Annals of Mathematics, **140** (1994), 295–329

Stably ergodic diffeomorphisms

By MATTHEW GRAYSON, CHARLES PUGH, and MICHAEL SHUB

Introduction

Ergodicity (time averages equal space averages) is the most basic, global, gross, statistical feature of conservative dynamical systems. Yet it was not until twenty-five years ago that D.V. Anosov [1] established the existence of open sets of ergodic systems on a wide range of compact manifolds. These systems, now called "Anosov", are globally uniformly hyperbolic. For diffeomorphisms this means that the derivative is the direct sum of a uniform expansion and a uniform contraction. Anosov systems are structurally stable—they are homeomorphically conjugate to their perturbations. All dynamical features of Anosov systems and their perturbations are the same at all scales. If, however, there are center directions (directions invariant under the diffeomorphism but neutral with respect to expansion and contraction) then the situation is rather different. Indeed, the work of Kolmogorov-Arnol'd-Moser (KAM) on invariant tori implies that ergodicity is often impossible.

Here we show that the KAM center theory may sometimes be irrelevant to ergodicity, even in the non-hyperbolic case—we construct examples of ergodic diffeomorphisms which are not hyperbolic, not structurally stable, but which remain ergodic after volume-preserving perturbations. They are **stably ergodic** but not structurally stable. Center directions exist but the fine scale structure of the dynamics in the center direction plays no role in the analysis. Instead, ergodicity results from the behavior of the strong unstable and stable directions.

The basic geometric Anosov system is the geodesic flow φ on M^3, the unit tangent bundle of a compact surface with constant negative Gaussian curvature. The flow φ is ergodic with respect to the natural Liouville measure on M^3, and so are the geodesic flows for perturbations of the Riemann structure. (This two-dimensional differential geometry result predates the work of Anosov and is due to Hedlund (1934) and Hopf (1939).) Our example is simply the time-one map

$$\varphi_1 : M^3 \to M^3.$$

As does any smooth flow, $T\varphi_1$ sends the vector field $\dot{\varphi}$ to itself. Thus, span $\dot{\varphi}$ is an invariant center direction, and φ_1 is non-hyperbolic. Although the flow φ is structurally stable as a flow, its time-one map is not structurally stable as a diffeomorphism. To see this, observe that φ_1 leaves invariant all the orbits of the flow, and some of these orbits are periodic—they are diffeomorphic copies of the circle. The restriction of φ_1 to a periodic orbit is a circle rotation, perturbations of which can dramatically alter the number and type of periodic points of φ_1. This precludes structural stability.

THEOREM 1. *The time-one map of the geodesic flow on the unit tangent bundle of a surface of constant negative curvature is stably ergodic—it is ergodic and so are all C^2 small volume-preserving perturbations of it.*

Let us say a few words about how we became interested in stably ergodic dynamics. Our work on the existence of open sets of non-Anosov topologically transitive diffeomorphisms on the 4-torus, announced in Shub ([12], [13]) and published in Hirsch, Pugh, and Shub [6], and the appearance in English of Anosov-Sinai [2] led us to wonder about the corresponding phenomenon for ergodicity. It was well known that a linear automorphism of the torus is ergodic if and only if none of its eigenvalues are roots of unity. William Parry [9] proved that toral automorphisms (or automorphisms of nil-manifolds) are ergodic if and only if their strong stable manifolds are dense. We asked whether nonlinear, non-measure-preserving perturbations of such non-hyperbolic ergodic automorphisms of the torus remain topologically transitive and ergodic. Theorem 1 answers the ergodicity question in the φ_1 context. The corresponding topological transitivity question for non-measure-preserving perturbations of φ_1 remains open.

Here is an example. Consider the 4×4 matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -6 \\ 0 & 0 & 1 & 8 \end{pmatrix}.$$

Peter Walters [15] proved that the eigenvalues of A are not roots of unity although two of them lie on the unit circle. It follows that the diffeomorphism $A\colon T^4 \to T^4$ is ergodic, topologically transitive, but not Anosov. Even for this example the following question is open:

> Is every small volume-preserving perturbation of A ergodic and/or topologically transitive?

The idea of the proof Theorem 1 is as follows. The diffeomorphism φ_1 is normally hyperbolic respecting the foliation of M^3 by the φ-orbits, and its unstable and stable manifolds foliate M^3. These are the horocycle foliations $\mathcal{H}^{\mathrm{u}}, \mathcal{H}^{\mathrm{s}}$. Let $f\colon M^3 \to M^3$ be a volume-preserving perturbation of φ_1. Stable manifold theory shows that f leaves invariant foliations $\mathcal{W}^{\mathrm{u}}, \mathcal{W}^{\mathrm{s}}$ similar (but perhaps not homeomorphic) to the horocycle foliations. It also leaves invariant a center foliation $\mathcal{W}^{\mathrm{c}}$ which is homeomorphic to the orbit foliation $\mathcal{H}^{\mathrm{c}}$ of φ. According to Pugh and Shub [10], the foliations $\mathcal{W}^{\mathrm{u}}, \mathcal{W}^{\mathrm{s}}$ are absolutely continuous and their Radon-Nikodym derivatives are uniformly bounded on reasonable transversals. See Lemma 2.2.

To analyze ergodicity of f we adapt the ideas from Hopf ([7]; see also Pugh and Shub [10, p. 20]), and consider the forward, backward, and bi-directional Birkhoff averages of functions $g\colon\ M^3 \to \mathbb{R}$ along f-orbits,

$$\mathcal{B}_N g(x) = \frac{1}{N+1}\sum_{n=0}^{N} g(f^n(x)), \quad \mathcal{B}_{-N} g(x) = \frac{1}{N+1}\sum_{n=0}^{-N} g(f^n(x)),$$

$$\mathcal{B}_{-N,N} g(x) = \frac{1}{2N+1}\sum_{n=-N}^{N} g(f^n(x)).$$

If g is any $\mathcal{L}^1$ function, these Birkhoff averages converge almost everywhere to limit functions $\mathcal{B}_+ g, \mathcal{B}_- g$, and $\mathcal{B}g$ as $N \to \infty$. These three functions are invariant under f and are equal almost everywhere. In fact the map

$$\mathcal{B}\colon\ g \mapsto \mathcal{B}g$$

is a continuous linear projection from $\mathcal{L}^1$ onto the closed linear subspace $\operatorname{Inv} f$ of f-invariant functions. See Mañé [8, pp. 89–100].

Suppose that f is not ergodic. Then $\operatorname{Inv} f$ includes essentially nonconstant functions—functions which are nonconstant on every set of full measure. The set $\mathcal{C}$ of continuous functions is dense in $\mathcal{L}^1$ and a continuous transformation carries dense subsets of the domain to dense subsets of the range. Thus, $\mathcal{B}(\mathcal{C})$ is dense in $\operatorname{Inv} f$, and for some continuous function g, $G = \mathcal{B}g$ is essentially nonconstant. However, except for a zero set, G is constant along stable leaves and along unstable leaves. For if y, y' belong to a common stable leaf then the distance between $f^n(y)$ and $f^n(y')$ is tiny when n is large. Continuity of g implies that $g(f^n(y)) - g(f^n(y'))$ is small. The forward Birkhoff average $\mathcal{B}_N g(y)$ is insensitive to the first few terms in the sum, so $\mathcal{B}_+ g(y)$ exists if and only if $\mathcal{B}_+ g(y')$ exists, and when they do exist they are equal. Since we know that $\mathcal{B}_+ g(y)$ exists almost everywhere, it follows from the absolute continuity of $\mathcal{W}^{\mathrm{s}}$ that for almost all $\mathcal{W}^{\mathrm{s}}$-leaves, $\mathrm{W}^{\mathrm{s}}(y)$, $\mathcal{B}_+ g$ exists everywhere on $W^{\mathrm{s}}(y)$ and is constant along the leaf. The same is valid for unstable leaves respecting backward Birkhoff averages. Since $\mathcal{B}_+ g(y) = \mathcal{B}_- g(y) = G$ almost everywhere, we deduce that our essentially nonconstant function G is essentially constant along stable and unstable leaves.

