SPHERES WITH POSITIVE CURVATURE AND NEARLY DENSE ORBITS FOR THE GEODESIC FLOW.

KEITH BURNS AND HOWARD WEISS

Northwestern University and Pennsylvania State University

June 2, 1999 10:24am

ABSTRACT. For any $\varepsilon > 0$, we construct an explicit smooth Riemannian metric on the sphere $S^n, n \geq 3$, that is within ε of the round metric and has a geodesic for which the corresponding orbit of the geodesic flow is ε-dense in the unit tangent bundle. Moreover, for any $\varepsilon > 0$, we construct a smooth Riemannian metric on $S^n, n \geq 3$, that is within ε of the round metric and has a geodesic for which the complement of the closure of the corresponding orbit of the geodesic flow has Liouville measure less than ε.

0. Introduction

It has long been known that the geodesic flow for a Riemannian metric of negative curvature possesses chaotic dynamics with the strongest possible stochastic behavior: the flow is not only ergodic but also has the Bernoulli property. A major open problem in ergodic theory and geometry is whether the geodesic flow of a Riemannian metric with everywhere positive sectional curvatures can exhibit such stochastic behavior.

Little is known about this question. Katok [Ka1,Z] gave examples of nonsymmetric Finsler metrics that are arbitrarily close to the round metric and have geodesic flows with only two ergodic components. But in the Riemannian case it is not even known if the geodesic flow for a metric of positive curvature can be topologically transitive. Topological transitivity for a flow is equivalent to the existence of a dense orbit of the flow and is the weaker topological analog of ergodicity.

If such an ergodic or topologically transitive metric exists close to the round metric on S^n, more precisely, if such a metric is 9/16-pinched $(9/16 \leq K \leq 1)$, then it must possess a non-hyperbolic closed geodesic. For, in [BTZ], the authors show that any metric satisfying this pinching condition must possess a non-hyperbolic closed geodesic. Generically these non-hyperbolic closed geodesics are non-degenerate elliptic [Kl], and it follows from the

1991 *Mathematics Subject Classification.* 58F05, 58F15, 58F30.

Key words and phrases. positive curvature, topological transitivity, topological mixing, geodesic flow, stochastic behavior.

The first author was partially supported by a National Science Foundation grant #DMS-9504760. The second author was partially supported by a National Science Foundation grant #DMS-9704913

Typeset by $\mathcal{AMS}$-TEX

KAM (Kolmogorov-Arnold-Moser) theorem [Ko, A1, A2] that the geodesic flow for a metric which possess a non-degenerate elliptic geodesic can not be ergodic. Thus any ergodic example which is 9/16-pinched must possess either a parabolic or degenerate elliptic closed geodesic.

The only positive result was obtained by Knieper and Weiss [KW] who constructed real analytic metrics with positive curvature and positive topological entropy on the sphere S^2. Their examples are small conformal perturbations of the standard metric on an ellipsoid whose three axes have different lengths and can be constructed arbitrarily close to the round metric. There are horseshoes in these examples on which the geodesic flow has strongly stochastic behaviour. However it is unknown whether these examples exhibit chaotic dynamics on sets of positive Liouville measure. The KAM theorem prevents these examples from having a dense orbit. The existence of the horseshoe implies that the geodesic flows of these surfaces have infinitely many hyperbolic closed geodesics, and in fact an exponential growth rate of closed geodesics.

Another measure of the complexity of a flow is the number and growth rate of closed orbits. It is well known that geodesic flows on negatively curved manifolds have an exponential growth rate of closed orbits, the growth rate being the topological entropy. Franks [F] has shown that the geodesic flow on every positively curved two-sphere possesses infinitely many closed orbits. One might think that the simple topology of the sphere could be an obstruction for the geodesic flow of g to have complicated dynamics on a large set. This is not the case. Donnay [D1] and Burns and Gerber [BG] have constructed smooth (and real analytic) metrics on the sphere whose geodesic flows are Bernoulli. Donnay, Burns, and Gerber construct their metrics by starting with a thrice punctured sphere and considering its complete hyperbolic metric. They then alter the metric far off into the cusps by cutting off the remainder of the cusps and gluing in reflecting caps. The geodesics leave the reflecting caps focused as they entered, and the cone family can be controlled in the caps. It is clear that these examples have "mostly" negative curvature, and that the negative curvature is the mechanism that causes the complicated dynamics. Although later examples by these authors require significantly less negative curvature, *some* negative curvature is essential for their constructions. In 1996, Lohkamp announced the construction of metrics with ergodic geodesic flow on *all* compact manifolds.

There are intriguing analogs between the geodesic flow for a Riemannian metric of positive curvature and the billiard flow on a smooth and strictly convex billiard table. In particular Lazutkin [L] showed that the billiard flow on any strictly convex billiard table (the obvious analog of positively curved surface) possesses a non-hyperbolic closed geodesic and that the billiard flow can not be topologically transitive. This follows from the existence of caustics near the boundary. These analogies have caused some people to speculate that the geodesic flow on a positively curved manifold may not be topologically transitive.

We now state our two main theorems. They show that rather complex dynamics can be achieved on rather large sets for the geodesic flows of metrics that are very close to the round metric on a sphere. For simplicity we shall consider the case of S^3, but the proofs easily extend to higher dimensions and thus the theorems are also true in dimension n for $n \geq 3$.

We first construct a metric close to the round metric on S^3 whose geodesic flow has a horseshoe that is nearly dense. More precisely:

Theorem 0.1. *For any $\epsilon > 0$ there exists a C^∞ Riemannian metric on the sphere S^3 that is within ε of the round metric and has a horseshoe in its geodesic flow which is ε-dense.*

To prove Theorem 0.1 we first perturb the round metric to create a set of ε-dense hyperbolic closed orbits $O_1, \dots, O_m$ (tangent to great circles $\gamma_1, \dots, \gamma_m$) with a heteroclinic connection between each successive pair of orbits, i.e., for each i there exists a bi-infinite orbit $O_{i,i+1}$ (tangent to a geodesic c_i) that is backwards asymptotic to O_i and forwards asymptotic to O_{i+1}. We then effect small localized metric perturbations to break all these heteroclinic connections so that the new stable and unstable manifolds of the perturbed hyperbolic closed geodesics (which remain hyperbolic) intersect transversely. It follows that this perturbed metric has an ε-dense horseshoe (locally maximal hyperbolic set) in the unit tangent bundle and thus will have an orbit which is ε-dense in the whole unit tangent bundle. The geodesic flow for this perturbed metric will have positive topological entropy becasue it contains a horseshoe.

By carefully iterating the construction in the proof of Theorem 0.1, we exhibit a metric on the sphere S^3 that is close to the round metric and has a geodesic flow with an orbit whose closure has almost full measure. More precisely:

Theorem 0.2. *Given $\varepsilon > 0$, there exists a C^∞ metric g on S^3 that is within ε of the round metric g_0 (in the C^∞ topology) with the property that there is an orbit of the geodesic flow ϕ_g^t whose closure has (normalized) Liouville measure at least $1 - \varepsilon$.*

It is not impossible that (at least some of) the metrics constructed in Theorem 0.2 actually have topologically transitive geodesic flows. However, we do not know how to show this. The geodesic flows constructed here do have positive topological entropy because they contain horseshoes.

The construction used to prove the above theorems requires three dimensions. However, on S^2, one can construct a metric whose geodesics approximate trajectories of the well known stadium billiard. Recall that the stadium is the C^1 convex curve formed by two semi circles joined by two parallel line segments (see Figure 1A). We can approximate the stadium by a table whose boundary is a smooth convex curve, and then form a smooth convex surface that contains parallel copies of this table separated by a narrow strip with very strong positive curvature, as shown in Figure 1B.

It is obvious that if the top and bottom of the surface are made close enough together, and each of them approximates the stadium well enough, long geodesic segments on the surface will closely follow trajectories from the stadium billiard. The geodesic will switch between the top and the bottom of the surface each time the billiard trajectory bounces off the edge of the table. The stadium billiard is ergodic, and hence there is a dense orbit of the billiard flow. We can construct our surface so that its geodesic flow has an orbit that approximates as long a piece of this dense orbit as closely as we wish. Thus we can produce a metric on S^2 with an ε-dense orbit of the geodesic flow for any $\varepsilon > 0$. These metrics clearly have non-negative curvature. However, an easy perturbation argument gives strictly-convex surfaces (metrics with positive curvature) with an ε-dense orbit of the geodesic flow for any $\varepsilon > 0$.

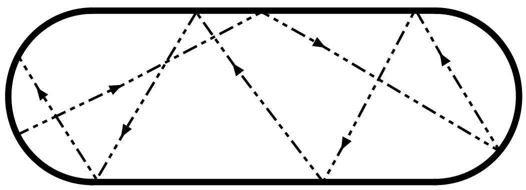

FIGURE 1A. STADIUM BILLIARD

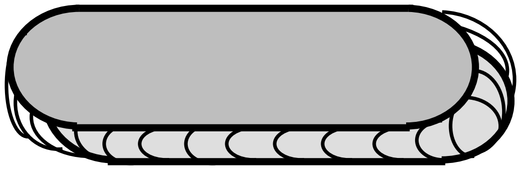

FIGURE 1B. OUR SURFACE

We thank Charles Pugh for his interest in this work and for convincing us to improve our exposition in Section 4.

