Geodesic flow

D. Blättler, A. Furlong, A. Sandamirskaya

November 2023

1 Recap: The Algebra-Geometry dictionary for H

The geodesic flow $g^t : T_1\mathbb{H} \to T_1\mathbb{H}$ is the movement by a distance t along the oriented geodesic tangent to a given vector v: for $v \in T_z\mathbb{H}, \phi(t) := \dot{\gamma}(t) \in T_{\gamma(z)}\mathbb{H}$ for γ the unit speed geodesic with $\dot{\gamma}(0) = v$. Noticing the special case at point (i, i) of the tangent bundle, we deduce that $g^t = \begin{pmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{pmatrix}$

Proof. For $v = i$, $\phi(t)$ is the unit vector on $i\mathbb{R}_+$ facing upwards at distance t from i. $d(i, e^{it}) = \int_1^{e^t} \frac{1}{y} dy = t$. $\phi_t(v) = e^{tv}$. Furthermore $g^t(i, v) = (e^{ti}, e^{tv})$ so they correspond. Given $\xi \in \mathbb{H}$ arbitrary, there is $g^\xi \in PSL_2(\mathbb{R})$ such that $\xi = vg^\xi$. Since $g^{z\xi}$ is an isometry, $\phi_t(\xi) = \phi_t(vg^{z\xi}) = \phi_t(v)g^{z\xi} = g^t vg^{z\xi} = g^t\xi$ □

The horocycle flow $h^s : T_1\mathbb{H} \to T_1\mathbb{H}$ is the movement by a distance s along the horocycle perpendicular to a given vector v, with v pointing inward. Under the identification of $T_1\mathbb{H} \cong PSL_2(\mathbb{R})$, the horocycle flow merely corresponds to right translations:$g \to g \cdot u(t)$ where $u(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ $\forall t \in \mathbb{R}$.

Proof. It suffices to consider the case where $z = i$ and ξ is the unit tangent vector at z pointing in the direction of the imaginary axis. $h^s(z, \xi) = (i + s, \xi)$, so $h^s(z, \xi) = u(s) \cdot (z, \xi)$ □

Geometrically, the horocycle flow through a point (x, y) is the set of points (u, v) such that $d(g^t(x, y), g^t(u, v)) \to 0$
These do indeed preserve the measure on $T_1\mathbb{H}$.
Since these flows are geometric, they commute with the action of $\mathbb{H}$. And they are connected to each other through the relations $g^t h^s g^{-t} = h^{s\cdot\exp(-t)}$. Notice also that $h^s \circ h^t = h^{s+t}$ and $g^s \circ g^t = g^{s+t}$.

2 Introduction to Ergodic theory

Measure-preserving maps

Remember:
For measure spaces $(X, \mathscr{B}, \mu)$ and $(Y, \mathscr{C}, \nu)$ a map
$\phi : X \to Y$ is <u>measurable</u> if $\phi^{-1}(B) \in \mathscr{B}$ for all $A \in \mathscr{C}$.

Definition (measure-preserving):
We call ϕ a <u>measure-preserving map</u> if ϕ is measurable and $\mu(\phi^{-1}(B)) = \nu(B)$ for all $B \in \mathscr{C}$.

Examples:

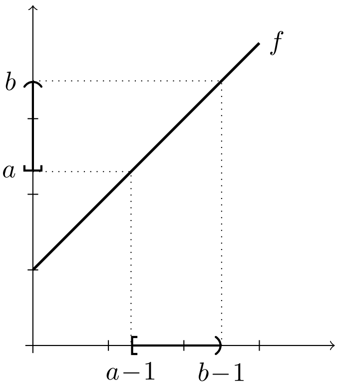
- The function $f : \mathbb{R} \to \mathbb{R}, x \mapsto x + 1$ preserves Lebesgue measure.

  Proof: It is sufficient to check the condition on intervals.
  So let $B = [a, b) \subseteq \mathbb{R}$. Then

  $m_{\mathbb{R}}(f^{-1}(B)) = m_{\mathbb{R}}(f^{-1}([a, b)))$

  $= m_{\mathbb{R}}([a - 1, b - 1)) = (b - 1) - (a - 1)$

  $= b - a = m_{\mathbb{R}}([a, b)) = m_{\mathbb{R}}(B) \quad \square$



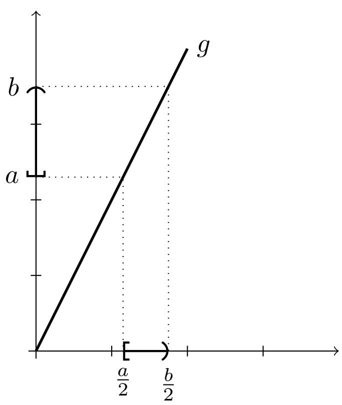
- The function $g : \mathbb{R} \to \mathbb{R}, x \mapsto 2x$ does <u>not</u> preserve Lebesgue measure.

  Proof: As a counterexample we choose
  $B = [0, 1] \subseteq \mathbb{R}$. Then

  $m_{\mathbb{R}}(g^{-1}(B)) = m_{\mathbb{R}}(g^{-1}([0, 1]))$

  $= m_{\mathbb{R}}([0, \frac{1}{2}]) = \frac{1}{2}$

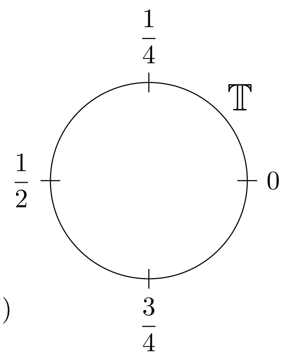
  $\neq 1 = m_{\mathbb{R}}([0, 1]) = m_{\mathbb{R}}(B) \quad \square$



- The circle-rotation map

  We observe the set $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with the metric $\mathrm{d}(r + \mathbb{Z}, s + \mathbb{Z}) = \min_{m \in \mathbb{Z}} |r - s + m|$.