For any value $v \in \mathbb{R}$, divide M^3, modulo a zero set, as $M^3 = A \sqcup B$ where $A = \{p \in M^3 \colon G(p) \le v\}$ and $B = \{p \in M^3 \colon G(p) > v\}$. Since G is $\mathcal{L}^1$, A and B are measurable, and since G is essentially nonconstant there exists a choice of v so that A and B both have positive measure. Since G is essentially constant along unstable and stable leaves, A and B consist (modulo zero sets) of essentially complete unstable and stable leaves. That is, A and B are **essentially $\mathcal{W}^{u,s}$-saturated.** (In contrast, a set is **completely** $\mathcal{W}^{u,s}$-saturated if it consists of complete (whole) unstable leaves and complete stable leaves.) If A is essentially $\mathcal{W}^{u,s}$-saturated one can replace it by its **unstable saturate** $\mathrm{Sat}^u A =$ the union of the unstable leaves it meets essentially. This does not affect its measure; however $\mathrm{Sat}^u A$ may fail to be essentially $\mathcal{W}^s$-saturated. In other words essential $\mathcal{W}^{u,s}$-saturation of A and B *cannot* be strengthened to complete $\mathcal{W}^{u,s}$-saturation.

The horocycle foliations have the property that any two points in M^3 are connected by a path of unit horocycle arcs, and the number of arcs in the path is uniformly bounded. This is due to the fact that the bracket of the unstable and stable horocycle directions is the geodesic direction,

$$[H^u, H^s] = -\dot{\varphi}.$$

Although the unstable and stable directions for f may not be of class C^1, and so their bracket makes no sense, and although the foliations $\mathcal{W}^u, \mathcal{W}^s$ need not be homeomorphic to the horocycle foliations, they do share the property that any two points of M^3 can be joined by a path of N unit unstable and stable arcs where N is fixed. We call such a path a **$\mathcal{W}^{u,s}$-path**. See Figure 1 and Lemma 1.1 below.

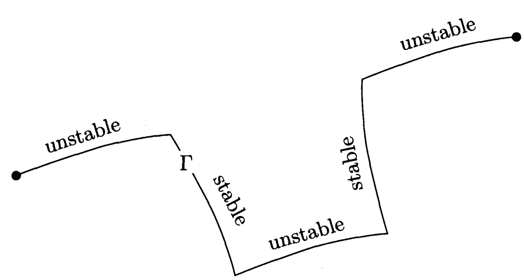

FIGURE 1. A $\mathcal{W}^{u,s}$ path.

Let a be a density point of A and b be a density of B. They are joined by a $\mathcal{W}^{u,s}$-path Γ, and we claim that this leads to a contradiction. Here are two similar, false proofs of this fact. Both rely on the correct intuition that sliding along Γ forces too much of A to exist near b. See Figure 2.

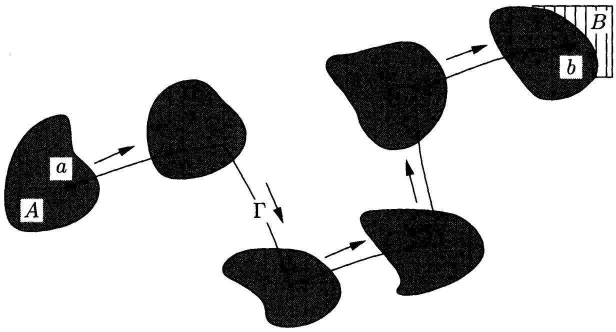

FIGURE 2. A slides to B along a $\mathcal{W}^{u,s}$ path.

Assume that the center foliation $\mathcal{W}^c$ is absolutely continuous. Since A and B have positive measure they meet many center leaves in sets of positive leaf measure. Thus, we may assume that a is a density point of $A \cap W^c(a)$ and b is a density point of $B \cap W^c(b)$. The natural thing to do is to start at a and slide along the foliations $\mathcal{W}^u, \mathcal{W}^s$ in the neighborhood of Γ, using $\mathcal{W}^u$ along the unstable arcs and $\mathcal{W}^s$ along the stable ones. Sliding along the unstable foliation sends center manifolds to center manifolds because the unstable manifolds foliate the center unstable manifolds. The same is true for the stable manifolds. See Figure 3.

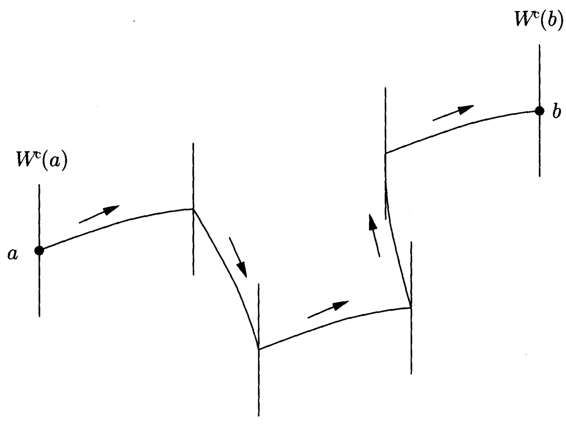

FIGURE 3. The concatenated holonomy map $\theta\colon W^c(a) \to W^c(b)$.

Thus, the concatenated holonomy map along Γ, $\theta\colon W^c(a) \to W^c(b)$, is absolutely continuous with bounded Radon-Nikodym derivative. It carries $A \cap W^c(a)$ onto $A \cap W^c(b)$ since A is essentially $\mathcal{W}^{u,s}$-saturated. The map θ is a C^1 diffeomorphism (see Lemma 2.3) and therefore does not affect density points, so we see that b is a density point of both $A \cap W^c(b)$ and $B \cap W^c(b)$. Since A and B are disjoint this is impossible.

The first defect in this proof is the assumption that $\mathcal{W}^c$ is absolutely continuous. Although the orbit foliation of φ is absolutely continuous, being smooth, there is no reason to expect that the same is true for its perturbation. Eventually we hope to find a dynamic example of this pathology. The second defect is that essential $\mathcal{W}^{u,s}$-saturation of A does not imply that θ carries $A \cap W^c(a)$ to $A \cap W^c(b)$. Complete saturation is necessary to deduce this.

Here is a second false proof. Instead of using the unstable and stable foliations to define holonomy maps on transverse center manifolds, one can use them to slide a three-dimensional neighborhood U of a along Γ. See Figure 2. (Points of U slide along unstable arcs neighboring the first leg of Γ, then along stable arcs neighboring the second leg of Γ, etc.) This gives a homeomorphism $\theta\colon U \to V$ with $\theta(a) = b$ and V a neighborhood of b. By absolute continuity of $\mathcal{W}^u, \mathcal{W}^s$, θ and θ^{-1} are absolutely continuous. Since A is essentially $\mathcal{W}^{u,s}$-saturated, $\theta(A \cap U) = A \cap V$ except for a zero set. Since $\theta(a) = b$, b is a density point of both A and B, a contradiction.

The defect of this proof is the tacit assumption that θ carries density points to density points. Although a Lipeomorphism between open sets of Euclidean space has this property,* the existence of a bounded Radon-Nikodym derivative does not imply Lipschitzness in dimensions > 1. Indeed, it is a sad fact that absolutely continuous homeomorphisms can destroy density points, as is shown in the following picture, Figure 4, of a smooth area-preserving

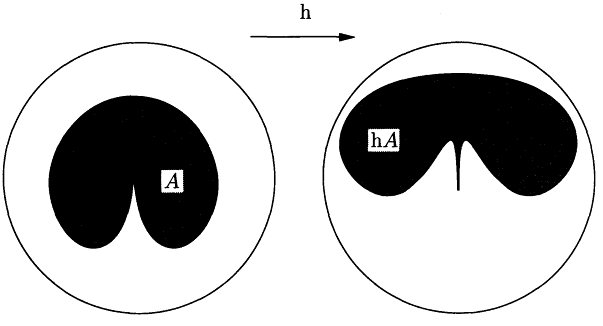

FIGURE 4. A smooth area-preserving homeomorphism which destroys a density point.

homeomorphism of the 2-disc which destroys the density of A at the cusp. Also, although it is clear that an absolutely continuous homeomorphism preserves almost all density points, this is not good enough for our purposes.

---

*Z. Buczolich [4] proved this is also true for non-open subsets of Euclidean space.

To get around the problem of density point destruction we resort to a finer description of the way A looks at the density point a. Density points of A have more regularity than density points of general measurable sets. The set A is not merely highly concentrated in small regular neighborhoods U of a (in terms of measure) but it is also spread through U thoroughly in terms of thin substrips, called "juliennes". See Section 3. It turns out that sliding along $\mathcal{W}^{\mathrm{u}}, \mathcal{W}^{\mathrm{s}}$ preserves this stronger type of density, and we do get a contradiction after all.

