LECTURES ON ERGODIC THEORY, GEODESIC FLOWS AND RELATED TOPICS.

Mark Pollicott

University of Manchester

Contents

I am grateful to my colleages in Manchester for their help in reducing the number of mistakes in these notes.

Typeset by $\mathcal{AMS}$-TEX

Lecture 1: Geodesic flows and symbolic dynamics

1.1 Compact surfaces and manifolds. Let M be a smooth compact n-manifold. We can conveniently think of it being covered by coordinate patches with C^∞ maps to regions in $\mathbb{R}^n$. Alternatively, we can think of M as a submanifold of $\mathbb{R}^{2n+1}$:

Embedding Theorem (Whitney, 1934). *There is always a smooth realization of M as a submanifold of $\mathbb{R}^{2n+1}$.*

We shall concentrate on surfaces (i.e., when $n = 2$). Fortunately, these don't come in too many different shapes:

Classification Theorem (Mobius, 1863). [1] *A compact orientable surface is either*

(1) *a sphere (genus zero);*
(2) *a doughnut or torus (genus one); or*
(3) *a g-holed doughnut (genus $g \geq 2$).*

The genus corresponds to the "number of holes"

1.2 Geodesics and geodesic flows. It is convenient to think of a geodesic $\gamma : \mathbb{R} \to M$ as a curve which (locally) minimizes the distance between points along its length (i.e., $\gamma(t, t + \epsilon])$ is the shortest curve from $\gamma(t)$ to $\gamma(t + \epsilon)$), for $\epsilon > 0$ sufficiently small.

Theorem (Hopf-Rinow, 1931). *Given any two points x, y there always exists a geodesic connecting them.*

Let SM denote the sphere bundle for M, i.e., the tangent vectors v to M which have unit length. The geodesic flow $\phi_t : SM \to SM$ moves a vector $v \in SM$ to $\phi_t(v) \in SM$ by "parallel transport". In particular, one chooses a unit speed geodesic $\gamma_v : \mathbb{R} \to SM$ such that $\dot\gamma_v(0) = v$ and then defines $\phi_t(v) = \dot\gamma(t)$.

Alternatively, one can write geodesics $\gamma(t)$ locally as solutions to differential equations

$$\frac{d^2\gamma_k}{dt^2} + \sum_{i,j=1}^{n} \Gamma^k_{ij} \frac{d\gamma_i}{dt} = 0, \text{ for } k = 1, \ldots, n,$$

where Γ^k_{ij} are the "Christoffel symbols" determined by the metric g_{ij} as:

$$\Gamma^k_{ij} = \frac{1}{2} \sum_s g^{ks} \left(\frac{\partial g_{si}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^s} + \frac{\partial g_{js}}{\partial x^i} \right)$$

[1] Interestingly, Mobius proved this when in his prime (aged 71) and submitted it for the Grand Prix of the Academie des Sciences de Paris. Alas, it was decided none of the submissions that year merited the prize being awarded.

1.3 Curvature. An intuitive definition of curvature on surfaces is the following. We can associate to a point $x \in M$ on a surface the value defined by

$$\kappa(x) = \lim_{r \to 0} \frac{12}{\pi} \left(\frac{\pi r^2 - \text{Area}(B(x,r))}{r^4} \right), \qquad \text{(Diquet's formula)}$$

where $B(x,r)$ is a ball of radius r about x.[2] κ measures how curved the surface is (e.g, $\kappa(S^2) = 1$, $\kappa(\mathbb{T}^2) = 0$).

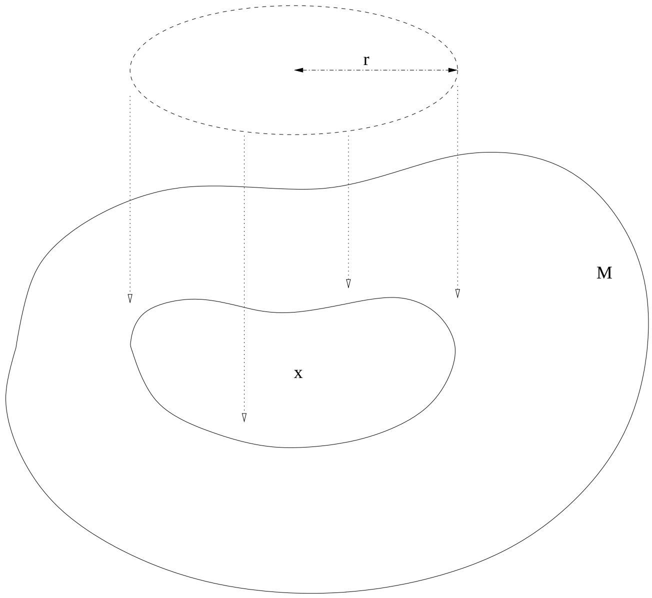

FIGURE 1. THE CURVATURE OF A SURFACE

For constant curvature surfaces there is a simple description of M.

Uniformization Theorem (Poincaré-Koebe, 1882 & 1907). *A compact orientable surface with constant curvature is covered by one of the following spaces*

1. *a sphere $\mathbb{S}^2$ (with $\kappa = 1$: genus zero);*
2. *the plane $\mathbb{R}^2$ (with $\kappa = 0$: genus one); or*
3. *the Poincaré disk $\mathbb{D}^2$ (with $\kappa = -1$:genus $g \geq 2$),*

with a suitable metric. In particular, M corresponds to the identification of this covering space by a discrete subgroup Γ of isometries. (The genus corresponds to the "number of holes".)

The 3-dimensional generalization of this is the Thurston Uniformization Conjecture (which may now be solved). The topology (i.e., the genus) forces some restrictions on the curvature:

[2] More formally, one can use Christoffel symbols to write

$$\kappa(x_1, x_2) = -\frac{1}{g_{11}} \left(\frac{\partial \Gamma_{12}^2}{x_1} - \frac{\partial \Gamma_{11}^2}{x_2} + \Gamma_{12}^1 \Gamma_{11}^2 + (\Gamma_{12}^2)^2 - \Gamma_{11}^2 \Gamma_{22}^2 - \Gamma_{11}^1 \Gamma_{12}^2 \right).$$

Theorem (Gauss-Bonnet, 1825 & 1848). *The total curvature of a surface M is related to the genus g by*

$$\int \kappa(x) d\,Vol(x) = -4\pi(g-1)$$

We say that M has negative curvature if $\kappa(x) < 0$, for all $x \in M$. By the Gauβ-Bonnet theorem only surfaces of genus $g \geq 2$ (i.e., "g-holed doughnuts") can have negative curvature. Small pieces of negatively curved surfaces can be isometrically embedded in $\mathbb{R}^3$ (as "saddles"). However:

Theorem (Hilbert-Efimov, 1901 & 1964). *A complete surface of negative curvature cannot be isometrically embedded in $\mathbb{R}^3$ (i.e, so that the metric on M matchs that on the surface).*

If $n \geq 3$, then one can consider the sectional curvature (i.e., the curvature in two-dimensional planes) in place of the Gaussian curvature for surfaces.

1.4 Anosov flows and examples. The geodesic flow in negative curvature has special dynamical features: Transverse to the flow direction the flow contracts exponentially quickly in one direction (stable direction) and expands exponentially quickly in another direction (unstable direction). Anosov extracted this dynamical property for the flows that bear his name (Anosov flows).

There is a splitting $T(SM) = E^0 \oplus E^s \oplus E^u$ such that

(i) E^0 is the (one dimensional) flow direction;
(ii) E^s is a contracting direction, i.e., $||D\phi_t v|| \leq Ce^{-\lambda t}||v||$ for $v \in E^s$
(iii) $||D\phi_{-t} v|| \leq Ce^{-\lambda t}||v||$ for $v \in E^u$

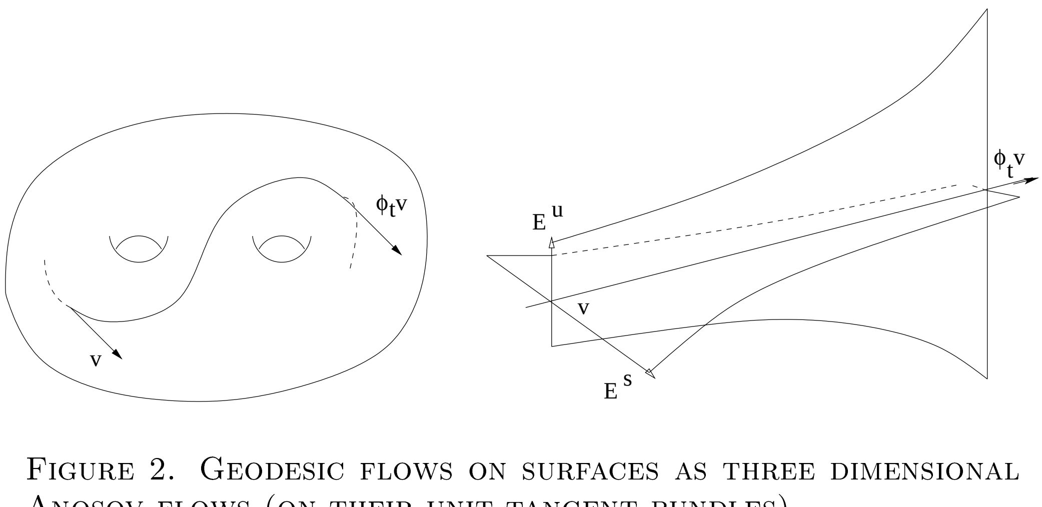

FIGURE 2. GEODESIC FLOWS ON SURFACES AS THREE DIMENSIONAL ANOSOV FLOWS (ON THEIR UNIT TANGENT BUNDLES)

For geodesic flows: E^0 is the tangent to the geodesic and E^s (rep. E^u) are horocycles, i.e., vectors which converge towards the geodesic as $t \to +\infty$ (resp. $t \to -\infty$).

Examples.

