

ERGODICITY OF THE GEODESIC FLOW ON NON-COMPLETE NEGATIVELY CURVED SURFACES*

MARK POLLICOTT[†] AND HOWIE WEISS[‡]

Key words. Ergodicity, geodesic flow, singular surface of revolution, Weil-Petersson metric, moduli space.

AMS subject classifications. 53D25, 53D25, 37D50

1. Introduction. We study the dynamics of the geodesic flow for a class of non-complete Riemannian metrics on a negatively curved surface M . These finite area surfaces are composed of finitely many “singular” surfaces of revolutions of the form $y = x^r$, $r > 1$ for $0 \leq x \leq 1$ (thin pieces), together with connecting surfaces of bounded negative curvature (thick pieces). The curvature at every point is negative and bounded away from zero, but is unbounded in the “cusps.” The geodesic flow is non-complete because geodesics corresponding to singular unit tangent vectors pointing into the origin hit the cusps in finite time and then cease to be defined. Such singular unit tangent vectors are dense in the unit tangent bundle. Our main result is ergodicity of the geodesic flow (Theorem 5.1). It immediately follows that these metrics have dense geodesics in the unit tangent bundle SM . We also prove that the closed geodesics are dense in the unit tangent bundle (Theorem 6.1)

The motivation for considering these metrics arises from studying the geodesic flow for the Weil-Petersson (WP) metric on two dimensional moduli spaces of Riemann surfaces., e.g., the moduli space for the once punctured torus. In several fundamental ways the geodesic flow we consider provides a good model for the WP geodesic flow. For example, Wolpert showed that the WP metric is also non-complete, has finite area, has negative curvature bounded away from zero, and is unbounded in the “cusps.” Also, various authors have shown that the cusps can be approximated by singular surface of revolutions. At present, there is only a single technical obstruction (additional estimates on the derivatives of the GF in the cusp) that prevent us from applying this general method to prove ergodicity of the WP geodesic flow. See Section 7.

The geometric properties of both types of flows imply that the geodesic flow is a uniformly hyperbolic dynamical system with singularities. The non-completeness causes pathologies in the stable and unstable manifolds, as for many billiard flows. For example, stable and unstable manifolds at a point, if they exist, may intersect a cusp point, and thus have only finite length. One needs extensions of “Pesin theory” to systems with singularities to study these geodesic flows. Another challenging aspect is that these surfaces may be simply connected, making the use of any of the traditional arguments involving boundaries of the covering spaces inapplicable. Studying the dynamics requires the development of new approaches.

Our strategy to prove ergodicity of the geodesic flow is to first establish non-uniform hyperbolicity, i.e., the Lyapunov exponents are non-zero at almost every

*Received August 4, 2009; accepted for publication October 7, 2009.

[†]Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom (mpollic@maths.warwick.ac.uk).

[‡]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA (weiss@math.gatech.edu).

point. We then prove that there are at most countably many ergodic components. The next step is to establish local ergodicity, i.e., to show that each ergodic component is an open set (modulo sets of measure zero). We use the axiomatic treatment of smooth ergodic theory with singularities presented in [8, 6, 2]. We then prove that the geodesic flow is epsilon-topologically transitive, i.e., for each $\epsilon > 0$ there is an ϵ -dense orbit in the unit tangent bundle. Ergodicity immediately follows.

2. The Geometry of a Singular Surface of Revolution. In this section we collect several key facts about the geometry of surfaces of revolution [4].

Let R denote the singular surface of revolution of the form $y = f(x) = x^r$, $r > 1$ for $0 \leq x \leq 1$. We introduce coordinates (u, v) on R where u is the distance along the axis of revolution from the cusp and $0 \leq v < 2\pi$ is the angle of rotation along a parallel. In particular, points on the surface of revolution can be written as $(u^r \cos v, u^r \sin v, u)$. The parallel circles correspond to $u = u_0$ and the meridian curves correspond to $v = v_0$.

The induced Riemannian metric on R can be written in as

$$g = (1 + f'(u)^2)du^2 + f(u)^2 dv^2 = (1 + r^2 u^{2r-2})du^2 + u^{2r} dv^2.$$

It follows that the area form on R has the form

$$dA(u) = 2\pi f(u) \sqrt{1 + f'(u)^2} du, \quad (1)$$

and hence R has finite area, since

$$\text{Area}(R) = 2\pi \int_0^1 f(u) \sqrt{1 + f'(u)^2} du = 2\pi \int_0^1 u^r \sqrt{1 + r^2 u^{2r-2}} du < \infty.$$

The Gaussian curvature can be expressed as

$$K(u) = -\frac{f''(u)}{(1 + f'(u)^2)^2} = -\frac{r(r-1)}{u^2(1 + (r^2 u^{r-1})^2)^2}. \quad (2)$$

Clearly, $K(u) \rightarrow -\infty$ as $u \rightarrow 0$. The total curvature of R is bounded, since,

$$-\int_0^1 K(u) dA(u) \leq -2\pi \int_0^1 \frac{f''(u)}{(1 + f'(u)^2)^{3/2}} dx = 2\pi \frac{r}{\sqrt{1 + r^2}} < \infty. \quad (3)$$

The equations of geodesics in (u, v) coordinates have a simple form:

$$f(u)^2 v' = c \quad (4)$$

$$(1 + f'(u)^2)u'^2 + c^2/f(u)^2 = 1, \quad (5)$$

where c is a constant, or first integral of the geodesic flow.