Brin and Pesin [3] also studied ergodic properties of partially hyperbolic systems and focused on the transitivity of the strong stable and unstable laminations. This transitivity is central to our analysis. Their techniques prove and generalize the ergodicity of the time-t map of the geodesic flow on surfaces of negative curvature and non-hyperbolic toral automorphisms with dense stable manifolds, but do not extend to perturbations. The false proofs we have given above have their correct counterparts in Brin and Pesin. We thank Tolya Katok for pointing out their work to us.

1. Transitivity

Recall that φ is the geodesic flow on the unit tangent bundle M of a compact surface with constant negative curvature. The manifold M is three-dimensional and φ is hyperbolic respecting the splitting $TM = H^{\mathrm{u}} \oplus H^{\mathrm{c}} \oplus H^{\mathrm{s}}$. The center direction H^{c} is the span of the vector field $\dot{\varphi}$. Tangent to H^{u} and H^{s} are the horocycle foliations $\mathcal{H}^{\mathrm{u}}$ and $\mathcal{H}^{\mathrm{s}}$; tangent to the H^{c} is the orbit foliation. All three foliations are φ-invariant and smooth, in fact, analytic. Similarly, if f C^2-approximates φ_1 then f leaves invariant a unique splitting $TM = E^{\mathrm{u}} \oplus E^{\mathrm{c}} \oplus E^{\mathrm{s}}$ which approximates the horocycle splitting. Tangent to E^{u}, E^{c}, and E^{s} are unique f-invariant foliations $\mathcal{W}^{\mathrm{u}}$, $\mathcal{W}^{\mathrm{c}}$, and $\mathcal{W}^{\mathrm{s}}$. A path consisting of arcs alternately in $\mathcal{W}^{\mathrm{u}}$-leaves and in $\mathcal{W}^{\mathrm{s}}$-leaves is a $\mathcal{W}^{\mathrm{u,s}}$ **path**.

1.1 LEMMA. *Any two points in M are connected by a $\mathcal{W}^{\mathrm{u,s}}$ path consisting of at most N arcs, each of length at most $\sqrt{2}$, where N depends only on the manifolds M.*

We give two proofs, the first geometric and the second topological.

Proof #1. We start with the horocycle foliations $\mathcal{H}^{\mathrm{u}}$ and $\mathcal{H}^{\mathrm{s}}$. Since the unit tangent bundle of the hyperbolic plane $T_1(\mathbb{H}^2)$ covers our manifold, we will work there and cover a fundamental domain.

Given a unit vector v in $T_1(\mathbb{H}^2)$, there are two points at infinity, v^+ and v^-, corresponding to the forward and backward limit points of the geodesic through v. The stable (unstable) leaf through v corresponds to the unit vectors normal to the horocycle through v centered about v^+ (v^-). Thus an $\mathcal{H}^{u,s}$-path in M^3 lifts to (the unit normal vectors to) a path of sequentially tangent horocycle arcs in $\mathbb{H}^2$.

To cover a large open set in $T_1(\mathbb{H}^2)$, we must be able to translate and rotate an arbitrary vector. We rotate via horocyclic triangles: Three distinct points at infinity in the hyperbolic plane define an ideal triangle. Centered on those ideal points are three pairwise tangent horocycles. By arranging the three ideal points in the upper half-plane model to lie at -1, 1, and ∞, we can calculate the side length of the horocyclic triangle which results in the center. See Figures 5a, 5b. The metric on the line $\mathrm{Im}(z) = 2$ is $ds = dx/2$. Therefore,

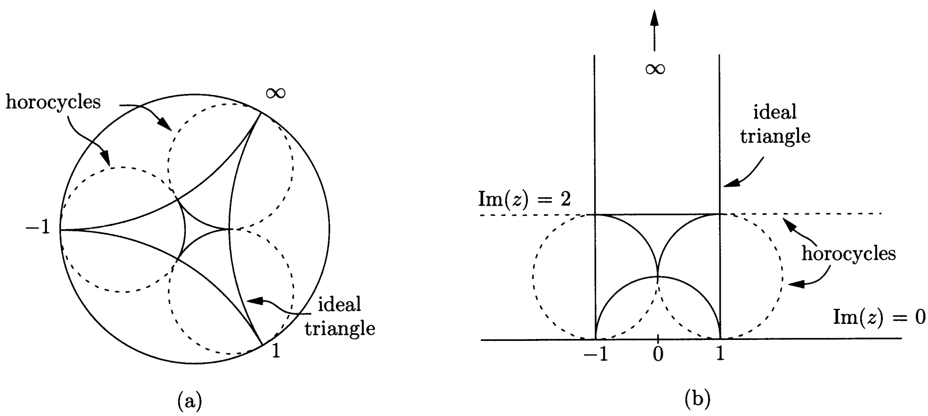

FIGURE 5. (a) Poincaré disk model. (b) Upper half-plane model

this equilateral horocyclic triangle has unit length sides. Following the sides of this triangle realizes a rotation by 180°. See Figure 6. Shortening the sides of

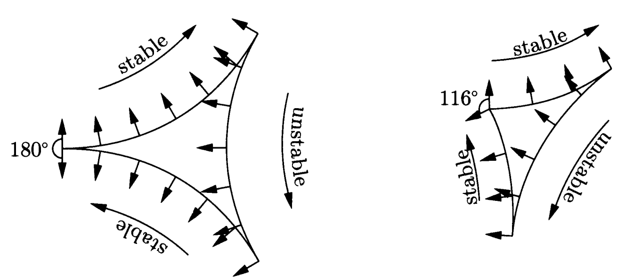

FIGURE 6. Following the sides of horocyclic triangles yields rotation.

the triangle, while retaining two of the tangencies, yields all smaller rotations. To achieve translation, we use a pair of horocycle arcs. Two of unit length produce a translation by over 1.9. See Figure 7. By now, the astute reader

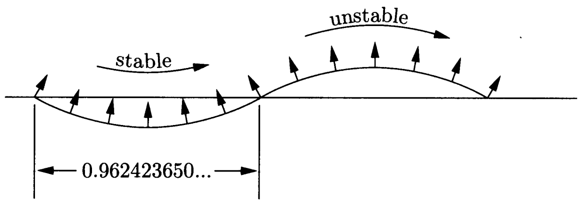

FIGURE 7. Following these horocyclic arcs yields translation.

should be annoyed by our lack of a metric on M. There is some ambiguity in choosing a metric, even if we require invariance under the action of the isometries of $\mathbb{H}^2$. The most natural metric assigns length 2π to the unit circle in the tangent plane at a point, and is the standard hyperbolic metric for parallel translations. Since horocycles have constant geodesic curvature $\equiv 1$, the translational and rotational parts are equal, and the length of an $\mathcal{H}^{u,s}$-path in M is precisely $\sqrt{2}$ times the hyperbolic length of its corresponding horocyclic path.

With the moves described above, it is possible to cover a ball of diameter R in $T_1(\mathbb{H}^2)$ with a three-parameter family of $\mathcal{H}^{u,s}$-paths, each made up of $\leq 1.1\text{R} + 12$ arcs of length $\leq \sqrt{2}$. The 12 comes from triangles at the beginning and end to aim and to change directions, each triangle being allowed two circuits to avoid trouble with the $-\pi = \pi$ identification; we cover $(-2\pi, 2\pi)$, and thus we cover the full $[-\pi, \pi]$ in an open fashion. See Figure 8. The

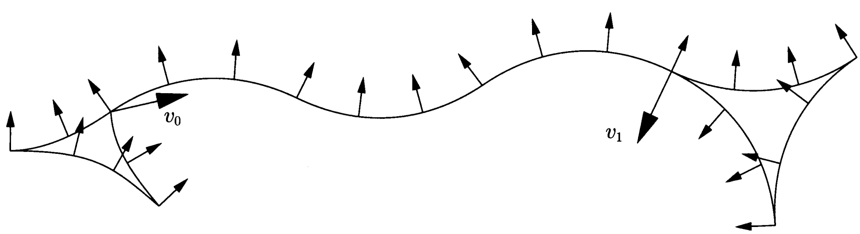

FIGURE 8. Horocyclic path from v_0 to v_1.

three parameters are the lengths, l_1 and l_2, of the edges of the two horocyclic triangles opposite the endpoint vectors and the length l_3 of each arc in the translational part. These are analogous to θ, z, and r in Euclidean cylindrical coordinates. The length of the first triangle base determines the angle of translation, the last triangle edge determines the point in the target fiber, and the translational length is, after a multiplicative correction between .9 and 1, the radial distance to the base point of the target fiber. With the open subset of the parameters $.1 < l_1, l_2, l_3 < .9$, the image of a fixed vector is an open subset of M, and the derivative map is surjective (the exact Jacobian is a nice exercise in plane hyperbolic geometry). Geometrically, we cover every fiber whose base point lies in an annulus with inner radius .2R and outer radius 1.6R. (The 1.6 is less than $2(.9^2)$.) If R is the diameter of a fundamental domain, then the preimage of this annulus contains at least one complete closed fundamental domain. An open map from the interior of one compact set, covering another compact set, which is a local homeomorphism on the preimage of the interior of the target is still onto after a continuous perturbation. Q.E.D.