1. Preliminaries from Dynamical Systems

A continuous flow ϕ^t on a compact metric space X is *topologically transitive* if given any two non empty open sets $U, V \subset X$, there exists $T > 0$ such that $\phi^T(U) \cap V \neq \emptyset$. This is equivalent to the existence of dense orbit of ϕ^t. We say that the flow is *ε-topologically transitive* if given any two open sets $U, V \subset X$ which contain balls of diameter ε, there exists $T > 0$ such that $\phi^T(U) \cap V \neq \emptyset$. This is equivalent to the existence of an ε-dense orbit of ϕ^t, i.e., there exists an orbit which intersects every ball of diameter at least ε.

A topological property of a flow which is stronger than topological transitivity is topological mixing. A continuous flow ϕ^t on a compact metric space X is *topologically mixing* if given any two non empty open sets $U, V \subset X$, there exists $T > 0$ such that for all $t \geq T$ we have $\phi^t(U) \cap V \neq \emptyset$. We say that the flow is *ε-topologically mixing* if given any two non empty open sets $U, V \subset X$ which contain balls of diameter ε, there exists $T > 0$ such that for any $t \geq T$ we have $\phi^t(U) \cap V \neq \emptyset$.

Let X be a C^1 manifold and $\phi^t : X \to X$ be a C^1 flow. Assume that the flow possesses a hyperbolic closed orbit O. Let U be a small neighborhood of O. We define the *stable manifold* $W^s(O)$ by

$$W^s(O) = \{x : \phi^t(x) \in U \text{ for all } t \geq 0\}$$

and the *unstable manifold* $W^u(O)$ by

$$W^u(O) = \{x : \phi^t(x) \in U \text{ for all } t \leq 0\}.$$

If follows from the stable manifold theorem [HPS] that $W^s(O)$ and $W^u(O)$ are immersed C^1 manifolds.

We require a version of Smale's Homoclinic Theorem [S] for flows possessing a heteroclinic connection. Let X be a C^1 manifold and $\phi^t : X \to X$ be a C^1 flow. Assume that the flow possesses hyperbolic closed orbits $O_1, \dots, O_m$ such that the stable manifold of O_i intersects the unstable manifold of O_{i+1} transversely at U_i for each $i = 1, \dots, m-1$ and the stable manifold of O_m intersects the unstable manifold of O_1 transversely at U_m. Then the flow has a locally maximal hyperbolic compact invariant set K (a horseshoe) containing $O_1 \cup \dots \cup O_m$ and $U_1 \cup \dots \cup U_m$. The periodic orbits are dense in K and the flow ϕ^t restricted to K is topologically mixing.

2. Small perturbations to the round metric

In this section we describe the types of perturbation which will be used in Section 3 to prove Theorems 0.1 and 0.2. Before doing so, we make some general remarks.

The geodesic flow ϕ_g^t of a Riemannian metric g on S^3 is usually thought of as acting on the bundle $T_g^1 S^3$ of vectors that have unit length with respect to g. This convention is inconvenient for us, because it would make the geodesic flows for different metrics act on different bundles. Instead we will use radial projection in the fibers of TS^3 to identify $T_g^1 S^3$ with the unit tangent bundle of the standard round metric g_0. The notation $T^1 S^3$ will mean $T_{g_0}^1 S^3$. It will be convenient to use the Sasaki metric and Liouville measure (normalized to be a probability measure) for g_0 to measure distances and volume in $T^1 S^3$. We interpret Theorems 0.1 and 0.2 in this way. It is easily seen that these theorems still hold if the distance and measure defined on $T^1 S^3$ by the perturbed metric g are used instead.

The basis of our construction is the following proposition, which allows us to perturb the round metric on S^2 so as to make given orbits of the geodesic flow hyperbolic and create a heteroclinic orbit connecting them.

Proposition 2.1. *Let γ_1 and γ_2 be geometrically distinct geodesics in S^2 with the standard metric. We can choose a metric g_1 on S^2, arbitrarily close (in the C^∞ topology) to the round metric, such that γ_1 and γ_2 are geodesics for g_1, the corresponding closed orbits O_1 and O_2 of the geodesic flow for g_1 are hyperbolic, and $W^u(O_1) \cap W^s(O_2) \neq \emptyset$. Furthermore this perturbation can be made in an arbitrarily small neighbourhood of one of the two intersection points of γ_1 and γ_2.*

Proof. Let p be one of the two antipodal points in which the geodesics γ_1 and γ_2 intersect. Let v_1 and v_2 be the vectors based at p that belong to O_1 and O_2 respectively. We shall use geodesic polar coordinates (for the round metric) on S^2 with p as center: let $r \in [0, \pi]$ be the radial coordinate and $\theta \in S^1$ the angular coordinate. We think of S^1 as $[0, 2\pi]$ with its endpoints identified. In these coordinates, the round metric on S^2 is $dr^2 + \sin^2 r \, d\theta^2$. Let θ_1 and θ_2 be the values of the θ coordinate on the geodesic rays emanating from p that are tangent to v_1 and v_2 respectively. The hypotheses imply that θ_1 and θ_2 are distinct and not antipodal to one another. We may assume that the direction in which we measure θ and the position of the ray $\theta = 0$ were chosen so that $0 < \theta_1 < \theta_2 < \pi$.

Choose a C^∞ flow α_t on S^1 such that

(1) the only fixed points in $[\theta_1, \theta_2]$ are a hyperbolic source at θ_1 and a hyperbolic sink at θ_2; and
(2) every point of the arc $[\pi + \theta_1, \pi + \theta_2]$ (which is antipodal to $[\theta_1, \theta_2]$) is fixed.

Choose $\rho \in (0, \pi)$ and a C^∞ function τ that is defined on $[0, \pi]$, 0 on $[0, \rho/3]$, nondecreasing on $[\rho/3, 2\rho/3]$, and constant and positive on $[2\rho/3, \pi]$. Let $f : S^2 \to S^2$ be the diffeomorphism that maps the point with coordinates (r, θ) to the point with coordinates $(r, \alpha_{\tau(r)}(\theta))$. It follows from (1), (2) and the definition of f that f fixes all points on γ_1 and γ_2. The circles $r = \text{const}$ are mapped to themselves. Outside the circle $r = 2\rho/3$, each ray $\theta = \text{const}$ is mapped to another ray of this form.

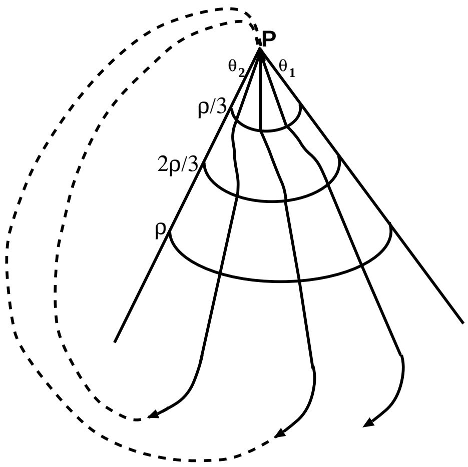

FIGURE 2. SLIGHTLY PERTURBED METRIC ON S^2

We are now ready to define the new metric g_1. Vectors tangent to the circles $r = \text{const}$ will have the same length as in the round metric. The images under f of the rays $\theta = \text{const}$ will be unit speed geodesics orthogonal to these circles. In terms of the coordinate vector fields, $\partial/\partial\theta$ and $\partial/\partial r$, this means that $\partial/\partial\theta$ has the same length in the new metric as in the old metric, and that the pushforward by f of $\partial/\partial r$ is a unit vector field orthogonal to $\partial/\partial\theta$. Notice that the pushforward by f of $\partial/\partial r$ coincides with $\partial/\partial r$ outside the circle $r = 2\rho/3$; it follows that g_1 coincides with the round metric outside the circle $r = 2\rho/3$. The metric g_1 has the form $a(r, \theta)\, dr^2 + \sin^2 r\, d\theta^2 + 2b(r, \theta)\, dr d\theta$, where $a(r, \theta)$ and $b(r, \theta)$ are smooth functions which depend on f.

Our construction ensures that the images under f of the level curves of θ, i.e. the images under f of the great circle arcs from p to its antipodal point, are geodesic segments for the new metric g_1. For $\theta \in (\theta_1, \theta_2)$ the image under f of such an arc is a geodesic for the new metric which behaves as shown in Figure 2. While $r < \rho/3$, this geodesic coincides with a great circle. But as the geodesic passes through the band where $\rho/3 \le r \le 2\rho/3$, the value of θ increases as the new geodesic passes through the band. After that the geodesic coincides with a great circle as it passes through the antipodal point of p and then returns to p through the sector where $\pi + \theta_1 < \theta < \pi + \theta_2$. This pattern is then repeated. It is easy to see from this that any geodesic of the new metric on S^2 that leaves p in the sector $\theta_1 < \theta < \theta_2$ is backwards asymptotic to γ_1 and forwards asymptotic to γ_2.

Since the great circles γ_1 and γ_2 are geodesics for the new metric g_1, O_1 and O_2 are periodic orbits for the geodesic flow of g_1 Since θ_1 and θ_2 were hyperbolic fixed points of the flow α_t, it is easily seen that the Poincaré map for each of these orbits has an eigenvalue that is not on the unit circle; and since the geodesic flow preserves the Liouville volume, it then follows that that O_1 and O_2 are hyperbolic closed orbits of the geodesic flow for the new metric. The discussion in the previous paragraph shows that $W^u(O_1) \cap W^s(O_2) \neq \emptyset$.

