S_t) \int \tilde{\alpha}(x)\, d\tilde{\nu}_t\, dt, \qquad (10) \end{aligned}$$

where $\tilde{\nu}_t := \pi_* \tilde{\lambda}_t$ is the uniform measure on the sphere $S_t := \partial B(p, t)$. As $t \to \infty$, the previous lemma implies that

$$\int \tilde{\alpha}(x)\, d\tilde{\nu}_t = \int \alpha(x)\, d\nu_t \to \frac{1}{\mathrm{vol}(M)} \int \alpha(x)\, d\mathrm{vol} = \frac{1}{\mathrm{vol}(M)}.$$

So, it follows that (10) is asymptotic to $\mathrm{vol}(B(p, R))/\mathrm{vol}(M)$. Since $\epsilon$ was arbitrary, one can then show the same asymptotics for $N(q, R)$. Namely,

$$N(q, R) \leq \int \alpha(x) N(x, R + \epsilon)\, d\mathrm{vol} \sim \mathrm{vol}(B(p, R + \epsilon))/\mathrm{vol}(M).$$

Since $\mathrm{vol}\, B(p, R)$ is asymptotically $Ce^{(n-1)R}$, for some $C$, the right hand side is asymptotic to $e^{c\epsilon} \cdot \mathrm{vol}\, B(p, R)$. Taking $\epsilon \to 0$, we get $N(q, R) \prec \mathrm{vol}(B(p, R))/\mathrm{vol}(M)$, and the other inequality follows similarly.

10.1. **Curve counting.** Suppose that $M = \Gamma \backslash \mathbb{H}^n$ is a closed hyperbolic $n$-manifold, with $\rho : \mathbb{H}^n \longrightarrow M$ the covering map. In the previous section, we fixed a point $p \in \mathbb{H}^n$ and counted (say) the number of points of the orbit $\Gamma p$ that lie in the ball $B(p, R)$. In the quotient, this corresponds to counting homotopy classes of loops based at the projection $\rho(q) \in M$ that have length less than $R$. What if instead we try to count closed geodesics of length at most $L$? This is a similar problem, but there are important differences: we're now counting elements of $\pi_1$ up to conjugacy, and we look at length of geodesic loops rather than loops based at $q$.

**Theorem 10.6** (Counting closed geodesics)**.** *Let $N_{geo}(L)$ be the number of closed geodesics in $M$ with length at most $L$. Then we have*

$$N_{geo}(L) \sim \frac{e^{(n-1)L}}{2(n-1)L}.$$

In 2-dimensions, this is a 1959 result of Huber [11]. The argument we sketch is due to Margulis, who in his thesis proved the theorem in the more general setting of compact manifolds with negative sectional curvature. In the theorem above, we identify two closed geodesics if they differ by a reparametrization. We also are not assuming our geodesics are primitive, i.e. if $\gamma$ is a closed geodesic then $\gamma^2$, obtained by running around $\gamma$ twice, is a different closed geodesic in the count. However, the asymptotics would be the same if we were only counting primitive closed geodesics. Indeed, there is some $\epsilon > 0$ that is less than the length of every closed geodesic in $M$, and then any non-primitive geodesic with length at most $L$ is the $n^{th}$ power of some geodesic with length less than $L/2$, where $n \leq L/\epsilon$, so the number of such non-primitive geodesics is at most

$$\frac{L}{\epsilon}\frac{e^{(n-1)L/2}}{(n-1)L} << \frac{e^{(n-1)L}}{2(n-1)L}.$$

Before starting the proof proper, we record the following lemma, which allows us to construct closed orbits from 'almost closed' orbits.

**Lemma 10.7** (Closing lemma)**.** *Given $\epsilon > 0$, there's some $\delta > 0$ as follows. Suppose $M$ is a hyperbolic n-manifold, and that for some $v \in T^1M$ and $L > 1$, say, we have $d(\phi_L(v), v) < \delta$. Then there's some $w \in T^1M$ with $d(v,w) < \epsilon$, such that $\phi_t(w) = w$ for some $t \in [L-\epsilon, L+\epsilon]$.*

*Sketchy proof sketch.* Working in the universal cover, suppose we have a vector $v \in T^1\mathbb{H}^n$ and an isometry $f : \mathbb{H}^n \longrightarrow \mathbb{H}^n$ such that $d(df(v), \phi_L(v)) < \delta$. Suppose for simplicity that $f$ is hyperbolic type with axis $\alpha$. If $\delta$ is small, then $v$ and $\phi_t(v)$ must lie very close to $\alpha$, as geodesic segments that don't start and end close to $\alpha$ bend toward $\alpha$ significantly, so that their initial and terminal velocities can't almost differ by $df$. So, $v$ lies close to some $w$ pointing along the axis of $\alpha$, which gives a nearby closed orbit in the quotient, and one can check that its period is almost $L$. We leave the details to the reader. See Figure 1. ☐

*Proof Sketch of Theorem 10.6.* Instead of counting closed geodesics with length at most $L$, we'll actually count closed orbits of the geodesic flow $(\phi_t)$ in $T^1M$ that have period at most $L$. If $\mathcal{O}(L)$ is the set of such orbits, then $|\mathcal{O}(L)| = 2N_{geo}(L)$, since a geodesic can be parametrized either forward or backward. So, we want

$$|\mathcal{O}(L)| \sim \frac{e^{(n-1)L}}{(n-1)L}.$$

Let's begin by isolating the dynamical ingredients in the proof. We'll using the fact that geodesic flow on $T^1M$ is mixing, plus the following statement about equidistribution of closed orbits. Fix $L, \epsilon > 0$ and let $\mathcal{O}(L,\epsilon)$ be the set of closed orbits of $(\phi_t)$ with period in $(L-\epsilon, L]$. Let

$$\mu_{L,\epsilon} := \frac{1}{|\mathcal{O}(L,\epsilon)|}\sum_{O \in \mathcal{O}(L,\epsilon)} \frac{1}{L}\nu_O.$$