The second proof of Lemma 1.1 is somewhat more general. It has two steps. Given smooth vector fields X, Y on a manifold M, an **X, Y path of length $\leq N$** consists of $\leq N$ concatenated trajectory arcs of the X- and Y-flows, each having time length < 1.

PROPOSITION 1.2. *If X,Y,Z are smooth, pointwise linearly independent vector fields on a compact 3-manifold M, $[X,Y] = Z$, and Z has a dense orbit, then there is an N such that any pair of points in M can be joined by an X,Y path of length $\leq N$.*

Proof. Let φ, ψ, and η be the X, Y, and Z flows. Fix $\mathrm{p} \in M$ and consider the map

$$\Phi: \ (r, s, t) \mapsto \varphi_r \circ \psi_s \circ \eta_t(\mathrm{p})$$

where r, s, t are small. Since X, Y, Z do not vanish at p, Φ is a diffeomorphism from a neighborhood of $(0,0,0)$ in $(-1,1)^3 \subset \mathbb{R}^3$ to a neighborhood U of p in M. By compactness of M, these neighborhoods U have inner radius ρ_{p} uniformly bounded away from zero, say $\rho_{\mathrm{p}} \geq \rho > 0$ for all $\mathrm{p} \in M$. Let U_{p} denote the open neighborhood of radius ρ at p, and let γ be a dense orbit of Z. There is a constant L such that any two of these neighborhoods U_{p} and U_{q} are joined by an arc of γ having time length $\leq L$. Otherwise, there is a sequence of pairs $\mathrm{p}_n, \mathrm{q}_n$ such that the shortest time for γ to transit from U_{p_n} to U_{q_n} is $\geq n$. By compactness, we may assume $\mathrm{p}_n \to \mathrm{p}$ and $\mathrm{q}_n \to \mathrm{q}$. Since γ is dense there is an arc $\gamma_0 \subset \gamma$ from U_{p} to U_{q}. Since U_{p_n}, U_{p}, U_{q_n}, and U_{q} all have the same radius ρ, γ_0 also joins U_{p_n} to U_{q_n}, contrary to the assumption that the shortest transit time is $\geq n$.

Since $[X, Y] = Z$ there is a constant $\mathrm{c} > 0$ such that if $\mathrm{q} = \eta_t(\mathrm{p})$ and $[t] \leq \mathrm{c}$, where η is the Z-flow, then there is an X, Y path of length 4 from p to q. This is the usual picture of the Lie bracket $[X, Y] = Z$, and becomes especially clear if drawn in a flowbox for X.

Now, given any pair $\mathrm{p}, \mathrm{q} \in M$, there is an X, Y path from p to q defined as follows. Some $\mathrm{p}_1 \in U_{\mathrm{p}}$ is joined to some $\mathrm{q}_1 \in U_{\mathrm{q}}$ by a Z-path of time length $\leq L$ and this path can be replaced by an X, Y path of length approximately $4L/\mathrm{c}$. Also there is an X, Y, Z path from p to p_1 of length 3, for $\mathrm{p}_1 \in U_{\mathrm{p}}$. The Z-arc of this path can again be replaced by an X, Y path of length $\leq N_1$

where $N_1 \doteq 4/c$. The same is true at q. Concatenation gives an X, Y path of length $N \doteq 4 + (8 + 4L)/c$ from p to q. Q.E.D.

COROLLARY 1.3. *There is an* $\mathcal{H}^{u,s}$ *path of length* $\leq N$ *joining any pair of points in* $M = T_1(\mathbb{H}^2)$.

Proof. The geodesic flow is ergodic, therefore topologically transitive, and therefore has a dense orbit. The formula $[H^u, H^s] = -H^c = -\dot{\varphi}$ completes the proof. Q.E.D.

PROPOSITION 1.4. *Suppose that* X,Y *are smooth vector fields on a compact m-manifold* M *and every pair of points* p,q $\in M$ *can be joined by an* X,Y *path of length* $\leq N$. *If* X',Y' *are uniquely integrable vector fields which* C^0*-approximate* X,Y *well enough, then every pair of points in* M *can be joined by an* X',Y' *path of length* $\leq 2N$.

Proof. Let φ and ψ be the X- and Y-flows. Let I^N be the unit cube in $\mathbb{R}^N$, where $\mathrm{I} = (-1, 1)$. Fix some $\mathrm{p} \in M$ and consider the map $\Gamma\colon \mathrm{I}^N \to M$ defined by

$$\Gamma\colon (t_1, \ldots, t_N) \mapsto \varphi_{t_1} \circ \psi_{t_2} \circ \cdots \circ \psi_{t_N}(\mathrm{p}).$$

(We tacitly assume that N is even.) The map Γ is smooth and onto. By Sard's Theorem, most of its values are regular. Choose one, say r. Since Γ is onto, $\Gamma^{-1}(r) \neq \emptyset$. Choose $\mathrm{t} = (\mathrm{t}_1, \ldots, \mathrm{t}_N) \in \Gamma^{-1}(r)$. Regularity implies that there is a small m-disc D at t which Γ carries diffeomorphically onto a neighborhood U of r.

Let q be an arbitrary point of M. There exists $\mathrm{s} = (\mathrm{s}_1, \ldots, \mathrm{s}_N) \in \mathrm{I}^N$ such that $\Phi(r) = \mathrm{q}$ where Φ is the composition

$$\Phi = \varphi_{\mathrm{s}_1} \circ \psi_{\mathrm{s}_2} \circ \ldots \circ \psi_{\mathrm{s}_N}.$$

The diffeomorphism Φ carries the neighborhood U onto a neighborhood V_{q} of q. This neighborhood V_{q} contains a compact sub-neighborhood W_{q} of q, and taking the appropriate orientation of the disc D, for all $w \in W_{\mathrm{q}}$ we have

$$\mathrm{index}(\Phi \circ \Gamma \,|_{\partial D}, w) = 1.$$

By equicontinuity and unique integrability, the perturbed vector fields X' and Y' generate flows φ' and ψ' which C^0 approximate φ and ψ uniformly on compact time intervals. Thus, for all $w \in W_{\mathrm{q}}$,

$$\mathrm{index}(\Phi' \circ \Gamma' \,|_{\partial D}, w) = 1,$$

where $\Phi' = \varphi'_{\mathrm{s}_1} \circ \psi'_{\mathrm{s}_2} \circ \cdots \circ \psi'_{\mathrm{s}_N}$ and $\Gamma' = \varphi'_{\mathrm{t}_1} \circ \psi'_{\mathrm{t}_2} \circ \cdots \psi'_{\mathrm{t}_N}(\mathrm{p})$. Consequently, $\Phi' \circ \Gamma'(D) \supset W_{\mathrm{q}}$.

By compactness, we cover M with finitely many such neighborhoods W_{q}. If X', Y' approximate X, Y well enough then we conclude that

$$\Phi' \circ \Gamma'(\mathrm{I}^N) \supset M.$$

That is, each pair $\mathrm{p}, \mathrm{q} \in M$ is joined by an X', Y' path of length $\leq 2N$. Q.E.D.

COROLLARY 1.5. *With reference to Lemma* 1.1, *if f approximates φ_1 well enough in the C^1 sense, then any two points in $M = T_1(\mathbb{H}^2)$ are connected by a $\mathcal{W}^{\mathrm{u,s}}$-path consisting of at most N arcs, each of length at most* 1, *where N depends only on the manifold M.*

Proof. By Corollary 1.3 there is an $\mathcal{H}^{\mathrm{u,s}}$ path of length $\leq N_0$ joining any two points in M. When f approximates φ_1 in the C^1 sense, then the line fields $E^{\mathrm{u}}, E^{\mathrm{s}}$ approximate $H^{\mathrm{u}}, H^{\mathrm{s}}$ in the C^0 sense and they are uniquely integrable. Since $H^{\mathrm{u}}, H^{\mathrm{s}}$ are orientable, so are $E^{\mathrm{u}}, E^{\mathrm{s}}$, and they support vector fields. By Proposition 1.4 any $\mathrm{p}, \mathrm{q} \in M$ are joined by a $\mathcal{W}^{\mathrm{u,s}}$ path of length $2N_0$. Q.E.D.

