Geodesic Flow on Hyperbolic Surfaces

Mitul Islam

January 17, 2019

Abstract

The target of this note is to prove the ergodicity of geodesic flow on closed hyperbolic surfaces. The main focus is on the celebrated Hopf argument. ***This is not an original work.*** It is a careful write up of the general ideas, specialized to the case of hyperbolic surfaces. This makes the exposition more easily accessible.

1 Ergodic Theory

It is a study of the long term behaviour of dynamical systems.

Motivating example - Rotation on a circle. Equip the circle with the usual measure - the nomralized Haar measure- $\mu = \dfrac{d\theta}{2\pi}$. Consider rotation map $R_\alpha : S^1 \to S^1$ by $R_\alpha(z) = e^{2\pi i\alpha}z$. Two cases arise:

1. α is rational. Then, the map is periodic. Take example, $\alpha = \pi/3$. Then, take set $A = [0, \pi/6] \cup [\pi/3, \pi/3 + \pi/6] \cup [2\pi/3, 2\pi/3 + \pi/6]$. Then, observe that $R_\alpha(A) = A$ and $0 < \mu(A) < 1$.

2. α is irrational. Following the above example, want to study A such that $R_\alpha(A) = A$. Here, studying behaviour of the set A geometrically gets complicated as points have dense orbits in S^1. However, we can look at ξ_A. Take Fourier expansion and use $\chi_A \circ R_\alpha = \chi_A$ given, χ_A =const. Thus, $\mu(A) = 0$ or 1.

The second kind - irrational rotations - are a motivating example for ergodic transformations.

For the following definitions (X, μ) will always denote a probability space.
Definition 1. *(Measure preserving map) Let (X, μ) be a probability space and $T : X \to X$ a measurable map. We call T measure preserving if $\mu \circ T^{-1} = \mu$.*

Instead of having a map, we could have a measure preserving flow on a probability space.
Definition 2. *(Measure preserving flow) A measure preserving flow $\{\phi_t\}_{t\in\mathbb{R}}$ on a probability space (X, μ) is a measurable mapping $\phi : \mathbb{R} \times X \to X$ satisfying :*

1. *$\phi(t + s, x) = \phi(t, (\phi(s, x))$ and $\phi(0, x) = x$.*

2. *Each measurable map* $\phi_t \equiv \phi(t,.)$ *is measure preserving*

Measure preserving maps have nice recurrence properties.
Theorem 1. *(Poincare's Recurrence) Let* $E \subset X$ *have positive measure. Then, for almost every point* x *of* E, $T^n(x) \in E$ *infinitely often.*
Definition 3. *(Ergodic maps) A measure preserving map* T *is called ergodic if any* $T-$*invariant measurable set* A*, that is,* $(T^{-1}(A) = A)$ *has* $\mu(A) = 0$ *or* $\mu(A) = 1$.
Definition 4. *(Ergodic flows) A measure preserving flow is called ergodic if any measurable set* A *satisfying* $\phi_t(A) = A$ *for all* $t \in \mathbb{R}$ *must have* $\mu(A) = 0$ *or* $\mu(A) = 1$.
Proposition 1. T *is ergodic if and only if for any* $f \in L^2(X,\mu)$, $f \circ T = f$ *(for flows,* $f \circ \phi_t = f$ *for all* t*)* $\implies$ f *is constant almost everywhere.*

The recurrence property is much nicer here. If T is ergodic, then almost every point in X returns to E and the average number of returns to E is asymptotically given by $\mu(E)$.

This follows from the following theorem:
Theorem 2. *(Birkhoff Ergodic Theorem) Let* $f \in L^1(X,\mu)$*. If* T *is measure preserving, then for* μ *almost every* x *in* X,

$$f^+(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} f(T^i x)$$

exists and

- $f^+(x) \in L^1(X,\mu)$
- $f^+(x)$ *is* T *invariant*
- $\int f^+ d\mu = \int f d\mu$

f^+ *is called the forward (ergodic) average.*

In addition, if T *is ergodic,* f^+ *is constant almost everywhere and equal to* $\int f d\mu$.

In addition, if we assume T is invertible and the inverse is measure preserving, then we can similarly define the backward ergodic average $f^-(x)$ which satisfies analogous properties.