The above construction can be performed with arbitrarily small ρ (hence it is enough to change the metric on S^2 in an arbitrarily small neighborhood of the point p) and f can be chosen as close to the identity in the C^∞ topology as we wish. Thus g_1 can be made as close to the round metric in the C^∞ topology as we wish; the new metric will have positive curvature provided we make the perturbation sufficiently small. ■

We now suppose that the S^2 considered in Proposition 2.1 is a great sphere Σ embedded in S^3. The next proposition shows that the change of metric on S^2 described above can be realized by a change of metric on S^3 which, in particular, leaves Σ totally geodesic.

Proposition 2.2. *Suppose that Σ is a great 2-sphere in S^3 (with the round metric) and Σ_1 and Σ_2 are great 2-spheres that intersect Σ orthogonally in distinct great circles σ_1 and σ_2. Let γ_1 and γ_2 be geodesics in S^3 obtained by choosing directions for σ_1 and σ_2 respectively. Then the change of metric g_1 on Σ described in Proposition 2.1 can be achieved by a change of metric in an arbitrarily small ball B in S^3 around one of the two points in $\Sigma \cap \Sigma_1 \cap \Sigma_2$, which leaves the three spheres Σ, Σ_1, and Σ_2 totally geodesic, and which does not change the metric on Σ_1 or Σ_2. This perturbation can effected arbitrarily close (in the C^∞ topology) to the round metric on S^3.*

Proof. Let p be one of the two antipodal points in $\Sigma \cap \Sigma_1 \cap \Sigma_2$. We extend the construction used in the proof of the previous proposition to a change of metric on a neighbourhood in S^3 of the point p. To this end we extend the coordinates used in the proof of Proposition 2.1 to a neighbourhood in S^3 of p. We introduce a third coordinate s such that $|s(q)|$ is the distance of a point q from Σ, measured along the great circles orthogonal to Σ. The other coordinates $r(q)$ and $\theta(q)$ are defined to be the r and θ coordinates of the orthogonal projection of q to Σ. These coordinates are illustrated in Figure 3.

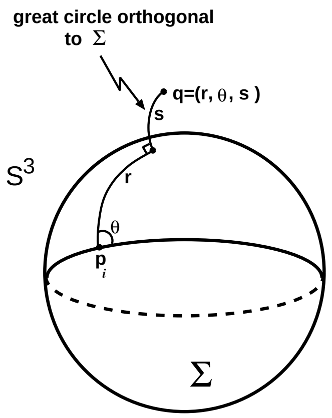

FIGURE 3. (θ, r, s) COORDINATE SYSTEM ON S^3

Choose $\rho_0 > 0$ sufficiently small such that the number ρ chosen during the construction of the metric g_1 on Σ in Propositon 2.1 satisfies $\rho < \rho_0$. The metric g_1 on Σ differs from the round metric only on $\Sigma \cap B(p, \rho_0)$.

Let $\beta(s)$ be a bump function such that $\beta(s) = 1$ if $|s| \leq \rho_0/3$, $0 \leq \beta(s) \leq 1$ if $\rho_0/3 \leq s \leq 2\rho_0/3$ and $\beta(s) = 0$ if $s \geq 2\rho_0/3$. Let F be the diffeomorphism of $B = B(p, 2\rho_0)$ that maps the point with coordinates (r, θ, s) to the point with coordinates $(r, \alpha_{\beta(s)\tau(r)}\theta, s)$, where α_t is the flow used in the construction of f in the proof of Proposition 2.1. Observe that F has properties analogous to those of f:

- The level surfaces of the coordinates r and s are mapped into themselves;
- Outside the region where $r \leq 2\rho/3$, one level surface of the coordinate θ is mapped to another level surface of θ.

Furthermore,

- The function F is the identity outside the region where $|s| \leq \rho_0$.

In the (r, θ, s) coordinates, the round metric on S^3 is $\cos^2 s\,(dr^2 + \sin^2 r\, d\theta^2) + ds^2$. We now define a new metric g_2 on B in S^3 which will coincide with the round metric on the complement of B. In B we decree that the coordinate vector fields $\partial/\partial\theta$ and $\partial/\partial s$ will have the same length as in the round metric and will still be orthogonal, while the unit vector field orthogonal to both $\partial/\partial\theta$ and $\partial/\partial s$ will now be the pushforward by F of $\partial/\partial r$. It follows from the above properties of F that the new metric agrees with the old metric except where $r \leq 2\rho/3$ or $|s| \leq \rho_0$. Near p (where $|s| < \rho_0/3$) this new metric takes the form

$$\cos^2 s\,(a(r,\theta)\, dr^2 + \sin^2 r\, d\theta^2 + 2b(r,\theta)\, dr d\theta) + ds^2, \tag{1}$$

where the functions $a(r,\theta)$ and $b(r,\theta)$ are functions introduced in the proof of Propositon 2.1, which give the metric g_1 in the (r,θ) coordinates. It is obvious that we can make g_2 as close to the round metric as desired by choosing ρ_0 sufficiently small.

It is easily seen that F maps $\Sigma \cap B$ into itself and maps $\Sigma_j \cap B$ into itself for $j = 1, 2$. Moreover F fixes each point of $\Sigma_j \cap B$ for $j = 1, 2$, and the restriction of F to $\Sigma \cap B$ is f.

The following lemma will be useful for showing that $\Sigma \cap B$ and $\Sigma_i \cap B$ are totally geodesic surfaces.

Lemma 2.3. *Let $x^1, \ldots, x^n$ be local coordinates on a Riemannian manifold (M, g) and let H be the hypersurface defined by $x^{\alpha_0} = c$, where c is a constant. Suppose that at all points in H we have $g_{\alpha_0\beta} = 0$ and $\partial g_{\beta\gamma}/\partial x^{\alpha_0} = 0$ unless $\beta = \alpha_0$ or $\gamma = \alpha_0$. Then H is totally geodesic.*

Proof. This is easily proved by computing Christoffel symbols. We sketch an alternative proof. Consider a curve σ that joins two points p and q of H and is a geodesic for the induced metric on H. We need to show that if we vary σ, keeping its endpoints fixed, then the energy integral is constant to first order. Since σ is a geodesic in H, all variations tangent to H have this property. Thus it suffices to consider a variation in the x^{α_0} direction.

Suppose $\psi(u,t)$ is such a variation. Let $v^\alpha(u,t)$ be the α'th component of $\partial\psi/\partial t(u,t)$ and $g_{\alpha\beta}(u,t)$ the metric at $\psi(u,t)$. When $u=0$ we have:

(1) $v^{\alpha_0}=0$;
(2) $g_{\alpha_0\beta}=0$ unless $\beta=\alpha_0$;
(3) $dv_\alpha/du=0$ unless $\alpha=\alpha_0$;
(4) $dg_{\alpha\beta}/du=0$ unless $\alpha=\alpha_0$ or $\beta=\alpha_0$.

Hence

$$\frac{d}{ds}\|\partial\psi/\partial t\|^2=\frac{d}{ds}\sum_{\alpha,\beta}g_{\alpha\beta}v^\alpha v^\beta=\sum_{\alpha,\beta}\left\{\frac{dg_{\alpha\beta}}{du}v^\alpha v^\beta+g_{\alpha\beta}\frac{dv^\alpha}{du}v^\beta+g_{\alpha\beta}v^\alpha\frac{dv^\beta}{du}\right\}$$

vanishes when $u=0$. Integrating with respect to t now shows that derivative at $u=0$ of the energy integral is 0. ∎

We now show that $\Sigma\cap B$ is totally geodesic. This set is the level surface $s=0$. Since the pushforward of $\partial/\partial r$ by F is a linear combination of $\partial/\partial r$ and $\partial/\partial\theta$, it is obvious that $\partial/\partial s$ is orthogonal to both $\partial/\partial r$ and $\partial/\partial\theta$ in the metric g. It follows immediately from the explicit form (1) of the metric g_2 that on Σ we have that

$$\frac{d}{ds}g(\partial/\partial r,\partial/\partial r)=\frac{d}{ds}g(\partial/\partial\theta,\partial/\partial\theta)=\frac{d}{ds}g(\partial/\partial r,\partial/\partial\theta)=0.$$

Finally, consider $\Sigma_1\cap B$ and $\Sigma_2\cap B$. Away from p, these are formed by level surfaces of the coordinate θ, namely $\theta=\theta_1$, $\theta=\theta_1+\pi$, $\theta=\theta_2$ and $\theta=\theta_2+\pi$. In order to be able to apply Lemma 2.3, we need to verify that $\partial/\partial\theta$ is orthogonal to $\partial/\partial r$ and $\partial/\partial s$ on $(\Sigma_1\cup\Sigma_2)\cap B$ and

$$\frac{d}{d\theta}g(\partial/\partial r,\partial/\partial r)=\frac{d}{d\theta}g(\partial/\partial r,\partial/\partial s)=\frac{d}{d\theta}g(\partial/\partial s,\partial/\partial s)=0 \tag{2}$$

on $(\Sigma_1\cup\Sigma_2)\cap B$. We saw in a previous paragraph that g makes $\partial/\partial\theta$ and $\partial/\partial s$ orthogonal. Recall that $\partial/\partial s$ and the pushforward $(F)_*\partial/\partial r$ of $\partial/\partial r$ by F are orthonormal with respect to the metric g, and observe that

$$(F)_*\frac{\partial}{\partial r}(r,\theta,s)=\frac{\partial}{\partial r}(r,\theta,s)+c(r,\theta,s)\frac{\partial}{\partial\theta}(r,\theta,s),$$

where $c(r,\theta,s)$ is a smooth function that vanishes when $\theta=\theta_1,\theta_1+\pi,\theta_2,\theta_2+\pi$. It follows immediately that $\partial/\partial r$ is orthogonal to $\partial/\partial\theta$ on $(\Sigma_1\cup\Sigma_2)\cap B$. It is straightforward to verify that (2) holds on $(\Sigma_1\cup\Sigma_2)\cap B$. We can now apply Lemma 2.3 to deduce that $\Sigma_1\cap B$ and $\Sigma_2\cap B$ are totally geodesic. ∎

The next proposition allows us to make the heteroclinic orbits created using the previous propositions into transverse intersections of the relevant stable and unstable manifolds. It will be used to obtain the horseshoe promised in Theorem 0.1.