Remark. Generalizing 1.2, 1.4 to more than two vector fields presents no problem. The hypothesis that X', Y' are uniquely integrable may be superfluous.

2. Uniformities

In this section we sharpen and refine several estimates involving perturbations of φ_1.

2.1 LEMMA. *The hyperbolic splitting for a perturbation of φ_1 is uniformly Hölder in the sense described below.*

Proof. Fix any θ, $0 \leq \theta < 1$. If $\mathcal{N}$ is a sufficiently small C^2-neighborhood of φ_1 and $f \in \mathcal{N}$ then we claim that the hyperbolic splitting $E^{\mathrm{u}} \oplus E^{\mathrm{c}} \oplus E^{\mathrm{s}}$ for f is θ-Hölder and that the θ-Hölder constant is uniformly bounded. θ-Hölderness of the splitting is standard (see Pugh and Shub, [10, p. 5], for example); the only issue is uniformity. The plane-field E^{cu} is found as the unique F-invariant section of a bundle of linear maps

$$\begin{array}{ccc} L & \xrightarrow{F} & L \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{f} & M. \end{array}$$

The fiber of L at $\mathrm{p} \in M$ is $L_{\mathrm{p}} = \{P \in L(H^{\mathrm{cu}}, H^{\mathrm{s}}) \colon \|P\| \leq 1\}$ and F_{p} sends L_{p} to $L_{f_{\mathrm{p}}}$ according to the natural Tf-action,

$$F_{\mathrm{p}} \colon P \mapsto (C + KP) \circ (A + BP)^{-1}$$

where

$$T_p f = \begin{pmatrix} A & B \\ C & K \end{pmatrix}, \quad \begin{matrix} A\colon H^{cu} \to H^{cu} \\ C\colon H^{u} \to H^{cs} \end{matrix}, \quad \begin{matrix} B\colon H^{s} \to H^{cu} \\ K\colon H^{s} \to H^{s} \end{matrix}$$

respecting $TM = H^{cu} \oplus H^{s}$. As the neighborhood $\mathcal{N}$ shrinks to φ_1 in the C^1-sense, $T_p f$ converges uniformly to

$$T_p \varphi_1 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda^{-1} \end{pmatrix}.$$

Thus, the fiber constant $k = \sup_p \operatorname{Lip} F_p$ tends uniformly to λ^{-1}, while the base constant $\mu = \operatorname{Lip} f^{-1}$ tends uniformly to λ. This gives the fiber dominance condition $k\mu^{\theta} < 1$. Besides, since $\mathcal{N}$ is a C^2-neighborhood, the C^2-size of $f \in \mathcal{N}$ is uniformly bounded. This implies that the C^1-size of F is uniformly bounded, and therefore that the θ-Hölder constant H of F is uniformly bounded. According to the Hölder Section Theorem in Shub [14, p. 46], the θ-Hölder constant of the uniquely F-invariant section is estimated as at most

$$H \frac{\mu^{\theta}}{1 - k\mu^{\theta}},$$

which is uniformly bounded. Q.E.D.

Next, we consider the holonomy along the unstable f-invariant lamination $\mathcal{W}^u$ of a C^2-small perturbation f of φ_1. Let D, D' be C^1 2-discs in M which are uniformly transverse to the unstable horocycle foliation $\mathcal{H}^u$ in the sense that the angle between H^u and the tangent planes to D, D' is always at least 45 degrees. Since f approximates φ_1, D and D' are transverse to $\mathcal{W}^u$ also.

A $\mathcal{W}^u$-**holonomy map** is a continuous function π^u defined on an open subset $U \subset D$ which satisfies $\pi^u(p) \in W^u(p) \cap D'$ for all $p \in U$. See Figure 9.

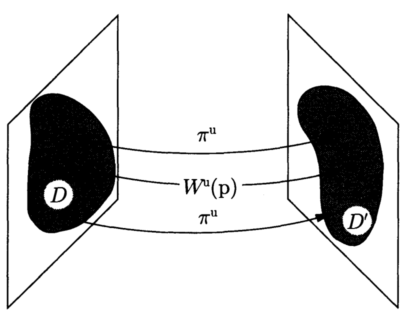

FIGURE 9.

We shrink D so that $U = D$. The fundamental fact about π^u is that it is a homeomorphism to its image and although π^u may fail to be differentiable, it is absolutely continuous: It sends sets of Lebesgue area zero in D to sets of Lebesgue area zero in D'. Any absolutely continuous function has a Radon-Nikodym derivative, a "generalized Jacobian", which is locally integrable. The Radon-Nikodym derivative $\mathrm{Jac}\,\pi^{\mathrm{u}}$ of π^{u} is not only locally integrable but actually continuous and positive. See Pugh and Shub [10, p. 8].

2.2 LEMMA. *If $f \in \mathcal{N}$ and $\mathcal{N}$ is a C^2-neighborhood of φ_1 which shrinks to φ_1 then π^{u} converges uniformly to the $\mathcal{H}^{\mathrm{u}}$-holonomy map h^{u} in the Radon-Nikodym sense described below. If $D \subset W^{\mathrm{cs}}_{\mathrm{q}}$ then* $\mathrm{Jac}\,\pi^{\mathrm{u}}$ *converges uniformly to* 1.

Proof. Consider $\mathrm{p} \in D$, $\pi^{\mathrm{u}}(\mathrm{p}) = \mathrm{q} \in W^{\mathrm{u}}(\mathrm{p}) \cap D'$, and assume that the arc length in $W^{\mathrm{u}}(\mathrm{p})$ from p to q is at most 1. With D, D' as above, we claim that $\pi^{\mathrm{u}} \rightrightarrows h^{\mathrm{u}}$ and $\mathrm{Jac}\,\pi^{\mathrm{u}} \rightrightarrows \mathrm{Jac}\,h^{\mathrm{u}}$ as $\mathcal{N}$ shrinks to φ_1. The fact that $\pi^{\mathrm{u}} \rightrightarrows h^{\mathrm{u}}$ amounts to the locally uniform convergence of $\mathcal{W}^{\mathrm{u}}$ to $\mathcal{H}^{\mathrm{u}}$, and this is a consequence of its construction via the graph transform method in Hirsch, Pugh, and Shub [6]. See also Lemma 2.4 below.

Anosov's formula for $\mathrm{Jac}\,\pi^{\mathrm{u}}$ as explained in Pugh and Shub [10, p. 8], is

$$\mathrm{Jac}_{\mathrm{p}}\,\pi^{\mathrm{u}} = \lim_{n\to\infty} \frac{\det T_{\mathrm{p}}(f^{-n}|_D)}{\det T_{\mathrm{q}}(f^{-n}|_{D'})}.$$

By the chain rule, this limit is the infinite product

$$\prod_{k=0}^{\infty} \frac{\det T_{\mathrm{p}_k}(f^{-1}|_{f^{-k}D})}{\det T_{\mathrm{q}_k}(f^{-1}|_{f^{-k}D'})} = \prod_{k=0}^{\infty} \frac{\det T_{\mathrm{p}_k}(f^{-1}|_{f^{-k}D})}{\det T^{\mathrm{cs}}_{\mathrm{p}_k} f^{-1}} \frac{\det T^{\mathrm{cs}}_{\mathrm{p}_k} f^{-1}}{\det T^{\mathrm{cs}}_{\mathrm{q}_k} f^{-1}} \frac{\det T^{\mathrm{cs}}_{\mathrm{q}_k} f^{-1}}{\det T_{\mathrm{q}_k}(f^{-1}|_{f^{-k}D'})},$$

where $f^{-k}\mathrm{p} = \mathrm{p}_k$ and $f^{-k}\mathrm{q} = \mathrm{q}_k$. We claim that the product converges uniformly exponentially. That is, its factors differ from 1 by at most $\mathrm{C}\beta^k$ where C, β are uniform constants and $\beta < 1$. Uniformity refers to all $f, \mathrm{p}, \mathrm{q}, D, D'$ as hypothesized. Since E^{cs} is a uniformly exponential attractor under Tf^{-k} as $k \to \infty$, the angle between the tangent plane $T_{\mathrm{p}_k} f^{-k} D$ and the plane $E^{\mathrm{cs}}_{\mathrm{p}_k}$ is at most $\mathrm{C}_1\beta_1^k$ for uniform constants C_1, β_1 with $\beta_1 < 1$. Since M is compact and Tf is continuous, the outer factors