(1) *Constant curvature $\kappa = -1$ surfaces M.* We recall from the Uniformization theorem that the (Universal) cover takes the form $\mathbb{D}^2 = \{x + iy \in \mathbb{C} : x^2 + y^2 < 1\}$ with the Poincaré metric

$$ds^2 = 4\frac{dx^2 + dy^2}{(1-(x^2+y^2))^2}.$$

The geodesics correspond to semi-circular arcs which meet the unit circle perpendicularly. The horocycles are circles which touch the real axis at a single point.

Let Γ be a discrete subgroup of isometries of $\mathbb{H}^2$ (i.e., maps which preserve distances). These can be identified with the action

$$\left(\begin{smallmatrix} a & b \\ \overline{b} & \overline{a} \end{smallmatrix}\right) : z \mapsto \frac{az+b}{\overline{b}z+\overline{a}}$$

of 2×2 matrices (with $|a|^2 - |b|^2 = 1$).[3]

(2) *Linkages.* Consider 3 pairs of double pendulum, each having one end attached to a vertex of an equilateral triangle and the other to a moving couple. If there is inertia for the outer rods then the associated motion is described by a geodesic flow. MacKay and Hunt showed that certain weights and lengths make the curvature negative (except at a few points) and thus the flow is still Anosov.

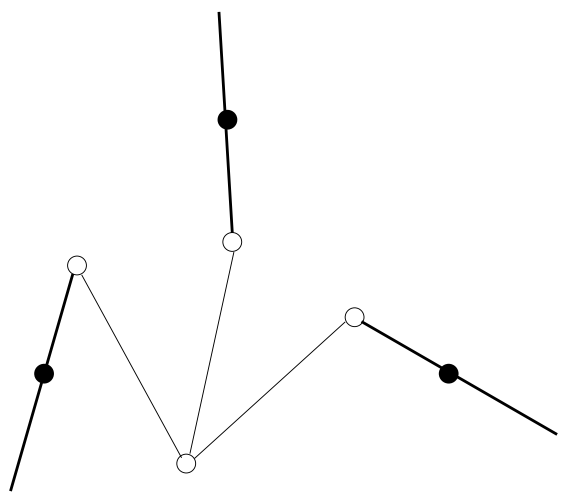

FIGURE 3. LINKAGES

1.5 . "Symbolic dynamics" for surfaces: Coding geodesics using the boundary. [4] Our approach in the next lectures to closed orbits of geodesic flows on negatively curved surfaces depends on symbolic dynamics.

Basic Principle. Closed orbits τ for the geodesic flow correspond to periodic orbits $\{x, f(x), f^2(x) \ldots, f^{(n-1)}(x)\}$ for some C^1 map $f : I \to I$ of an interval I.[5]

For geodesic flows on surfaces of constant curvature such coding was first introduced by Morse and Hedlund. We shall discuss these ideas because they are more geometric. Additionally, in this case the length of the orbit is typically given by

[3] A quite general characterization of negative curvature is the *thin triangle property*: There exists $\delta > 0$ such that any triangle Δ has the property that any point on the triangle is δ-close to a point on one of the other sides. (This leads to a natural extension to the idea of $CAT(-1)$ spaces - such as graphs and trees.)

[4] Actually, we will usually stop one step shy of the full blown symbolic dynamics, by focusing on expanding maps rather than sequence spaces.

[5] More accurately, I may be a union of intervals I_i and f may be a piecewise C^1 map which is Markov (i.e., each image $f(I_i)$ is a union of other intervals)

$\lambda(\gamma) = \log |(f^n)'(x)|$ (corresponding to the "expansion along the unstable bundle around the orbit").

A special example: "Modular surface". It is notationally more convenient to work with $\mathbb{H}^2 = \{x + iy : y > 0\}$ rather than $\mathbb{D}^2$. The boundary of $\mathbb{H}^2$ then becomes $\mathbb{R} \cup \infty$. Geodesics are semi-circles which meet the boundary perpendicularly, or else vertical straight lines. The corresponding group of isometries is $SL(2, \mathbb{Z})$[6] and the familiar fundamental domain for the Modular surface has three sides which are identified by $S = S^{-1} = -1/z$ and $T(z) = z + 1$.

Given any geodesic γ on $M = \mathbb{H}^2/\Gamma$ we can lift it to a geodesic $\widetilde{\gamma}$ on $\mathbb{H}^2$ which is completely determined by its endpoints $\xi = \widetilde{\gamma}(-\infty)$ and $\eta = \widetilde{\gamma}(+\infty) \in \mathbb{R} \cup \infty$. Moreover, we can arrange that either

(0) $-1 < \xi < 0$ and $\eta > 1$, and then we can "code" these endpoints by their continued fractions

$$\xi = -\frac{1}{n_0 + \frac{1}{n_{-1} + \frac{1}{\dots}}} \text{ and } \eta = n_1 + \frac{1}{n_2 + \frac{1}{n_3 + \frac{1}{\dots}}} \quad (\text{with } n_i \in \mathbb{N},\ i \in \mathbb{Z})$$

or else

(1) $\eta < -1$ and $0 < \xi < 1$ and we can "code" these endpoints by their continued fractions

$$-\eta = n_1 + \frac{1}{n_2 + \frac{1}{n_3 + \frac{1}{\dots}}} \text{ and } \xi = \frac{1}{n_0 + \frac{1}{n_{-1} + \frac{1}{\dots}}} \quad (\text{with } n_i \in \mathbb{N},\ i \in \mathbb{Z})$$

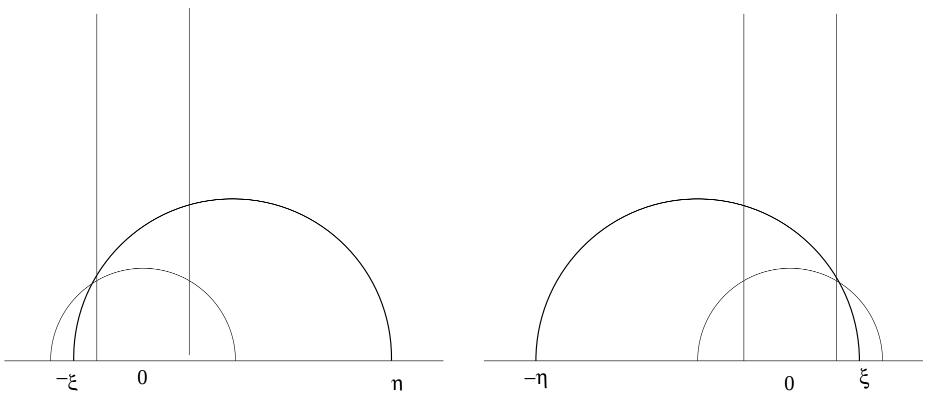

FIGURE 4. ENDPOINTS OF THE LIFTS OF GEODESICS TO THE UNIVERSAL WITH END POINTS EITHER: (0) $-1 < -\xi < 0$, $\eta > 1$; OR (1) $-\eta < -1$, $0 < \eta < 1$

In particular, Adler-Flatto and Series showed that moving along the geodesic, crossing the side of the Fundamental domain and moving the geodesic $\widehat{\gamma}$ back into position (by replacing it by $g\widetilde{\gamma}$, for some $g \in \Gamma$) corresponds to shifting the sequence $((n_i)_{i\in\mathbb{Z}}, j) \in \mathbb{N}^{\mathbb{Z}} \times \{0, 1\}$ to $((n_{i-1})_{i\in\mathbb{Z}}, 1-j)$. The end points of the geodesics move according to $j = 0$ or $j = 1$ in the two cases above.

[6]Of course, matrices A and $-A$ correspond to the isometries. But let us overlook this, rather than introduce $SL(2, \mathbb{Z})/\{\pm I\}$.

In particular, the periodic orbits correspond to periodic sequences and are described by the periodic points for the Gauβ (continued fraction) transformation:

$$T : (0,1) \to (0,1)$$
$$T : x \mapsto \frac{1}{x} \pmod 1$$

A general construction for compact surfaces (Adler-Flatto and Series).

Step 1. Choose a fundamental domain in the universal cover $\mathbb{D}^2$ (a copy of the surface) which has boundaries which are geodesics and corners which are right angles. Every geodesic on $\mathbb{D}^2$ (which passes through the fundamental domain) is determined precisely by its end points.

Step 2. There are a finite set of generators $\{g_1, \dots, g_n\} \subset \Gamma \subset SL(2, \mathbb{R})$ which identify pairs of sides of the fundamental domain.

Step 3. Each geodesic arc meets the boundary circle at two points. The collection of these points neatly divides the boundary circle into finitely many arcs $\{I_i\}$.

Step 4. The coding comes about naturally using the (induced) action of the g_i on the boundary. We associated to each of the arcs I_i an (expanding) map $f : I_i \to I := \coprod_i I_i$ defined by $f(x) = g_i(x)$, for a corresponding g_i

By restricting the action of the generators to appropriate parts of the boundary we get an expanding (Markov) map $f : I \to I$ on $I = \coprod_i I_i$.

Appendix (Ia): Counting geodesic arcs. The Hopf-Rinow gives that there is a closed geodesic linking any two points x, y. If we assume that a surface has $\kappa < 0$ then we can count geodesic arcs between any fixed points on the surface. More precisely:

Theorem (Margulis, 1969). *Let M be a compact manifold and let $x, y \in M$. Let $A(T)$ denotes the number of geodesic arcs from x to y of least period at most T. Then there exists $C, h > 0$ such that $A(T) \sim Ce^{hT}$, as $T \to +\infty$.*

This is very similar to the ideas in the next lectures on counting closed geodesics, except one complex function in the proof (the zeta function) is replaced by another (the Poincaré series).