Proposition 2.4. *Let O_1 and O_2 be hyperbolic periodic orbits for the geodesic flow of a Riemannian manifold (M, g) that are connected by a heteroclinic orbit $O_{1,2} \subset W^u(O_1) \cap W^s(O_2)$. Then we can make an arbitrarily small (in the C^∞ topology) perturbation of g in an arbitrarily small neighbourhood of any given point on the geodesic corresponding to $O_{1,2}$ so that O_1 and O_2 are hyperbolic closed orbits for the new geodesic flow and $W^u(O_1)$ and $W^s(O_2)$ now intersect transversally along $O_{1,2}$.*

Proof. The method for making the intersection transverse has been described by Donnay [D2] in the two dimensional case, and in general by Petroll [P]. Since Petroll's paper is not widely available, we give a brief sketch of the idea; [G] and [E] are good references for background material.

Let γ be the geodesic to which $O_{1,2}$ is tangent, parametrized so that the vectors $\dot{\gamma}(t)$ belong to $O_{1,2}$. Let $W^{ss}(t)$ and $W^{uu}(t)$ be respectively the strong stable manifold for O_i and the strong unstable manifold for O_{i+1} that contain $\dot{\gamma}(t)$. It is well known that for all t, except for a discrete subset, there are neighbourhoods N^{ss} and N^{uu} of $\dot{\gamma}(t)$ in $W^{ss}(t)$ and $W^{uu}(t)$ respectively that project to smooth hypersurfaces $H^{ss}(t)$ and $H^{uu}(t)$; the neighbourhoods N^{ss} and N^{uu} consist of the unit normals to $H^{ss}(t)$ and $H^{uu}(t)$ that point in the right direction.

Let $U^{ss}(t)$ and $U^{uu}(t)$ denote the second fundamental forms of $H^{-}(t)$ and $H^{uu}(t)$ respectively with respect to $\dot{\gamma}(t)$. The intersection of $W^s(O_i)$ with $W^u(O_{i+1})$ along $O_{1,2}$ is transversal if and only if $W^{ss}(t)$ and $W^{uu}(t)$ intersect transversally at $\dot{\gamma}(t)$; if this property holds for one time t_0 it will hold for all times t. The intersection of $W^{ss}(t_0)$ and $W^{uu}(t_0)$ at $\dot{\gamma}(t_0)$ is transversal if and only if 0 is not an eigenvalue of the quadratic form $U^{uu}(t_0) - U^{ss}(t_0)$.

Now $U^{ss}(t)$ and $U^{uu}(t)$ are solutions of a differential equation. Suppose that $E_i(t)$ are covariantly constant vector fields along γ that form an orthonormal basis for the orthogonal complement of $\dot{\gamma}(t)$ for every t. Let $R(t)$ be the matrix whose ij'th entry is $\langle R(\dot{\gamma}(t), E_i(t))\dot{\gamma}(t), E_j(t)\rangle$, where R is the curvature tensor. Then the matrices that express $U^{ss}(t)$ and $U^{uu}(t)$ in terms of this basis satisfy the following matrix Riccati equation:

$$U'(t) + U^2(t) + R(t) = 0.$$

What is crucial is not the exact form of the equation, but the fact that $U^{ss}(t_0)$ is determined by $R(t)$ for $t < t_0$ and $U^{uu}(t_0)$ is determined by $R(t)$ for $t > t_0$. The orbit $O_{1,2}$ is nonrecurrent, because it is forwards and backwards asymptotic to the closed orbits O_{i+1} and O_i respectively. It follows from this that the times at which the geodesic γ crosses itself are isolated. Hence there exist t_1 and t_2 such that $t_0 < t_1 < t_2$ and γ does cross itself at any point $\gamma(t)$ with $t_1 \leq t \leq t_2$. Petroll uses Fermi coordinates along $\gamma(t)$ for $t_1 < t < t_2$ to perturb the metric so that the curvature tensor changes but the curve γ remains a geodesic. This will affect $U^{uu}(t_0)$ but not $U^{ss}(t_0)$. It is then easy to choose a perturbation such that 0 is not an eigenvalue of $U^{uu}(t_0) - U^{ss}(t_0)$. *We emphasize that this perturbation can be made arbitrarily small in the C^∞ topology and can be localized in an arbitrarily small neighborhood of the point $\gamma(t_0)$.* ■

The final proposition allows us to glue together geodesics that are forward and backwards asymptotic to a hyperbolic closed geodesic.

Proposition 2.5. *Let γ be a closed geodesic in a Riemannian manifold (M,g) for which the corresponding orbit O of the geodesic flow is hyperbolic. Let $v \in W^s(O)$ and $w \in W^u(O)$. Then we can make an arbitrarily small (in the C^∞ topology) perturbation of g in an arbitrarily small neighbourhood of any given point on γ so that v and w lie on the same orbit of the new geodesic flow.*

Proof. Let N be a neighbourhood of the given point on γ. We can choose $t_1 > 0 > t_2$ and $\delta > 0$ such that, if N_1 and N_2 are the closed δ-discs around $p_1 = \gamma_v(t_1)$ and $p_2 = \gamma_w(t_2)$ respectively, then

(1) $N_1 \cap N_2 = \emptyset$ and $N_1 \cup N_2 \subset N$;
(2) $\gamma_v(t) \notin N_1 \cup N_2$ for $t \geq 0$ except for $t_1 - \delta \leq t \leq t_1 + \delta$ when $\gamma_v(t) \in N_1$;
(3) $\gamma_w(t) \notin N_1 \cup N_2$ for $t \leq 0$ except for $t_2 - \delta \leq t \leq t_2 + \delta$ when $\gamma_w(t) \in N_2$.

Since the closed orbit O is hyperbolic, we can find vectors v' and w' as close as we wish to $\dot\gamma_v(t_1)$ and $\dot\gamma_w(t_2)$ respectively such that w' lies on the forward orbit of v' and the orbit segment O' between v' and w' lies as close as we wish to the union of the forward orbit of v and the backward orbit of w. In particular, we can ensure that $\gamma_{v'}(\delta)$ and $\gamma_{w'}(-\delta)$ lie on the boundaries of N_1 and N_2 respectively and the segment of $\gamma_{v'}$ that lies between these points does not enter $N_1 \cup N_2$.

We now show that it is possible to perturb the metric inside N_1 so that the geodesic which enters N_1 tangent to $\dot\gamma_v(t_1 - \delta)$ exits N_1 tangent to $\dot\gamma_{v'}(\delta)$. In order to do this, choose a sphere S that is close to the geodesic sphere of radius $\delta/2$ around $p_1 = \gamma_v(t_1)$, passes through $\gamma_v(t_1 - \delta/2)$ and $\gamma_{v'}(\delta/2)$, and is orthogonal to $\dot\gamma_v(t_1 - \delta/2)$ and $\dot\gamma_{v'}(\delta/2)$. Let D_r be the closed r-disc and S_r the sphere of radius r around the origin in $T_{p_1}M$ with the geometry given by the inner product $g_{p_1}(\cdot,\cdot)$ defined on $T_{p_1}M$ by g. Choose a diffeomorphism $\psi : D_{\delta/2} \to N_1$ that is close to the exponential map and has the following properties:

(1) The initial value $\psi(0) = p_1$.
(2) The map ψ satisfies $\psi(\partial D_{\delta/2}) = S$ and $\psi(\operatorname{Int} D_{\delta/2})$ lies inside S.
(3) The map ψ maps one diagonal of $D_{\delta/2}$ to a curve joining $\gamma_v(t_1 - \delta/2)$ and $\gamma_{v'}(\delta/2)$.
(4) The map $D\psi$ maps the inward unit normal vector field on ∂D to the inward unit normal vector field on S.
(5) If $z \in \partial D$, then the curve $t \mapsto \psi((1-t)z)$, $0 \leq t \leq 1/2$ is the geodesic segment in N_1, parametrized with speed $\delta/2$, that starts orthogonally from S at $\psi(z)$ and goes distance $\delta/4$ into the interior of S.

Let V be the vector field on $D_{\delta/2} \setminus \{0\}$ that points radially outward and has unit length with respect to g_{p_1}. Properties (4) and (5) of ψ ensure that if $z \in D_{\delta/2} \setminus D_{\delta/4}$, then $D\psi(V(z))$ is a unit vector orthogonal to $\psi(S_{\|z\|})$.

We can define a new Riemannian metric on $\psi(D_{\delta/2} \setminus \{0\})$ by leaving the lengths of vectors tangent to the spheres $\psi(S_r)$ unchanged and decreeing that $D\psi(V(z))$ is a unit vector field that is orthogonal to these spheres. This metric agrees with the original metric outside the image of $\psi(D_{\delta/4})$ and extends smoothly to a metric on $\psi(D_{\delta/2})$. It follows from Gauss' Lemma (see e.g. Lemma 3.3.5 in [DoC]) that ψ maps the diagonals of D to unit speed geodesics in N_1 with its new metric.