$$\frac{\det T_{\mathrm{p}_k}(f^{-1}|_{f^{-k}D})}{\det T^{\mathrm{cs}}_{\mathrm{p}_k} f^{-1}} \quad \text{and} \quad \frac{\det T^{\mathrm{cs}}_{\mathrm{q}_k} f^{-1}}{\det T_{\mathrm{q}_k}(f^{-1}|_{f^{-k}D'})}$$

differ from 1 by at most $\mathrm{C}_2\beta_2^k$. The same is true of the middle factor. For by Lemma 2.1, the planes $E^{\mathrm{cs}}_{\mathrm{p}_k}, E^{\mathrm{cs}}_{\mathrm{q}_k}$ differ by at most $Hd(\mathrm{p}_k, \mathrm{q}_k)^\theta$, d being the distance in M. Since f is uniformly C^2, the determinants $\det T^{\mathrm{cs}}_{\mathrm{p}_k} f^{-1}, \det T^{\mathrm{cs}}_{\mathrm{q}_k} f^{-1}$ differ by at most $C_3 d(\mathrm{p}_k, \mathrm{q}_k)^\theta$ where C_3 is a uniform constant. The determinants themselves are bounded away from 0 and ∞ since M is compact. Thus, their ratio differs from 1 by at most $\mathrm{C}_4 d(\mathrm{p}_k, \mathrm{q}_k)^\theta$ where C_4 is another uniform constant. Since $\mathrm{q} \in W^{\mathrm{u}}(\mathrm{p})$ and the arc length from p to q in $W^{\mathrm{u}}(\mathrm{p})$ is uniformly bounded, $d(\mathrm{p}_k, \mathrm{q}_k) \leq \mathrm{C}_5(\lambda - \varepsilon)^{-k}$ where $\varepsilon \to 0$ as the neighborhood $\mathcal{N}$ shrinks to φ_1, and C_5 is a uniform constant. Thus,

$$\left| \frac{\det T^{\mathrm{cs}}_{\mathrm{p}_k} f^{-1}}{\det T^{\mathrm{cs}}_{\mathrm{q}_k} f^{-1}} - 1 \right| \leq C_6\big((\lambda - \varepsilon)^{\theta}\big)^{-k}.$$

Because each of the three factors in the infinite product converges uniformly exponentially to 1, the infinite product converges uniformly to a uniformly bounded limit.

Since the infinite product expression for $\operatorname{Jac} \pi^{\mathrm{u}}$ is dominated by a fixed absolutely convergent product, $\prod(1 + \mathrm{C}\beta^k)$, and since for each fixed k,

$$\frac{\det T_{\mathrm{p}_k}(f^{-1}|_{f^{-k}D})}{\det T_{\mathrm{q}_k}(f^{-1}|_{f^{-k}D'})} \rightrightarrows \frac{\det T_{\varphi_{-k}(\mathrm{p})}(\varphi_{-1}|_{\varphi_{-k}D})}{\det T_{\varphi_{-k}(\mathrm{q})}(\varphi_{-1}|_{\varphi_{-k}D'})},$$

the Lebesgue dominated convergence theorem implies that the limit of the products is the product of the limits as $\mathcal{N}$ shrinks to φ_1. That is,

$$\operatorname{Jac} \pi^{\mathrm{u}} \rightrightarrows \operatorname{Jac} h^{\mathrm{u}}.$$

If $D \subset W^{\mathrm{cs}}_{\mathrm{p}}$ and $D' \subset W^{\mathrm{cs}}_{\mathrm{q}}$ then $\operatorname{Jac} \pi^{\mathrm{u}} \rightrightarrows 1$ since $\operatorname{Jac} h^{\mathrm{u}} = 1$ on center stable transversals. Q.E.D.

This last is easily seen in the $\mathrm{SL}(2, \mathbb{R})$ model for the geodesic flow. Here the flow is $g \mapsto \left[\begin{smallmatrix} e^t & 0 \\ 0 & e^{-t} \end{smallmatrix}\right] g$. The center stable manifold through g is given by $\left[\begin{smallmatrix} e^t & 0 \\ v & e^{-t} \end{smallmatrix}\right] g$ and the h^{u}-holonomy from g to $\left[\begin{smallmatrix} 1 & u \\ 0 & 1 \end{smallmatrix}\right] g$ is $h^{\mathrm{u}}(t, v) = (s, w)$ when there exists an x such that

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^t & 0 \\ v & e^{-t} \end{bmatrix} g = \begin{bmatrix} e^{s} & 0 \\ w & e^{-s} \end{bmatrix} \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix} g.$$

The matrix of partial derivatives at $t = 0 = v$ is

$$\begin{bmatrix} \frac{\partial s}{\partial t} & \frac{\partial s}{\partial v} \\ \frac{\partial w}{\partial t} & \frac{\partial w}{\partial v} \end{bmatrix} = \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix},$$

which has determinant 1. The frames in which the matrix calculation is done are orthonormal, so the Jacobian of h^{u} is 1.

Remark. Here is a second way to prove Lemma 2.2. Returning to the proof of Anosov's Jacobian formula in Pugh and Shub [10, pp. 5–11], note that all constructions of pre-foliations and pre-foliation Jacobians depend continuously on $f \in \mathcal{N}$. The natural action of Tf on these Jacobians is also continuous with respect to f, and this action can be viewed as a contraction mapping on the space of trial Jacobian functions. The fixed point of a family of uniform contractions which depend continuously on a parameter also depends continuously on that parameter. In this case the fixed point is the actual Jacobian of the $\mathcal{W}^{\mathrm{u}}$-holonomy and the parameter is the diffeomorphism f.

The unstable horocycle foliation $\mathcal{H}^{\mathrm{u}}$ is subordinate to the center unstable horocycle foliation $\mathcal{H}^{\mathrm{cu}}$, so the $\mathcal{H}^{\mathrm{u}}$-holonomy map h^{u} preserves the $\mathcal{H}^{\mathrm{cu}}$-leaves. It sends the φ-trajectory $H^{\mathrm{c}}_{\mathrm{p}}$ to the φ-trajectory $H^{\mathrm{c}}_{\mathrm{q}}$ when $\mathrm{q} \in H^{\mathrm{u}}_{\mathrm{p}}$, and its restriction to $H^{\mathrm{c}}_{\mathrm{p}}$ is an isometry,

$$h^{\mathrm{u}} : H^{\mathrm{c}}_{\mathrm{p}} \longrightarrow H^{\mathrm{c}}_{\mathrm{q}}.$$

The next result states that the same is nearly true after perturbation.

2.3 LEMMA. *The $\mathcal{W}^{\mathrm{u}}$-holonomy map π^{u} preserves the center foliation $\mathcal{W}^{\mathrm{c}}$. If $\mathrm{q} \in W^{\mathrm{u}}(\mathrm{p})$ then π^{u} sends $W^{\mathrm{c}}(\mathrm{p})$ to $W^{\mathrm{c}}(\mathrm{q})$, is C^1, and is uniformly nearly isometric.*

Proof. The $\mathcal{W}^{\mathrm{u}}$-foliation is subordinate to the $\mathcal{W}^{\mathrm{cu}}$-foliation, so the $\mathcal{W}^{\mathrm{u}}$-holonomy map π^{u} automatically carries $W^{\mathrm{c}}(\mathrm{p})$ to $W^{\mathrm{c}}(\mathrm{q})$. We will show that the line field E^{u}, restricted to each $\mathcal{W}^{\mathrm{cu}}$-leaf, is C^1. It follows that the $\mathcal{W}^{\mathrm{u}}$-foliation and its holonomy maps along $\mathcal{W}^{\mathrm{cu}}$ are C^1. In the proof of Lemma 2.2, we exhibited E^{cu} as an invariant section of a bundle of linear maps. Similarly, to construct E^{u} directly one considers the new bundle map

$$\begin{array}{ccc} L & \xrightarrow{F} & L \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{f} & M \end{array}$$

where the fiber of L at $\mathrm{p} \in W$ is $L_{\mathrm{p}} = \{P \in L(H^{\mathrm{u}}, H^{\mathrm{cs}}) : \|P\| \le 1\}$, F is the natural map $L_{\mathrm{p}} \to L_{f\mathrm{p}}$ sending P to $F(P) = (C + KP) \circ (A + BP)^{-1}$, and