Appendix (Ib): Symbolic dynamics for general Anosov flows. For general three dimensional Anosov flows, one has a far less canonical construction. We can still a C^1 expanding map $f : I \to I$ on a disjoint union of intervals $I = \cup_{i=1}^k I_i$. The idea is:

1. (1) Choose a finite number of 2-dimensional sections transverse to the flow;
2. (2) Consider the discrete Poincaré map induced on sections;
3. (3) Collapse the 2-dimensional sections along the stable directions to get a one-dimensional expanding map $f : L \to I$ (and the time to flow between them reduces to a function $f : I \to \mathbb{R}$).

Finally, the hyperbolicity of the Anosov flow allows us to shuffle the sections in such a way that every thing is well-defined.

The system $f : I \to I$ and $r : I \to \mathbb{R}$ is:

- (i) real analytic if $\kappa = -1$; or
- (ii) C^1 if $\kappa < 1$.

This all depends on the regularity of the "collapsing" in the stable direction (i.e., the regularity of the splitting for the Anosov flow).

Lecture 2: Closed orbits and zeta functions

2.1 General results for surfaces. We begin with a simple but important observation.

Trivial Observation. A closed orbit τ for the geodesic flow $\phi_t : SM \to SM$ of least period $\lambda(\tau)$ corresponds to a (directed) closed geodesic γ on M of length $\lambda(\gamma)$.

Clearly, the standard two dimensional sphere $\mathbb{S}^2$ has infinitely many closed geodesics. But for any metric we have the following remarkable result.

Theorem (Klingenberg Conjecture = Franks-Bangert Theorem, 1992). *Let M be a two dimensional sphere with any C^∞ Riemannian metric. Then there are infinitely many closed geodesics γ.*

Interestingly, the proof uses the geodesic flow. It begins with a standard result that there is at always at least one closed geodesic. The Poincaré map for geodesics crossing this one gives a map of an (open) annulus and Franks then used this to reformulate the result in terms of periodic points of an annulus *and* shows they are infinite under some additional assumptions. When these assumptions don't hold, the result still holds by more geometric results of Bangert.

In fact, any other orientable surface (with genus $g \geq 1$) will have a countable infinitely of closed geodesics (since every conjugacy class in the fundamental group contains a closed geodesic, and there are infinitely many of these). However, the next result gives a qualitative estimate when $g \geq 2$.

Theorem (Katok, 1980). *Let M be a surface of genus $g \geq 2$ with any C^∞ Riemannian metric. Then there exist $h > 0$ and $0 < T_1 < T_2 < \cdots < T_n < \cdots \nearrow +\infty$ such that*

$$\mathrm{Card}\{\gamma : \lambda(\gamma) \leq T_n\} \geq e^{hT_n}, \ \textit{for } n \geq 1.$$

The proof uses ideas from Pesin Theory (essentially an extension of the Anosov theory to where the hyperbolicity is no longer uniform in the constant $C = C(x)$).

Application to linkages. In the case that the phase space of the linkage described in lecture 1 is a surface of genus at least 2, we see that Katok's Theorem applies. In particular, the number of (unstable) periodic configurations for a coupling grows at an exponential rate as it evolves in its phase space. This happens, for example, with the regular pentagon asnd holds without any assumption on the masses.

2.2. Negative curvature and closed orbits. For negatively curved manifolds there are better results. To begin with, the following growth rates exist and are the same:

$$h = \lim_{T\to+\infty} \frac{1}{T} \log \ \mathrm{Card}\{\gamma : \lambda(\gamma) \leq T\} \qquad \text{(Sinai, 1966)}$$

$$h = \lim_{T\to+\infty} \frac{1}{T} \log \mathrm{Volume}\{x \in \widetilde{M} : \ \mathrm{dist}\ (x, x_0) \leq T\} \qquad \text{(Manning, 1979)}$$

(where $x_0 \in \widetilde{M}$ is any fixed point in the Universal Cover $\widetilde{M}$ for M).[7]

The constant h is called the *(topological) entropy* (of geodesic flow).

[7] A result close to that of Manning was apparently contained in Margulis' unpublished thesis

Interesting Fact. [8] *Any metric of negative curvature on a surface can be continuously changed to another metric of negative curvature on the same surface through the family of metrics of negative curvature.*

In any dimension, Katok-Knieper-Pollicott-Weiss (1989) showed that the entropy changes smoothly over metrics with negative sectional curvature and the minimum over metrics with the same total volume is attained at one with constant curvature metric (when available). This was called the Gromov-Katok Conjecture and was solved by Katok (1982) for surfaces and by Besson-Courtois-Gallot (1990) in any dimension.

The Sinai estimate extends to an asymptotic result. Let us denote $\pi(T) = \mathrm{Card}\{\tau : \lambda(\tau) \leq T\}$.

Theorem (Prime Orbit Theorem). *We have the asymptotic formula*

$$\pi(T) \sim \frac{e^{hT}}{hT}, \text{ as } T \to +\infty,$$

(i.e., $\lim_{T\to\infty} \pi(T)/\frac{e^{hT}}{hT} = 1$*).*

For constant curvature surfaces this proved by Huber (1959) with $h = \sqrt{|\kappa|}$ using the Selberg trace formula ideas. For variable curvature this was proved by Margulis (1969), up to an unknown constant[9], using a direct (but unpublished) method rather than zeta functions. This was reconstructed in the thesis of Toll (1983) (but using, in addition, Bowen's results on equidistribution of closed geodesics, but with the final form above). Parry and Pollicott (1983) proved a more general result using zeta functions.

2.3. Counting with error terms. How can we get better estimates in the variable curvature case? (Huber did this for constant curvature) We first need the "correct" version of the principle asymptotic expression:

$$\mathrm{li}(e^{hx}) = \int_2^{e^{hx}} \frac{du}{\log u} \sim \frac{e^{hx}}{hx}.$$

The strengthening of the previous theorem becomes:

Theorem (Prime Orbit Theorem with error term). *There exists $h > 0$ such that*

$$\pi(T) = \mathrm{li}(e^{hT}) \left(1 + O\left(e^{-\epsilon T}\right)\right), \text{ as } T \to +\infty,$$

(i.e., there exists $C > 0$ such that $|\pi(T)/\mathrm{li}(e^{hT}) - 1| \leq Ce^{-\epsilon T}$).

For constant curvature this was part of Huber's result. For variable curvature this was done by Pollicott and Sharp (1998), using estimates of Dolgopyat (1997).

A basic approach to proving such theorems is to use ζ-functions:

---

[8] Using this and structural stability we see that the constant curvature coding can alos be applied to variable curvature.

[9] Which from a modern perspective is easier to identify

2.4 . Dynamical zeta functions and their properties. Following Ruelle, we can formally define a complex function

$$\zeta(s) = \prod_{\gamma} \left(1 - e^{-s\lambda(\gamma)}\right)^{-1} \qquad \textit{(Ruelle dynamical } \zeta\textit{-function)}$$

where $s \in \mathbb{C}$. Sinai's characterization of the entropy h as a growth rate of closed geodesics immediately shows that:

(i) ζ actually converges to a non-zero analytic complex function providing $Re(s) > h$.

We are guided by the following ...

Philosophy. The more we know about the poles for $\zeta(s)$ the more we know about $\pi(T) = \mathrm{Card}\{\gamma : \lambda(\gamma) \le T\}$.

(ii) This function has a simple pole at $s = h$ (Ruelle, 1978 & 2002).[10]
(iii) Otherwise the function is non-zero and analytic on $s = h + it$ (Parry and Pollicott, 1983).

This is enough for us to prove the Prime Orbit Theorem. Of some independent interest are:

(iv) $\zeta(s)$ always has a non-zero meromorphic extension to a larger half-plane $Re(s) > h - \epsilon$ (Pollicott, 1985).
(v) This function always has a non-zero meromorphic extension to $s \in \mathbb{C}$ (Kitaev, 1999).

The extra ingredient to prove the error terms in the prime orbit theorem was the following improvement to (iv):

(vi) $\zeta(s)$ always has a non-zero *analytic* extension to a larger half-plane $Re(s) > h - \epsilon$, except for the simple pole at $s = h$ (with suitable bounds).

Remark. Any resemblance of the Dynamical Zeta function and Prime Orbit Theorem (with error term) to the Riemann Zeta function and Prime Number Theorem (with Riemann Hypothesis), respectively, is not coincidental and has been discussed at length by more eloquent authors.

2.5. The Selberg zeta function. The (more) famous Selberg Zeta function

$$Z(s) = \prod_{n=0}^{\infty} \prod_{\gamma} \left(1 - e^{-(s+n)\lambda(\gamma)}\right) \qquad \textit{(Selberg } \zeta\textit{-function)}$$

is more natural to study for constant curvature. From the definitions there is a simple formal relationship:

$$\zeta(s) = \frac{Z(s+1)}{Z(s)} \text{ and } Z(s) = \prod_{n=0}^{\infty} \zeta(s+n)^{-1}.$$

[10]Actually, this is an exercise which appears in Ruelle's book

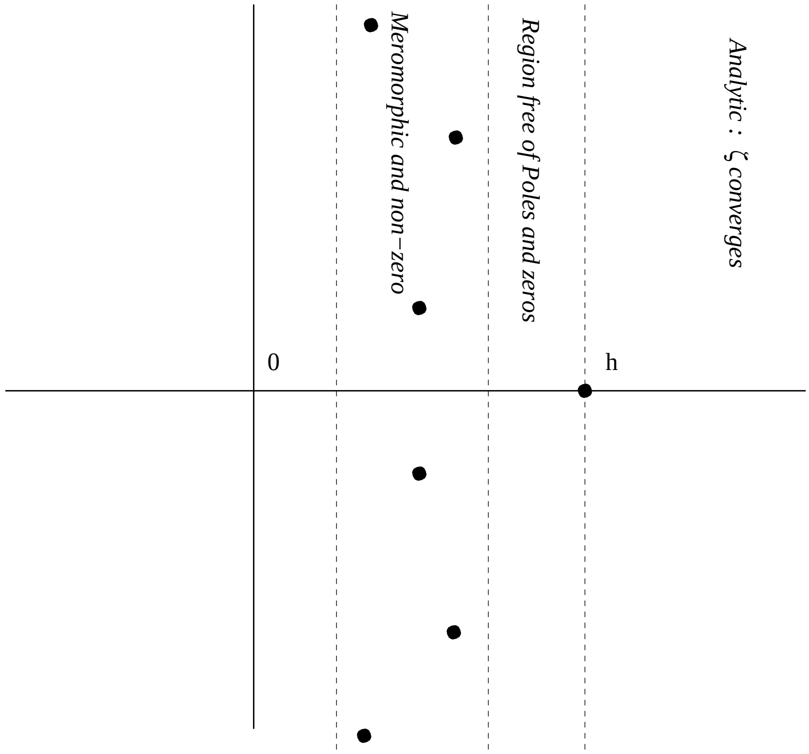

Figure 5. The domain of the zeta function

The zeros of $Z(s)$ are closely related to the spectrum of the laplacian $\Delta\phi_n = -\lambda_n\phi_n$, where $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \nearrow +\infty$.