There is a version of Birkhoff's Ergodic Theorem that holds true for flows, the only difference being the definition of forward ergodic averages :

$$f^+(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f \circ \phi_s ds.$$

2 Geodesic Flow on Hyperbolic Surfaces

2.1 Geodesic Flow on Hyperbolic Surfaces

Flow ϕ_t defined on SM via : For any $(p,v) \in SM$, $\phi_t(p,v) = (\gamma(t), \dot{\gamma}(t))$. Assumption : M is complete so that ϕ_t is defined for all $t \in \mathbb{R}$.

Recall that $S\mathbb{H}^2 = PSL(2,\mathbb{R})$. Any finite volume hyperbolic surface M is given by $\Gamma\backslash\mathbb{H}^2$, where $\Gamma = \pi_1(M)$, is a lattice $PSL(2,\mathbb{R})$. Look at $SM = T^1(\Gamma\backslash\mathbb{H}^2)$. Observe that $PSL(2,\mathbb{R})$ acts transitively on SM by $g.(\bar{p},\bar{v}) = (\bar{g(p)}, dg(\bar{v}))$. It can be verified that if $\gamma \in \Gamma$, $d\gamma(\bar{v}) = \bar{v}$ and thus, by Orbit-Stabilizer theorem, $\Gamma\backslash PSL(2,\mathbb{R}) \equiv SM$.

We can now define the geodesic flow on SM. Geodesic flow on $S\mathbb{H}^2$ does not have interesting dynamical properties since it runs off to infinity. But in a finite volume hyperbolic surface, the unit tangent bundle is a finite measure space and the geodesic flow has rich dynamics.

Geodesic flow in $PSL(2,\mathbb{R})$ is defined by right multiplication by the "diagonal" subgroup :

$$A = \{g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} : t \in \mathbb{R}\},$$

that is,

$$\phi_t(g) = gg_t.$$

Reason :

Geodesic flow on $\Gamma\backslash PSL(2,\mathbb{R})$ is given by

$$\phi_t(\Gamma g) = \Gamma g g_t.$$

Reason : Geodesics in M lift to geodesics in $\mathbb{H}^2$. So, the geodesic flow in the quotient of $\mathbb{H}^2$ is a quotient of the geodesic flow in $\mathbb{H}^2$.

Theorem 3. *(Key Theorem) For a compact hyperbolic surface M, the geodesic flow on SM is ergodic.*

Remarks

1. Geodesic flow is ergodic for any compact surface of variable negative curvature. (Proved by E. Hopf, Ref:Ballmann)
2. Ergodicity holds for any compact hyperbolic manifold (not just surfaces). (Ref :Ballmann)
3. Ergodicity result is true for any manifold of negative sectional curvature with finite volume, bounded and bounded away from zero curvature and bounded first derivatives of the curvature. (Ref : Ballmann)
4. The proof goes through for any manifold whose geodesic flow is Anosov. But the issue is that for Anosov flows, the stable and unstable foliations are only Holder continuous. In the following, we will see that life is easier in compact hyperbolic setting as foliations are smooth. This implies smoothness of holonomy map which in turn leads to the local product structure of measure on SM. For Anosov flows (with Holder continuous foliations), Anosov proved absolute continuity to show that the local product structure exists. This was one of his key ideas in the proof of ergodicity of Anosov flows.

2.2 Measure on SM

A measure could be put on SM in several ways, all of them producing the same one up to rescaling.

The cleanest among them is putting the Haar measure on $PSL(2,\mathbb{R})$ and looking at the measure on the quotient $\Gamma\backslash PSL(2,\mathbb{R})$. As $PSL(2,\mathbb{R})$ is semi-simple, this is in fact a bi-invariant measure on SM.

Alternatively, pick up a left invariant top-degree form on $PSL(2,\mathbb{R})$. This will produce a measure on $PSL(2,\mathbb{R})$ and hence descend to $SM = \Gamma\backslash PSL(2,\mathbb{R})$. But the volume form being left-invariant, this measure is also left -invariant. By uniqueness of Haar measure, this measure coincides with the previously defined measure on M. Moreover, as any two volume forms are scalar multiple of each other, any Riemannian volume form will also produce the same measure on SM.

2.3 Metric on $S\mathbb{H}^2$ and SM

In the following $G = PSL(2,\mathbb{R})$, $\mathfrak{g} = T_eG$ and Γ is a lattice in G, so that $SM = \Gamma\backslash G$.