If t denotes the time parametrization of the geodesic $t \mapsto (u(t), v(t))$, then integrating (4) and (5) yield

$$t = \int_{u(0)}^{u(t)} f(u) \sqrt{\frac{1 + f'(u)}{f(u)^2 - c^2}} du \quad (6)$$

and

$$v(t_1) - v(t_0) = \int_{t_0}^{t_1} \frac{c}{f(u(t))^2} dt. \tag{7}$$

Dividing (4) by (5) and integrating yields

$$v = \int \frac{c}{f(u)} \sqrt{\frac{1 + f'(u)^2}{f(u)^2 - c^2}} du. \tag{8}$$

It follows from (4) and (5) that if $\theta(u_1)$ denotes the angle that the geodesic makes with the parallel circle at $u = u_1$, then the quantity

$$u_1 \sin \theta(u_1) \tag{9}$$

is a constant. This is the famous Clairaut integral.

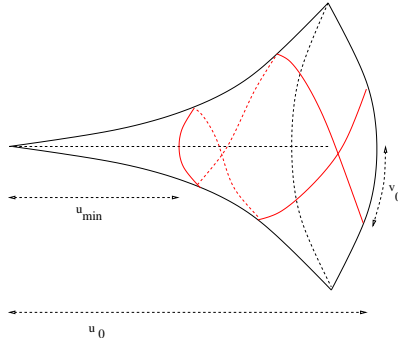


FIG. 1. A geodesic segment in the cusp

3. Local ergodicity.

3.1. Hyperbolicity. Let M denote a finite area surface isometric to the union of finitely many “singular” surfaces of revolutions of the form $y = x^r, r > 1$ for $0 \leq x \leq 1$, together with connecting surfaces of bounded negative curvature. We call a geodesic on M *complete* if it does not hit a cusp (which corresponds to the point $x = 0$ on the surfaces of revolution), and *singular* if reaches a cusp in finite time. We call a unit tangent vector complete or singular if the corresponding geodesic is complete or singular, respectively.

Let SM denote the unit tangent bundle, which is undefined at the finite set of cusp points. Let $\phi : SM \rightarrow SM$ denote the geodesic flow, which is undefined on the singular unit tangent vectors. Let $X \subset SM$ be the set of complete unit tangent vectors. This set is a flow invariant and dense Baire set with full (normalized) Liouville measure μ_L . Since we are considering ergodicity, the geodesic flow only needs to be defined on a set of full Liouville measure, and thus the measure zero set where the geodesic flow is not defined plays no measure theoretic role.

We say that a flow is *Anosov* (as an incomplete flow) if it is uniformly hyperbolic on the set of complete tangent vectors, i.e., there is a splitting of the unit tangent bundle $S_X M = E^0 \oplus E^s \oplus E^u$ and constants $C, \lambda > 0$ such that $\|D\phi_t|E^s\| \leq Ce^{-\lambda t}$, $\|D\phi_{-t}|E^s\| \leq Ce^{-\lambda t}$ (for $t \geq 0$) and E^0 is tangent to the flow direction.

THEOREM 3.1. *The geodesic flow $\phi : SM \rightarrow SM$ is Anosov (as an incomplete flow).*

Proof. The geodesic flow preserves Louville measure, since the set of singular tangent vectors has zero measure. This is easily seen since the tangent vectors pointing towards the cusp describe two dimensional subspaces in the three dimensional manifolds.

We modify the usual Jacobi field argument for complete surfaces and construct two uniform cone families that yield an Anosov splitting for this incomplete flow. The negative curvature ensures that the standard cone family $\mathcal{C}^u = \{J \cdot J'\} \geq 0$ is uniformly contracting, since if the curvature $K \leq -b^2 < 0$, then along a complete geodesic

$$\frac{d}{dt}(J(t)J'(t)) = J(t)J''(t) + J'(t)^2 = -k(t)J(t)^2 + J'(t)^2 \geq b^2J(t)^2 + J'(t)^2 > 0,$$

unless $J \equiv 0$. It follows that

$$\frac{d}{dt}(J(t)J'(t)) \geq b^2J(t)^2 + J'(t)^2 > 2b(J(t)J'(t)).$$

Since $J(0)J'(0) > 0$ for any nontrivial $J \in \mathcal{C}^u$, it follows that $J(t)J'(t)$ grows exponentially, and hence $J(t)^2 + J'(t)^2$ grows exponentially, since $J^2(t) + J'(t)^2 \geq 2J(t)J'(t)$. The unstable direction at v corresponds to the intersection $\cap_{t \geq 0} D\phi_t \mathcal{C}^u(\phi_{-t}v)$ of the images $D\phi_t \mathcal{C}^u(\phi_{-t}v)$ of the cones $\mathcal{C}^u(\phi_{-t}v)$ above $\phi_{-t}v$. The same holds for the stable cone $\mathcal{C}^s = \{J \cdot J'\} \leq 0$. \square

The following corollary stating that the geodesic flow is non-uniformly hyperbolic is immediate.

COROLLARY 3.2. *The geodesic flow has non-zero Lyapunov exponents almost everywhere.*

3.2. Countably many ergodic components. In light of the need to restrict the dynamics to the finite area, non-compact set X , we cannot directly apply the theory of Anosov systems on compact manifolds, and thus need to apply ideas from the theory of hyperbolic systems with singularities. There is a well developed theory for hyperbolic dynamics for with singularities, described in [6] and extended in [8] which provides an axiomatic approach to establishing countably many ergodic components and local ergodicity. Local ergodicity means that each ergodic component is contained in an open set (modulo a set of measure zero). Together, they would imply the existence of countably many open ergodic components.

We begin by verifying conditions 1.1–1.3 on pages 2–3 in [6] required to construct local stable and unstable manifolds almost everywhere, establish absolute continuity, establish the existence of at most countably many ergodic components, and establish a Pesin entropy formula.