  The function
  $R_\alpha : \mathbb{T} \to \mathbb{T}, x \mapsto x + \alpha \pmod 1$
  preserves Lebesgue measure.



  Proof: It is sufficient to check the condition on intervals.

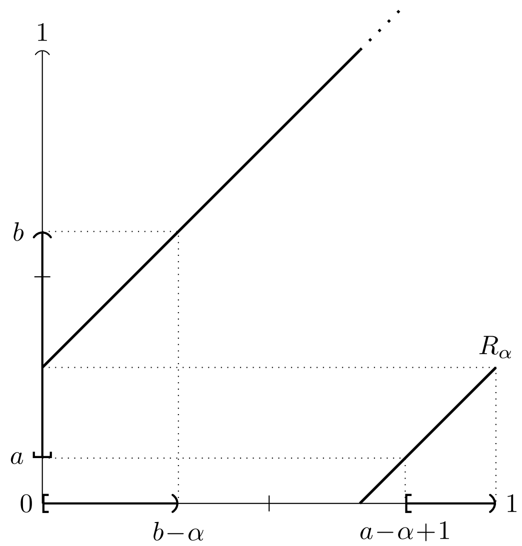
  So let $B = [a, b) \subseteq \mathbb{R}$. Then

  $m_{\mathbb{T}}(R_\alpha^{-1}(B)) = m_{\mathbb{T}}(R_\alpha^{-1}([a, b)))$

  $= m_{\mathbb{T}}([\min(a - \alpha, 0), b - \alpha) \cup [a - \alpha + 1, \max(b - \alpha + 1, 1)))$

  $= ((b - \alpha) - \min(a - \alpha, 0)) - (\max(b - \alpha + 1, 1) - (a - \alpha + 1))$

  $= b - a = m_{\mathbb{R}}([a, b)) = m_{\mathbb{R}}(B) \quad \square$



- The circle-doubling map

  The function $T_2 : \mathbb{T} \to \mathbb{T}, t \mapsto 2t \pmod 1$
  preserves Lebesgue measure.

  Proof: It is sufficient to check the condition on intervals.
  So let $B = [a, b) \subseteq [0, 1)$. Then

  $m_{\mathbb{T}}(T_2^{-1}(B)) = m_{\mathbb{T}}(T_2^{-1}([a, b]))$

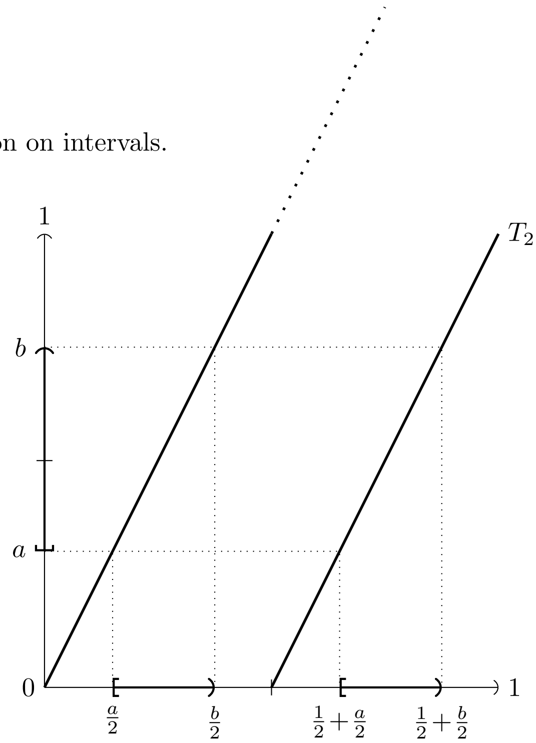
  $= m_{\mathbb{T}}([\frac{a}{2}, \frac{b}{2}) \cup [\frac{1}{2} + \frac{a}{2}, \frac{1}{2} + \frac{b}{2}))$

  $= m_{\mathbb{T}}([\frac{a}{2}, \frac{b}{2})) + m_{\mathbb{T}}([\frac{1}{2} + \frac{a}{2}, \frac{1}{2} + \frac{b}{2}))$

  $= (\frac{b}{2} - \frac{a}{2}) + ((\frac{1}{2} + \frac{b}{2}) - (\frac{1}{2} + \frac{a}{2}))$

  $= \frac{1}{2}(b - a) + \frac{1}{2}(b - a) = b - a$

  $= m_{\mathbb{T}}([a, b)) = m_{\mathbb{T}}(B) \quad \square$



- The left shift map on $\{0, 1\}^{\mathbb{N}}$

  We define the measure $\mu_{(\frac{1}{2},\frac{1}{2})}(0) = \mu_{(\frac{1}{2},\frac{1}{2})}(1) = \dfrac{1}{2}$ on the set $\{0, 1\}$.

  Now let $X = \{0, 1\}^{\mathbb{N}}$ with the infinite product measure $\mu = \prod_{\mathbb{N}} \mu_{(\frac{1}{2},\frac{1}{2})}$.

  This can be thought of as a model for outcomes of infinitely repeated tosses of a fair coin.