We will begin by putting a left-invariant Riemannian metric on G. Pick any positive definite inner product $< .,. >$ on $\mathfrak{g}$ and define the Riemannian metric $< .,. >_g$ on T_gG by left-invariance : $< .,. >_g =< .,. >$. We want a left invariant metric on G to retain the homogeneity of the Lie group.

Then, for any $g_1, g_2 \in G$ (lying in the same connected component of G), if $\psi : [0,1] \to G$ is a path connecting them, then we can determine length of the path by $L(\psi) = \int_0^1 ||D\psi(t)||dt$. So, define the metric by

$$d(g_1, g_2) = \inf_{\psi} L(\psi).$$

Proposition 2. *This is a left invariant metric on $S\mathbb{H}^2$ in the sense that $d(gg_1, gg_2) = d(g_1, g_2)$.*

We can now define a metric on $\Gamma\backslash G$ by

$$d(\Gamma g_1, \Gamma g_2) = \inf_{\gamma\in\Gamma} d(g_1, \gamma g_2).$$

This metric is again left invariant.

2.4 Stable and Unstable manifold of geodesic flow

Let f_t be a flow on a compact metric space X. Then, the stable set at a point $p \in X$ is

$$W^s(p) = \{x \in X : d(f_t(x), f_t(p) \to 0 \text{ as } t \to \infty)\}$$

and the unstable set at p is

$$W^u(p) = \{x \in X : d(f_{-t}(x), f_{-t}(p) \to 0 \text{ as } t \to \infty)\}.$$

When the flow is on a manifold X and it satisfies the conditions of an Anosov flow, these stable and unstable sets become submanifolds of X and enjoy certain invariance properties.

Definition 5. *(Anosov flow) A smooth flow f_t on a Riemannian manifold X is called Anosov if there exists constants $C > 0$ and $0 < \lambda < 1$ and a continuous splitting of the tangent space for all $p \in X$ and all $t \geq 0$ satisfying*

1. $T_pX = E^s(p) \bigoplus E^u(p) \bigoplus \mathbb{R}\dot{\phi}_t$
2. $||d\phi_t(v^s)|| \leq C\lambda^t||v^s||$ *for all* $v^s \in E^s(p)$
3. $||d\phi_{-t}(v^u)|| \leq C\lambda^t||v^u||$ *for all* $v^u \in E^u(p)$
4. $E^s(\phi_t(p)) = d\phi_t(E^s(p))$ *and similarly for the* E^u

For Anosov flows, the sets $W^s(p), W^u(p)$ are integral submanifolds for the distributions $E^s(p), E^u(p)$ respectively. Clearly, these submanifolds are transversal to each other and both are transversal to the flow direction $\phi_t(p)$. Also, $W^s(f(p)) = f(W^s(p))$ and $W^u(f^{-1}(p)) = f^{-1}(W^u(p))$.

We now look at $S\mathbb{H}^2 = PSL(2, \mathbb{R})$ and show that geodesic flow on the compact manifold SM is Anosov. This is where we will use the metric on $S\mathbb{H}^2$ and SM. For any g in $S\mathbb{H}^2$, $d(\phi_t(e), \phi_t(g)) = d(g_t, gg_t) = d(e, g_{-t}gg_t)$. Observe that $g_{-t}gg_t = \begin{pmatrix} a & be^{-t} \\ ce^t & d \end{pmatrix}$. Thus,

$$W^s(e) = \{h_s^+ = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R}\}$$

and

$$W^u(e) = \{h_s^- = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} : s \in \mathbb{R}\}.$$

h_s^+ and h_s^- are called the orbit of the positive and the negative horocycle flow respectively.

Recall that

$$psl(2, R) = \mathbb{R}X \bigoplus \mathbb{R}Y \bigoplus \mathbb{R}H$$

where

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Observe that

$$W^s(e) = \exp(\mathbb{R}X), W^u(e) = \exp(\mathbb{R}Y), \dot{\phi}_t(e) = H.$$

Moreover, for any $A = rX \in \mathbb{R}X$,

$$||d\phi_t(A)|| = ||\frac{d}{ds}|_{s=0} \begin{pmatrix} e^{t/2} & sre^{-t/2} \\ 0 & e^{-t/2} \end{pmatrix} || = || \begin{pmatrix} 0 & re^{-t/2} \\ 0 & 0 \end{pmatrix} || = ||A||e^{-t/2},$$

where $||.||$ is any norm (coming from an inner product) chosen on $psl(2, R)$.