PROPOSITION 3.3. *Let $\phi : SM \rightarrow SM$ denote the geodesic flow on the unit tangent bundle of M . The following properties hold.*

1. *(Small area in cusp hypothesis) Let U_ϵ be the ϵ -neighborhood of one of the cusps in M . There exists a positive constant C_1 such that $\text{Area}(U_\epsilon) \leq C_1\epsilon^{r+1}$.*

2. (Oseledec-Pesin integrability condition) The integral

$$\int_{SM} \log \|d\phi_1(v)\| d\mu_L(v) < \infty.$$

3. (Controlled growth of the differential of the geodesic flow in the cusps condition) If an orbit segment $\phi[0, 1](v)$ is at least distance ϵ from a cusp, then

$$\|D\phi_{t=1}(v)\| = O(\epsilon^{-(2r+1)}) \text{ and } \|D^2\phi_{t=1}(v)\| = O(\epsilon^{-(2r+2)}).$$

REMARK 3.4. It immediately follows from 1. that M and thus X have finite area.

Proof.

1. Let $B(\epsilon)$ denote an ϵ -neighbourhood of the cusp in the surface of revolution. A direct integration yields that

$$\text{Area}(B(\epsilon)) = 2\pi \int_0^\epsilon x^r \sqrt{1 + rx^{r-1}} dx = \frac{2\pi}{r+1} \epsilon^{r+1} + O(\epsilon^{2r}).$$

□

2. It follows from Lemma 5.1 of [1] that

$$\int_{SM} \log \|d\phi_1(v)\| d\mu_L(v) \leq \frac{1}{2} \mu_L(SM) - \pi \int_M K(p) dA(p).$$

Thus we need to verify that the integral on the right hand side is finite.

Let $\epsilon > 0$ and define $M_\epsilon = \{p \in M : K(p) \geq -1/\epsilon\}$. We can write

$$-\int_M K(p) dA(p) = -\int_{M_\epsilon} K(p) dA(p) - \int_{M \setminus M_\epsilon} K(p) dA(p).$$

It follows from the compactness of M_ϵ that the second integral on the right hand side is finite. An easy calculation using (1) and (2) gives that

$$-\int_{M_\epsilon} K(p) dA(p) = O(\epsilon^{r-1}),$$

and thus the first integral on the right hand side is finite. □

3. Clearly, for segments of the geodesic arcs that are outside of the cusp, the first and second derivatives are bounded by compactness. Let us assume that a geodesic enters the cusp on a reference meridian (corresponding to $u = u_0$, say) with a corresponding angular coordinate $u = u_0$.

We want to consider a subsequent point $(u, v) = (u(t, u_0), v(t, v_0))$ along the geodesic in the cusp. It is easy to see from the definitions that $\frac{\partial u}{\partial v_0} = 0$ and $\frac{\partial v}{\partial v_0} = 1$.

Using (6), we can characterize $u = u(t, u_0)$ implicitly by $F(u, u_0) = 0$, where

$$F(u, u_0) = t - \int_{u_0}^u u^r \sqrt{\frac{1 + ru^{r-1}}{u^{2r} - c^2}} du$$

where $c = u_{\min}^r$, with $u_{\min} \geq \epsilon$ denoting smallest value of u in the cusp. We can apply the implicit function theorem to this equation to write that

$$\frac{\partial u}{\partial u_0} = -\frac{\frac{\partial F}{\partial u_0}}{\frac{\partial F}{\partial u}} = \left(\frac{u_0}{u}\right)^r \sqrt{\frac{(1+r^2u_0^{2r-2})(u^{2r}-c^2)}{(1+r^2u^{2r-2})(u_0^{2r}-c^2)}}.$$

This expression is clearly bounded.

Using (5), we can write

$$v = v(1, v_0) = \int_0^1 \frac{c}{u(t, u_0)^{2r}} dt$$

and thus

$$\frac{\partial v}{\partial u_0} = \int_0^1 \frac{c}{2ru(t, u_0)^{2r+1}} \frac{\partial u}{\partial u_0} dt = O(\epsilon^{-(2r+1)}).$$

One can similarly treat the second derivatives. One easily obtains that

$$\frac{\partial^2 v}{\partial v_0^2} = \frac{\partial^2 v}{\partial v_0 \partial u_0} = \frac{\partial^2 u}{\partial v_0^2} = \frac{\partial^2 u}{\partial v_0 \partial u_0} = 0$$

and it is easy to check that

$$\frac{\partial^2 v}{\partial v_0^2} = O(\epsilon^{-(2r+2)}) \text{ and } \frac{\partial^2 u}{\partial u_0^2} = O(\epsilon^{-(2r+1)}).$$

The foliations by local stable and unstable manifolds are absolutely continuous if there exists sets X_ϵ of measure at least $(1-\epsilon)$ upon which the stable manifolds W^s have a length bounded below by a uniform positive constant, and for nearby transverse sections to this lamination the holonomy map between them is absolutely continuous with respect to the induced measure on the sections. Absolute continuity is required to invoke the results from Part II of [6]. This notion of absolute continuity is similar to the usual notion of absolute continuity for complete Anosov flows, except for some complications arising from local stable and unstable manifolds only existing at almost every point and not having uniform length.

Combining Corollary 2.2, Proposition 2.3, and results in [6], we obtain the following result.

PROPOSITION 3.5. *Let $\phi : SM \rightarrow SM$ denote the geodesic flow on the unit tangent bundle of a surface M composed of finitely many surfaces of revolutions of the form $y = x^r, r > 1$ for $0 \leq x \leq 1$ together with connecting surfaces of bounded negative curvature.*

1. *There exist local stable and unstable manifolds at almost every point.*
2. *The local stable and unstable manifolds are absolute continuous.*
3. *The geodesic flow has countably many ergodic components.*
4. *The Liouville entropy of the geodesic flow satisfies*

$$h_L = \int_{SM} \chi^+(v) d\mu_L(v) > 0,$$

where $\chi^+(v)$ denotes the positive Lyapunov exponent at $v \in X$.

It is useful to have more control on the lengths of stable and unstable manifolds are almost all point, and this is given by the following lemma.