Let us now restrict to the constant curvature case $\kappa = -1$. An interesting connection is with the *determinant of the laplacian* defined by

$$\det(\Delta) := \exp(-\eta'(0)) \text{ where } \eta(s) := \sum_{n=1}^{\infty} \lambda_n^{-s}.$$

We have the identity

Lemma (D'Hoker-Phong, Sarnak, 1986&1987). *There exists $C = C(g)$ depending only on the genus $g \geq 2$ of the surface:*

$$\det(\Delta) = CZ'(0).$$

In particular, we can find a formal expression for $\det(\Delta)$ as a series defined in terms of lengths of closed geodesics using the definition of the Selberg Zeta function. It is an interesting aspect of the Ruelle approach that this series actually converges quickly and gives a route to numerical computation (Pollicott-Rocha, 1997).

Appendix II. Counting self-intersections. Rather than counting all closed geodesics we might only count those which never intersect themselves. This is substantially smaller than the total number.

Proposition (Rees, 1980). *The number of closed geodesics of length at most T without any self-intersections grows at only polynomial speed (i.e., it is bounded by CT^ρ, for some $C, \rho > 0$).*

If we associate to a typical closed geodesic γ then number of times $i(\gamma)$ that it intersects with itself then we have the following "average".

Proposition (Pollicott, Lalley). *There exists $C > 0$ such*

$$\lim_{T \to +\infty} \frac{1}{\pi(T)} \sum_{\gamma : \lambda(\gamma) \leq T} \frac{i(\Gamma)}{T^2} = C$$

One can refine these sorts of results to deal with the angle of self-intersection (Sieber, Pollicott-Sharp).

We can look beyond just closed geodesics and consider any geodesic on the surface which doesn't intersect itself. We can consider the lifts of these to $\mathbb{D}^2$ and let $X \subset S^1$ be the set of endpoints of such geodesics.

Proposition (Birman-Series, 1985). *The set X has zero measure (and zero Hausdorff dimension).*

Lecture 3: Transfer operators and zeta functions

3.1 "Symbolic" dynamics and ζ-function. For $\kappa = -1$ the Selberg Trace Formula is the most successful approach to study ζ-functions. For variable curvature, symbolic dynamics still applies. In particular, can associate a C^1 one-dimensional expanding map $f : I \to I$ and a function $r : I \to \mathbb{R}^+$.

"Symbolic" model. The idea is that the "complicated" flow $\phi_t : SM \to SM$ can be modeled by a simpler (semi)-flow $\psi_t : \Omega \to \Omega$ on the area under the graph $\Omega = \{(w, u) \in I \times \mathbb{R} : 0 \leq u \leq r(w)\}$. The (semi)-flow is given by flowing vertically (i.e., $\psi_t(w, u) = (w, u + t)$ for $t \geq 0$) with the identification at the top that that $(w, r(w)) = (f(w), 0)$.

A key point of symbolic dynamics. The closed orbits τ on SM correspond to periodic orbits $\{x, fx, \cdots, f^{n-1}x\}$ with $\lambda(\gamma) = r^n(x) := \sum_{i=0}^{n-1} r(f^i x)$.[11]

We denote $r^n(x) := \sum_{i=0}^{n-1} r(f^i x)$. Since we are using "symbolic dynamics" we can consider a *"symbolic" ζ-function* defined by

$$\zeta(s) = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} Z_n(s)\right) \text{ where } Z_n(s) = \sum_{f^n x = x} \exp\left(-s r^n(x)\right),$$

is a sum over all periodic points $f^n x = x$ of period n.

By elementary series expansions of the series and the above "key ingredient" we can see that for an orbit τ:

$$\begin{aligned}(1 - e^{-s\lambda(\tau)})^{-1} &= \exp\left(-\log(1 - e^{-s\lambda(\tau)})\right) \\ &= \exp\left(\sum_{k=1}^{\infty} \frac{e^{-sk\lambda(\tau)}}{k}\right) = \exp\left(\sum_{k=1}^{\infty} \frac{e^{-skr^n(x)}}{k}\right),\end{aligned}$$

and thus since the ζ-function for the geodesic flow is the product of all such terms one can see ...

[11] For geodesic flows on surfaces there may be finitely many exceptional primitive orbits - but these don't spoil the results.

Easy calculation. *The ζ-function for the geodesic flow agrees with the "symbolic" ζ-function.*[12]

Thus the ζ-function for geodesic flows is replaced by one for interval maps. Symbolic dynamics is good for us because we can introduce ...

3.2. The Ruelle Transfer operator. Given a C^1 expanding (Markov) map $f : I \to I$ and a C^1 function $g : I \to \mathbb{R}$ we can associate a bounded linear *(Ruelle) transfer operator* $\mathcal{L}_g : C^1(I) \to C^1(I)$ defined by

$$\mathcal{L}_g w(x) = \sum_{y \,:\, fy=x} w(y)e^{g(y)}.$$

on the Banach space of C^1 functions (with norm $||w|| = |w|_\infty + |w'|_\infty$, where $|w|_\infty = \sup_{x\in I} |w(x)|$ is the supremum norm). The nice thing about this operator is its spectral properties:

Theorem (Ruelle Operator Theorem, Ruelle 1968, Bowen, 1975. *The operator $\mathcal{L}_g$ has $e^{P(g)}$ as a simple maximal positive eigenvalue. The corresponding eigenfunction w satisfying $\mathcal{L}_g w = e^{P(g)} w$ is positive. The rest of the spectrum is in a strictly smaller disk.*[13]

The spectral properties of transfer operators are intimately related to those of ζ-functions ...

3.3 Families of Transfer operators as a route to ζ-functions. To describe a zeta function $\zeta(s)$ depending on the variable $s \in \mathbb{C}$ we need to associate a *family* of complex transfer operators $\mathcal{L}_s : C^1(I) \to C^1(I)$ defined by

$$\mathcal{L}_{-sr} w(x) = \sum_{y: fy=x} w(y)e^{-sr(y)}$$

(i.e., we consider only complex functions $g = -sr$, $s \in \mathbb{C}$).

Example: The Gauß continued fraction map (revisited). Let $f : (0,1) \to (0,1)$ be defined by $f(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$. We associate $r : (0,1) \to \mathbb{R}$ defined by $r(x) = \log|f'(x)| = -2\log x$. The Ruelle operator takes the form[14]

$$(\mathcal{L}_{-sr} w)(x) = \sum_{n=1}^{\infty} \frac{1}{(x+n)^{2s}} w\left(\frac{1}{x+n}\right)$$

Let us choose $0 < \theta < 1$ such that $1 < \theta^{-1} \le |f'(x)|$, for any $x \in I$. Given $s \in \mathbb{C}$, the function $-Re(s)r : I \to \mathbb{R}$ is real valued and so by the Ruelle operator theorem $\mathcal{L}_{-Res(s)r}$ has maximal eigenvalue $e^{P(-Re(s)r)}$. If $s \notin \mathbb{R}$ then we don't necessarily expect that $\mathcal{L}_{-sr}$ has a maximal positive eigenvalue. However, we still have the following restriction on the spectrum.

[12] Again, this is up to a finite number of exceptional closed orbits.

[13] The value $P(g)$ is usually called the *pressure* of the function. It actually has lots of equivalent definitions, including the well known variational principle.

[14] Actually, this operator even preserves analytic functions. However, in this special case with we should restrict to $Re(s) > 1/2$ to make sure the operator is well-defined

Theorem (on spectrum). *The spectrum of* $\mathcal{L}_{-sr}$ *on* C^1 *functions consists of:*

(1) *isolated eigenvalues* λ_i *in the annulus* $\{z \in \mathbb{C} : \theta e^{P(-Re(s)r)} < |z| \leq e^{P(-Re(s)r)}\}$; *and*
(2) *the rest contained in a ball* $\{z \in \mathbb{C} : |z| \leq \theta e^{P(-Re(s)r)}\}$.

This sort of result was proved for interval maps by Keller (1984) and for subshifts by Pollicott (1984). The connection between the poles of the ζ-function and the spectrum of the transfer operator(s) is the following:

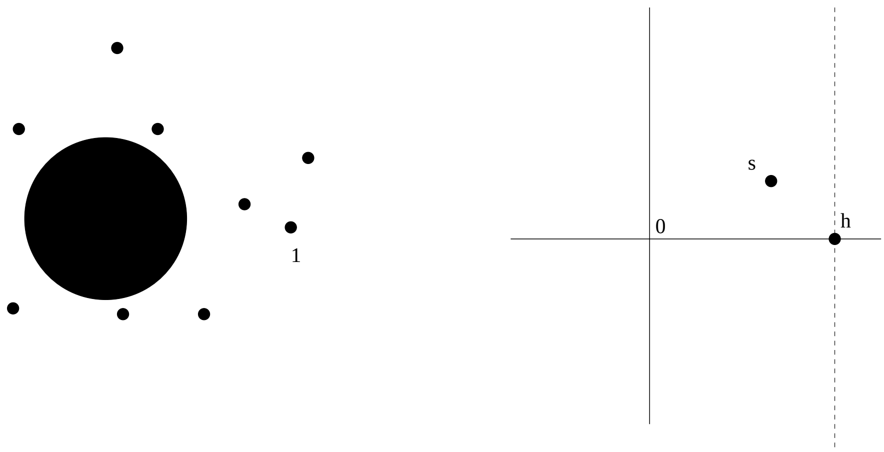

Figure 6. (i) Spectra of $\mathcal{L}_{-sr}$ contains 1; implies (ii) s is a pole for $\zeta(s)$.