  Then the function $\sigma : X \to X, (x_0, x_1, ...) \mapsto (x_1, x_2, ...)$ preserves the measure μ.

Proof: We will show that (X, μ, σ) is measurably isomorphic to $(\mathbb{T}, m_{\mathbb{T}}, T_2)$.

We define the measure-preserving map $\phi : X \to \mathbb{T}$ as

$$\phi(x_0, x_1, ...) = \sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}}$$

Then we will show that for any $x = (x_0, x_1, ...) \in X : T_2(\phi(x)) = \phi(\sigma(x))$:

$$T_2(\phi(x)) = T_2(\phi(x_0, x_1, ...)) = T_2\left(\sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}}\right) = 2\sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}} \pmod 1$$

$$= \sum_{n=0}^{\infty} \frac{x_n}{2^n} \pmod 1 = \underbrace{x_0}_{\in\{0,1\}} + \sum_{n=1}^{\infty} \frac{x_n}{2^n} \pmod 1 = \sum_{n=0}^{\infty} \frac{x_{n+1}}{2^{n+1}} \pmod 1$$

$$= \phi(x_1, x_2, ...) = \phi(\sigma(x_0, x_1, ...)) = \phi(\sigma(x))$$

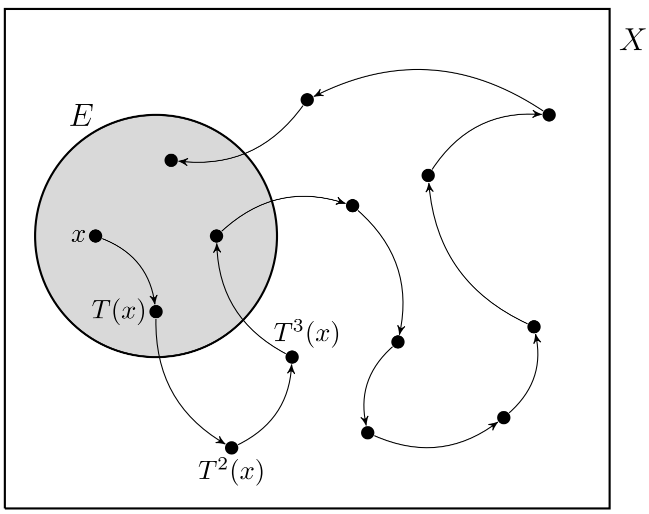
We have already shown that T_2 is measure-preserving on $\mathbb{T}$ with the measure $m_{\mathbb{T}}$, thus σ is measure-preserving on X with the measure μ. $\square$

Poincaré Recurrence

Theorem: Let $T : X \to X$ be measure-preserving on a probability space $(X, \mathscr{B}, \mu)$ and $E \subseteq X$ be measurable.

Then for almost every $x \in E$, x returns to E infinitely often. That means

$\exists F \subseteq E$ measurable with $\mu(F) = \mu(E)$ such that for any $x \in F$
there exist $0 < n_1 < n_2 < ...$ with $T^{n_i}(x) \in E$ for all $i \geq 1$.



Proof: Let $B = \{x \in E \mid T^n(x) \notin E \text{ for any } n \geq 1\}$ be the elements of E which never return to E. This set can be rewritten as

$$B = \left\{x \in E \wedge T(x) \notin E \wedge T^2(x) \notin E \wedge ...\right\}$$

$$= E \cup \left\{x \in X \mid T^2(x) \notin E\right\} \cup \left\{x \in X \mid T^2(x) \notin E\right\} \cup ...$$

$$= E \cup T^{-1}(X \setminus E) \cup T^{-2}(X \setminus E) \cup ...$$

Because E is measurable and $T^{-n}(X \setminus E)$ is measurable for any $n \geq 1$,
B is also measurable as a union of measurable sets.

For any $n \geq 1$ we can write

$$T^{-n}(B) = T^{-n}\left(E \cup T^{-1}(X \setminus E) \cup T^{-2}(X \setminus E) \cup ...\right)$$

$$= T^{-n}(E) \cup T^{-n-1}(X \setminus E) \cup T^{-n-2}(X \setminus E) \cup ...$$

Thus the sets $B, T^{-1}(B), T^{-2}(B), ...$ must be disjoint.

As a consequence, we get that

$$\infty > \mu(X) \geq \mu\left(B \cup T^{-1}(B) \cup T^{-2}(B) \cup ...\right)$$

$$= \mu(B) + \mu(T^{-1}(B)) + \mu(T^{-2}(B)) + ...$$

$$= \mu(B) + \mu(B) + \mu(B) + ... \quad \text{because } T \text{ is measure-preserving}$$

Hence we get $\mu(B) = 0$.

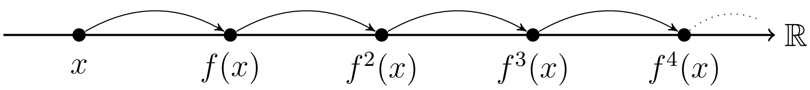
So there exists the set $F_1 = E \setminus B$ with $\mu(F_1) = \mu(E)$ for which every point in F_1 returns to E at least once.

Iteratively repeating the same construction with the maps $T^2, T^3, ...$ we get sets $F_2, F_3, ...$ where for any $n \geq 1$, we have $F_{n+1} \subseteq F_n$, $\mu(F_n) = \mu(E)$ and for which each element of F_n returns to E at least once under iteration of the function T^n.

We define the set $F = \bigcap_{n \geq 1} F_n \subseteq E$.