LEMMA 3.6. *There exists a measurable function $\epsilon : M \rightarrow \mathbb{R}^+$ so that almost every point x has a stable manifold $W^s(x)$ of length $\epsilon(x) > 0$.*

Proof. The construction of local stable manifolds is based on a series of approximations [8], [6]. Since many geometers are likely to be unaware of the strategy of the proof, we provide a short sketch.

Let $W_1^s(x)$ be a small submanifold through $x \in X$ in the general direction of the stable bundle, and define $W_n^s(x) = \phi_n W_1^s(\phi_{-n}x)$. For complete hyperbolic flows, the sequence $\{W_n^s(x)\}$ converges C^∞ to the required stable manifold through x . However, in the present case, this construction formally breaks down whenever the forward orbit of any point $y \in W_n^s(\phi_{-n}x)$ meets a cusp point. However, since the stable manifolds converge exponentially quickly we can modify the standard argument to give pieces of local stable manifold at almost every point in X .

Let $B(\epsilon)$ denote an ϵ -neighbourhood of the cusp. By property (1) in Proposition 3.3, the area $\text{Area}(B(\epsilon)) = O(\epsilon^{r+1})$. If we let $\epsilon_n = 1/n$, then since the flow preserves the Liouville measure μ , we have that $\mu(\phi_{-t}B(\epsilon_n)) = \mu(B(\epsilon_n)) = O(1/n^{r+1})$, for every $t \in \mathbb{R}$. In particular, we have that

$$\sum_n \mu(B(\epsilon_n)) = \sum_n \mu(\phi_{-t_n}B(\epsilon_n)) < \infty,$$

for the sequence of times $t_n \rightarrow \infty$ which are nearest approaches made by the orbit to the cusp, each time it enters the region based on the surface of revolution.

By the Borel-Cantelli Lemma we see that for a.e. $(\mu) x$, $\phi_{t_n}x \in B(\epsilon_n)$ for at most finitely many t_n . It is clear from the geometry that there exists $D > 0$ such that $t_n \geq Dn$, for all $n \geq 1$. Since the lengths of the unstable manifolds contract at an exponential rate we see that for such x we can indeed construct an unstable manifold (since the exponential convergence of the orbits keeps nearby points away from the cusp too). \square

3.3. Local ergodicity. The next step is to establish local ergodicity, i.e., to show that each ergodic component is an open set (modulo sets of measure zero). The proof of local ergodicity follows the lines of the traditional Hopf argument, with the help of a particularly simple form of the Chernov-Sinai Ansatz to deal with the singularity set. We employ the axiomatic framework stated in Section 7 of [8] and [2].

The basic idea in the Hopf argument is that any two points that lie on the same stable manifold must necessarily share the same forward Birkhoff averages for any continuous function. Similarly, any two points that lie on the same unstable manifold must necessarily share the same backward Birkhoff averages. Thus, the ergodic averages will agree at almost all points which can be connected by paths consisting of pieces of stable and unstable manifolds. Since the local foliations are absolutely continuous, one would like to deduce that the points thus connected by such paths (called Hopf chains) share the same ergodic averages. In particular, this would show that the ergodic components are open sets (modulo sets of zero measure).

In the present context, an additional complication arises from those geodesics starting or ending at a cusp. Let $S_- \subset SM$ denote the set of geodesics which start at a cusp (and have finite length 1, say) and let $S_+ \subset SM$ denote the set of geodesics ending at a cusp (and have finite length 1, say). The sets $S_-^N = \cup_{k=1}^n \phi_k S_-$ and $S_+^N = \cup_{k=1}^n \phi_n S_+$ are finite unions of two dimensional sheets in SM . The sets $S_-^\infty =$

$\bigcup_{k=0}^{\infty} \phi_n S_-$ and $S_+^{\infty} = \bigcup_{n=0}^{\infty} \phi_{-n} S_+$ denote the set of all geodesics which start or finish (respectively) in a cusp. Each of these have zero Liouville measure, and the negative curvature ensures that both are dense in SM . These observations cover hypothesis A and B in Section 7 of [8]. The uniformly contracting and expanding cone families constructed in the proof of Theorem 3.1 insure that hypotheses C (Monotonicity) and E (Noncontraction) are satisfied. Because of the hyperbolicity, $\phi_n S_-$ becomes more closely aligned with the unstable manifolds (in the sense that they are contained in cones that typically converge) and similarly $\phi_n S_+$ becomes more closely aligned with the stable manifolds. This is hypothesis D, Proper Alignment of the Singularity Set.

Hypothesis E is the analog of the Chernov-Sinai Ansatz. The double singularity set $S^{\infty} = S_-^{\infty} \cap S_+^{\infty}$ consists of geodesic arcs which start and end at a cusp, and has zero induced measure on S_- and S_+ . Since the geodesic flow is uniformly hyperbolic, for every unit tangent vector $v \in S_-^{\infty} \setminus S^{\infty}$ (vectors which never hits a cusp in positive time), the norm of the differential of the geodesic flow at v grows exponentially, and thus becomes infinite, i.e., $\|D\phi^t v\| \rightarrow \infty$ as $t \rightarrow \infty$. This is the Chernov-Sinai Ansatz.

The following proposition follows from the Main Theorem on page 39 of [8].

PROPOSITION 3.7. *Let $\phi : SM \rightarrow SM$ denote the geodesic flow on the unit tangent bundle of a surface M composed of finitely many surfaces of revolutions of the form $y = x^r, r > 1$ for $0 \leq x \leq 1$ together with connecting surfaces of bounded negative curvature. Then the geodesic flow is locally ergodic.*

REMARK 3.8. The underlying idea in the Chernov-Sinai Ansatz is that provided $N > 0$ is sufficiently large, the set of local unstable manifolds that intersect S_-^N has an arbitrarily large proportion of the measure of X . Similarly, provided $N > 0$ is sufficiently large, the set of local stable manifolds that intersect S_+^N has an arbitrarily large proportion of the measure of X . One can then consider small parallelepipeds constructed from local stable manifolds, unstable manifolds and flow segments disjoint from $S_-^N \cup S_+^N$ (cf. Remark 3.6 and Theorem 9.7 in [8]). Within this region, a standard argument (originally due to Sinai) shows that there are sufficiently many long local stable and unstable manifolds to apply a variant of the Hopf argument (cf. pp. 14-17 and §11 in [8]).