$$T_{\mathrm{p}}f = \begin{pmatrix} A & B \\ C & K \end{pmatrix} \quad \begin{array}{ll} A: H^{\mathrm{u}} \to H^{\mathrm{u}} & B: H^{\mathrm{cs}} \to H^{\mathrm{u}} \\ C: H^{\mathrm{u}} \to H^{\mathrm{cs}} & K: H^{\mathrm{cs}} \to H^{\mathrm{cs}} \end{array}$$

respecting $TM = H^{\mathrm{u}} \oplus H^{\mathrm{cs}}$. Thus, F is a fiber contraction with fiber constant $k \rightrightarrows \lambda^{-1}$ and base constant $\mu = \operatorname{Lip} f^{-1} \rightrightarrows \lambda$. (So far, this tells us nothing new because the fiber constant only dominates the base constant at θ-Hölder scales, $\theta < 1$.) However, we can replace the base manifold M with a manifold where the base constant tends to 1. Namely, let W be the non-separable C^2 manifold consisting of the disjoint union of all the $\mathcal{W}^{\mathrm{cu}}$-leaves. Since $\mathcal{W}^{\mathrm{cu}}$ is f-invariant, f is a diffeomorphism of W to itself and F becomes a fiber contraction

$$\begin{array}{ccc} L & \xrightarrow{F} & L \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ W & \xrightarrow{f} & W. \end{array}$$

Now the fiber constant k dominates the base constant $\mu = \mathrm{Lip}(f\,|_W)^{-1}$ since $\mu \rightrightarrows 1$ as $\mathcal{N}$ shrinks to φ_1. The C^r Section Theorem of Hirsch, Pugh and Shub [6] implies that the unique F-invariant section of L, namely E^{u}, is C^1. The C^1-uniformity of the invariant section follows from the facts that the $\mathcal{W}^{\mathrm{cu}}$-leaves are uniformly C^2, the map f is uniformly C^2, the splitting $H^{\mathrm{u}} \oplus H^{\mathrm{cs}}$ is fixed and C^1, and the fiber constant uniformly dominates the base constant. Since E^{u} is uniformly C^1 on W, the $\mathcal{W}^{\mathrm{u}}$-holonomy maps on W are also locally uniformly C^1. As $\mathcal{N}$ shrinks to φ_1, the $\mathcal{W}^{\mathrm{cu}}$-leaves tend locally C^2-uniformly to the $\mathcal{H}^{\mathrm{cu}}$-leaves and the line field E^{u}, considered on W, tends locally C^1-uniformly to H^{u}. Thus, the holonomy maps π^{u} restricted to the center leaves tend locally C^1-uniformly to the horocycle isometry $H^{\mathrm{c}}_{\mathrm{p}} \to H^{\mathrm{c}}_{\mathrm{q}}$. Q.E.D.

2.4 LEMMA. *Given θ, $0 \le \theta < 1$, the unit $\mathcal{W}^{\mathrm{u}}$-holonomy maps for any f in a small C^2-neighborhood of φ_1 are uniformly θ-Hölder.*

Remark. By the proof of Lemma 2.3, the line field E^{u} is Hölder, but in general the integral curves of a Hölder line field need not form a Hölder foliation. By a unit holonomy map we indicate the assumption that the $\mathcal{W}^{\mathrm{u}}$-arcs along which we slide have length at most 1.

Proof. Let p, q, D, D' be points and transversal discs as in Lemma 2.2. We know that the Jacobian of the holonomy map $D \to D'$ is uniformly bounded and we also know by Lemma 2.3 that the holonomy map restricted to the center leaves is C^1. Thus, if the holonomy map *were* differentiable, its derivative would be a 2×2 matrix $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ respecting the splitting $E^{\mathrm{c}} \oplus E^{\mathrm{s}}$, while the entry a and the determinant would be bounded and bounded away from 0. Thus d would be bounded and bounded away from 0. However, this would still give us no control over the shear entry b.

To prove that the $\mathcal{W}^{\mathrm{u}}$-holonomy map is uniformly θ-Hölder, we exhibit the unit unstable manifolds $W^{\mathrm{u}}(\mathrm{p})$ as graphs of maps which are Hölder functions of p. To simplify the estimates, we introduce uniform coordinate systems adapted to φ. Given $\mathrm{p} \in M$, we define $\omega_{\mathrm{p}} \colon \mathbb{R}^3 \to M$ as follows.

(1) ω_{p} sends the x-axis to $H^{\mathrm{u}}(\mathrm{p})$, the unstable horocycle at p. The parametrization is an orientation-preserving isometry.

(2) ω_{p} sends the (x, y)-plane to $H^{\mathrm{s}}(H^{\mathrm{u}}(\mathrm{p})) = \bigcup_{\mathrm{q} \in H^{\mathrm{u}}\mathrm{p}} H^{\mathrm{s}}(\mathrm{q})$, and for each fixed x, the parametrization $(x, y, 0) \mapsto \omega_{\mathrm{p}}(x, y, 0)$ is an orientation-preserving isometry from the y-line $x \times \mathbb{R} \times \{0\}$ onto $H^{\mathrm{s}}(\omega_{\mathrm{p}}(x, 0, 0))$.

(3) $\omega_{\mathrm{p}}(x, y, t) = \varphi_1(\omega_{\mathrm{p}}(x, y, 0))$.

See Figure 10.

Since the horocycle foliations for the geodesic flow φ are smooth, each ω_{p} is C^∞. In fact $\omega \colon M \times \mathbb{R}^3 \to M \times M$ defined by $(\mathrm{p}, w) \mapsto (\mathrm{p}, \omega_{\mathrm{p}}(w))$ is C^∞ too. The ω_{p}-coordinate system is a special type of flowbox for φ. Thus, ω_{p} is

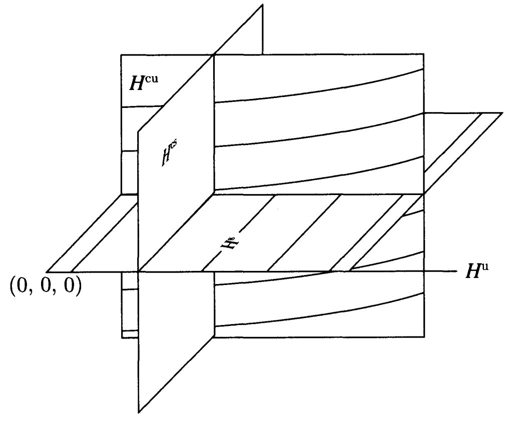

FIGURE 10. The horocycle foliations at p.

uniquely specified: (1), (2) specify it on the (x,y)-plane through $\mathrm{p} = (0,0,0)$ and (3) specifies elsewhere. Besides, the center unstable horocycles are the (x,z)-planes and the center stable horocycles are the (y,z)-planes. However, since $[H^{\mathrm{u}}, H^{\mathrm{s}}] \neq 0$, the (x,y)-planes have no dynamical significance.

On $\mathbb{R}^3$ we choose the maximum coordinate norm $\|(x,y,z)\| = \max(|x|,|y|,|z|)$ and the corresponding metric.

Let the time-one map φ_1 be called ϕ and represent ϕ in ω-coordinates at p and ϕp as

$$\phi_{\mathrm{p}} = \omega_{\phi\mathrm{p}}^{-1} \circ \phi \circ \omega_{\mathrm{p}}.$$

More precisely, restrict ω_{p} to a large cube $Q(R)$ in $\mathbb{R}^3$ such that $\omega_{\phi\mathrm{p}} \circ \phi \circ \omega_{\mathrm{p}}(Q(R))$ makes sense. In these coordinates ϕ is linear,

$$\phi_{\mathrm{p}}(x,y,z) = (\lambda x, \lambda^{-1} y, z)$$

and $\lambda > 1$. Next, consider a C^2-approximation f to ϕ. Its representation in the ω-coordinates is

$$f_{\mathrm{p}} = \omega_{f\mathrm{p}}^{-1} \circ f \circ \omega_{\mathrm{p}}.$$

Since ω is C^∞, f_{p} uniformly C^2-approximates ϕ_{p} when f C^2-approximates φ_1.