Then $\mu(F) = \mu(E)$ and every element of F returns to E infinitely often. So we have proven the theorem. $\square$

Note: As seen in the proof, it is necessary that $\mu(X) < \infty$.
The function $f : \mathbb{R} \to \mathbb{R}, x \mapsto x + 1$ seen in our first example of measure-preserving maps satisfies all other requirements of the theorem but clearly the theorem doesn't hold for it.



Ergodicity

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

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

Proposition: The following are equivalent properties for a measure-preserving transformation T of $(X, \mathscr{B}, \mu)$:

(1) T is ergodic.

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

(3) For $A \in \mathscr{B}$, $\mu(A) > 0$ implies that $\mu\left(\bigcup_{n=1}^{\infty} T^{-1}(A)\right) = 1$.

(4) For $A, B \in \mathscr{B}$, $\mu(A)\mu(B) > 0$ implies that there exists $n \geq 1$ with $\mu(T^{-n}(A) \cap B) > 0$.

(5) For $f : X \to \mathbb{C}$ measurable, $f \circ T = f$ almost everywhere implies that f is equal to a constant almost everywhere.

Proof: We will only prove the equivalence between (1), (2) and (5).

(1) $\Longrightarrow$ (2): Assume T is ergodic.

Choose $B \in \mathscr{B}$ such that $\mu\left(T^{-1} \triangle B\right) = 0$.

We will construct a T-invariant set with the same measure as B.

So let $C = \limsup_{n\to\infty} T^{-n}(B) = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}(B)$. Then for any $N \geq 0$,

$$\begin{aligned}
B \triangle \bigcup_{n=N}^{\infty} T^{-n}(B) &= \left(B \setminus \bigcup_{n=N}^{\infty} T^{-n}(B)\right) \cup \left(\left(\bigcup_{n=N}^{\infty} T^{-n}(B)\right) \setminus B\right) \\
&\subseteq \left(\bigcup_{n=N}^{\infty} B \setminus T^{-n}(B)\right) \cup \left(\bigcup_{n=N}^{\infty} T^{-n}(B) \setminus B\right) \\
&= \bigcup_{n=N}^{\infty} \left(B \setminus T^{-n}(B) \cup T^{-n}(B) \setminus B\right) = \bigcup_{n=N}^{\infty} B \triangle T^{-n}(B)
\end{aligned}$$

So for every $n \geq 1$,

$$\begin{aligned}
\mu\left(B \triangle T^{-n}(B)\right) &\leq \mu\left(\bigcup_{i=0}^{n-1} T^{-i}(B) \triangle T^{-(i+1)}(B)\right) \\
&= \mu\left(\bigcup_{i=0}^{n-1} T^{-i}\left(B \triangle T^{-1}(B)\right)\right) \\
&\leq \sum_{i=0}^{n-1} \mu\left(T^{-i}\left(B \triangle T^{-1}(B)\right)\right) \\
&= \sum_{i=0}^{n-1} \mu\left(B \triangle T^{-1}(B)\right) = n \cdot 0 = 0
\end{aligned}$$

Let $C_N = \bigcup_{n=N}^{\infty} T^{-n}(B)$

Then $C_0 \supseteq C_1 \supseteq ...$ and as shown $\mu(C_N \triangle B) = 0$ for every $N \geq 0$.
$\mu(C \triangle B) = 0$ which implies that $\mu(C) = \mu(B)$.

$$\begin{aligned}
\text{Moreover, } T^{-1}(C) = T^{-1}\left(\bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}(B)\right) &= \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-(n+1)}(B) \\
&= \bigcap_{N=0}^{\infty} \bigcup_{n=N+1}^{\infty} T^{-n}(B) = 0
\end{aligned}$$

Thus by ergodicity, we know that $\mu(C) = 0$ or 1 and thus $\mu(B) = 0$ or 1. $\square$

(2) $\implies$ (3):

Let $f : X \to \mathbb{R}$ with $f \circ T = f$ almost everywhere. We define

$A_n^k = \left[\frac{k}{n}, \frac{k+1}{n}\right]$ and $B_n^k = f^{-1}(A_n^k) = \left\{x \in X : f(x) \in A_n^k\right\}$

Then we see that

$$T^{-1}(B_m^k) = T^{-1} \circ f^{-1}(A_n^k) = (f \circ T)^{-1}(A_n^k)$$

which is equal to $f(A_n^k) = B_n^k$ almost everywhere as $f \circ T = f$ almost everywhere.

Thus we can use (2) to get that $\mu(B_n^k) = 0$ or 1. Because $(X, \mathscr{B}, \mu)$ is a probability space and all sets B_n^k are disjoint for a fixed n, we get that

$$1 = \mu(X) = \mu\left(\bigcup_{k \in \mathbb{Z}} B_n^k\right) = \sum_{k \in \mathbb{Z}} \mu(B_n^k)$$

This is only the case if for each n there exists exactly one k with $\mu(B_n^k) = 1$. Thus we get the sequence $(k_n)_n \in \mathbb{N}$ with $\mu(B_n^{k_n} = 1$ for every n.

Because $\lim_{n\to\infty} \mu(A_n^k) = 0$ there is only one number $c \in \limsup_{n\to\infty} A_n^{k_n}$ and thus

$\mu(f^{-1}(c)) = 1 \implies f(x) = c$ almost everywhere. $\square$

(3) $\implies$ (1):

Let $B \in \mathscr{B}$ such that $T^{-1}(B) = B$

The function $\chi_B : X \to \{0, 1\}$ is T-invariant because B itself is T-invariant.