4. ϵ -Topological transitivity. In this section we prove that the geodesic flow is ϵ -topologically transitive. It will follow from ergodicity of the geodesic flow that the geodesic flow is topologically transitive.

PROPOSITION 4.1. *For each $\epsilon > 0$ one can find an ϵ -dense orbit segment in SM .*

Proof. We begin with the following simple lemma. For each point $x \in M$ which is not a cusp point, we let $\mathcal{B}_x \subset S_x \subset SM$ be the set of unit tangent vectors $v \in SM$ for which the geodesics originating at x eventually hit a point cusp.

LEMMA 4.2. $\mathcal{B}_x \subset S_x M$ is dense.

Proof. This is a consequence of the negative curvature. Consider an arc $U \subset S_x$ and assume for a contradiction that $U \cap \mathcal{B}_x = \emptyset$. The image $\phi_t U$ under the geodesic flow grows in length exponentially and sweeps out an area $\phi_{[0,t]} U$, say, which eventually covers all of M . However, this must include the cusp points, giving a contradiction. \square

It is a consequence of the Clairaut Integral (9) that on a singular surface of revolution, the only geodesics which reach the cusp point are meridian curves. Any

other geodesic will have a nearest approach c , say, to the cusp. From (8) we observe that twice the integral from 1 to c gives the total angle that the geodesic arc makes around the cusp between successive transitions across the meridian corresponding to $u = 1$. It is easy to see that this diverges as c approaches 0, i.e, the number of times the geodesic winds around the cusp tends to infinity as c tends to zero.

Given $\epsilon > 0$, we now construct an ϵ -dense orbit segment by a (finite) iterative procedure.

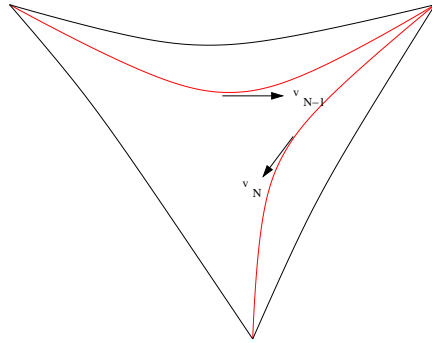


FIG. 2. Singular geodesics

Step 1. We consider the dense set $\mathcal{B} = \cup_{x \in X} \mathcal{B}_x$ in the unit tangent bundle SM . By considering pairs of vectors in \mathcal{B}_x with the same base points, but with vectors $v_1, v_2 \in \mathcal{B}_x$ pointing in approximately opposite directions, we can associate a geodesic curve passing arbitrarily to v_1 , hitting a cusp point both in the future and past. (We can consider the two piece geodesic segment associated to v_1 and v_2 , and contract this to a genuine geodesic. If we encounter a cusp point during this process we can simply replace the endpoint by this one). Moreover, we can assume that these new unit tangent vectors $\{v_n\}_{n=1}^N \subset SM$ are a $\epsilon/2$ -dense set.

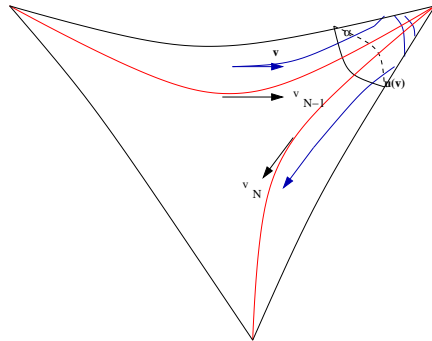


FIG. 3. Perturbing the vector v_N

Step 2. We can consider first the vector v_N , and assume for simplicity that the geodesic for the vector $v_{N-1} \in \mathcal{B}$ ends in the same cusp as v_N starts. The associated geodesic segments for v_N and v_{N-1} cross a fixed meridian for their common cusp

perpendicularly (one entering the other leaving). We want to change the vector v_{N-1} continuously through the geodesics which start from the same cusp (i.e., a geodesic spray). More precisely, as v tends to v_{N-1} , the angle $\alpha(v)$ that the associated geodesic makes with the meridian (which is less than $\frac{\pi}{2}$) tends to zero. Of course, because of the geometry of surfaces of the revolution the geodesic associated to v now no longer ends at the cusp point, but emerges from the cusp, crossing the meridian circle at the same angle. Let us denote by $u(v) \in [0, 2\pi)$ the position on the meridian at which the geodesic associated to v reemerges.

LEMMA 4.3 (Sub-lemma 1). *As $\alpha(v)$ tends to zero, then the natural lift of $u(v)$ tends to infinity. In particular, given any $0 \leq u_0 < 2\pi$ we have a sequence of vectors $\{v_{N-1}^{(n)}\}$ tending to v_{N-1} , such that the associated geodesics re-emerge from the cusp at the point $u(v_{N-1}^{(n)}) = u_0$.*

Proof. We observe that if the meridian circle is given by $u = 1$, say, then

$$2c \int_c^1 \frac{\sqrt{1 + (pv^{p-1})} dv}{v^p \sqrt{v^{2p} - c^2}}$$

gives the total angle that the geodesic arc makes around the cusp between successive transitions across the meridian circle corresponding to $v = 1$. It is easy to see that this integral diverges as c approaches 0, i.e, the number of times the geodesic winds around the cusp tends to infinity as c tends to zero. This in turn follows if $\alpha(v) \rightarrow 0$, as assumed. \square

We want to choose $v'_{N-1} = v_{N-1}^{(n)}$, for n sufficiently large. Indeed, since the geodesic associated to v'_{N-1} emerges at precisely the point where the geodesic from v_N meets the meridian, we deduce that the two geodesics are arbitrarily close providing their respective angles, $u(v'_{N-1})$ and $\pi/2$ are close. In particular, by the sub-lemma, we can choose n sufficiently large that there is a vector v'_N on the geodesic through v'_{N-1} which is $\epsilon/2$ -close to v_N .