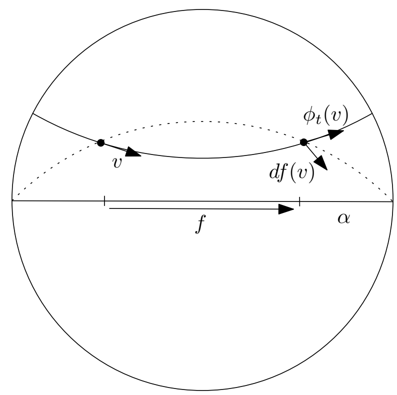



FIGURE 1. A sketchy proof of the closing lemma. If $v$ doesn't lie close to $\alpha$, then $df(v)$ isn't close to $\phi_t(v)$. The dotted line is (part of) the set of points at a certain constant distance from $\alpha$, on which $v$ and $df(v)$ lie. The map $f$ translates along the dotted line, so $v, df(v)$ are as pictured, while $\phi_t(v)$ points more upwards.

**Fact 10.8** (Equidistribution of closed orbits)**.** *The measures $\mu_{L,\epsilon}$ converge weakly to the normalized Liouville measure* $\mathrm{vol}/\mathrm{vol}(T^1M)$ *on* $T^1M$*.*

We'll accept this without proof, but the basic point is that given any subsequence of the measure above, after passing to a further subsequence, they weakly converge to some flow invariant probability measure on $T^1M$, and one can show that this measure has maximum possible entropy for the geodesic flow, and therefore must be the normalized Liouville measure. See [15, §20.1] and [6] for details.

Let $B \subset T^1M$ be a 'flow box', that is a subset with a diffeomorphism

$$\Phi : [0,\epsilon] \times [0,\epsilon] \times [0,\epsilon] \longrightarrow B$$

such that the three coordinate foliations are the images of $S_-, S_+$ and the foliation by flow lines, and where $\Phi(x,y,t) = \phi_t(x,y,0)$ for all $x,y,t$. Briefly, the idea is as follows. To estimate $|\mathcal{O}(L)|$, we'll show it suffices to estimate $|\mathcal{O}(L,\epsilon)|$, since then we can divide $[0,L]$ into intervals of length $\epsilon$, and sum (or rather integrate, as $\epsilon \to 0$) these estimates. Equidistribution implies that on average, closed orbits with period in $(L-\epsilon, L]$ spend a proportionate amount of time running through $B$, so estimating the number of such orbits boils down to estimating the number of times they run through $B$. The closing lemma says that this is essentially the same as counting the number of (essentially different) 'almost closed' orbits that pass through $B$, or in other words, the number of 'essentially different' ways you can take a vector in $B$, flow it for time $t \in (L-\epsilon, L]$, and end up back in $B$. But this is trying to understand something like the intersection $\phi_t(B) \cap B$, which you can understand using the mixing of the geodesic flow.

A bit more rigorously, first note that

$$\mu_{L,\epsilon}(B) = \frac{\epsilon \cdot (\#\text{ of transits})}{L \cdot |\mathcal{O}(L,\epsilon)|},$$

where here a 'transit' is a connected component of $O \cap B$ for some $O \in \mathcal{O}(L,\epsilon)$, which necessarily has the form $I_{x,y} := \{\Phi(x,y,t) \mid t \in [0,\epsilon]\}$ for some $x,y$. Note that as $L \to \infty$, weak convergence $\mu_{L,\epsilon} \to \mathrm{vol}/\mathrm{vol}(T^1M)$ implies that

$$\frac{\epsilon \cdot (\#\text{ of transits})}{L \cdot |\mathcal{O}(L,\epsilon)|} \to \frac{\mathrm{vol}(B)}{\mathrm{vol}(T^1M)}. \tag{11}$$

So, how do we estimate the number of transits? Let $A \subset B$ be a 'slab' of the form $\Phi([0,1]^{n-1} \times [0,1]^{n-1} \times [0,\delta])$, for small $0 < \delta << \epsilon$. We claim:

**Claim 10.9.** *The number of transits of $B$ by elements of $\mathcal{O}(L,\epsilon)$ is approximately the same as the number $N_L$ of components of $\phi_L(A) \cap B$.*

Note that $N_L$ really depends on $A, B, \epsilon$, not just $L$.

*Proof Idea.* If $O \in \mathcal{O}(L,\epsilon)$ and we're given a component $I_{x,y} \subset O \cap B$, set $v = \Phi(x,y,0)$, and then $\phi_L(v) \in O \cap B \subset B$ again, so in particular it lies in some component of $\phi_L(A) \cap B$. Conversely, say we have a component $U \subset \phi_L(A) \cap B$. Then picking a vector $v \in A$ with $\phi_L(v) \in U$, the Closing Lemma gives a closed orbit of geodesic flow with period near the interval $(L-\epsilon, L]$ that passes near $v$. □