Thus we can use (3) and get that χ_B =constant almost everywhere. As χ_B maps onto the set $\{0, 1\}$, said constant is either 0 or 1. With this we can calculate the measure of B:

$$\mu(B) = \int_X \chi_B \mathrm{d}\mu = \begin{cases} \int_X 0 \mathrm{d}\mu = 0 \text{ or} \\ \\ \int_X 1 \mathrm{d}\mu = \mu(X) = 1 \qquad \square \end{cases}$$

Associated Unitary Operators and Unitary representations

In general, consider a Hilbert space $\mathcal{H}$ and a continuous linear operator $U : \mathcal{H} \to \mathcal{H}$. Then the relation $\langle Uf, g\rangle = \langle f, U^*g\rangle$ defines another continuous linear operator $U^* : \mathcal{H} \to \mathcal{H}$ called the **adjoint of** U. Note that U is an isometry iff $U^*U = id_{\mathcal{H}}$ and $UU^* = Proj_{Im(U}$.
We call U **unitary** iff it is invertible and $U^{-1} = U^*$, this is equivalent to $\forall f, g \in \mathcal{H}$: $\langle Uf, Ug\rangle = \langle f, g\rangle$.

Let T be a measure preserving map and recall that L^2_μ is a Hilbert space. Define the **associated operator** or **Koopman operator** of T by:

$$U_T : L^2_\mu \to L^2_\mu, \quad f \mapsto f \circ T$$

It turns out that U_T is actually a unitary operator, as the following quick calculations shows: $\langle U_T f, U_T g\rangle = \int f \circ T \cdot \overline{g \circ T} d\mu = \int f \cdot \overline{g} d\mu = \langle f, g\rangle$
Where we used that μ is T-invariant, so performing the substitution $x \mapsto T(x)$ cancels out the change to the push-out measure T_μ. Now we can see some equivalent statements to ergodicity.

Proposition: T is ergodic $\Leftrightarrow$ 1 is an eigenvalue of U_T with multiplicity one. Thus we say that ergodicity is a unitary property

Proof. Recall that: T ergodic $\Leftrightarrow \forall f : X \to \mathbb{C} \quad \mu$-meas. $U_T f = f \circ T = f$ a.e. $\Rightarrow$ f is constant a.e.
Note that $f \equiv 1$ a.e. is an eigenfuction with eigenvalue 1 of U_T. Let T be ergodic. If we have another eigenfunction g for the eigenvalue 1, then since T is ergodic g must be constant a.e. and is in particular a multiple of f a.e.
Let 1 be an eigenvalue with multiplicity one of U_T. Again, all other functions which satisfy $U_T g = g \circ T = g$ must be a multiple of $f \equiv 1$ a.e. and are hence constant a.e., so T is ergodic. □

The next equivalent formulation requires the theory of representations. We will start with a few definitions.

A **representation of a group G onto a vector space V** is a group homomorphism $\rho : G \to GL(V)$. Let $\langle ., .\rangle$ be a scalar product on V. We say that ρ is **unitary** if $\forall g \in G : \quad \rho(g)$ is unitary.
In our case G is a metric group, and V=$\mathcal{H}$ is a Hilbert space, and we impose the additional requirement that for a fixed $v \in \mathcal{H}$ the map $G \to \mathcal{H}, \quad g \mapsto g(v)$ must be continuous with respect to the metric on G, and the induced metric by the scalar product on $\mathcal{H}$.
We define a **character** of a representation ρ to be the map $\chi_\rho : G \to \mathbb{C}, \quad g \mapsto \text{tr}(\rho(g))$.

Example: An example of a representation is the **regular representation**. Let G be a group and V $= \{f : G \to \mathbb{C}\}$ the vector space of all linear functionals, also denoted by $\hat{V}$. Note the characters are elements of V. The regular representation is given by:

$$\rho_{reg}(g)f : h \mapsto f(g^{-1}h)$$

It is easy to check that this is indeed a representation.
We can also consider $X = \Gamma\backslash G$ and μ the Haar-measure. Then the action of G onto

$L^2_\mu(X)$ given by $f \mapsto f \circ g^{-1}$ is a unitary representation.

Example: A even simpler example of a unitary representation is matrix multiplication. Let G = U(n, $\mathbb{C}$) be the group of all $n \times n$ unitary matrices, and let $\mathcal{H} = \mathbb{C}^n$. Then the group action of matrix multiplication is a unitary representation.

Theorem: Let X be a compact abelian group, and $T : X \to X$ a continuous surjective homomorphism. Then T is ergodic with respect to the Haar measure m_X iff for a $n > 0$ and a character $\chi : X \to \mathbb{C}$ the equality $\chi \circ T^n = \chi$ implies that χ is the trivial character, i.e. $\chi \equiv 1$.

Proof. Assume there exists a non-trivial character χ with $\chi(T^n(x)) = \chi(x)\, \forall x \in X$ for some $n > 0$, chosen to be minimal with this property. Define the function:

$$f(x) = \chi(x) + \chi(T(x)) + \cdots + \chi(T^{n-1}(x)$$

Observe that it is invariant under T and non-constant, since it is a sum of distinct, non-trivial characters. Thus, T is not ergodic.