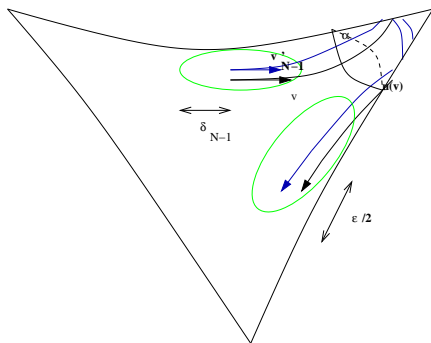


FIG. 4. *Choosing N sufficiently large*

By uniform continuity of the geodesic flow (in a suitably small neighbourhood of the geodesic arc associated to v'_{N-1}) we see that if we choose $\delta_{N-1} < \frac{\epsilon}{2}$ sufficiently small, then for any vector with $v \in B(v_{N-1}, \delta_{N-1})$ the associated geodesic passes within $\epsilon/2$ of v'_N and thus, by the triangle inequality, within ϵ of v_N .

Step 3. We can next consider the vector v_{N-2} , and again assume for simplicity that the geodesic for the vector $v_{N-2} \in \mathcal{B}$ ends in the same cusp as v_{N-1} (and v'_{N-1} starts).

The associated geodesic segments for v_{N-2} and v'_{N-1} cross a fixed meridian for their common cusp perpendicularly (the first entering the second leaving). As before we can change v_{N-2} continuously through the geodesics which start from the same cusp (i.e., a geodesic spray). As v tends to v_{N-2} , the angle $\alpha(v)$ that the associated geodesic makes with the meridian (which is less than $\frac{\pi}{2}$) again tends to zero. We have the following version of Sublemma 1.

LEMMA 4.4 (Sub-lemma 2). *Given any arc $U \subset [0, 2\pi)$ we have a sequence of intervals $\{I_{N-2}^{(n)}\}$ in $[0, 2\pi)$ tending to the entry point on the meridian of v_{N-1} , such that the associated geodesics re-emerge from the cusp in the arc U .*

Proof. As observed in the proof of the previous sublemma, the number of times the geodesic winds around the cusp tends to infinity as c tends to zero thus by continuity we can choose suitable intervals I_n . \square

We want to let U correspond to those geodesics which cross the meridian perpendicularly and then pass through $B(v'_{N-1}, \delta_{N-1})$. We next want to choose v'_{N-2} lying on a geodesic which starts on the same cusp point as v_{N-2} and meets the meridian at $I_{N-2}^{(n)}$, for n sufficiently large. The geodesic associated to v'_{N-2} then emerges from the cusp and meets the neighbourhood $B(v'_{N-1}, \delta_{N-1})$. By uniform continuity of the geodesic flow in a neighbourhood of the geodesic through v'_{N-2} we can choose $0 < \delta_{N-2} < \frac{\epsilon}{2}$ sufficiently small that for any $v \in B(v'_{N-2}, \delta_{N-2})$ the associated forward geodesic passes through $B(v'_{N-1}, \delta_{N-1})$ (and thus by the triangle inequality within a distance ϵ of v_{N-1}) and then, by construction, later it passes within ϵ of v_N .

Step 4. We proceed inductively. Assume that we have constructed vectors $v'_k, v'_{k+1}, \dots, v'_N$ and a sequence $0 < \delta_k, \delta_{k+1}, \dots, \delta_N < \frac{\epsilon}{2}$.

We can next consider the vector v_{k-1} , and again assume for simplicity that the geodesic for the vector $v_{k-1} \in \mathcal{B}$ ends in the same cusp as v_k (and v'_k) starts. The associated geodesic segments for v_{k-1} and v'_k cross a fixed meridian for their common cusp perpendicularly (the first entering the second leaving). As before we can change v_{k-1} continuously through the geodesics which start from the same cusp (the geodesic spray). As v tends to v_{N-2} , the angle $\alpha(v)$ that the associated geodesic makes with the meridian again tends to zero.

We want to apply Sublemma 2 again. We want to choose v'_{k-1} lying on a geodesic which starts on the same cusp point as v_{k-1} and meets the meridian in an arc $I_{k-1}^{(n)}$, for n sufficiently large. The geodesic associated to v'_{k-1} emerges from the cusp and meets the neighbourhood $B(v'_k, \delta_{N-1})$. By uniform continuity of the geodesic flow in a neighbourhood of the geodesic through v'_{k-1} we can choose $\delta_{k-1} > 0$ sufficiently small that for any $v \in B(v'_{k-1}, \delta_{k-1})$ the associated forward geodesic passes through $B(v'_k, \delta_k)$. Then by construction, the geodesic passes through $B(v'_{k+1}, \delta_{k+1}), \dots, B(v'_{N-1}, \delta_{N-1})$ and within ϵ of v_N .