So, how do we estimate $N_L$? First, note that $\phi_L$ stretches in the direction of $S_-$ and contracts in the direction of $S_+$, both by exponential factors, so it stretches/contracts $(n-1)$-dimensional volume in those directions by $e^{(n-1)t}$ and $e^{-(n-1)t}$. If we take the width $\delta$ of our slab to be really small, it basically approximates the face $\{\Phi(x,y,0) \mid x,y \in [0,\epsilon]\}$, so we can understand how it is stretched by $\phi_L$ just in terms of what happens in the directions of $S_-, S_+$. Namely, if we look at a single component $U \subset \phi_L(A) \cap B$, for large positive $L$, it'll be skinny in the direction of $S_+$, but in the direction of $S_-$ it still just traverses $B$, so it's only the contraction that matters, not the expansion. So, we get

$$\mathrm{vol}(U) \approx e^{-(n-1)L}\mathrm{vol}(A), \quad \Longrightarrow \mathrm{vol}(\phi_L(A) \cap B) \approx N_L \cdot e^{-(n-1)L} \cdot \mathrm{vol}(A).$$

But since geodesic flow is mixing, we have

$$\mathrm{vol}(\phi_L(A) \cap B)/\mathrm{vol}(A) \to \mathrm{vol}(B)/\mathrm{vol}(T^1M)$$

as $L \to \infty$, which implies that

$$N_L \cdot e^{-(n-1)L} \to \mathrm{vol}(B)/\mathrm{vol}(T^1M). \tag{12}$$

Combining (11) and (12), we get that

$$\frac{\epsilon \cdot e^{(n-1)L}}{L \cdot |\mathcal{O}(L,\epsilon)|} \to 1, \quad \Longrightarrow |\mathcal{O}(L,\epsilon)| \sim \epsilon \cdot e^{(n-1)L}/L.$$

Dividing the interval $[0,L]$ into segments of length $\epsilon$ and summing, and then letting $\epsilon \to 0$, we get that the number of closed orbits of $(\phi_t)$ with period at most $L$ is

$$\sim \int_0^L e^{(n-1)t}/t\,dt \sim \frac{1}{L}\int_0^L e^{(n-1)t}\,dt \sim \frac{e^{(n-1)L}}{(n-1)L}. \quad \square$$

## 11. The surface subgroup theorem

As a further application of dynamical properties of the geodesic flow, we'll sketch in this section a proof of the following theorem of Kahn-Markovic [13].

**Theorem 11.1** (Surface Subgroup Theorem)**.** *Let $M$ be a closed hyperbolic* 3*-manifold. Then $\pi_1 M$ contains a subgroup isomorphic to $\pi_1 S$, where $S$ is a closed orientable surface with genus at least* 2*.*

Since $M$ has contractible universal cover, the conclusion above is equivalent to saying that there is a $\pi_1$-injective map $S \longrightarrow M$, which one can even take to be an immersion. We'll say that an immersed surface is *incompressible* if the associated map is $\pi_1$-injective; so, the theorem says that any closed hyperbolic 3-manifold admits an incompressible immersed surface with genus at least 2.

Let's now present some motivation for the theorem, and discuss some applications of the theorem and its proof techniques.

11.1. **Haken manifolds and the Virtual Haken Conjecture.** A 3-manifold $M$, possibly with boundary, is called *irreducible* if every embedded $S^2 \hookrightarrow M$ bounds a ball. If there is a 2-sphere in $M$ that doesn't bound a ball, you can cut along that sphere and glue balls onto the two resulting 2-sphere boundary components, thus 'reducing' $M$ either as a connected sum in the case that the 2-sphere is separating, or as a sort of 'self-sum' if the 2-sphere is nonseparating. (One says $M$ is *prime* if $M$ is not a nontrivial connected sum; irreducible implies prime, but not the other way around, since $M = S^2 \times S^1$ is prime but not irreducible.)

Every hyperbolic 3-manifold is irreducible, since any embedding $f : S^2 \hookrightarrow M$ lifts to an embedding $\tilde{f} : S^2 \hookrightarrow \mathbb{H}^3$, whose image bounds a ball $B \subset \mathbb{H}^3$ by the Schoenflies theorem, and then $B$ projects to a ball bounded by the image of $f$. (To check that the projection is embedded in $M$, use the Brouwer Fixed Point Theorem and the fact that the deck group of $M$ acts freely on $\mathbb{H}^3$.)

An orientable, irreducible compact 3-manifold with boundary is called *Haken* if it has an incompressible properly embedded orientable surface that isn't a sphere.

**Fact 11.2.** *If $M$ is a compact 3-manifold with boundary and the first Betti number $b_1(M) := H_1(M;\mathbb{R})$ is positive, then $M$ is Haken.*

*Proof Sketch.* By Poincaré duality, $H_2(M, \partial M; \mathbb{R})$ is nontrivial. Any integral class is represented by an orientable embedded surface $S \subset M$. If $S$ is compressible, then the Loop Theorem implies that there's an essential simple closed curve on $S$ that bounds a disk in $M$, and then doing surgery on that curve gives us a (possibly disconnected) surface of lower complexity that represents the same homology class. One of its components is nontrivial in homology, so we can repeat this process, eventually ending up with an incompressible embedded surface in $M$. □

For example, suppose $M$ is a compact 3-manifold with a boundary component that's not a 2-sphere. The 'Half-Lives-Half-Dies' theorem says that the image of

$$H_1(\partial M; \mathbb{R}) \longrightarrow H_1(M; \mathbb{R})$$

has dimension half that of the domain. So, $b_1(M)$ is nontrivial, and $M$ is Haken. This is an important observations, since it allows one to start with an arbitrary Haken 3-manifold $M$, cut it along some incompressible surface, creating a 3-manifold with boundary that is Haken by the argument above, and then cut that along another incompressible surface, etc..., continuing until you end up with a collection of 3-balls. This decomposition of $M$ is called a *Haken heirarchy*, and then one can prove lots of theorems above Haken 3-manifolds by an inductive argument, where the base case is when $M$ is a ball, and the inductive case involves showing that when the theorem is true for the pieces one obtains by cutting a manifold along an incompressible surfaces, then the theorem is true for the manifold itself.