Theorem (relating spectrum to ζ-function). *$\zeta(s)$ has a meromorphic extension to the domain* $\{s : \theta e^{P(-Re(s))} < 1\}$. *Moreover, s is a pole for $\zeta(s)$ in this region if and only if* 1 *is an isolated eigenvalue for* $\mathcal{L}_{-sr} : C^1(I) \to C^1(I)$.

This was proved by Hofbauer-Keller for interval maps (1984) and for subshifts by Pollicott (1984).

A little retroactive motivation. Instead of transfer operators we might consider a simpler finite rank operator: a $k \times k$ (probability) matrix:

$$\mathcal{M} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix} \text{ and } \mathcal{M}_s = \begin{pmatrix} p_{11}^s & p_{12}^s & \cdots & p_{1k}^s \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}^s & p_{k2}^s & \cdots & p_{kk}^s \end{pmatrix} \text{ for } s \in \mathbb{C}.$$

The analogue of the ζ-function would be

$$\begin{aligned} \zeta(s) &= \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} \sum_{i_0,\cdots,i_{n-1}} (p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_0})^s\right) \\ &= \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} \mathrm{tr}(\mathcal{M}_s^n)\right) = 1/\det(I - \mathcal{M}_s). \end{aligned}$$

In particular, in this case we have the analogue of the last Theorem:

(a) $\zeta(s)$ has a non-zero meromorphic extension to $\mathbb{C}$; and
(b) s is a pole iff $\mathcal{M}_s$ has 1 as an eigenvalue.

When it comes to the nitty-gritty of applying transfer operators to ζ-functions, the connection between transfer operators and zeta functions can be made in several ways. One of the nicest is the following (due to Ruelle): For each interval I_i we fix a choice $x_i \in I_i$ We want to connect $\mathcal{L}^n_{-sr}$ to $Z_n(s)$ by:

Technical Lemma. *There exists* $0 < \theta < 1$ *such that We can expand*

$$Z_n(s) = (\mathcal{L}^n_{-sr}\chi_{I_i})(x_i)(1 + O(n\theta^n))$$

This particular version is due to Ruelle (1990). Previous versions were due to Ruelle (1978) and Hayden (1990).

3.4 A special case: Constant curvature. If we were in the constant curvature case then we can look at the smaller class of analytic functions then we get a better result.

Theorem (Ruelle, 1976). *On analytic functions* $\mathcal{L}_{-sr} : C^\omega(I) \to C^\omega(I)$ *is a nuclear (trace class) operator.*

If $U \supset I$ is a fixed open neighbourhood in $\mathbb{C}$ then we can understand $C^\omega(I) = \{w : I \to \mathbb{C} \text{ is analytic }\}$ with norm $||w|| = \sup_{z\in\overline{U}} |w(z)|$. In particular, we can bypass the technical lemma and one *almost* gets $Z_n(s)$ as the trace of $\mathcal{L}^n_{-sr}$:

Theorem (Ruelle, 1976). *For* $n \geq 1$,

$$\mathrm{tr}(\mathcal{L}^n_{-sr}) = \sum_{f^n x = x} \frac{\exp\left(-sr^n(x)\right)}{1 - \frac{1}{|(f^n)'(x)|}}.$$

However, this isn't so bad, since it makes a direct connection with the Selberg ζ-function:

$$Z(s) = \prod_{n=0}^{\infty} \prod_{\gamma} \left(1 - e^{-(s+n)\lambda(\gamma)}\right) \qquad \textit{(Selberg } \zeta\textit{-function)}$$

In particular:

Corollary. *We can identify*

$$Z(s) = \det(I - \mathcal{L}_{-sr}) \left(:= \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} \mathrm{trace}(\mathcal{L}^n_{-sr}) \right) \right)$$

3.5 Dolgopyat's results. Finally, the most remarkable result in recent years for the case transfer operators associated to surfaces of (variable) negative curvature is the following:

Theorem (Dolgopyat, 1998). *There exists* $\epsilon' > 0$ *such that for* $s = \sigma + it$ *with* $|t| \geq 1$ *and* $\sigma > h - \epsilon'$ *then the spectral radius of* $\mathcal{L}_{-sr}$ *strictly smaller than* 1.

In particular, comparing with the Theorem (relating spectrum to ζ-function) this shows that:

(vi) $\zeta(s)$ always has a non-zero *analytic* extension to a larger half-plane $Re(s) > h - \epsilon$, except for the simple pole at $s = h$.

Remark on dimensions $n \geq 3$. For higher dimensions Dolgopyat's original proof only holds providing there is some pinching condition of the curvature (for example, if the sectional curvatures are all close enough to -1, say). Recently, Liverani (2003) has removed this extra condition in relation to decay of correlations (cf. next lecture) and one *expects* (vi) to hold without the pinching condition.

Appendix III. Word length:. The length of a geodesic represents the geometry, but the topology is better represented by the *word length*. One chooses a finite set of generators $\gamma_1, \dots, \gamma_k$ for the fundamental group. The word length of a curve in the fundamental group is the smallest number of generators needed to represent it. A closed geodesic γ corresponds to a conjugacy class in the fundamental group and its word length $|\gamma|$ is the shortest over all elements of the conjugacy class.

In the symbolic dynamics is often corresponds to the n, i.e., the period of the point. The simpler analogue of Margulis' theorem is the following: $\exists h_0 > 0$ such that

$$\begin{aligned} \mathrm{Card}\{\gamma : |\gamma| = n\} &\left(= \frac{1}{n}\mathrm{Card}\{x \in I : f^n x = x\}\right) \\ &\sim \frac{e^{h_0 n}}{n} \text{ as } n \to +\infty. \end{aligned}$$

Somewhere between these two different types of asymptotic results one can show the following:

Theorem (Pollicott-Sharp, 2003). *Let M be a compact surface of negative curvature. There exists $C > 0$ such that provided $\epsilon_n \searrow 0$ sufficiently slowly we have that*

$$\mathrm{Card}\{(\gamma, \gamma') : |\gamma|, |\gamma'| \leq n, 0 \leq \lambda(\gamma) - \lambda(\gamma') \leq \epsilon_n\} \sim C\epsilon_n \frac{e^{2h_0}}{n^{5/2}}, \ \textit{as } n \to \infty.$$

These results are in the tradition of earlier work by Steiner, et al.

Lecture 4: Mixing and resonances

4.1 Ergodicity. The natural invariant measure on SM is the *Liouville measure* μ. This is an invariant probability measure which is equivalent to the volume, i.e., if $B \subset SM$ is a ball then $\mu(\phi_t B) = \mu(B)$.

Theorem (Moser, 1966). *By changing the metric by a coordinate change (e.g., applying a diffeomorphism) we can assume that any given smooth measure is the volume.*

The measure μ is *ergodic* if for almost every point x (with respect to μ) the time averages equal the spatial averages, i.e., for any continuous function $F : SM \to \mathbb{R}$

$$\frac{1}{T}\int_0^T F(\phi_t v)dt \to \int_{SM} F(\xi)d\mu(\xi), \text{ as } T \to \infty \ (\textit{Birkhoff's Theorem})$$

for almost every vector $v \in SM$.

Theorem (Hopf, 1939). *For surfaces of negative curvature the geodesic flow is ergodic*

The idea of the proof is so good that it is still one of the mains tools for establishing ergodicity of other systems:

(1) For almost all vectors the forward and backward averages exist and are equal:

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T F(\phi_t v)dt \;=\lim_{T\to-\infty}\frac{1}{T}\int_0^T F(\phi_t v)dt$$

(but might conceivably vary with v ...)

(2) If v_1, v_2 lie on the same stable manifold (i.e., $d(\phi_t v_1, \phi_t v_2) \to 0$ as $t \to \infty$) then

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T F(\phi_t v_1) = \lim_{T\to+\infty}\frac{1}{T}\int_0^T F(\phi_t v_2)dt.$$

Similarly, if v_2, v_3 lie on the same unstable manifold (i.e., $d(\phi_t v_2, \phi_t v_3) \to 0$ as $t \to -\infty$) then

$$\lim_{T\to-\infty}\frac{1}{T}\int_0^T F(\phi_t v_2) = \lim_{T\to-\infty}\frac{1}{T}\int_0^T F(\phi_t v_3)dt.$$

(3) For every pair $v, w \in SM$ we can find a sequence of points $v = v_0, v_1, v_2, \ldots$, $v_k = w$ such that v_{2i}, v_{2i+1} lie on the same stable manifold and v_{2i+1}, v_{2i+2} lie on the same unstable manifold. Moreover, since the foliations by stable and unstable manifolds are absolutely continuous (i.e., no sets of non-zero measure are lost under holonomy along them) we deduce that for almost all (with respect to μ) $v, w \in SM$ the limits in Birkhoff's Theorem are equal, and therefore the flow is ergodic.