Conversely, assume that only the trivial character is invariant under a non-zero power of T, and let $f \in L^2_\mu(X)$ be a T-invariant function. We want to show that f is constant. Consider the Fourier expansion $f = \sum_{\chi \text{character}} c_\chi \chi$, where $||f||_2^2 = \sum_{\chi \text{ character}} |c_\chi|^2 < \infty$.
Since f is invariant under T, it follows that $c_\chi = c_{\chi \circ T^n} \quad \forall n \in \mathbb{N}$. Hence either $c_\chi = 0$ or there are only finitely many distinct characters among the $\chi \circ T^i$ otherwise $\sum_{\chi \text{character}} |c_\chi|^2$ would be infinite. This means there exist integers p¿q such that $\chi \circ T^p = \chi \circ T^q$. Since the map $\chi \mapsto \chi \circ T$ is injective (due to T being sujrective), it follows that χ is invariant under T^{p-q}. By hypothesis, χ is trivial. This means that the Fourier expansion of f is a constant, so f is a constant a.e.. Hence T is ergodic. □

In particular, we can apply this theorem together with the previous Proposition to the torus:
Corollary: Let $A \in GL_d(Z)$ be an invertible integer matrix. A induces a surjective homomorphism $\mathbb{R}^d/\mathbb{Z}^d \to \mathbb{R}^d/\mathbb{Z}^d$ given by matrix multiplication, which preserves the Lebesgue measure. The transformation T_A is ergodic iff no eigenvalue of A is a root of unity.

Actions on the space

Let Γ be a discrete subgroup of $SL_2(\mathbb{R})$. We define the geodesic and horocycle flow on the space $X = \Gamma \backslash SL_2(\mathbb{R})$ as follows:
geodesic flow: $R_{a_t}(x) = x \begin{bmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{bmatrix}$ horocycle flow: $R_{u_t}(x) = x \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$

This motivates us to consider the sets $A = \left\{ \begin{bmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{bmatrix} |t \in \mathbb{R} \right\}$ and $U = \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} |s \in \mathbb{R} \right\}$ more closely.

Proposition: Every $g \in SL_2(\mathbb{R})$ is conjugate to an element of $\pm A$, $\pm U$ or $SO_2\mathbb{R}$. This is known as Iwasawa decomposition, and you can think of it as just Cartesian coordinates over $\mathbb{H}$.

Proof. We consider three cases:

- If g is diagonalisable over $\mathbb{R}$. Then we can diagonalise it and get that g is conjugate to a matrix of the form $\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$ where $d_2 = \frac{1}{d_1} \in \mathbb{R}$ since the determinant of g is 1. Define t such that $|d_1| = \mathrm{e}^{\frac{-t}{2}}$. We get that g is conjugate to an element of A or -A and call such a g **hyperbolic**.
- If g if diagonalisable over $\mathbb{C}$. We again diagonalise it and get the eigenvalues λ and λ^{-1}, their absolute value is 1, hence $\lambda = \mathrm{e}^{i\theta}$ for some $\theta \in \mathbb{R}$. This corresponds to a rotatory matrix which all lie in $SO_2\mathbb{R}$. We call g **elliptic**.
- If g is not diagonalisable, then it only has one eigenvalue and its Jordan Normal Form is $\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$ which is an element of U. We call this kind of g **parabolic**.

The considered all possible cases and this concludes the proof. □

This Proposition is useful because of the following result:

Proposition: Let Γ be a discrete subgroup of a closed linear group G. Let $g_1, g_2, h \in G$ such that $g_2 = hg_1h^{-1}$. Then $R_{g_2} = R_h R_{g_1} R_h^{-1}$ holds as maps on $\Gamma\backslash G$. In particular, if Γ is a lattice, then the measure preserving systems $(X, \mathcal{B}_X, m_X, R_{g_1})$ and $(X, \mathcal{B}_X, m_X, R_{g_2})$ are conjugate, i.e. measurably isomorphic.

Proof. The first part immediately follows from the definition of R_g. If Γ is a lattice, then the map R_h preserves the finite measure m_X which proves the second part of the statement. □

In the next section we want to prove the Ergodic theorem which states the following:

Theorem: (Ergodic Theorem) Let Γ be a subgroup of $SL_2(\mathbb{R})$ and a lattice. Define $X = \Gamma\backslash SL_2(\mathbb{R})$. Let $g \in SL_2(\mathbb{R})$ be an element not conjugate to $SO_2(\mathbb{R})$. Then R_g acts ergodically on $(X, \mathcal{B}_X, m_X)$.

3 Putting things together

There are a couple ways to show that the geodesic flow is ergodic. Behind each proof lies the basic idea of contraction.

Lemma: $PSL_2(\mathbb{R})$ is generated by the matrices in $\langle N^+, A, N^- \rangle$. Recall that A is just the set of matrices representing the geodesic flow. It turns out this decomposition of the space is just Cartesian coordinates (N corresponds to the x-axis, and A the y-axis). We want to show the geodesic flow is ergodic for every quotient of $\mathbb{H}$ by a lattice H/Γ, because we need a finite measure to talk about ergodicity. Note: generally this space is not compact.
It turns out that being ergodic for a topological group is quite a weak property, since we demand invariance under the whole group: X is G-ergodic if $\mu(\{x : gx = x, \forall g \in G\}) \in \{0, 1\}$. Our equivalence shown above still applies, so it is equivalent to proving that any A-invariant function f is constant almost everywhere. It will turn out that A-invariant $\implies PSL_2(\mathbb{R})$-invariant, and so if f is in L^2, then it must be constant, therefore ergodic. We need the following contraction lemma:

Lemma: For $g_a \in A$ and $h \in N^+$ if $a < 1$, in N^- otherwise, we have that $\lim_{n\to\infty} g_a^n h g_a^{-n} = e$. Essentially, when we conjugate by g_a, we **contract** h.