This completes the proof. We need only observe that the hypothesis that v_n enters the cusp which v_{n+1} leaves can be achieved by suitably organizing (and repeating if necessary) the choice of vectors. \square

5. Ergodicity. Combining the results of the last two sections gives the following result.

THEOREM 5.1. *The geodesic flow is ergodic.*

Proof. We need to show there is a single ergodic component by showing that for every $\epsilon > 0$ there is an ϵ -dense orbit (segment). Thus if we were to assume for a

contradiction there were two distinct ergodic components then we could choose open sets U and V in each and an orbit segment connecting U to V . In particular, for some $t \in \mathbb{R}$ we have that $\phi_t U \cap V \neq \emptyset$, and since this intersection has non-zero Liouville measure and the ergodic components are invariant, we are done. \square

COROLLARY 5.2. *The geodesic flow is topologically transitive.*

6. More topological properties. We claim the following result holds for closed geodesics on the surface M .

THEOREM 6.1. *The union of closed geodesics are dense on M (and their tangent vectors are dense in SM).*

Proof. Given $v \in S_1M$ and $\epsilon > 0$ sufficiently small we can use the Poincaré recurrence to find $T > 0$ such that $d(\phi_T v, v) < \epsilon$. Let $\pi : SM \rightarrow M$ be the canonical projection from the unit tangent bundle onto the surface M . We begin by joining the projections $\pi(v), \pi(\phi_T v) \in M$ of $v, \phi_T v$ onto M by a short geodesic segment. Moreover, we can join the projections $\pi(\phi_t v), \pi(\phi_{T-t} v) \in M$ of the tangent vectors $\phi_t v, \phi_{T-t} v$ by a continuously changing geodesic arc $\gamma_t^{(1)}$, for $0 \leq t \leq T/2$. In particular, the geodesic arc $\gamma^{(1)} = \gamma_{T/2}^{(1)}$ is based at the point $x^{(1)} = \gamma(T/2)$. If the geodesic arc $\gamma^{(1)}$ has length $T^{(1)}$ then we can write $\gamma^{(1)}(0) = \gamma^{(1)}(T^{(1)}) = x^{(1)}$.

We shall use a method which is a version of the standard Birkhoff curve shortening method. We next consider a sequence of continuously varying geodesics $\gamma_t^{(2)}$ joining $\gamma^{(1)}(t)$ and $\gamma^{(1)}(T^{(1)} - t)$, for $0 \leq t \leq T^{(1)}/2$. In particular, $\gamma^{(2)} = \gamma_{T^{(1)}/2}^{(2)}$ is a closed geodesic based at the point $x^{(2)} = \gamma^{(1)}(T^{(1)}/2)$. Let us denote the length of this curve by $\gamma^{(2)}$.

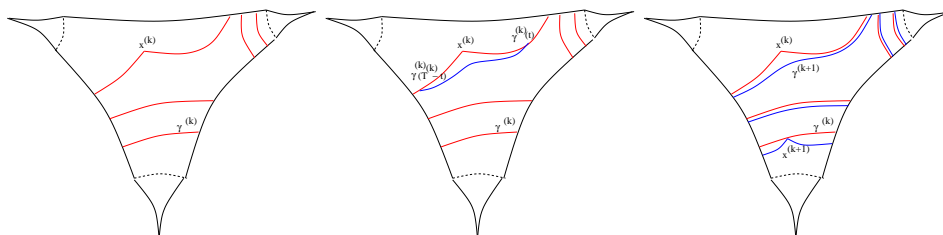


FIG. 5. *Piecewise Geodesic Smoothing*

Continuing inductively gives a sequence $(\gamma_n)_{n=1}^\infty$ of geodesics arcs in the same free homotopy curve. Moreover, the curves converge to a single curve γ_∞ . Furthermore, the angle between the tangents $\gamma_n'(0)$ and $\gamma_n'(T^{(n)})$, at either end of the geodesic arcs, converge to zero. Thus the limiting curve γ_∞ is a closed geodesic.

The only point that that it remains to check in this article is that none of the curves passes through a cusp point. However, within the cusp region (corresponding to the surface of revolution) the only way for a geodesic to approach closer to a cusp point is for the curve to wind an increasing number of times around the cusp, as we see from the Clairaut results. However, this contradicts that the curves (given by composing the curves $\gamma^{(k)}$ and $\gamma_t^{(k-1)}$, between $\gamma^{(k-1)}(t)$ and $\gamma^{(k-1)}(T^{(k-1)} - t)$) lie in the same free homotopy class. \square

However, in contrast to the situation for complete surfaces we have the following:

THEOREM 6.2. *There exist (infinitely many distinct) values $L > 0$ such that we can sequences of lengths of closed geodesics whose lengths accumulate to L .*

Proof. Given any closed geodesic on M we can increase the number of times the geodesic spirals around a given cusp. A simple computation shows that the limit of the (distinct) lengths of these geodesics converges. \square

Let $\pi(T)$ denote the number of closed geodesics of least period less than T .

COROLLARY 6.3. *Then there exists $L > 0$ such that $\pi(L) = \infty$.*

REMARK 6.4. We could attempt to define the closed geodesic entropy as the upper bound on the rate of growth of closed geodesics

$$h_{per} = \limsup_{T \rightarrow \infty} \frac{1}{T} \log \pi(T).$$

However, we immediately see this is infinite by the previous result.

We could also attempt define the volume entropy as upper bound on the growth rate of the volume of a ball $B(x_0, R)$ on the universal cover for $M - \{\text{cusps}\}$. Equivalently, we let $\text{Vol}(B(x_0, R))$ be the volume counted with multiplicity for the over laps. We then define

$$h_{vol} = \limsup_{T \rightarrow \infty} \frac{1}{T} \log \pi(T).$$

As an application of the above ideas, we again observe that the volume entropy is infinite.

7. Application to Weil-Petersson geodesic flows. Wolpert showed that the WP metric is also non-complete, has finite area, has negative curvature bounded away from zero, and is unbounded in the “cusps [11].” Thus most of the preceding analysis also applies to the geodesic flow for the Weil-Petersson (WP) metric on two dimensional moduli spaces of Riemann surfaces. These include the moduli spaces for the once punctured torus $M_{1,1}$ and sphere with four punctures $M_{0,4}$. We now argue that with the exception of the estimates in Proposition 3.3 (c), all of the stated results carry over. Thus, the only ingredient missing from the proof of ergodicity of the geodesic flow in these cases is the verification of the estimates on the derivatives of the time one map in the cusp.