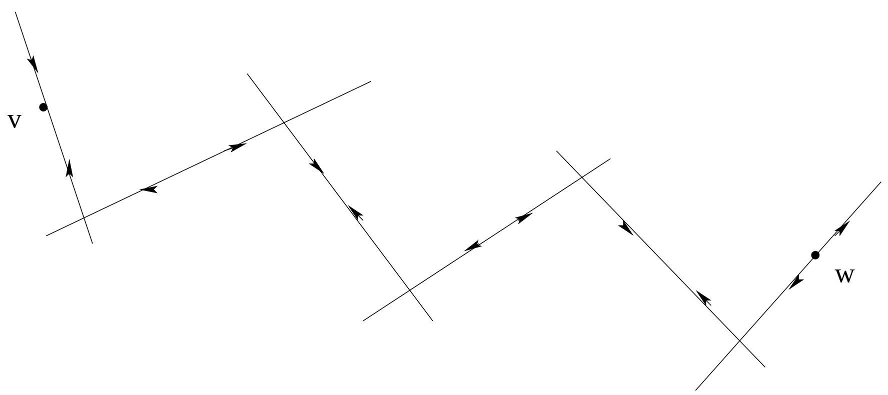

Figure 7. Points $v, w \in SM$ are joined by paths made up of pieces of stable and unstable manifolds

Remark. The usual metrics on the 2-sphere or $\mathbb{T}^2$ are certainly not ergodic, for example. However, lightly suprisingly, there are some metrics on the 2-sphere for which the geodesic flow is ergodic (Donnay, Burns-Gerber).

4.2 Mixing. We consider mixing ("decay of correlations") for the measure μ by introducing C^∞ test functions $F, G : SM \to \mathbb{R}$ and defining, for each $t \in \mathbb{R}$,

$$\rho(t) = \int F(\phi_t v).G(v)d\mu(v) - \int F(v)d\mu(v) \int G(v)d\mu(v).$$

We say that ϕ is *(strong) mixing* (with respect to μ) if $\rho(t) \to 0$ as $t \to \infty$.

Proposition (Ornstein-Weiss, Ratner; 1973&1974). *Geodesic flows are mixing (with respect to μ).*[15]

Every mixing flow is automatically ergodic, so this improves on Hopf's result.

4.3 Exponential Mixing. We say that the geodesic flow $\phi_t : SM \to SM$ *mixes exponentially* (with respect to μ) if there exists $C > 0$ and $\delta > 0$ such that $|\rho(t)| \leq Ce^{-\delta|t|}$.

Theorem (Dolgopyat, 1998). *For surfaces with negative curvature one has that the geodesic flow $\phi_t : SM \to SM$ mixes exponentially (with repect to μ).*

For constant curvature surfaces these results were know using representation theory (e.g., Moore (1986)). At exactly the same time as Dolgopyat proved his result, Chernov proved the slightly weaker stretched exponential mixing (i.e. $|\rho(t)| = O(e^{-C\sqrt{n}})$) by a very different method.

4.4 Consequences of faster mixing. One application of the exponential mixing is that it leads to error estimates for the Birkhoff ergodic theorem:

Theorem (Error terms in Birkhoff theorem). *Assume that $F : SM \to \mathbb{R}$ is a C^∞ function. Then, for almost every $v \in SM$ (with respect to μ) one has that for every $\delta > 0$:*

$$\frac{1}{T}\int_0^T F(\phi_t v)dt = \int_{SM} F(\xi)d\mu(\xi) + O\left(T^{-1/2+\epsilon}\right)$$

An even stronger result than ergodicity is that the following estimate holds.

Theorem (Central Limit Theorem, Ratner, 1969+1973). *Let $F : SM \to \mathbb{R}$ be a C^∞ function.*[16] *There exists $\sigma = \sigma(F)$ such that for each $y \in \mathbb{R}$:*

$$\lim_{T\to+\infty} \mu\left\{v \in SM : \frac{1}{\sqrt{T}}\left[\int_0^T F(\phi_t v)dt - T\int_{SM} F(\xi)d\mu(\xi)\right] < y\right\}$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^y e^{-\sigma^2 t}dt.$$

This was later improved by Denker and Philipp (1984) and is one of a heirachy of statistical properties one can show for geodesic flows and other hyperbolic systems.

Although it is not necessary to have exponential mixing to show the Central Limit Theorem, it allows the variance σ^2 to be identified with $\sigma^2 := \int_{-\infty}^{\infty} \rho(t)dt$.

[15] In fact, these authors prove the stronger result that the flows are Bernoulli

[16] Strictly speaking, we should assume we cannot write $F(v) = \frac{d}{dt}U(\phi_t)$, for some $U : SM \to \mathbb{R}$.

4.5 Fourier transforms and transfer operators (again). For variable curvature, a natural way to show this property for $\rho(t)$ is to use the Fourier transform:

$$\widehat{\rho}(s) = \int_{-\infty}^{\infty} e^{-ist}\rho(t)dt, \quad s \in \mathbb{R}.$$

In particular, one wants to apply the following classical result.

Lemma (Paley-Wiener, 1933). [17] *Assume that there exists $\delta > 0$ such that :*

(i) *$\widehat{\rho}(s)$ has an analytic extension from $\mathbb{R}$ to the strip $|Im(s)| < \delta$; and*

(ii) *the function $\mathbb{R} \ni \sigma \mapsto \widehat{\rho}(\sigma + it)$ is L^1 for $|t| < \delta$.*

Then $\rho(t)$ tends to zero exponentially (i.e., $\exists C, \delta > 0$ such that $|\rho(t)| \leq Ce^{-\delta t}$).

Using symbolic dynamics we can again reduce this to a situation where we can apply transfer operators (again).

The key to understanding the Fourier transform is to relate it to a transfer operators appropriate to this setting. In this case, if we choose the weight functions $g = -\log|f'|$. Then we have a new family of transfer operators

$$\mathcal{L}_{-\log|f'|} : C^1(I) \to C^1(I)$$
$$\mathcal{L}_{-\log|f'|}w(x) = \sum_{y\,:\,fy=x} \frac{w(y)}{|f'(y)|}.$$

The results on the Ruelle Operator Theorem from before imply:

Theorem (Lasota-Yorke,1973 [or Corollary of Ruelle Operator Theorem]). [18] *The operator $\mathcal{L}_{-\log|f'|}$ has 1 as a maximal eigenvalue. The corresponding eigenfunction $\mathcal{L}_{-\log|f'|}w = w$ is positive. The rest of the spectrum is in a strictly smaller disk.*

A very natural question to ask here is: *What was wrong with the earlier transfer operator?* The short answer is nothing: except that if we use that operator it is more appropriate if we replace the Liouville measure μ by another invariant measure: the Bowen-Margulis measure μ_{BM} (described in the next lecture).

Another useful idea from symbolic dynamics. Both $\rho(t)$ and $\widehat{\rho}(s)$ can be studied "symbolically" in terms of the semi-flow $\psi_t : \Omega \to \Omega$ under the suspension of $r : I \to \mathbb{R}^+$ over $f : I \to I$.

In this case, we choose the weight functions $g = -\log|f'| - isr$, $s \in \mathbb{C}$. Then we have a new family of transfer operators

$$\mathcal{L}_{-\log|f'|-isr} : C^1(I) \to C^1(I)$$
$$\mathcal{L}_{-\log|f'|-isr}w(x) = \sum_{y\,:\,fy=x} \frac{w(y)e^{-isr(y)}}{|f'(y)|},$$

for each $s \in \mathbb{C}$.

[17] Sadly, Paley died in a skiing accident in 1933 at the age of 26.

[18] This is a well known result. In particular, it can be used to show that $d\nu(x) = w(x)dx$ is an absolutely continuous invariant measure for $f : I \to I$

Corollary (of Theorem on spectrum). *The spectrum of operator* $\mathcal{L}_{-\log|f'|-isr} : C^1(I) \to C^1(I)$ *consists of:*

(1) *isolated eigenvalues in the annulus* $\{z \in \mathbb{C} : \theta e^{P(-\log|f'|+Im(s)r)} < |z| \leq e^{P(-\log|f'|+Im(s)r)}\}$*; and*
(2) *the rest contained in the ball* $\{z \in \mathbb{C} : e^{P(-\log|f'|+Im(s)r)}\}$.

As for the ζ-function, we can relate the extension of $\widehat{\rho}(s)$ to the spectrum of such transfer operators. Consider the analogous problem for the symbolic semi-flow $\psi_t : \Omega \to \Omega$ with functions $F, G : \Omega \to \mathbb{R}$ and a measure $d\nu \times du$. A rearrangement of integrals gives, for example:

$$\int_0^\infty e^{-ist}\left(\int_\Omega (F \circ \psi_t)(w,u)G(w,u)d\nu(w)du\right)dt$$
$$= \sum_{n=0}^\infty \int f_s(w)\left[\mathcal{L}^n_{-\log|f'|-isr}g_{-s}\right](w)d\nu(w),$$

where $f_s(w) = \int_0^{r(w)} e^{is}F(w,u)du$. The general case involves more complicated manipulations - but the outcome is similar. In this particular case, we get the following:

Theorem (relating spectrum to $\widehat{\rho}(s)$). *The Fourier transform* $\widehat{\rho}(s)$ *has a meromorphic extension to a domain* $\{s \in \mathbb{C} : |Im(s)| < \delta\}$[19] *Moreover,* $\pm s$ *is a pole for* $\widehat{\rho}(s)$ *in this strip if and only if* 1 *is an isolated eigenvalue for (either)* $\mathcal{L}_{-\log|f'|\pm isr} : C^1(I) \to C^1(I)$.

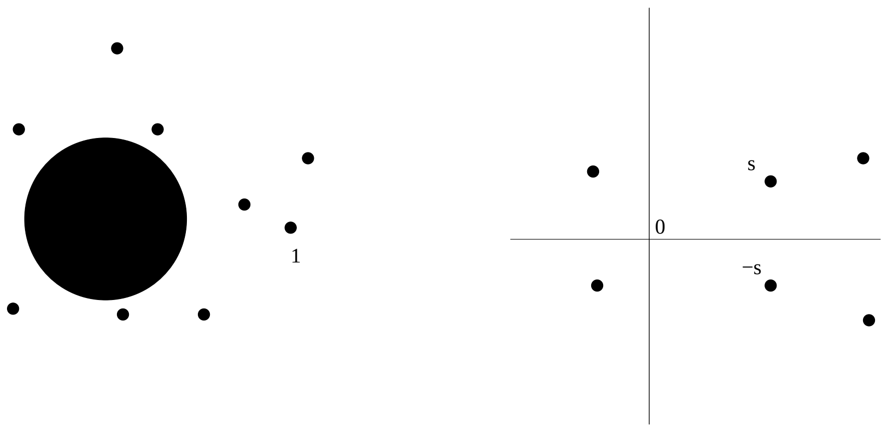

FIGURE 8. (I) SPECTRA OF $\mathcal{L}_{-\log|T'|\pm sr}$ CONTAINS 1; IMPLIES (II) $\pm s$ IS A POLE FOR $\widehat{\rho}(s)$.