Finally we make an analogous change of metric inside N_2 so that the geodesic which enters N_2 tangent to $\dot\gamma_{w'}(-\delta)$ exits N_2 tangent to $\dot\gamma_w(t_2 + \delta)$. ∎

3. Construction of an arbitrarily dense horseshoe

We begin by reminding the reader of some facts about the geometry of the round sphere S^3, which we collect into Lemma 3.1.

Lemma 3.1. *Two distinct closed geodesics on S^3 intersect if and only if they lie in a common great sphere. If they intersect, they lie on a unique 2-sphere, intersect at a pair of antipodal points and have a common normal direction.*

Proof. The proof is an immediate consequence of the facts that a closed geodesic is the intersection of a plane through the origin with S^3 and a great sphere is the intersection of a hyperplane through the origin with S^3. ∎

It is evident from the proof of this lemma that it is exceptional for two geodesics to intersect.

By a **sequence of orthogonal spheres**, we will mean a sequence of great 2- spheres in S^3 with the property that consecutive terms are orthogonal. Such a sequence provides an environment in which we can apply the propositions developed in the previous section. We say that a sequence of orthogonal spheres is **nondegenerate** if all its terms are distinct and the intersection of any four of the spheres is empty[1].

Let $\{\Sigma_i\}_{i\in I}$ be a sequence of orthogonal spheres. The sequence may be finite, infinite or cyclic. In the cyclic case, we interpret I as the integers modulo m, where m is the number of terms in the sequence. Let σ_i be the great circle in which Σ_{i-1} and Σ_i intersect. The circles formed in this way will be called the **great circles associated with** $\{\Sigma_i\}_{i\in I}$. If the sequence is nondegenerate, then

(1) σ_i and σ_j are distinct unless unless $i = j$.
(2) If $i \neq j$, then σ_i intersects σ_j if and only if i and j are consecutive integers.

A sequence of orthogonal spheres in which all terms are distinct can be made nondegenerate (without destroying the orthogonality property) by an arbitrarily small perturbation. In order to see this, observe that after making any small perturbation to one term Σ_i, we can maintain the orthogonality of consecutive terms by perturbing only the two adjacent terms Σ_{i-1} and Σ_{i+1}. Since all terms in the sequence are distinct, $\Sigma_i \cap \Sigma_{i-2}$ and $\Sigma_i \cap \Sigma_{i+2}$ are both great circles. When we make a small enough perturbation to Σ_i, these great circles move to nearby great circles, and we can choose the new Σ_{i-1} and Σ_{i+1} to be orthogonal to these new great circles (which ensures that the sequence still has the orthogonality property) and close to the old Σ_{i-1} and Σ_{i+1}. We can make the sequence nondegenerate by iterating this construction.

In the proofs of Theorems 0.1 and 0.2 we need to able to start with a given collection of great circles and construct a sequence of orthogonal spheres whose associated great circles include the given collection. To this end, we introduce the notion of a **cross** centered at σ, which is an ordered pair of great spheres $X = (\Sigma^-, \Sigma^+)$ that intersect orthogonally at the great circle σ. If we are given a sequence of crosses $X_1, \dots, X_n$, we can always make an arbitrarily small perturbation so that no two of the crosses contain the same sphere and any four spheres that belong to the crosses have empty intersection; we shall

[1]If the chain has $k \leq 3$ terms, we require that they intersect in a great sphere of dimension $3 - k$.

always assume that a sequence of crosses has this property. In this case, each pair Σ_i^+ and Σ_{i+1}^- will intersect in a great circle, and any great sphere orthogonal to this circle will be orthogonal to both Σ_i^+ and Σ_{i+1}^-. We say that such a sphere **links** X_i and X_{i+1}. Linking all consecutive pairs of terms from a sequence of crosses creates a sequence of orthogonal spheres. All terms in this sequence will be distinct if the linking spheres are chosen suitably. If necessary, it can then be made nondegenerate by an arbitrarily small perturbation, as described above.

Proof of Theorem 0.1. We begin by choosing a finite number of (necessarily closed) orbits of the geodesic flow of the usual round metric on S^3 whose union is ε-dense in the unit tangent bundle (with respect to the round metric). Let $c_1, \ldots, c_m$ be the great circles to which these orbits are tangent. We may assume that the orbits are all tangent to distinct circles. Choose a cross X_k centered at each c_k so that the $2m$ spheres in the crosses are all distinct. By linking consecutive terms of the cyclic sequence $X_1, \ldots, X_m$, we obtain a cyclic sequence of orthogonal spheres with $3m$ terms $\Sigma_1, \ldots, \Sigma_{3m}$. It is clear that we can choose the c_k, the X_k and the linking spheres so that $\Sigma_1, \ldots, \Sigma_{3m}$ is nondegenerate. The great circles $\sigma_1, \ldots, \sigma_{3m}$ associated to this sequence include the c_i. (Recall that $\sigma_i = \Sigma_{i-1} \cap \Sigma_i$. Every third σ_i is one of the original c_k's.) For each i, let O_i be one the two orbits of the geodesic flow formed by unit vectors tangent to σ_i. We choose these directions of $O_1, \ldots, O_{3m}$ so that these orbits include the m orbits with which we began the construction.

For each $i \in \mathbb{Z}/3m$, let p_i be one of the two antipodal points in which the great circles σ_i and σ_{i+1} intersect. Then $p_i \in \Sigma_{i-1} \cap \Sigma_i \cap \Sigma_{i+1}$ and it follows from (2) above that p_i is the center of a ball B_i which does not intersect any other Σ_j. In particular σ_i and σ_{i+1} are the only σ_j's that enter B_i.

We now describe the sequence of perturbations that produces the desired metric. Inside each B_i we apply Proposition 2.2 to make O_i and O_{i+1} hyperbolic as orbits of the geodesic flow for Σ_i and create a heteroclinic orbit connecting them in the unit tangent bundle of Σ_i. Note that the change of metric in Proposition 2.2 leaves Σ_i totally geodesic, so the set of unit vectors tangent to Σ_i is an invariant subset for the geodesic flow on T^1S^3.

Observe that each O_i is hyperbolic as an orbit of the geodesic flows of *both* Σ_{i-1} and Σ_i. It follows that the derivative (at the fixed point corresponding to the closed orbit) of the Poincaré map for O_i has two expanding eigenvectors and two contracting eigenvectors. Hence each O_i is a hyperbolic closed orbit for the geodesic flow of the new metric on S^3.

Now we can apply Proposition 2.4 to the heteroclinic orbits that connect O_i to O_{i+1} for each $i \in \mathbb{Z}/m$. We can arrange that these new perturbations also take place inside the B_i's and have supports disjoint from any of the previous perturbations. After these perturbations, each of the orbits O_i is hyperbolic and $W^u(O_i)$ and $W^s(O_{i+1})$ have a transverse intersection for each $i \in \mathbb{Z}/3m$. It follows immediately from Smale's theorem (see §1) that there is a horseshoe containing $O_1, \ldots, O_{3m}$. Since these orbits include the orbits with which we began the proof, the horseshoe that we have created is ε-dense in T^1S^3, in the sense explained at the beginning of Section 2. Theorem 0.1 now follows, since the restriction of the geodesic flow to the horseshoe is topologically transitive. ■

4. Topological transitivity except on a set of arbitrarily small measure

In this section we prove Theorem 0.2. The proof is similar to that of Theorem 0.1. Again we create a sequence of orthogonal spheres and apply Proposition 2.2 to create a sequence of hyperbolic closed orbits linked by heteroclinic orbits. Then we will use Proposition 2.5 to join all of these heteroclinic orbits into a single orbit, rather than applying Proposition 2.4.

The major difference from Theorem 0.1, however, is that the sequence of orthogonal spheres is infinite rather than cyclic. When there are only finitely many spheres, we can choose the small balls in which to perturb the metric, *after* we have chosen the sequence of spheres, simply by ensuring that the sequence of spheres is in general position and choosing the balls small enough so that each ball intersects only the three spheres that pass through its center. Now we must choose the spheres and balls in batches. Each ball will contain a smaller ball, which we call its core, and each core will contain a still smaller ball, which we call the inner core. The perturbations to the metric will eventually be made in the inner cores of the balls. We ensure that the spheres in each new batch do not intersect the inner cores of the previously chosen balls. Then we perturb the new spheres so that they are in general position and choose a small ball around one of the intersection points of each triple of consecutive spheres in the batch.

We choose each batch of spheres as we did in the proof of Theorem 0.1, by first choosing a sequence of crosses and then linking the crosses. It is straightforward to choose the crosses so that they miss the cores of the pre- existing balls, but choosing the linking spheres so that they miss the inner cores involves a subtlety. The spheres which can link two crosses (Σ_1^-, Σ_1^+) and (Σ_2^-, Σ_2^+) are the spheres orthogonal to the great circle $\Sigma_1^+ \cap \Sigma_2^-$. These spheres all intersect in a common great circle. In order to be able to choose the linking sphere so that it misses the inner cores, we must ensure that this common great circle misses the outer cores of the pre-existing balls.

We now introduce the geometrical tools which enable us to overcome this difficulty. Although these ideas come from projective geometry, it will be more convenient for us to express them in terms of the geometry of S^3 with the round metric.