Since the WP metric has negative curvature bounded away from zero, the proof of Theorem 3.1 shows that the WP geodesic flows are Anosov (as incomplete flows) and thus have non-zero Lyapunov exponents almost everywhere.

The authors of [3, 12, 13] have shown that in a C^0 sense, the WP metric in a cusps coincides with the induced metric on a singular surfaces of revolution of the form $y = x^r$, up to an error which approaches zero as the cusp point is approached. The analog of Proposition 3.3 (1) follows immediately.

To establish the analog of Proposition 3.3 (2) for the moduli space for the once-punctured torus, we estimate the total curvature in the cusp using the following special coordinates. A once-punctured torus can be uniformized by a Fuchsian group Γ generated by hyperbolic elements $A, B \in PSL(2, \mathbb{R})$ with $ABA^{-1}B^{-1}$ parabolic [7]. With standard normalizations, the Teichmüller space for the once punctured torus can be identified with the locus $x^2 + y^2 + z^2 = 3xyz$ satisfying $x, y, z > 2$, where $x = \text{tr}(A)$ and $y = \text{tr}(B)$, and $z = \text{tr}(AB)$. Letting $a = x/yz, b = y/xz$ and

$z = z/xy$, the Teichmüller space is the simplex $a + b + c = 1$ with $a, b, c > 0$. We can write $x = x(a, b, c)$, etc. The Moduli space corresponds to choosing $a, b, c < 1/2$. The area element on this simplex takes the form $dadb/ab(1 - a - b)$. To integrate the curvature we first note that a corresponds to a closed geodesic of length l such that $x = e^{l/2} + e^{-l/2} = 2 + l^2/4 + o(l^2)$.

We apply Huang's estimate [5] on the curvature $|K(a, b)|$ to bound it by C/l as $l \rightarrow 0$. In particular, in the cusp we can bound the integral of the curvature by an expression of the form

$$\int_0^{1/2} \int_{1/2-b}^{1/2} K(a, b) \frac{dadb}{ab(1 - a - b)} = 5.93 \dots$$

In the general two-dimensional case (which includes the moduli space for the four times punctured sphere), we can use results in [10] and obtain that the area of the cusp corresponding to surfaces with shortest geodesic length l is a polynomial in l^2 . Thus we can bound the integral of the curvature in the cusp by $\int l^{-1} d(l^2) < \infty$.

If we were able to establish the analog of Proposition 3.3 (3), then we could apply the above-mentioned results from [8] and conclude that the WP geodesic flows have countably many ergodic components.

The next step would be to prove local ergodicity of the WP geodesic flow. Since the WP geodesic flows is Anosov (as an incomplete flow), all the hypothesis in Section 3.3 are satisfied, with identical proofs. Finally, in place of local topological transitivity, the authors, along with Scott Wolpert, have shown that the WP geodesic flow is transitive [9]. This would finish the proof of ergodicity of WP geodesic flows.

8. Acknowledgements. The authors would like to thank Scott Wolpert for many helpful discussions related to the WP metric.

REFERENCES

- [1] W. BALLMANN, M. BRIN, AND K. BURNS, *On surfaces with no conjugate points*, J. Differential Geom., 25:2 (1987), pp. 249–273.
- [2] N. I. CHERNOV, *Local ergodicity of hyperbolic systems with singularities*, Funktsional. Anal. i Prilozhen., 27:1 (1993), pp. 60–64.
- [3] G. DASKALOPOULOS AND R. WENTWORTH, *Classification of Weil-Petersson isometries*, Amer. J. Math., 125:4 (2003), pp. 941–975.
- [4] N. HITCHIN, *Geometry of surfaces*, Lecture Notes b3, 2004.
- [5] Z. HUANG, *On asymptotic Weil-Petersson geometry of Teichmüller space of Riemann surfaces*, Asian J. Math., 11:3 (2007), pp. 459–484.
- [6] A. KATOK, J.-M. STRELCCYN, F. LEDRAPPIER AND F. PRZYTYCKI, *Invariant manifolds, entropy and billiards; smooth maps with singularities*, volume 1222 of “Lecture Notes in Mathematics”, Springer-Verlag, Berlin, 1986.
- [7] L. KEEN, *On fundamental domains and the Teichmüller modular group*, in “Contributions to analysis (a collection of papers dedicated to Lipman Bers)”, pp. 185–194, Academic Press, New York, 1974.
- [8] C. LIVERANI AND M. P. WOJTKOWSKI, *Ergodicity in hamiltonian systems*, in “Dynamics reported”, volume 4 of “Dynam. Report. Expositions Dynam. Systems (N.S.)”, pp. 130–202. Springer, Berlin, 1995.
- [9] S. WOLPERT M. POLLICOTT AND H. WEISS, *Topological dynamics of the weil-petersson geodesic flow*, submitted for publication, 2007.
- [10] M. MIRZAKHANI, *Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces*, Invent. Math., 167:1 (2007), pp. 179–222.
- [11] S. A. WOLPERT, *Noncompleteness of the Weil-Petersson metric for Teichmüller space*, Pacific J. Math., 61:2 (1975), pp. 573–577.

- [12] S. A. WOLPERT, *Geometry of the Weil-Petersson completion of Teichmüller space*, in “Surveys in Differential Geometry VIII: Papers in Honor of Calabi, Lawson, Siu and Uhlenbeck”, pp. 357–393. Intl. Press, Cambridge, MA, 2003.
- [13] S. A. WOLPERT, *Behavior of geodesic-length functions on Teichmüller space*, to appear J. Differential Geom., 2008, [arxiv:math/0701556](https://arxiv.org/abs/math/0701556).