Proof. We prove this by just computing the relevant matrices.

$$\begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & a^2x \\ 0 & 1 \end{bmatrix}$$

So,

$$\begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix}^n \cdot \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix}^{-n} = \begin{bmatrix} 1 & a^{2n}x \\ 0 & 1 \end{bmatrix}$$

for $a < 1$ so we see this does indeed tend to 0. If instead $a > 1$, then by taking the transpose of our h, so it is now in N^-, we get

$$\begin{bmatrix} 1 & 0 \\ a^{-2n}x & 1 \end{bmatrix}$$

So we have proved the lemma. Geometrically this is just saying that we can always find some element g_a that will contract h along the horocycle. □

We can then use Mautner's lemma, which translates this result in terms of representations.

Mautner's lemma: Consider a unitary representation on $PSL_2(\mathbb{R})/\Gamma$, and suppose we have $g, h \in SL_2(\mathbb{R})$ as above, i.e. $\lim_{n\to\infty} g^n h g^- n = 1$. Then all $f \in L^2(PSL_2(\mathbb{R})/\Gamma)$ invariant under the action of g are also invariant under the action of h.

Proof. We use the associated operators given by the representation: $||T_h f - f|| = ||T_h T_{g^{-n}} f - T_{g^{-n}} f|| = ||T_{g^n} T_h T_{g^{-n}} f - T_{g^n} T_{g^{-n}} f||$ and so

$$\lim_{n\to\infty} ||T_{g^n} T_h T_{g^{-n}} f - f|| \leq || \lim_{n\to\infty} T_{g^n} T_h T_{g^{-n}} f - f|| = 0$$

. Therefore f is invariant under h. □

We now have a 2-line proof of ergodicity:
Suppose $T_a f = f$ for all $a \in A$. Then by the first lemma we have for all $h \in N^+ \cup N^-$ we have a g_a that contracts h; this allows us to apply Mautner, and we have that for all f invariant under g is invariant under h. But since $PSL_2(\mathbb{R})$ is generated by these matrices, we must have that f is invariant under any matrix $M \in PSL_2(\mathbb{R})$: $T_M f = f$. So $f \in L^2(PSL_2(\mathbb{R})/\Gamma)$ invariant under $PSL_2(\mathbb{R})$, it must be essentially constant, which is equivalent to saying A is ergodic.

This same idea, with a small twist can be used to prove the ergodicity of the horocycle flow (N^+).

Proof 2: Hopf's original idea

This was the first proof of ergodicity of the geodesic flow. It uses the Pointwise Ergodic Theorem, which just says that time averages = space averages, and which is discussed in the appendix. Hopf's proof is all about making the following idea rigorous:
Given any $x \in \mathbb{H}/\Gamma, u \in N$ we want to show that $f(x) = f(xu)$. By A-invariance, $f(xu) = f(xua_n) = f(xa_n a_n^{-1} u a_n)$. By the contraction principle, the RHS tends to $f(xa_n) = f(x)$ as required. The problem with this is the convergence: f is only assumed to be measurable, not continuous, and even continuity isn't enough. Remember from Analysis I that the image of a Cauchy sequence is Cauchy if f is uniformly continuous. We have no such restrictions. So we will restrict to compact sets of arbitrarily large size and use measure theory magic, also known as Luzin's theorem. Luzin's theorem states that given $f : X \to Y$ with $\mu(X) = 1$, for all $\epsilon > 0$ there exists K compact with $\mu(K) = 1 - \epsilon$ and f restricted to K is continuous, and so uniformly continuous.

Proof. Let $f : X \to \mathbb{R}$ be a measurable g^t-invariant function, for non-zero t. By Lusin's theorem, for any $\epsilon > 0$ there exists a compact set K such that $m(K) > 1 - \epsilon$ and $f|_K$ is continuous. Define $B := \{x : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_K(g_t^l x) > \frac{1}{2}\}$. Intuitively this set should be quite large if everything is ergodic, so we want to show it has large measure. The limit exists almost everywhere and belongs in $[0,1]$ and $\int h := \int \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_K(g_t^l x) = m(K) \geq 1 - \epsilon$ by the pointwise Ergodic Theorem. So

$$1 - \epsilon \leq \int_B h + \int_{X/B} h \leq m(K) + \frac{1}{2} m(X/K) = \frac{1}{2} m(B) + \frac{1}{2}$$

This implies $m(B) \geq 1 - 2\epsilon$.

We now look to apply the contraction. Suppose $x \in B, y = h^s \cdot x \in B$ for some s. Then $f(x) = f(g_t^l x), f(y) = f(g_t^l y)$ for all $l \geq 1$ (by g_t-invariance), and so $d_X(f(g_t^l x), f(g_t^l y)) = d_X(xa_t^{-l}, x \cdot h^s a_t^{-l}) \leq d_{PSL_2\mathbb{R}}(I_2, a_t^l h^s a_t^{-l}) \to_{l\to 0} 0$. Since x, y spend more than half their time in the set K (by definition of Y), there exists a common sequence l_n such that the orbits of x, y become arbitrarily close, and in K. By compactness, $f|_K$ is uniformly continuous, so $f(g_t^{l_n} x)$ and $f(g_t^{l_n} y)$ converge along a subsequence, which gives us that $f(x) = f(h^s x)$ whenever both arguments are in B. We now build a sequence of bigger sets; for $\epsilon_1 < \epsilon$, there exists compact $K_1 \subset X$ such that $m(K_1) > 1 - \epsilon_1$ on which f is continuous. Wlog $K \subset K_1$. We can define B_1 as before, with $B \subset B_1$. ϵ arbitrary $\implies \exists A, m(A) = 1, \forall x, y = h^s x \in A$ we have

$f(x) = f(y)$.