The intersection of the sphere S^3 with a k-dimensional linear subspace of $\mathbb{R}^4$ is a **great** $(k-1)$**-sphere**. The case $k = 0$ gives us the empty set, which has topological dimension -1 and is the intersection of S^3 with the space $\{0\}$. The cases $k = 1, 2, 3$ give us pairs of antipodal points, great circles and great spheres respectively. If α is a great k-sphere, let V_α denote the corresponding $(k+1)$-dimensional subspace of $\mathbb{R}^4$. The **dual** $\alpha^\perp$ of α is the great $(2-k)$-sphere corresponding to the orthogonal complement in $\mathbb{R}^4$ of V_α. If α and A are great spheres with $\alpha \subset A$ and $\dim \alpha \le k \le \dim A$, we define $\mathcal{S}^k(\alpha, A)$ to be the collection of all great k-spheres S with $\alpha \subseteq S \subseteq A$. If α is a great circle, then $\mathcal{S}^2(\alpha^\perp, S^3)$ is the space of all great 2-spheres that intersect α orthogonally. If $\alpha \subset A$, then $A^\perp \subset \alpha^\perp$.

The set $\mathcal{S}^k(\alpha, A)$ carries a canonical probability measure $\mu_{\mathcal{S}^k(\alpha,A)}$ induced by Haar measure on the group $SO(\alpha, A)$, which consists of orientation preserving orthogonal linear maps of $\mathbb{R}^4$ that map V_α and V_A into themselves. The measure $\mu_{\mathcal{S}^k(\alpha,A)}$ is the unique probability measure on $\mathcal{S}^k(\alpha, A)$ that is invariant under the natural action of $SO(\alpha, A)$. Since $SO(A^\perp, \alpha^\perp) = SO(\alpha, A)$, it is easy to prove

Lemma 4.1. *The bijection $S \mapsto S^\perp$ from $\mathcal{S}^k(\alpha, A)$ to $\mathcal{S}^{2-k}(A^\perp, \alpha^\perp)$ carries $\mu_{\mathcal{S}^k(\alpha,A)}$ to $\mu_{\mathcal{S}^{k-2}(A^\perp,\alpha^\perp)}$.*

Now suppose that we have a collection of balls in S^3. The closures of the balls are disjoint and the sum of their radii is less than a number $\rho < \pi/1000$. Define the **core** and **inner core** of a ball from the collection with radius r to be the concentric balls with radius r^2 and r^3 respectively. Let $\mathcal{B}$ denote the union of the balls, $\mathcal{C}$ the union of their cores, and $\mathcal{IC}$ the union of their inner cores.

Proposition 4.2. *There is $\rho_0 > 0$ such that, if the sum of the radii of the balls is less than ρ_0, then, for any space $\mathcal{S}^k(\alpha, A)$, where α and A are a pair of great spheres with $\alpha \subset A$ and k is an integer with $\dim \alpha < k < \dim A$, we have:*

(1) *If $\alpha \cap \mathcal{B} = \emptyset$, then $\mu_{\mathcal{S}^k(\alpha,A)}\{S \in \mathcal{S}^k(\alpha, A) : S \cap \mathcal{C} = \emptyset\} > .99$.*
(2) *If $A^\perp \cap \mathcal{B} = \emptyset$, then $\mu_{\mathcal{S}^k(\alpha,A)}\{S \in \mathcal{S}^k(\alpha, A) : S^\perp \cap \mathcal{C} = \emptyset\} > .99$.*
(3) *If $\alpha \cap \mathcal{C} = \emptyset$, then $\mu_{\mathcal{S}^k(\alpha,A)}\{S \in \mathcal{S}^k(\alpha, A) : S \cap \mathcal{IC} = \emptyset\} > .99$.*
(4) *If $A^\perp \cap \mathcal{C} = \emptyset$, then $\mu_{\mathcal{S}^k(\alpha,A)}\{S \in \mathcal{S}^k(\alpha, A) : S^\perp \cap \mathcal{IC} = \emptyset\} > .99$.*

Proof. We prove (1) and (3); (2) and (4) then follow easily by applying (1) and (3) to $\mathcal{S}^{2-k}(A^\perp, \alpha^\perp)$ and using Lemma 4.1. Let $\beta = \alpha^\perp \cap A$ and let B be a ball of radius r in S^3. Consider the radial projection of $A \setminus \{\alpha\}$ onto β that is defined by mapping $p \in A \setminus \alpha$ to the intersection with β of the unique sphere in $\mathcal{S}^{1+\dim \alpha}(\alpha, A)$ that contains p.

If $\alpha \cap B = \emptyset$, then this projection maps the core C of the ball B to the union of a ball in β with radius at most r and the antipodal ball in β; and the same is true of the inner core of B if $\alpha \cap C = \emptyset$. There is a measure preserving bijection between $\mathcal{S}^k(\alpha, A)$ and $\mathcal{S}^{k-1}(\emptyset, \beta)$. The proposition follows from the next lemma. ■

Lemma 4.3. *There is a constant $b > 0$ such that, if $0 \le j < l \le 3$ and S is a great l-sphere, then the probability (with respect to $\mu_{\mathcal{S}^j(\emptyset,S)}$) that an element of $\mathcal{S}^j(\emptyset, S)$ intersects a given ball in S of radius $r < \pi/1000$ is at most br.*

Proof. We may assume without loss of generality that the ball in question is the r-neighbourhood in S^l of the north pole $(0, \dots, 0, 1)$. Let $E = S^l \cap (\mathbb{R}^l \times \{0\})$ be the equator. The probability that we wish to estimate is the same as the probability that the first $j+1$ vectors of a randomly chosen positively oriented orthonormal basis of $\mathbb{R}^{l+1}$ span a subspace that does not intersect the cone subtended at the origin by the given ball. If the subspace does intersect this cone, then the last element of the basis, which is orthogonal to the subspace, must lie in the r-neighbourhood N_r of E. The probability of this is the measure of N_r with respect to the Lebesgue measure on S^l (normalized to be a probability measure). It is clear the volume of N_r is $O(r)$. ■

Let us call a great circle σ **good** if $\sigma \cap \mathcal{B} = \emptyset$ and $\sigma^\perp \cap \mathcal{B} = \emptyset$.

Proposition 4.4. *Given $\varepsilon > 0$, we can choose $\rho_\varepsilon > 0$ such that, if the sum of the radii of the balls is less than ρ_ε, then the set $\mathcal{G}$ of good great circles is a subset of $\mathcal{S}^1(\emptyset, S^3)$ satisfying $\mu_{\mathcal{S}^1(\emptyset,S^3)}(\mathcal{G}) \ge 1 - \varepsilon/2$.*

Proof. This follows from Lemmas 4.1 and 4.3. ■

Proposition 4.5. *Assume that the balls satisfy (1)–(4) from Proposition 4.2. Let $X_1 = (\Sigma_1^-, \Sigma_1^+)$ be a cross such that the dual of Σ_1^+ does not intersect the balls. Let σ_2 be a good* *great circle that makes an angle greater than $\pi/4$ with Σ_1^+. Let σ_3 be another great circle. Then there is a cross $X_2 = (\Sigma_2^-, \Sigma_2^+)$ centered at σ_2 such that:*

1. *Σ_2^- and Σ_2^+ do not meet the cores of the balls.*
2. *X_1 and X_2 can be linked by a great sphere Σ that does not intersect the inner cores of the balls.*
3. *The duals of Σ_2^- and Σ_2^+ do not intersect the balls.*
4. *The angle between Σ_2^+ and σ_3 is at least $\pi/4$.*

Proof. Since X_2 is centered at the good great circle σ, the duals of Σ_2^- and Σ_2^+ both lie in $\sigma_2^\perp$, which misses the balls because σ_2 is good. Thus (3) will hold for any choice of X_2.

The cross X_2 is uniquely determined by Σ_2^-. We can think of the choice of Σ_2^- in two different ways.

On the one hand, $\Sigma_2^- \in \mathcal{S}^2(\sigma_2, S^3)$. Since σ_2 is good, it does not meet the balls, and we can apply (1) of Proposition 4.2 to show that the set $\mathcal{G}_1$ of spheres in $\mathcal{S}^2(\sigma_2, S^3)$ that do not meet the cores of the balls satisfies $\mu_{\mathcal{S}^2(\sigma_2,S^3)}(\mathcal{G}_1) \geq .99$. Rotation by $\pi/2$ about σ_2, which moves a choice for Σ_2^- to the corresponding choice for Σ_2^+, is a measure preserving map of $\mathcal{S}^2(\sigma_2, S^3)$. Hence the set $\mathcal{G}_2 \subset \mathcal{G}_1$ consisting of choices for Σ_2^- such that both Σ_2^- and the corresponding Σ_2^+ miss the cores of the balls satisfies $\mu_{\mathcal{S}^2(\sigma_2,S^3)}(\mathcal{G}_2) \geq .98$. Thus most X_2 satisfy (1): more precisely X_2 satisfies (1) if $\Sigma_2^- \in \mathcal{G}_2$.

It is not difficult to show that the set of spheres in $\mathcal{S}^2(\sigma_2, S^3)$ that make angle less than $\pi/4$ with σ_3 has measure at most $1/2$. Hence the set $\mathcal{G}_3 \subset \mathcal{G}_2$ consisting of choices of Σ_2^- such that both (1) and (4) hold satisfies $\mu_{\mathcal{S}^2(\sigma_2,S^3)}(\mathcal{G}_3) \geq .48$.