We could now get rid of the transfer operators and use their common occurrence for $\widehat{\rho}(s)$ and (suitable) ζ-functions to get the following:

Theorem (Pollicott-Ruelle, 1986&1987). *Let us define a new zeta function*

$$\zeta_u(\xi) = \prod_\tau (1 - \exp(-\lambda^u(\tau) - \xi\lambda(\tau)))^{-1}, \xi \in \mathbb{C},$$

[19] where $\delta > 0$ satisfies $\theta e^{P(-\log|f'|+\delta r)} = 1$

where $\lambda^u(\tau)$ measures the expansion in the unstable direction for τ. There is a correspondence between the poles ξ_j for the ζ-function $\zeta_u(\xi)$ and the poles $s_j = \pm\xi_j$ for the Fourier transform $\widehat{\rho}(s)$.

Finally, in this context Dolgopyat's results on the spectrum of transfer operators take the following form.

Theorem (Dolgopyat, 1998). *There exists $\delta' > 0$ such that for $s = \sigma + it$ with $|\sigma| \geq 1$ and $0 \leq t < \delta'$ then the spectral radius of $\mathcal{L}_{-\log|f'|-isr}$ strictly smaller than 1.*

This can be used to show that $\widehat{\rho}(s)$ has an *analytic* extension to a strip $\{s \in \mathbb{C} : |Im(s)| < \delta'\}$. In particular, with a little more work, the Paley-Wiener theorem applies to show the promised exponential mixing.

Appendix IVa): Stable ergodicity. A modern take on ergodicity is *stable ergodicity.* Consider another flow $\psi_t : SM \to SM$ which is close to the geodesic flow $\phi_t : SM \to SM$. The following result is well known:

Theorem (Structural stability: Anosov & Moser 1967&1969). *Provided that a C^∞ flow ψ is sufficiently close to ϕ (in the C^1 topology) then*

(1) *the orbits of ϕ and ψ correspond (by a homeomorphism).*
(2) *if ψ preserves μ then it too must be ergodic.*

A more difficult problem is to consider analogous problems for the discrete time-one map $f := \phi|_{t=1} : SM \to SM$. This map also preserves the measure μ. Consider a nearby diffeomorphism $g : SM \to SM$ that also preserves the measure μ.

Theorem (Grayson-Pugh-Shub, 1994). *Provided that a C^∞ diffeomorphism g is sufficiently close to f (in the C^2 topology) then it too must be ergodic.*

Appendix IVb) Dependence of poles. One might reasonably ask how the poles change with the metric. In this direction we have the following:

Proposition. *The poles s_j change smoothly as the metric on M changes.*

This is not so difficult to show (Pollicott, 2003). An approach for certain Anosov diffeomorphisms which doesn't use symbolic dynamics but only establishes continuity of the rates of mixing (poles) is due to Blank-Keller-Liverani.

Lecture 5: Related flows (Frame flows, Horocycle flows, etc.)

5.1 Horocycle flows on compact surfaces. A second flow associated to a compact negatively curved surface is the horocycle flow $h_t : SM \to SM$. This is defined by moving vectors v along the unstable manifolds (horocycles) they lie on and gives a perfectly good flow[20] However, they have somewhat different features. For example:

Lemma. *Let M be a compact surface of negative curvature.*

(1) *There are no closed orbits for the horocycle flow.*
(2) *Every orbit of the horocycle flow is dense in SM.*

[20] Modulo deciding on the parameterization.

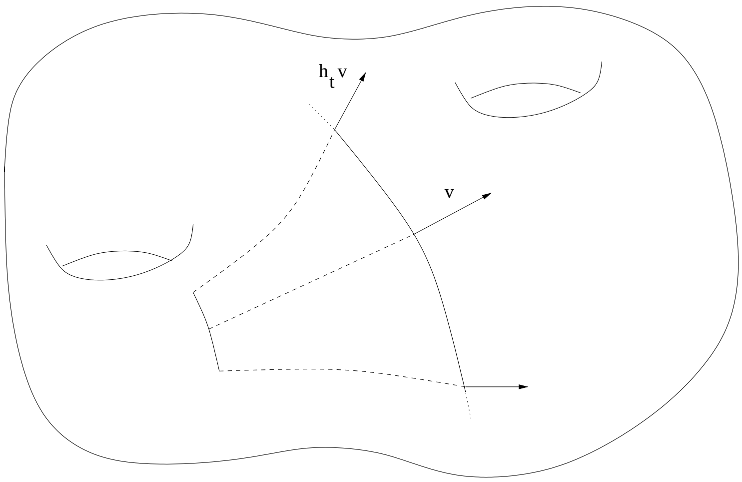

FIGURE 9. THE HOROCYCLE FLOW ON A SURFACE.

Whereas as geodesic flows have many different invariant (probability) measures, horocycle flows have precisely one (i.e, they are uniquely ergodic). This requires an appropriate parameterization of the flow h_t.

Theorem (Hedlund-Marcus, 1934 & 1987). *The horocycle flow on a compact surface of negative curvature has precisely one invariant probability measure* μ_{BM}*.*

In the case of non-compact surfaces the situation can be a little different. For example, the modular surface has closed horocycles which circle the cusp. Each of these can support an invariant measure.

In constant curvature, the measure μ_{BM} is the usual Liouville measure μ. However, in general it is not and is called the *Bowen-Margulis measure.* Interestingly, this measue has a simple characterization in terms of closed orbits for the geodesic flow.

Theorem (Bowen, 1972). *Let* μ_τ *be the invariant probability measure supported on the closed orbit* τ *for the geodesic flow* $\phi_t : SM \to SM$*. Then* μ_{BM} *is the unique (invariant) probability measure* SM *such that*

$$\frac{\sum_{\tau\,:\,\lambda(\tau)\le T}\int F d\mu_\tau}{\mathrm{Card}\{\tau\,:\,\lambda(\tau)\le T\}} \to \int F d\mu_{BM}, \ \ \textit{as } T \to +\infty,$$

for any continuous function $F : SM \to \mathbb{R}$*.*

This can be improved to say that increasing proportions of closed orbits are distributed according to μ_{BM} (i.e., large deviation results of Kifer (1990)).

5.2 Horocycle flows on infinite area surfaces. A particularly surprising extension of these results occurs when one looks at a (non-compact) $\mathbb{Z}^d$-cover $\widehat{M}$ of the surface M. Let $\widehat{h}_t : S\widehat{M} \to S\widehat{M}$ be the associated horocycle flow.

Theorem (Babillot-Ledrappier, 1996). [21] *The horocycle flow on the* $\mathbb{Z}^d$ *cover* $\widehat{M}$ *of a compact surface* M *of negative curvature is ergodic with respect to the lift*

[21] Sadly, Martine Babillot died earlier this year.

$\widehat{\mu}_{BM}$ of the measure μ_{BM}. However, it is not uniquely ergodic (i.e., there exist other non-equivalent invariant measures).

Here ergodicity means that for any $\widehat{h}_t$-invariant (Borel) set $B \subset SM$ we have that $\widehat{\mu}_{BM}(B) = 0$ or $\widehat{\mu}_{BM}(\widehat{SM} - B) = 0$

Their original method was more geometric and used the relationship (in constant curvature) between the geodesic flow ϕ_t and horocycle flow h_T given by $\phi_{-t} h_T \phi_t = h_{Te^{-ht}}$. There are other proofs by Coudene, Kaimanovich-Schmidt, etc.

Babillot and Ledrappier actually constructed large families of different invariant measures. Recently, Sarig (preprint) showed that these were essentially all of the invariant measures.

In higher dimensions one can recast all of these questions in terms of transverse measures for the horocyclic foliations - in place of the horocycle flow.

5.3 Linear actions and distribution of orbits. One particularly nice way in which to recast ergodicity of the horocycle flow for constant curvature surfaces in terms of simple linear actions of 2×2 matrices on the plane $\mathbb{R}^2$.

Assume that we write the surface as $\mathbb{H}^2/\Gamma$, where $\Gamma \subset SL(2,\mathbb{R})$. One can consider the simplest possible "linear action" of matrices $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ given by

$$\begin{aligned} g &: \mathbb{R}^2 \mapsto \mathbb{R}^2 \\ g &: (x,y) \mapsto (ax+by, cx+dy) \end{aligned}$$

In this case, ergodicity has a very simple interpretation. If we have any (Borel) set $B \subset \mathbb{R}^2$ which is invariant under the linear action of Γ (i.e., $\Gamma B := \cup_{g\in\Gamma} gB$) then either B or $\mathbb{R}^2 - B$ must have zero Lebesgue measure.

Theorem (Babillot-Ledrappier, 1996 &1997). *Let $\Gamma \subset SL(2,\mathbb{R})$ be a cocompact subgroup (i.e., $SL(2,\mathbb{R})/\Gamma$ is compact).*

(i) *The action Γ is ergodic;*
(ii) *For any $x = (x_1, x_2) \in \mathbb{R}^2 - \{(0,0)\}$ orbits of Γ are distributed according to:*

$$\frac{1}{T}\sum_{||g||\leq T} F(gx) \to \frac{1}{||x||}\int_{\mathbb{R}^2} \frac{F(y)}{||y||} dy \text{ as } T \to +\infty,$$

where $F : \mathbb{R}^2 - \{(0,0)\} \to \mathbb{R}$ is compactly supported[22] *and $||g|| = \sqrt{|a|^2+|b|^2+|c|^2+|d|^2}$*

The connection between these two results is that we can identify the space of horocycles with

$$SL(2,\mathbb{R})/\left\{\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}\right\} = \mathbb{R}^2,$$

and the proof involves formalizing this with sections, etc.