So if f not constant a.e., we have two disjoint intervals I_1, I_2 such that $C_i = \{h \in PSL_2(\mathbb{R}) | f(\Gamma h) \in I_i\}$ has measure in $(0,1) \implies \exists g : m(C_1 \cap C_2 g) > 0 \implies C_1 \cap C_2 \cap \{h : \Gamma h \in X_g\} \neq \emptyset$. Contradiction since $I_1 \ni f(\Gamma h) = f(\Gamma h g^{-1}) \in I_2$. □

Right at the end we used the following small lemma: if for two Borel sets B_1, B_2: $\mu(B_1)\mu(B_2) > 0$, then $\{g : \mu(gB_1 \cap B_2) > 0\}$ is open and non-empty. A full proof can be found in 8.3 of Einsiedler and Ward's book.

Proof using the full power of representations, Chapter 11

Lemma: Let X be a locally compact metric space, and let µ be a probability measure on X. Assume that G is a metrizable group that acts continuously on X (see p. 229) and preserves the measure μ. Then the action of G on $L^2_\mu(X)$ defined by $g : f \to f \circ g^{-1}$ is a unitary representation.

Proof. Since g preserves μ we know this is unitary. The continuity requirement was proved above.

□

Proposition 11.18: Let H be a Hilbert space carrying a unitary representation of a metric group G. Suppose that $v_0 \in H$ is fixed by some subgroup $L \subset G$. Then v_0 is also fixed by any other element $h \in G$.
Using representations gives a less geometric proof than the original. If $f \in L^2$ is R_g-invariant, then by Proposition 11.18 it is also $R_{U_g^-}$ and R_{U^+}-invariant. In the case $g = a_t \in G = SL2(R)$, the subgroups U generate all of $SL_2(\mathbb{R})$, so the function f is $SL_2(\mathbb{R})$-invariant and therefore equal to a constant almost everywhere. Note that here the invariance is always understood in L^2.

Corollary 11.19. Let $\Gamma \leq G$ be a lattice in a closed linear group and let X be the homogeneous space $\Gamma \backslash G$. If $g \in G$ has the property that G is generated by $U^+ \cup U^-$ then R_g is ergodic with respect to the Haar measure.

Proof. Prop 18: Without loss of generality we may assume that $||v_0|| = 1$. We define the auxiliary function (also called a matrix coefficient, in analogy with the Gram matrix) $p(h) = \langle h(v_0), v_0 \rangle$ for $h \in G$. Notice that by the continuity requirement in the definition of a unitary representation, $p(h)$ depends continuously on h. Moreover, for $g_1, g_2 \in L$ and $h \in G$,

$$p(g_1 h g_2) = \langle g_1 h(g_2(v_0)), v_0 \rangle = \langle h(v_0), g_1^{-1} v_0 \rangle = p(h)$$

(since v_0 is fixed by g_1, g_2 in L). Now $h \in G$ acts unitarily, so $||h(v_0)|| = 1$. We claim that $p(h) = 1$ implies $h(v_0) = v_0$. This may be seen as a consequence of the fact that equality in the Cauchy–Schwartz inequality $|\langle v, w \rangle| \leq ||v|| \cdot ||w||$ only occurs if v and w are linearly dependent. Now let $h \in G$ be as in the statement of the proposition, and choose sequences $g_n \to e$ (the identity in G), $l_n, l'_n \in L$ with $l_n g_n l'_n \to h$. Then, on the one hand, by the equation (inner product up above) we have $p(l_n g_n l'_n) = p(g_n) \to p(e) = ||v_0||^2 = 1$, while $p(l_n g_n l'_n) \to p(h)$ by continuity. It follows that $p(h) = 1$, and so $h(v_0) = v_0$ by the claim above □

4 Appendix: Pointwise Ergodic Theorem

Following on from Poincarre recurrence, we have that the space of L^2 functions can be decomposed orhtogonally into the space of functions that are T-invariant ($f = f \circ T$) which we will call H_{inv} and the closure of the space of functions that are coboundaries $H_{erg} = \overline{\{f : \exists g, f = g - g \circ T\}}$. One can easily show, just using the properties of the inner product, that $L^2 = H_{inv} \oplus H_{erg}$. This gives us the mean ergodic theorem:

This states that taking the limit of averages converges in L_2-norm to the projection onto the invariant subspace: $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} U_T^n f = f_{inv}$. The proof is straightforward. We use the decomposition up above, the invariant bit is invariant which is what we want, so we only need to show that f_{erg} disappears. Since we are doing all this in norm, we can ignore the fact that we took the closure and just assume $f_{erg} = g - g \circ T$. This gives a telescoping sum, which tends to 0.

If we now want to prove this in the pointwise case, it becomes a lot more involved, because we now cannot ignore the closure. So we have to use a technical lemma called the maximal inequality. A full proof can be found in Section 2 of Einsiedler's book. Supposing we now have the theorem up above for pointwise convergence, you may be wondering where the integral comes from. Well it turns out that $f_{inv} = \int f$ when the system is ergodic, and that comes from the fact that invariant functions are constant almost everywhere. This is all we need.