As one particular example, Thurston proved the following famous theorem:

**Theorem 11.3** (Haken hyperbolization theorem)**.** *Suppose that $M$ is a closed, orientable, irreducible Haken 3-manifold, and that $\pi_1 M$ is infinite, but does not contain any $\mathbb{Z}^2$ subgroup. Then $M$ admits a hyperbolic metric.*

In 2003, Perelman [24, 25, 26] removed the assumption that $M$ is Haken, but while Thurston's proof uses techniques squarely in hyperbolic geometry, Perelman's proof starts with an arbitrary Riemannian metric on $M$, flows the metric in a way satisfying a differential equation involving its curvature, and shows that in the limit you get a hyperbolic metric. This proof requires a lot of analysis. It's unclear whether there's a more hyperbolic geometric proof for non-Haken manifolds.

In some sense, Haken manifolds are easier to understand because the incompressible surface gives you a 'place to start' in investigating their topology. The Surface Subgroup Theorem shows that every closed hyperbolic 3-manifold contains an incompressible *immersed* surface with genus at least 2. You might wonder, then, if this can be upgraded to give an incompressible embedded surface. It's known that not every closed hyperbolic 3-manifold is Haken: for instance, Thurston [34] showed that all but finitely many Dehn fillings of the figure eight knot complement are hyperbolic and non-Haken. However, using the Surface Subgroup Theorem, Agol [1] proved the following, previously conjectured by Waldhausen.

**Theorem 11.4** (The Virtual Haken Conjecture)**.** *Suppose that $M$ is a closed hyperbolic 3-manifold. Then $M$ has a finite cover that is Haken.*

In fact, combining the above with some more 3-manifold topology, Agol shows that every closed aspherical 3-manifold is virtually Haken, where 'aspherical' means the universal cover is contractible.

The philosophy of the theorem above is that often, immersed objects can be lifted to embedded objects in finite covers. As an example, draw a figure eight $\gamma$ on closed surface $S$, say, and then construct a finite cover of $S$ where the figure eight lifts to a simple closed curve. In general, resolving self-intersections in a finite cover is really a group-theoretic condition.

**Definition 11.5.** If $G$ is a group and $H \leq G$ is a subgroup, one says that $H$ is *separable* in $G$ if for every $g \in G \setminus H$, there's a finite index subgroup $G' \leq G$ such that $H \leq G'$ but $g \notin G'$.

Here's the connection with the embedding problem.

**Lemma 11.6.** *Suppose that $M$ is a manifold with universal cover $\tilde{M}$ and deck group $G$. If $H \subset G$ is separable and $C \subset H \backslash \tilde{M}$ is compact, then there's a finite index subgroup $G' \subset G$ that contains $H$, and where $C$ embeds under the projection $G' \backslash \tilde{M} \longrightarrow H \backslash \tilde{M}$.*

*Proof.* Let $\pi_H : \tilde{M} \longrightarrow H \backslash \tilde{M}$ be the covering map, and let $\tilde{C} \subset \tilde{M}$ be a compact subset such that $\pi_H(\tilde{C}) = C$. Then the set $S = \{g \in G \setminus \{id\} \mid g(\tilde{C}) \cap \tilde{C} \neq \emptyset\}$ is finite, by proper discontinuity. Since $H$ is separable, there's some finite index subset $G' \subset G$ that contains $H$, and where $S \subset G \setminus G'$. Then given $x \neq y \in C$, pick $\tilde{x} \neq \tilde{y} \in \tilde{C}$ that project to them, and note that $\tilde{x}, \tilde{y}$ can't differ by an element of $G'$, since all nontrivial elements of $G'$ translate $\tilde{C}$ off itself. So, $x, y$ project to distinct elements in $G' \backslash \tilde{M}$. □

To apply the lemma, say we have a 3-manifold $M$ and an incompressible immersion $f : S \longrightarrow M$ such that $f_*(\pi_1 S)$ is separable. Let $\hat{M}$ be the cover of $M$

corresponding to $f_*(\pi_1 S)$. Then $f$ lifts to a map $\hat{f} : S \longrightarrow M$, and using some 3-manifold topology you can show that $\hat{f}$ is homotopic to an embedding. By the lemma, there's a finite intermediate cover $\hat{M} \xrightarrow{\pi} M' \longrightarrow M$ such that $\pi \circ \hat{f}$ is an embedding, so this $M'$ is Haken. To prove Theorem 11.4, then, the goal is to show that the subgroups of closed hyperbolic 3-manifold groups provided by the surface subgroup theorem are separable.

So what are some examples of separable and nonseparable subgroups? A group $G$ is called *residually finite* if the trivial subgroup $1 \subset G$ is separable, i.e. if every nontrivial element of $G$ lies outside some finite index subgroup. As long as $G$ is finitely generated, it has only finitely many subgroups of a given index. So if $g \in G$ lies outside of a finite index subgroup $G'$, then it also lies outside the finite index normal subgroup that is the intersection of all the (finitely many) conjugates of $G'$, and hence there is a homomorphism to a finite group $\phi : G \longrightarrow F$ with $\phi(g) \neq 1$. More generally, $G$ is *residually (blah)* if for every $g \in G \setminus 1$, there is a homomorphism $\phi : G \longrightarrow F$ with $\phi(g) \neq 1$, where $F$ is a (blah) group. It's a theorem of Malcev [21] that every finitely generated subgroup of $GL(n, \mathbb{R})$, say, is residually finite.