5.4 Frame flows. We can consider the geodesic flow on any n-dimensional manifold with negative sectional curvatures. Furthermore, given any unit tangent vector $v_1 \in SM$ one can also choose $(n-1)$ other orthonormal vectors (over the same point

[22] Actually, we also need to assume $-I \in \Gamma$

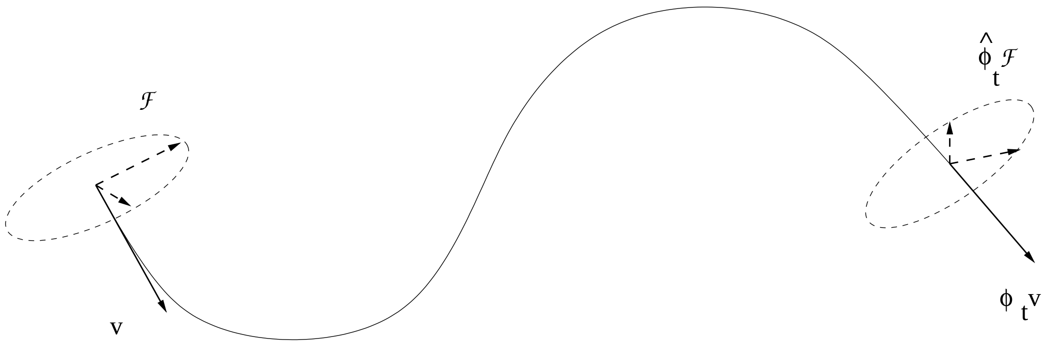

Figure 10. The frame flow on three dimensional manifold.

in M) to give a *frame* $\mathcal{F} = (v_1, v_2, \ldots, v_n)$. Using the geodesic flow $\phi_t : SM \to SM$ we can move the first vector v_1 to $\phi_t(v_1)$. By parallel transporting the other orthonormal vectors along this geodesic we get another frame F' (the first vector of which is $\phi_t(v_1)$).

Let $\mathcal{F}M$ denote the (oriented) n-frames on M. We define the frame flow $\widehat{\phi}_t$: $\mathcal{F}M \to \mathcal{F}M$ by $\widehat{\phi}_t(\mathcal{F}) = \mathcal{F}'$. There is a natural invariant (probability) measure on $\mathcal{F}M$ given by $\widehat{\mu} = \mu \times \lambda_{SO(n-1)}$, where $\lambda_{SO(n-1)}$ is the Haar measure on (rotations of) individual frames.

As one might hope:

Theorem. *If M is a compact manifold with constant sectional curvatures -1 then the frame flow $\widehat{\phi}_t : \mathcal{F}M \to \mathcal{F}M$ is ergodic, i.e.,*

$$\frac{1}{T}\int_0^T F(\widehat{\phi}_t\mathcal{F})dt \to \int \widehat{F} d\widehat{\mu}(\mathcal{F}), \text{ as } T \to \infty,$$

for almost every ($\widehat{\mu}$) frame $\mathcal{F}$, where $F : \mathcal{F}M \to \mathbb{R}$ is a continuous function.

Unfortunately, in contrast to the case for geodesic flows it is *not* the case that frame flows on any manifold with negative sectional curvatures is ergodic:

Proposition. *There exist compact manifolds*[23] *with negative sectional curvatures, for which the associated frame flow $\widehat{\phi}_t : \mathcal{F}M \to \mathcal{F}M$ is not ergodic.*

Let us assume that the sectional curvatures are bounded by $-\Lambda \leq \kappa_{ij}(x) \leq -\lambda < 0$. The situation is *far* from satisfactory, but the best known positive results are the following:

Theorem (Ergodicity of frame flows). *The frame flow is known to be ergodic under any of the following conditions:*

(1) *if n is odd, but not equal to 7 (Brin-Gromov);*
(2) *$n = 7$ and $\lambda/\Lambda > 0.999714\ldots$ (Burns-Pollicott);*
(3) *if n is even, but not equal to 8, and $\lambda/\Lambda > 0.93$ (Brin-Karcher); or*
(4) *$n = 8$ and $\lambda/\Lambda > 0.999785\ldots$ (Burns-Pollicott).*

In fact, the frame flow is also stably ergodic (in the sense described before).

[23] Even dimensional examples which are quotients of complex hyperbolic space $\mathbb{CH}^n$

5.5 Frame flows and stable horocycles flows. A natural question about the linear actions is what happens for, say, complex matrices? Consider a discrete subgroup $\Gamma \subset SL(2,\mathbb{C})$. The linear action of matrices $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ given by

$$g : \mathbb{C}^2 \mapsto \mathbb{C}^2$$
$$g : (x,y) \mapsto (ax+by, cx+dy)$$

The analogue of the result on $\mathbb{R}^2$ is now:

Theorem (Ledrappier-Pollicott). *Let $\Gamma \subset SL(2,\mathbb{C})$ be a cocompact subgroup (i.e., $SL(2,\mathbb{C})/\Gamma$ is compact).*

(i) *The action Γ is ergodic;*

(ii) *For any $x = (x_1, x_2) \in \mathbb{C}^2$ orbits of Γ are distributed according to:*

$$\frac{1}{T^2} \sum_{||g|| \leq T} F(gx) \to \frac{1}{||x||^2} \int_{\mathbb{C}^2} \frac{F(y)}{||y||^2} dy \text{ as } T \to +\infty,$$

where $F : \mathbb{C}^2 - \{(0,0)\} \to \mathbb{R}$ is compactly supported.[24]

The proof uses stable manifolds (horocycles) for frame flows. Let us denote

$$\mathbb{H}^3 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \, : \, x_3 > 0\}$$

equipped with the Poincaré metric $ds^2 = (dx_1^2 + dx_2^2 + dx_3^2)/x_3^2$. The discrete groups Γ is identified with discrete subgroups of orientation preserving isometries $\mathrm{Isom}(\mathbb{H}^3)$ of $\mathbb{H}^{n+1}$ and we can associate to the group Γ the manifold $V = \mathbb{H}^{n+1}/\Gamma$. In the same way that geodesic flows have stable horocycles, frame flows too have stable manifolds the space of which is identified with

$$SL(2,\mathbb{C}) / \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\} = \mathbb{C}^2.$$

References

1. Baladi, *Positive transfer operators and decay of correlations*, World Scientific, Singapore, 2000.

   [This book has a particularly comprehensive list of references]

2. T. Bedford, M. Keane and C. Series, *Ergodic Theory, Symbolic dynamics and Hyperbolic geometry*, O.U.P., Oxford, 1989.

   [A collection of surveys and a good place to find references for the coding of geodesics. The icing on the cake is the beautiful cover drawing.]

3. M. Berger, *A panorama of Riemannian Geometry*, Springer, Berlin, 2003.

   [This is a very readable (and often witty) account of geometry.]

4. M. Brin, *Ergodic theory of frame flows*, Ergodic theory and dynamical systems, II, Progr. Math., 21,, Birkhuser, Boston, Mass., 1982, pp. 163–183,.

   [This is probably remains the best introduction to frame flows.]

5. Dolgopyat, *On decay of correlations in Anosov flows*, Annals of Math. (1998).

   [This remarkable paper solved the long standing problem of showing exponential decay of correlations for suitable geodesic flows.]

[24] We should assume that $-I \in \Gamma$.

6. A. Kachurovskii, *Rates of converence in ergodic theorems*, Russ. Math. Surv. **51** (1996), 5653-703.

[An interesting article which is perhaps not as well known as it deserves to be.]

7. A. Katok and B. Hasselblatt, *An introduction tio the modern theory of dynamical systems*, C.U.P., Cambridge, 1995.

[This is a very good reference for the Moser, Hopf and structural stability results. Moreover, the appendix of Katok-Mendoza has a nice account of the machinery of Katok's result. It was originally to be an independent book, but Mendoza gave up mathematics to become a venezualan businessman. It had existed as an unpublished manuscript since the mid-1980s.]

8. G. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, Fun. Anal. Appl **3** (1969), 89-90.

[The original proof of his theorem by Margulis was never widely circulated and only recently translated (and will be published soon by Springer, along with an update by Sharp).]

9. Mayer, *On a ζ function related to the continued fraction transformation*, Bull. Soc. Math. France **104** (1976), 195-203.

[A very nice concrete setting to start if one wants to understand the Ruelle-Grothendeick approach to zeta functions. This paper concentrates on the case of the Gauss map.]

10. W. Parry and M. Pollicott, *Zeta functions and the closed orbit structure of hyperbolic flows*, Asterisque **187-188** (1990), 1-268.

[An early (and slightly dated) reference for the zeta function stuff. It actually began as a Master's course at Warwick in Spring of 1984. Originally due to be given by Parry, it was actually give by Pollicott.]

11. M. Pollicott and R. Sharp, *Exponential error terms for growth functions on negatively curved surfaces*, Amer. J. Math **120** (1998), 1019-1042.

[There are very few articles containing independent accounts of Dolgopyat's ideas. This is one]

12. Ratner, *The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature*, Israel J. Math. **16** (1973), 181-197.
13. D. Ruelle, *Resonances for Axiom A flows*, J. Diff. Geom. **25** (1987), 99-116.
14. D. Ruelle, *Dynamical zeta functions and transfer operators*, Notices of the Amer. Math. Soc. **49** (2002), 887-895.

[A amusing general article. Worth reading just for the anecdote about Parry and Pollicott.]

15. M. Spivak, *An introduction to Differential Geometry*, Volumes 1,2,3,4 and 5, Publish or Perish, Wilmington, 1975.

[A long and comprehensive classic.]

Department of Mathematics, Manchester University, Oxford Road, Manchester, M13 9PL, UK.