On the other hand, Σ_2^- is uniquely determined by the great circle $\Sigma_2^- \cap \Sigma_1^+$ in which it intersects $\Sigma+_1$. Let $q = \{q', q''\} = \sigma_2 \cap \Sigma_1^+$. Then $\Sigma_2^- \cap \Sigma_1^+$ lies in $\mathcal{S}^1(q, \Sigma_1^+)$. Since the dual of Σ_1^+ does not meet the balls, we can apply (2) of Proposition 4.2 to show that the set $\mathcal{G}_4$ of $\sigma \in \mathcal{S}^1(q, \Sigma_1^+)$ such that $\sigma^\perp$ misses the cores of the balls satisfies $\mu_{\mathcal{S}^1(q,\Sigma_1^+)}(\mathcal{G}_4) \geq .99$. By (3) of Proposition 4.2, we see that for any $\sigma \in \mathcal{G}_4$ there are spheres in $\mathcal{S}^2(\sigma^\perp, S^3)$ that miss the inner cores of the balls. But $\mathcal{S}^2(\sigma^\perp, S^3)$ is the space of great spheres orthogonal to σ; any of these spheres can be used to link X_1 to the cross X_2 determined by σ. Thus most X_2 satisfy (2): more precisely X_2 satisfies (2) if $\Sigma_2^- \cap \Sigma_1^+ \in \mathcal{G}_4$.

It remains to show that the two notions of "most" in the above statements are sufficiently compatible to enable us to choose Σ_2^- so that $\Sigma_2^- \in \mathcal{G}_3$ and $\Sigma_2^- \cap \Sigma_1^+ \in \mathcal{G}_4$. Thus we wish to show that $\phi^{-1}(\mathcal{G}_4) \cap \mathcal{G}_3 \neq \emptyset$, where $\phi : \mathcal{S}^2(\sigma_2, S^3) \to \mathcal{S}^1(q, \Sigma_1^+)$ takes $S \in \mathcal{S}^2(\sigma_2, S^3)$ to $S \cap \Sigma_1^+$. In order to this, we need to estimate the Jacobian of ϕ with respect to the measures $\mu_{\mathcal{S}^2(\sigma_2,S^3)}$ and $\mu_{\mathcal{S}^1(q,\Sigma_1^+)}$.

Observe that there is a natural identification of $\mathcal{S}^2(\sigma_2, S^3)$ with the set of antipodal pairs in the circle C_1 of unit vectors at q' that are orthogonal to σ_2. With this identification, $\mu_{\mathcal{S}^2(\sigma_2,S^3)}$ is the measure induced on the set of antipodal pairs by Lebesgue measure on C_1. Similarly there is a natural identification of $\mathcal{S}^1(q, \Sigma_1^+)$ with the set of antipodal pairs in the circle C_2 of unit vectors tangent to Σ_1^+ at q'. The measure $\mu_{\mathcal{S}^1(q,\Sigma_1^+)}$ becomes the measure induced on the space of antipodal pairs by Lebesgue measure on C_2.

The Jacobian of ϕ that we wish to estimate is the same as the Jacobian (with respect to the Lebesgue measures) of the map $\widehat{\phi} : C_1 \to C_2$ defined by first projecting vectors in C_1 to the plane $T_{q'}\Sigma_1^+$ along the lines in $T_{q'}S^3$ that are parallel to $T_{q'}\sigma_2$, and then normalizing the resulting vectors. An elementary calculation shows that the Jacobian of $\widehat{\phi}$ lies between $\sin\alpha$ and $\csc\alpha$, where α is the angle between σ_2 and Σ_1^+. Since we have assumed that $\alpha \geq \pi/4$, it follows that the Jacobian of ϕ lies between $1/\sqrt{2}$ and $\sqrt{2}$. Using this and our earlier estimates, we see that $\mu_{\mathcal{S}^2(\sigma_2,S^3)}(\phi^{-1}(\mathcal{G}_4)) + \mu_{\mathcal{S}^2(\sigma_2,S^3)}(\mathcal{G}_3) > 1$. Hence $\phi^{-1}(\mathcal{G}_4) \cap \mathcal{G}_3 \neq \emptyset$, as desired. ■

Proof of Theorem 0.2. Let $\varepsilon > 0$ be given. Choose $\rho > 0$ small enough so that the conclusions of Propositions 4.2, 4.4 and 4.5 hold for any collection of balls whose diameters have sum less than ρ. We shall construct an infinite sequence of orthogonal spheres

$$\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2, \Sigma_1^3, \ldots.$$

The sequence will be nondegenerate, so all of its terms are distinct, any three spheres from the sequence intersect in a pair of antipodal points, and at most three spheres from the sequence can intersect at any point of S^3. These balls will be pairwise disjoint, and the inner core of each of them will intersect with only three spheres from the sequence, namely the three spheres that intersect at its center.

For each triple of consecutive spheres from the sequence, we shall choose a ball centered at one of the antipodal points where the three spheres intersect. We emphasize that the choice of the balls is part of the process of choosing the sequence; we cannot try to choose the balls after first choosing the sequence.

Let $\sigma_1^1 = \Sigma_1^0 \cap \Sigma_1^1$; for $j \geq 2$, let $\sigma_1^j = \Sigma_{m_{j-1}}^{j-1} \cap \Sigma_1^j$; and for $j \geq 1$ and $2 \leq i \leq m_j$, let $\sigma_i^j = \Sigma_{i-1}^j \cap \Sigma_i^j$. Then

$$\sigma_1^1, \ldots, \sigma_{m_1}^1, \sigma_1^2, \ldots, \sigma_{m_2}^2, \sigma_1^3, \ldots$$

is the sequence of great circles associated with our sequence of orthogonal spheres. For each σ_i^j, O_i^j will be one of the two orbits of the geodesic flow formed by unit vectors tangent to σ_i^j. Our construction will ensure that there is a sequence of sets

$$T^1S^3 \supset G_1 \supset G_2 \supset \cdots$$

such that for each $j \geq 1$ the measure of $T^1S^3 \setminus G_j$ is at most $\varepsilon - \varepsilon/2^j$ and every element of G_j will lie within distance $\varepsilon/2^j$ of $O_1^j \cup \cdots \cup O_{m_j}^j$. (As explained at the beginning of Section 2, we use the distance and volume induced on T^1S^3 by the round metric g_0.)

Once we have the sequences of spheres and orbits satisfying the above properties, the proof of Theorem 0.2 is simple. As in the proof of Theorem 0.1, we see from Proposition 2.2 that we can make perturbations of the metric that are supported in the inner cores of the balls so as to make each of the orbits in the sequence

$$\mathcal{O} = O_1^1, \ldots, O_{m_1}^1, O_1^2, \ldots, O_{m_2}^2, O_1^3, \ldots$$

hyperbolic. At the same time we can create heteroclinic orbits joining each pair of consecutive terms of $\mathcal{O}$. More precisely, for each pair of consecutive terms in $\mathcal{O}$, we construct an orbit in the intersection of the unstable manifold of the earlier term and the stable manifold of the later term. We now apply Proposition 2.5 to each orbit from $\mathcal{O}$ (except

O_1^1). We see that, by making perturbations to the metric in the inner cores of the balls, we can create a single orbit O which contains long pieces from each of the heteroclinic orbits. We can arrange that, for each $j \geq 1$, the orbit O passes within distance $\varepsilon/2^j$ of each element of $O_1^j \cup \cdots \cup O_{m_j}^j$. Hence, for each $j \geq 1$, the orbit O passes within distance $\varepsilon/2^{j-1}$ of each element of G_j. Thus the closure of the orbit O contains the set $\bigcap_{j=1}^{\infty} G_j$, which has measure at least $1-\varepsilon$.

It remains to construct sequences of orbits, balls and orthogonal spheres that satisfy all of the above properties. This is done by iterating a construction very similar to that used in the proof of Theorem 0.1. We shall outline the first two steps.

In the first step, we choose orbits $O_1^1, \ldots, O_{m_1}^1$, great spheres $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$, and great circles $\sigma_1^1, \ldots, \sigma_{m_1}^1$ so that:

(1) $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$ is a nondegenerate sequence of orthogonal spheres.
(2) $\sigma_1^1, \ldots, \sigma_{m_1}^1$ is the sequence of great circles associated with $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$.
(3) The vectors in O_i are tangent to σ_i^1.
(4) $O_1^1 \cup \cdots \cup O_{m_1}^1$ is $\varepsilon/2$ dense in $G_0 := T^1S^3$.
(5) The pair of antipodal points dual to $\Sigma_{m_1}^1$ is disjoint from all the spheres $\Sigma_1^0, \ldots, \Sigma_{m_1}^1$ except for $\Sigma_{m_1-1}^1$.

The above can be achieved by following the construction used in the proof of Theorem 0.1. We start with a set of orbits whose union is $\varepsilon/2$ dense in T^1S^3, choose a sequence of crosses centered at the great circles to which these orbits are tangent and then link the crosses to obtain a sequence of orthogonal spheres; these are the spheres $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$. As we explained in Section 3, we can ensure that the sequence is nondegenerate. After having done this, we perturb $\Sigma_{m_1}^1$ among the great spheres orthogonal to $\Sigma_{m_1-1}^1$ so that the dual of $\Sigma_{m_1}^1$ does not coincide with any of the pairs of antipodal points where three spheres from the sequence intersect. This is possible because each pair of antipodal points in $\Sigma_{m_1-1}^1$ is the dual of a great sphere orthogonal to $\Sigma_{m_1-1}^1$. Finally the σ_i^1 are chosen to be the great circles associated with $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$ and the orbits O_i^1 are chosen so that the vectors in O_i^1 are tangent to σ_i^1 and the orbits O_i^1 include the orbits with which we began.