On the other hand, we have:

**Fact 11.7.** *The Baumslag-Solitar group* $BS(2,3) := \langle a, t \mid ta^2t^{-1} = a^3 \rangle$ *is not residually finite.*

*Proof.* A group $G$ is called *Hopfian* if every surjective homomorphism $G \longrightarrow G$ is injective. Every finitely generated residually finite group $G$ is Hopfian. Indeed, if $G$ is finitely generated and $f : G \longrightarrow G$ is surjective and not an isomorphism, take some $g \in G$ in the kernel, and some surjection $\phi : G \longrightarrow F$ onto a finite group. There are only finitely many homomorphisms from $G$ to a given finite group, so for some $m < n$ we have $\phi \circ f^n = \phi \circ f^m$. But if we take $h$ such that $f^m(h) = g$, then

$$1 \neq \phi \circ f^m(h), \quad \text{but} \quad \phi \circ f^n(h) = \phi \circ f^{n-m}(g) = \phi(1) = 1,$$

a contradiction. For $BS(2,3)$, though, you can check that the homomorphism $f$ defined by $f(a) = a^2$ and $f(t) = t$ is surjective but not injective. $\square$

A group $G$ is *extended residually finite (ERF)* if all its subgroups are separable. For example, all finitely generated abelian groups are ERF. This is a really strong property, though. For instance, free groups aren't ERF: the kernel of a surjection $F_2 \longrightarrow BS(2,3)$ isn't separable, since if it were $BS(2,3)$ would be residually finite. It's more useful for us to restrict to finitely generated subgroups.

A group $G$ is *locally extended residually finite (LERF)* if any finitely generated subgroup of $G$ is separable. Here, the 'locally' refers to the finitely generated assumption. Note that LERF implies residually finite, so $BS(2,3)$ isn't LERF. However, there are non-LERF residually finite groups, e.g. $F_2 \times F_2$, see [2]. All nilpotent groups are LERF. Hall proved in 1949 that free groups are LERF, and Scott proved in [30] that fundamental groups of closed surfaces are LERF. Note that Scott's result implies, for instance, that any closed curve on a surface can be lifted to a simple closed curve in some finite cover.

To prove Theorem 11.4, then, Agol shows:

**Theorem 11.8.** *Fundamental groups of closed hyperbolic* 3*-manifolds are LERF.*

Consequently, you the immersed surfaces provided by Kahn-Markovic can be lifted to embedded surfaces in a finite cover, proving the Virtual Haken Conjecture.

Theorem 11.8 is entirely group theoretic. Essentially, the idea is as follows. Kahn-Markovic really produce *many* immersed surfaces $S \longrightarrow M$. Lifting to the universal cover, you get a collection of planes that chop up $\mathbb{H}^3$ into compact pieces, say. Following Sageev and Bergeron-Wise, 'dual' to this collection is *cube complex* which is invariant under the action of the deck group $\Gamma$. (One dimension down, imagine taking a bunch of lines in general position in the plane, and making a dual square complex, with one square for each intersection point of two lines, and where each polygonal complementary region has one interior point that's a corner of all the squares corresponding to such intersection points on its boundary.) Taking the quotient, $\Gamma$ is the fundamental group of a 'nonpositively curved cube complex'. Agol shows that this cube complex is 'virtually special', meaning that up to taking a finite index subgroup it has nice combinatorial properties. Then previous work of Wise says that $\Gamma$ is LERF.

In fact, the same work of Wise implies that $\Gamma$ is RFRS (residually finite rationally solvable), a property previously shown by Agol to imply the following theorem, originally posed as a question by Thurston.

**Theorem 11.9** (Virtual Fiber Conjecture)**.** *Any closed hyperbolic* 3*-manifold has a finite cover that fibers over the circle.*

11.2. **The proof of the surface subgroup theorem.** Our goal is to show that for every closed hyperbolic 3-manifold $M$, there's a $\pi_1$-injective map $S \longrightarrow M$, where $S$ is a closed, orientable surface with genus at least 2. We'll see that these surfaces will be 'almost hyperbolic' in some sense, so it makes sense to start with a discussion of how to build hyperbolic metrics on surfaces.

A *pair of pants* is a compact, orientable surface $P$ with genus zero and three boundary components. A *pants decomposition* of a surface $S$ is a multicurve $\Gamma \subset S$ that cuts $S$ into a collection of pairs of pants. If $S$ has a hyperbolic metric, after a homotopy we can assume $\Gamma$ is a union of simple closed geodesics, in which case it cuts $S$ into hyperbolic pairs of pants with geodesic boundary. Conversely, different hyperbolic metrics on $S$ can be constructed by varying the geometries of the pants and the way they're glued together.

**Lemma 11.10.** *If $P$ is a pair of pants with boundary components $\gamma_1, \gamma_2, \gamma_3$, and we're given $l_1, l_2, l_3 > 0$, there's a hyperbolic metric on $P$ with geodesic boundary such that $\gamma_i$ has length $2l_i$. Furthermore, this metric is unique up to isometry isotopic to the identity.*

*Proof Sketch.* Given $l_1, l_2, l_3$, there's a right-angled hexagon (RAH) in $\mathbb{H}^2$ such that the numbers $l_1, l_2, l_3$ are lengths of 3 nonadjacent sides. See Figure 2. This hexagon is unique given a labeling of the sides. Glue two copies of the hexagon together along the remaining three sides, to give a hyperbolic pair of pants as desired.