Since tangent space to G and $\Gamma\backslash G$ are the same, this clearly shows that the geodesic flow on compact hyperbolic surfaces is Anosov.

Proposition 3. *The stable and unstable manifold for geodesic flow in SM at the point Γ is given by*

$$W^s(\Gamma) = \{\Gamma h_t^+\} \text{ and } W^u(\Gamma) = \{\Gamma h_t^-\}.$$

Special Note : Consider $\pi : SM \to M$ is the canonical projection map. Then, $\pi(\Gamma h_s^+)$ and $\pi(\Gamma h_s^-)$ are the horocycle orbits. Hedlund's theorem tells us that orbits of horocycle flow on compact hyperbolic surfaces have to be dense. Thus, the stable and unstable manifolds densely foliate SM. In other words, both of these submanifolds are immersed in SM.

The above characterization clearly shows that the stable and unstable manifolds and the flow lines foliate SM. Here, leaves of the foliation are smooth.

For the discussion in this Section, we will pretend that the flow direction in the splitting of SM is not there. This will be done for the brevity of the discussion. In the end, I will bring back the flow direction and show how everything works out even in the presence of the flow direction.

2.5 Local product structure of measure on SM

Consider the foliations W^s and W^u of SM, that are transversal to each other. Let (U, F) be a local foliation chart. We will write everything in local coordinates in the following. Let $W^s(., y)$ be the vertical lines denoting leaves of the stable foliation and $W^u(x, .)$ be the horizontal lines denoting the unstable foliation.

Let m be the measure induced by the Riemannian volume on SM and m_N be the measure on the (immersed) submanifold N obtained by pulling back the volume form on SM. Then, we have measures $m_{W^s(.,y)}$ and $m_{W^u(x,.)}$ along the stable and the unstable leaves.

Local product structure of the measure means : There exists "conditional density functions" $\delta_y : W^s(., y) \to \mathbb{R}$ for each stable leaf $W^s(., y)$ such that for any measurable subset $A \subset F(U)$,

$$m(A) = \int_{W^u(p,.)} \int_{W^s(.,y)} 1_A(x, y)\delta_y(x) dm_{W^s(.,y)} dm_{W^u(p,.)}$$

The same definition can be written down by interchanging the roles of W^s and W^u. The local product structure is vital as it allows us to use Fubini-like theorems locally, as in the following

Proposition 4. *Suppose the measure m has local product structure. Let $m(A) = 0$. Then $m_{W^s(.,y)}(A \cap W^s(., y)) = 0$ for $m_{W^u(p,.)}$ a.e. y for any unstable leaf $W^u(p, .)$.*

The big question now is whether or not the measure on SM has this local product structure along stable and unstable manifolds. It is a deep theorem that if a foliation has a property called "transversal absolute continuity", then the foliation has this local product structure. For the foliation W^s, consider any two transversals $L_1 = W^u(p_1, .)$ and $L_2 = W^u(p_2, .)$ in the foliation local chart. Let $p : L_1 \to L_2$ be the map between the transversals (called holonomy map). The foliation W^s will be said to be "transversally absolutely continuous" if for any two transversals L_1 and L_2, there exists a positive measurable function $q : L_1 \to \mathbb{R}$ such that

$$m_{L_2}(p(A)) = \int_{L_1} 1_A(.)q(.)dm_{L_1(.)}.$$

In our case, the foliation W^s and its transversal W^u are both smooth and the map p is obtained by flowing from one transversal to the other using the smooth horocycle flow. So, p is a smooth function. Also, the measures along the transversals are smooth (obtained by integrating smooth functions - since the measures along the leaves are induced by pulling back the smooth volume form).