We now choose a sequence of balls. Let p_i^1, $1 \leq i \leq m_1 - 1$, be one of the two points where σ_i^1 and σ_{i+1}^1 intersect. Thus $p_1^1 \in \Sigma_1^0 \cap \Sigma_1^1 \cap \Sigma_2^1$ and $p_i^1 \in \Sigma_{i-1}^1 \cap \Sigma_i^1 \cap \Sigma_{i+1}^1$ for $2 \leq i \leq m_1 - 1$. We choose a ball B_i^1 centered at each point p_i^1 such that:

(1) The balls are pairwise disjoint;
(2) Exactly three of the spheres $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1$ enter the inner core of each ball.
(3) The dual of $\Sigma_{m_1}^1$ is outside the balls.
(4) The sum of the diameters of $B_1^1, \ldots, B_{m_1-1}^1$ is at most $\rho/2$.
(5) The set G_1 of vectors in T^1S^3 that are good with respect to $B_1^1, \ldots, B_{m_1-1}^1$ has measure at least $1 - \varepsilon/2$.

These properties can be achieved simply by choosing the balls small enough. This completes the first step.

In the second step, we choose $O_1^2, \ldots, O_{m_2}^2$, great spheres $\Sigma_1^2, \ldots, \Sigma_{m_1}^2$, and great circles $\sigma_1^2, \ldots, \sigma_{m_2}^2$ so that:

(1) None of the spheres $\Sigma_1^2, \ldots, \Sigma_{m_2}^2$ intersect with the inner cores of any of the balls

$B_1^1, \ldots, B_{m_1-1}^1$.
(2) $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2$ is a nondegenerate sequence of orthogonal spheres.
(3) $\sigma_1^2, \ldots, \sigma_{m_1}^2$ is the sequence of great circles associated with the sequence of orthogonal spheres $\Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2$.
(4) σ_i^2 is the great circle to which the unit vectors belonging to the orbit O_i^2 are tangent.
(5) $O_1^2 \cup \cdots \cup O_{m_2}^2$ is $\varepsilon/4$ dense in G_1.
(6) The pair of antipodal points dual to $\Sigma_{m_2}^2$ is disjoint from all the spheres $\Sigma_1^0, \ldots, \Sigma_{m_2}^2$ except for $\Sigma_{m_2-1}^2$.

The process of choosing these objects is a little more complicated than the equivalent part of step one because of the extra constraints involving the balls. We start with a sequence of orbits whose union is an $\varepsilon/4$ dense subset of G_1. We also ensure that the great circle to which the first of these orbits is tangent makes angle greater than $\pi/4$ with the sphere $\Sigma_{m_1}^1$. Now we can repeatedly apply Proposition 4.5 to choose a sequence of crosses centered at the great circles to which these orbits are tangent so that:

(1) None of spheres in the crosses enters the cores of the balls $B_1^1, \ldots, B_{m_1-1}^1$.
(2) The first cross (which will be called (Σ_2^2, Σ_3^2)) can be linked to $(\Sigma_{m_1-1}^1, \Sigma_{m_1}^1)$ by a great sphere Σ_1^2 that does not enter the inner cores of $B_1^1, \ldots, B_{m_1-1}^1$.
(3) Each succeeding cross can be linked to the previous cross by a great sphere that does not enter the inner cores of $B_1^1, \ldots, B_{m_1-1}^1$.

In this way we choose great spheres $\Sigma_1^2, \ldots, \Sigma_{m_2}^2$ such that

$$\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2$$

is a sequence of orthogonal spheres. As in the first step, this sequence can be made nondegenerate by arbitrarily small perturbations, and we can perturb $\Sigma_{m_2}^2$ among the great spheres orthogonal to $\Sigma_{m_2-1}^2$ so that the dual of $\Sigma_{m_2}^2$ does not coincide with any of the pairs of antipodal points where three spheres from the sequence intersect. The σ_i^2 are chosen to be the great circles associated with $\Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2$ and the orbits O_i^2 are chosen so that the vectors in O_i^2 are tangent to σ_i^2 and the orbits O_i^2 include the orbits with which we began step 2.

We conclude step 2 by extending the sequence of balls. Let $p_{m_1}^1$ be one of the two points where $\sigma_{m_1}^1$ and σ_1^2 intersect, and for $1 \leq i \leq m_2 - 1$ let p_i^2 be one of the two points where σ_i^2 and σ_{i+1}^2 intersect. Thus $p_{m_1}^1 \in \Sigma_{m_1-1}^1 \cap \Sigma_{m_1}^1 \cap \Sigma_1^2$, $p_1^2 \in \Sigma_{m_1}^1 \cap \Sigma_1^2 \cap \Sigma_2^2$ and $p_i^2 \in \Sigma_{i-1}^2 \cap \Sigma_i^2 \cap \Sigma_{i+1}^2$ for $2 \leq i \leq m_2 - 1$. We choose a ball $B_{m_1}^1$ centered at $p_{m_1}^1$ and a ball B_i^2 centered at each point p_i^2 such that:

(1) The balls $B_1^1, \ldots, B_{m_1}^1, B_1^2, \ldots, B_{m_2-1}^2$ are pairwise disjoint;
(2) Exactly three of the spheres $\Sigma_1^0, \Sigma_1^1, \ldots, \Sigma_{m_1}^1, \Sigma_1^2, \ldots, \Sigma_{m_2}^2$ enter the inner core of each ball.
(3) The dual of $\Sigma_{m_2}^2$ is outside the balls $B_1^1, \ldots, B_{m_1}^1, B_1^2, \ldots, B_{m_2-1}^2$.
(4) The sum of the diameters of $B_1^1, \ldots, B_{m_1}^1, B_1^2, \ldots, B_{m_2-1}^2$ is at most $\rho/2 + \rho/4$.
(5) The set G_2 of vectors in T^1S^3 that are good with respect to the sequence of balls $B_1^1, \ldots, B_{m_1}^1, B_1^2, \ldots, B_{m_2-1}^2$ has measure at least $1 - \varepsilon/2 - \varepsilon/4$.

These properties are obtained in part from our choice of the spheres Σ_i^2 and in part by choosing the new balls sufficiently small.

The proof of Theorem 0.2 is completed by iterating the above construction, and then perturbing the metric as described above. ■

References

[A1] V. Arnold, *Proof of A. N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian*, Russian Math. Surveys **18** (1963), no. 5 pages 9–36; English translation.

[A2] V. Arnold, *Small denominators and problems of stability of motion in classical and celestial mechanics* **18** (1963), no. 6, 85–192; English translation.

[BTZ] W. Ballmann, G. Thorbergsson, and W. Ziller, *Closed geodesics on positively curved manifolds*, Ann. of Math. **116** (1982), 213–247.

[BG] K. Burns and M. Gerber, *Real analytic Bernoulli geodesic flows on S^2*, Ergod. Th. & Dynam. Sys. **9** (1989), 27–45.

[D1] V. Donnay, *Geodesic flow on the two–sphere, Part I: Positive measure entropy*, Ergod. Th. & Dynam. Sys. **8** (1989), 531–553.

[D2] V. Donnay, *Transverse homoclinic connections for Geodesic Flows*, IMA Vol. Math. Appl. **63** (1995), Springer Verlag, 115–125.

[DoC] M. Do Carmo, *Riemannian Geometry*, Birkhäuser, Boston - Basel - Berlin, 1992.

[E] P. Eberlein, *When is a geodesic flow of Anosov type? I*, J. Differential Geometry **8** (1973), 437–463.

[F] J. Franks, *Geodesics on S^2 and periodic points of annulus homeomorphisms*, Invent. Math. **108** (1992), 403–418.

[G] L. Green, *A theorem of E. Hopf*, Michigan Math. J. **7** (1956), 31–34.

[HPS] M. Hirsch, C. Pugh and M. Shub, *Invariant Manifolds*, Springer Lecture Notes in Mathematics, vol. 583, Springer Verlag, New York-Heidelberg-Berlin, 1977.

[Ka1] A. Katok, *Ergodic perturbations of degenerate integrable Hamiltonian systems*, Izv. Akad. Nauk. SSSR **37** (1973), 539–576; English translation in Math USSR-Izv. **7** (1973), 535–572.

[Ka2] A. Katok, *Dynamical Systems with Hyperbolic Structure*, Am. Math. Soc. Transl. **116** (1981), 43–96.

[Kl] W. Klingenberg, *Lectures on Closed Geodesics*, Grundlehren Math. Wiss. Bd. 230, Springer Verlag, New York - Heidelberg - Berlin, 1982.

[Ko] A. Kolmogorov, *On conservation of conditionally periodic motions under small perturbations of the Hamiltonian*, Dokl. Akad. Nauk SSSSR **98** (1954), 527–530.

[KW] G. Knieper and H. Weiss, *A surface with positive curvature and positive topological entropy*, J. Diff. Geometry **39** (1994), 229–249.

[L] V. F. Lazutkin, *The existence of caustics for a billiard problem in a convex domain*, Math. USSR Isvestija **7** (1973), 185–214.

[P] D. Petroll, *Existenz und Transversalität von homoklinen und heteroklinen Orbits beim geodätischen Fluss*, Universität Freiburg Thesis (1996).

[S] S. Smale, *Diffeomorphisms with many periodic points*, Differential and Combinatorical Topology (edited by S. S. Cairnes) (1965), Princeton University Press, 63-80.

[Z] W. Ziller, *Geometry of the Katok examples*, Ergod. Th. & Dynam. Sys. **3** (1983), 135–157.

Keith Burns
Department of Mathematics
Northwestern University
Evanston, IL 60208
U.S.A.
Email: burns@math.nwu.edu

Howard Weiss
Department of Mathematics
Pennsylvania State University
University Park, PA 16802
U.S.A.
Email: weiss@math.psu.edu