Conversely, if we're given such a hyperbolic metric, consider the shortest paths between the boundary components of $P$, which are geodesics perpendicular to the boundary; we call these the *orthogeodesics* of $P$. Cutting along the orthogeodesics decomposes $P$ into two RAH's, which are isometric since the sidelengths agree on 3 nonadjacent sides. So, $P$ is obtained via the construction above. $\square$

So, to produce hyperbolic metrics on $S$, we can fix a pants decomposition $\Gamma$ along with desired lengths $l_\gamma$ for each component $\gamma \in \Gamma$, then hyperbolize each pair of pants so that its boundary components have the desired lengths, then glue all these

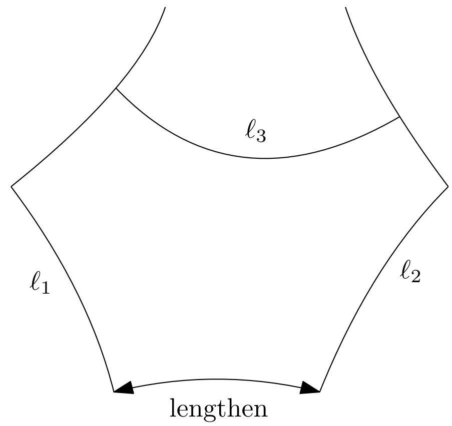



FIGURE 2. Start with a geodesic segment adjacent to two perpendicular segments of length $\ell_1, \ell_2$, and draw two additional perpendicular geodesic rays to those segments. Then vary the length of the original segment until the shortest path between the two rays has length $\ell_3$.

pants together to get a hyperbolic metric on $S$. Note that there's an additional degree of freedom in this construction, since we get to choose how much to twist when we glue each pair of pants to another along a curve in $\Gamma$.

Let's try to repeat some of this up one dimension; it'll be largely the same if we use complex distances instead of real distances as follows. If $\gamma$ is an oriented geodesic in $\mathbb{H}^3$ and $v, w \in N(\gamma)$, the unit normal bundle, then the *complex distance*

$$d_\gamma(v, w) \in \mathbb{C}/2\pi i\mathbb{Z}$$

is defined by setting its real part to be the distance along $\gamma$ from $v$ to $w$, and the imaginary part to be the angle from the plane through $\gamma, v$ to $w$. A *skew right angled hexagon* $H$ in $\mathbb{H}^3$ is just a cyclic concatenation of 6 geodesic segments that meet at right angles, say which we fill in arbitrarily with a 2-cell, if desired. The *complex length* of a side $\gamma$ of $H$ is the distance $d_\gamma(v, w)$, where $v, w$ are the unit normals in the directions of the adjacent sides. Then as in the 2-dimensional case, (the boundary of) any skew RAH in $\mathbb{H}^3$ is uniquely determined up to isometry by three non-adjacent complex side lengths, which can be specified freely.

Now fix a hyperbolic 3-manifold $M$. A *skew pants* in $M$ is a $\pi_1$-injective map $P \longrightarrow M$, where $P$ is an oriented pair of pants, such that each component of $\partial P$ maps to a closed geodesic in $M$. We usually only consider skew pants up to homotopy. There's no restriction on how $int(P)$ is mapped into $M$, but we'll usually suppress the map in notation and pretend that $P$ is embedded in $M$. We always consider $\partial P$ as oriented, with the boundary orientation.

Given a skew pants $P \subset M$, the *orthogeodesics* of $P$ are the shortest paths in $M$ between the boundary components of $P$ that are homotopic rel $\partial P$ to paths on $P$; after a homotopy, we can assume the orthogeodesics all lie in $P$. The *feet* of $P$ are the unit normals to $\partial P$ in the direction of the orthogeodesics, so there are two feet on each component of $\partial P$. The *half length* $hl(\gamma)$ of a component $\gamma \subset \partial P$ is the complex distance from one foot to the other along $\gamma$. Since the orthogeodesics cut up $P$ into two skew RAH's that share 3 nonadjacent complex side lengths (those of

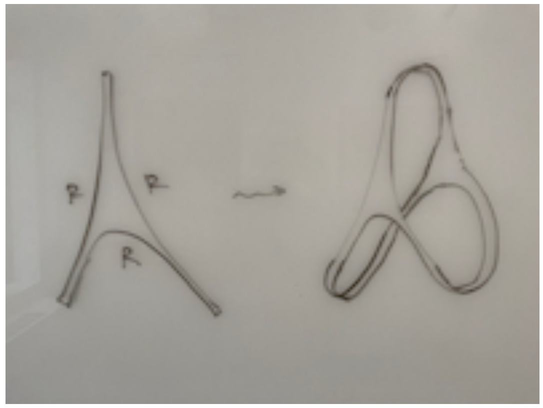



FIGURE 3. Pants where all three half lengths are large and roughly equal look as above, two 'fat vertices' joined by 'thin strips'.

the orthogeodesics), the half lengths are well defined independent of the ordering of the two feet, and $2hl(\gamma) \in \mathbb{C}/2\pi i\mathbb{Z}$ is the complex translation length of $\gamma$.

An *assembly of skew pants in* $M$ is a finite collection $\{P_i\}$ of skew pants $P_i \longrightarrow M$, together with a pairing of the boundary components of the pants, such that paired boundary components map to the same closed geodesic in $M$, with opposite orientations. Gluing paired boundary components gives a closed surface $S = S(\{P_i\})$, together with a natural map $S \longrightarrow M$.