3 Proof of Ergodicity : Hopf Argument

Proposition 5. *A continuous function ϕ_t invariant function f on SM is constant along the leaves of the stable and unstable foliation.*

Proof. Fix x and pick $y \in W^s(x)$. Then,

$$|f(x) - f(y)| = |f(\phi_t y) - f(\phi_t x)| \to 0$$

as $t \to \infty$, hence f is constant along $W^s(x)$. Similarly for $W^u(x)$. □

Proposition 6. *Let $g \in L^2(SM, m)$ be a ϕ_t invariant function. Then, g is constant a.e. along the stable and unstable foliations, that is, there exists $N_s, N_u \subset SM$ with $m(N_s) = m(N_u) = 0$ such that for any $x, y \in SM \setminus N_s(respectively, SM \setminus N_u)$ and $y \in W^s(x)(respectively, W^u(x))$, $g(x) = g(y)$.*

Proof. Approximate g in L^2 by a sequence of continuous functions f_k. For each continuous function f_k, define its forward ergodic average f_k^+ (exists by Birkhoff ergodic theorem). The approximating functions f_k might not be ϕ_t invariant, but the ergodic average f_k^+ is. So they are constant along each leaf of the stable and unstable foliation (note that these functions f_k^+ are possibly defined a.e.). Similarly, take the forward ergodic average of g, say g^+. But since g is ϕ_t invariant, $g^+ = g$. It is easy to show that f_k^+ converges to g^+ in L^2 (requires the property that m is preserved by geodesic flow). Then, passing to a subsequence, f_k^+ converges to g^+ almost everywhere. f_k^+ is constant a.e. along leaves of stable and unstable foliation. Hence, g is a.e. - constant along each leaf of the stable and the unstable foliation. □

Proposition 7. *The function g is locally constant.*

Proof. Using local foliation charts, we will now write everything in coordinates. For visualizing the rest, think of a neighbourhood of origin where the lines parallel to x-axis are the leave of the stable foliation and the lines parallel to the y-axis are the unstable foliation.

Observation 1: Appealing to Proposition 4, we can say that $m_{W^s(.,y)}(N_s \cap W^s(.,y)) = 0$ for a.e. $y (\in W_u(p,.)$ for any p. So, $g(x,y)$ depends only on y for almost every y as we move along $W^u(p,.)$.

Observation 2: But appealing to Proposition 4, we observe that $m_{W^u(x,.)}(N_u \cap W^u(x,.)) = 0$ for a.e. $x(\in W^s(.,q))$ for any q. Then, $g(x,y)$ is a function of x only $W^s(.,q)$. Pick such a x_0.

Observation 1 tells us that $g(x,y)$ is a function of y alone, that is, $g(x,y) = h(y)$. So, look at $W^u(x_0,.)$. Along this unstable leaf, $g(x,y)$ should be a constant, say c_{x_0} (Observation 2). That is, $g(x,y) = h(y) = c_{x_0}$. □

Proposition 8. *Assume that M is connected. The function g is then constant. Hence, the geodesic flow is ergodic.*

Proof. Suppose $\{U_i : i = 1, ..., n\}$ is an open cover of M with $g = c_i$ on U_i. Observe that if $U_i \cap U_j \neq \phi$, then $c_i = c_j$. If there is a U_i that doesn't intersect any other U_j, then that U_i produces a disconnection of M. Hence, all U_i-s must intersect. So, g must be constant. □

4 Alternative Proof of Ergodicity : Mautner Phenomenon

Consider the left regular representation of G on $L^2(G)$.

Consider $f \in L^2(\Gamma\backslash G)$ that is invariant under g_t, that is, $g_t.f(\Gamma g) = f(\Gamma g g_t) = f(\Gamma g)$. We will show that f is invariant under $N^+ = \{h_s^+ : s \in \mathbb{R}\}$ and $N^- = \{h_s^- : s \in \mathbb{R}\}$. This will prove that f is G invariant since G is generated by N^+ and N^-.

It is straightforward to check that $g_t h_s g_{-t} = h_{se^{2t}}$. Then

$$\lim_{n\to\infty} g_{-n} h_s g_n = e$$

. Let $\pi : G \to L^2(\Gamma\backslash G)$ denote the action of G on L^2.

Observe that

$$||h_s.f - f||^2 = ||h_s g_n.f - f||^2 = ||g_{-n} h_s g_n.f - g_{-n}.f||^2 = ||g_{-n} h_s g_n.f - f||^2 \to 0.$$

Thus $h_s^+.f = f$ and similarly, $h_s^-.f = f$ and hence, f is G invariant.

References

[1] Ballmann, W., Lectures on Spaces of Nonpositive Curvature.

[2] Einseidler Ward, Ergodic Theory with a view towards Number Theory

[3] Brin, M., and Stuck, G., Introduction to Dynamical Systems.