Now fix $R, \epsilon > 0$. A skew pants $P \longrightarrow M$ is $(R,\epsilon)$-*good* if for each component $\gamma \subset \partial P$, we have $|hl(\gamma) - R| < \epsilon$. An assembly of skew pants $\{P_i\}$ is $(R,\epsilon)$-*good* if:

(1) for each $i$, the pants $P_i$ is $(R,\epsilon)$-good, and
(2) if $P_i, P_j$ have boundary components that are paired and map to $\gamma \subset M$, there are feet $v_i, v_j$ of $P_i, P_j$ on $\gamma$ such that

$$|d_\gamma(v_i, v_j) - (1 + \pi i)| < \epsilon/R.$$

Here, the first condition implies that for each $i$, the pants $P_i$ is the union of two skew RAH's with a triple of non-adjacent complex side lengths that are almost equal to $R$, so in particular are almost real, and hence $P_i$ can be taken to be a nearly totally geodesic pants in $M$ with geodesic boundary components of length near $R$. Note that if $R$ is large, then the pair of pants looks roughly like in Figure 3. The second condition says that if we use feet as a reference, adjacent pants are glued along $\gamma$ with nearly no bend, and with a twist of 1.

**Theorem 11.11** (Kahn-Markovic). *For small $\epsilon > 0$ and large $R > 0$, if $\{P_i\}$ is an $(R,\epsilon)$-good assembly of skew pants, with $S = S(\{P_i\})$ the corresponding surface, then the associated map $S \longrightarrow M$ is $\pi_1$-injective.*

The basic intuition is as follows.

**Lemma 11.12** (Local/global principle). *Given $\epsilon > 0$, there's some $l > 0$ such that no concatenation of length at least $l$ geodesic segments in $\mathbb{H}^n$, connected with bends of at most $\epsilon$, can be a closed path.*

Small bends means that the concatenation is locally close to being geodesic, whereas not being a closed path is a global notion, hence the name of the lemma.

Note that an analogous result is not true in $\mathbb{R}^2$: for any $l$, the bends at the vertices of a regular $n$-gon with side lengths $l$ are small if $n$ is large.

*Proof.* Suppose we have a concatenation of segments $\gamma_0, \gamma_1, \ldots, \gamma_n$. Let $P_i$ be the perpendicular plane to the midpoint of $\gamma_i$. If $l$ is large enough relative to $\epsilon$, each $P_i$ is disjoint from $P_{i+1}$. So, the path can't be closed, since $P_0$ separates the initial point of $\gamma_0$ from the terminal point of $\gamma_n$. □

Really, you can prove a stronger result: if you have a (say, piecewise smooth) path in $\mathbb{H}^n$ and you know that the total curvature of the path (say, obtained by integrating the geodesic curvature and then adding on the bends at any corners) is at most $\epsilon$ along every subpath of length $R$, then the path can't close up.

Now let $S \longrightarrow M$ be the surface in the theorem statement above. Let $\tilde{S}$ be the universal cover, so that the map $S \longrightarrow M$ lifts to a map $\tilde{S} \longrightarrow \mathbb{H}^3$. If $S$ isn't $\pi_1$-injective, then there's an arc in $\tilde{S}$ that maps to a closed loop in $\mathbb{H}^3$. Here, $\hat{S}$ is a union of universal covers of the individual pants, which look like trees of thin strips and fat vertices, and which map nearly totally geodesically to $M$. Let's pretend they map totally geodesically. Then the arc in $\tilde{S}$ maps piecewise geodesically to $M$, and in light of condition (2) in the definition of $(R, \epsilon)$-good, all bends are at most $\epsilon/R$. Since the image closes up, in light of the 'stronger version' of the lemma above, it has to accumulate a fair amount of total bend in a short time. The thing you might worry about, then, is that a large number of thin strips all line up one after another, so that the loop can accumulate a large amount of bend in a small amount of length. However, *because in condition (2) the adjacent pants are glued with a twist of around* 1, if you travel straight through a bunch of these strips, after at most $R$ steps you're guaranteed to be twisted into a fat vertex, therefore accumulating a reasonable amount of length with no bend.

So, how do we build $(R, \epsilon)$-good assemblies of pants in $M$? This is where dynamics comes in. Via similar arguments to the previous section, closed geodesics $\gamma$ in $M$ with complex length within $\epsilon$ of $2R$ are equidistributed in $M$. Moreover, if we fix such a $\gamma$, and let $\mathcal{P}_{R,\epsilon}(\gamma)$ be the set of all (unoriented) $(R, \epsilon)$-good skew pants $P$ for which $\gamma$ is a boundary curve, then the 2-element subsets

$$\mathrm{feet}(P, \gamma) \subset N(\gamma), \;\; P \in \mathcal{P}_{R,\epsilon}(\gamma)$$

are almost equidistributed in $N(\gamma)$ if $R$ is large, i.e. the Dirac measure on their union approximates Lebesgue measure on the torus $N(\gamma)$. Since Lebesgue measure is invariant under translation, if we're given a foot $v \in N(\gamma)$, there's roughly the same number of feet near the point $w \in N(\gamma)$ with $d_\gamma(v, w) = 1 + \pi i$ as there are near $v$. Setting $\mathcal{P}^{\pm}_{R,\epsilon}(\gamma)$ to be copies of $\mathcal{P}_{R,\epsilon}(\gamma)$ where the pants are oriented to that $\gamma$ inherits a plus or minus orientation, we can then use the Hall Marriage Theorem to pair up the pants $P \in \mathcal{P}^{-}_{R,\epsilon}(\gamma)$ with those in $P \in \mathcal{P}^{+}_{R,\epsilon}(\gamma)$ so that paired pants have feet $v, w$ such that $|d_\gamma(v, w) - (1 + \pi i)| < \epsilon/R$. Doing this for every $(R, \epsilon)$-good $\gamma$, we construct a $(R, \epsilon)$-good assembly, and hence a $\pi_1$-injective closed surface.

## References

1. I. Agol, D. Groves, and J. Manning, *The virtual Haken conjecture*, ArXiv e-prints (2012).
2. RBJT Allenby and Robert John Gregorac, *On locally extended residually finite groups*, Conference on Group Theory: University of Wisconsin-Parkside 1972, Springer, 2006, pp. 9–17.
3. Werner Ballmann, *Lectures on spaces of nonpositive curvature*, vol. 25, Birkhäuser, 2012.

4. M Bachir Bekka and Matthias Mayer, *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, vol. 269, Cambridge University Press, 2000.
5. Vladimir Igorevich Bogachev and Maria Aparecida Soares Ruas, *Measure theory*, vol. 2, Springer, 2007.
6. Rufus Bowen, *The equidistribution of closed geodesics*, American journal of mathematics **94** (1972), no. 2, 413–423.
7. Alex Eskin and Curt McMullen, *Mixing, counting, and equidistribution in lie groups*, (1993).
8. Harry Furstenberg, *Ergodic behavior of diagonal measures and a theorem of szemerédi on arithmetic progressions*, Journal d'Analyse Mathématique **31** (1977), no. 1, 204–256.
9. ———, *The unique ergodigity of the horocycle flow*, Recent Advances in Topological Dynamics: Proceedings of the Conference on Topological Dynamics, held at Yale University, June 19–23, 1972, in honor of Professor Gustav Arnold Hedlund on the occasion of his retirement, Springer, 2006, pp. 95–115.
10. Gustav A Hedlund, *The dynamics of geodesic flows*, (1939).
11. Heinz Huber, *Zur analytischen theorie hyperbolischer raumformen und bewegungsgruppen*, Mathematische Annalen **138** (1959), no. 1, 1–26.
12. MN Huxley, *Integer points, exponential sums and the riemann zeta function*, Surveys in Number Theory: Papers from the Millennial Conference on Number Theory, AK Peters/CRC Press, 2002, p. 109.
13. Jeremy Kahn and Vladimir Markovic, *Immersing almost geodesic surfaces in a closed hyperbolic three manifold*, Annals of Mathematics (2012), 1127–1190.
14. Anatole Katok, *Interval exchange transformations and some special flows are not mixing*, Israel Journal of Mathematics **35** (1980), no. 4, 301–310.
15. Anatole Katok, AB Katok, and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, no. 54, Cambridge university press, 1995.
16. Michael Keane, *The essence of the law of large numbers*, Algorithms, Fractals, and Dynamics (1995), 125–129.
17. Harvey B Keynes and Dan Newton, *A ?minimal?, non-uniquely ergodic interval exchange transformation*, Mathematische Zeitschrift **148** (1976), 101–105.
18. John M Lee, *Introduction to riemannian manifolds*, vol. 2, Springer, 2018.
19. Jialun Li and Wenyu Pan, *Exponential mixing of geodesic flows for geometrically finite hyperbolic manifolds with cusps*, Inventiones mathematicae **231** (2023), no. 3, 931–1021.
20. Elon Lindenstrauss, *Pointwise theorems for amenable groups*, Inventiones mathematicae **146** (2001), no. 2, 259–295.
21. AI Mal et al., *On the faithful representation of infinite groups by matrices*, Fifteen Papers on Algebra **45** (1965), 1–18.
22. Howard Masur, *Interval exchange transformations and measured foliations*, Annals of Mathematics **115** (1982), no. 1, 169–200.
23. Amos Nevo, *Harmonic analysis and pointwise ergodic theorems for noncommuting transformations*, Journal of the American Mathematical Society **7** (1994), no. 4, 875–902.
24. G. Perelman, *The entropy formula for the ricci flow and its geometric applications*, arXiv:math.GT/0405568.
25. ———, *Finite extinction time for the solutions to the ricci flow on certain three-manifolds*, arXiv:math.GT/0405568.
26. ———, *Ricci flow with surgery on three-manifolds*, arXiv:math.GT/0405568.
27. Mark Pollicott, *Exponential mixing for the geodesic flow on hyperbolic three-manifolds*, Journal of statistical physics **67** (1992), 667–673.
28. Burton Randol, *The behavior under projection of dilating sets in a covering space*, Transactions of the American Mathematical Society **285** (1984), no. 2, 855–859.
29. Marina Ratner, *The rate of mixing for geodesic and horocycle flows*, Ergodic theory and dynamical systems **7** (1987), no. 2, 267–288.
30. Peter Scott, *Subgroups of surface groups are almost geometric*, Journal of the London Mathematical Society **2** (1978), no. 3, 555–565.
31. Sashi Mohan Srivastava, *A course on Borel sets*, vol. 180, Springer Science & Business Media, 2008.
32. Terence Tao, *Failure of the pointwise and maximal ergodic theorems for the free group*, Forum of Mathematics, Sigma, vol. 3, Cambridge University Press, 2015, p. e27.

33. Arcady Aleksandrovich Tempel'man, *Ergodic theorems for general dynamical systems*, Trudy Moskovskogo Matematicheskogo Obshchestva **26** (1972), 95–132.
34. William Thurston, *The geometry and topology of 3-manifolds*, Lecture notes at Princeton University, 1980.
35. William P. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. (N.S.) **19** (1988), no. 2, 417–431. MR MR956596 (89k:57023)
36. William A Veech, *Interval exchange transformations*, Journal d?Analyse Mathématique **33** (1978), no. 1, 222–272.