Consider the space $\mathcal{G}$ of maps $g\colon [-R,R] \to \mathbb{R}^2$ such that $g(0) = 0$ and $\operatorname{Lip} g \leq 1$. The natural action of ϕ_{p} on $\mathcal{G}$ is given by the graph transform,

$$\phi_{\mathrm{p}\#}\colon g \mapsto \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} g(\lambda^{-1} x).$$

The graph of $\phi_{\mathrm{p}\#} g$ is the ϕ_{p}-image of the graph of g. The graph transform is a contraction respecting the special metric

$$d(g,g') = \sup \frac{|gx - g'x|}{|x|}.$$

The sup is taken over $x \in [-R, R]$. In fact, since we use the maximum coordinate norm in $\mathbb{R}^3$,

$$\begin{aligned} d(\phi_{\mathrm{p}\#}g, \phi_{\mathrm{p}\#}g') &\leq \left\| \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \right\| \sup \frac{|g(\lambda^{-1}x) - g'(\lambda^{-1}x)|}{|x|} \\ &= 1 \cdot \sup \frac{|g(\lambda^{-1}x) - g'(\lambda^{-1}x)|}{|\lambda^{-1}x|} \frac{|\lambda^{-1}x|}{|x|} \leq \lambda^{-1} d(g, g'). \end{aligned}$$

Now f_{p} C^1-approximates ϕ_{p} and it too defines a graph transform $f_{\mathrm{p}\#}: \mathcal{G} \to \mathcal{G}$. The formula for $f_{\mathrm{p}\#}$ is

$$f_{\mathrm{p}\#}g(x) = \pi_{23} \circ f_{\mathrm{p}} \circ G \circ (\pi_1 f_{\mathrm{p}} G)^{-1}(x)$$

where $G\colon x \mapsto (x, gs)$, while π_1 and π_{23} are the projections of $\mathbb{R}^3$ onto the x-axis and the (y, z)-plane. Thus, G is the function whose image is the graph of g. Since f_{p} C^1-approximates ϕ_{p}, $f_{\mathrm{p}\#}$ is a contraction of $\mathcal{G}$ with nearly the same contraction constant λ^{-1} as has $\phi_{\mathrm{p}\#}$. Since M is compact these contractions are uniform over M and we get a fiber contraction

$$\begin{array}{ccc} M \times \mathcal{G} & \xrightarrow{F} & M \times \mathcal{G} \\ \pi \downarrow & & \downarrow \pi \\ M & \xrightarrow{f} & M. \end{array} \qquad F\colon (\mathrm{p}, g) \mapsto (\mathrm{p}, f_{\mathrm{p}\#}g)$$

We claim that the Hölder Section Theorem implies that the unique F-invariant section of this bundle is θ-Hölder. The unique invariant section is of course the unstable manifold at p. To apply the theorem we must check that F is a θ-Hölder map of the bundle $M \times \mathcal{G}$ and that the fiber constant appropriately dominates the base constant. In fact we prove that F is Lipschitz. We first establish

(4) $|f_{\mathrm{p}}(w) - f_{\mathrm{p}'}(w)| \leq C|w| d(\mathrm{p}, \mathrm{p}')$

for a uniform constant C. We know that $f_{\mathrm{p}}(0) = 0$ for all $\mathrm{p} \in M$ and that $w \mapsto f_{\mathrm{p}}(w)$ is C^2. Thus, $\frac{\partial f_{\mathrm{p}}(w)}{\partial \mathrm{p}}$ is C^1 and vanishes identically when $w = 0$. By the mean value theorem in local coordinates when $d(\mathrm{p}, \mathrm{p}')$ is small,

$$f_{\mathrm{p}}(w) - f_{\mathrm{p}'}(w) = \int_0^1 \left(\frac{\partial f_{\mathrm{q}}}{\partial \mathrm{q}} \right)_{\mathrm{q} = t\mathrm{p}' + (1-t)\mathrm{p}} dt (\mathrm{p} - \mathrm{p}'),$$

which implies that $|f_{\mathrm{p}}(w) - f_{\mathrm{p}'}(w)| \leq C|w|\,|\mathrm{p} - \mathrm{p}'|$. Interpreted in the metric of M this gives (4).

Now we deduce that F is Lipschitz. We know that F contracts the fiber $\mathrm{p} \times \mathcal{G}$ into the fiber $f\mathrm{p} \times \mathcal{G}$, and also that it is Lipschitz on the base, being just f there, so we must merely estimate the shear term. We claim that

$$|f_{\mathrm{p}\#}g(x) - f_{\mathrm{p}'\#}g(x)| \leq C d(\mathrm{p}, \mathrm{p}')$$

for a uniform constant C. For brevity, write $h(x) = \pi_1 \circ f_{\mathrm{p}} \circ Gx$, and $h'(x) = \pi_1 \circ f_{\mathrm{p}'} \circ Gx$. Thus h, h' are overflowing, expanding Lipeomorphisms:

$$\operatorname{Lip} h^{-1} < 1,$$

$\operatorname{Lip} h'^{-1} \leq 1$. Then $f_{\mathrm{p}\#}g(x) = \pi_{23} \circ f_{\mathrm{p}} \circ G \circ h^{-1}(x)$, and so by (4):

$$\begin{aligned} |f_{\mathrm{p}\#}g(x) - f_{\mathrm{p}'\#}g(x)| &\leq |\pi_{23} \circ f_{\mathrm{p}} \circ G \circ h^{-1}(x) - \pi_{23} \circ f_{\mathrm{p}'} \circ G \circ h^{-1}(x)| \\ &\quad + |\pi_{23} \circ f_{\mathrm{p}'} \circ G \circ h^{-1}(x) - \pi_{23} \circ f_{\mathrm{p}'} \circ G \circ h'^{-1}(x)| \\ &\leq C|G \circ h^{-1}(x)| d(\mathrm{p}, \mathrm{p}') \\ &\quad + \operatorname{Lip}(\pi_{23} \circ f_{\mathrm{p}'}) \operatorname{Lip} G |h^{-1}(x) - h'^{-1}(x)| \\ &\leq C|h^{-1}(x)| d(\mathrm{p}, \mathrm{p}') \\ &\quad + 1 \cdot |h^{-1} \circ h' \circ h'^{-1}(x) - h^{-1} \circ h \circ h'^{-1}(x)| \\ &\leq C|x| d(\mathrm{p}, \mathrm{p}') + \operatorname{Lip} h^{-1} |h' \circ h'^{-1}(x) - h \circ h'^{-1}(x)| \\ &\leq C|x| d(\mathrm{p}, \mathrm{p}') + |h'(x') - h(x')| \end{aligned}$$

where $x' = h'^{-1}(x)$. By (4),

$$\begin{aligned} |h'(x') - h(x')| = |\pi_1 \circ f_{\mathrm{p}} \circ Gx' - \pi_1 \circ f_{\mathrm{p}'} \circ Gx'| &\leq C|Gx'| d(\mathrm{p}, \mathrm{p}') \\ &\leq C|x'| d(\mathrm{p}, \mathrm{p}') \leq C|x| d(\mathrm{p}, \mathrm{p}'). \end{aligned}$$

Thus, $|f_{\mathrm{p}\#}g(x) - f_{\mathrm{p}'\#}g(x)| \leq 2C|x| d(\mathrm{p}, \mathrm{p}')$, which completes the proof that F is Lipschitz, since

$$d(f_{\mathrm{p}\#}g, f_{\mathrm{p}'\#}g) = \sup \frac{|f_{\mathrm{p}\#}g(x) - f_{\mathrm{p}'\#}g(x)|}{|x|} \leq 2Cd(\mathrm{p}, \mathrm{p}').$$

The Hölder Section Theorem therefore applies and we deduce that $M \times \mathcal{G}$ has a unique F-invariant section $\sigma\colon M \to M \times \mathcal{G}$, which is θ-Hölder provided that $k\mu^\theta < 1$, k being the fiber constant of F and $\mu = \operatorname{Lip} f^{-1}$ being the base constant. If $f \in \mathcal{N}$ and $\mathcal{N}$ shrinks to φ_1, k tends to λ^{-1} and μ tends to λ. Thus, when $\mathcal{N}$ is small, $k\mu^\theta < 1$ and so σ is θ-Hölder. Besides, the θ-Hölder constant of σ is at most $H\mu^\theta/(1 - k\mu^\theta)$ where $H = \operatorname{Lip} F$. It is uniformly bounded.

Since the graph of $\sigma(\mathrm{p})$ is the unit unstable manifold at p, the $\mathcal{W}^{\mathrm{u}}$-foliation is uniformly θ-Hölder. Q.E.D.

3. Juliennes

We are going to study some geometric objects which could be described as "small, narrow, nonlinear prisms", or "juliennes" since they resemble slivered vegetables. First we introduce coordinates adapted to a perturbation f of φ_1. Let f be such a perturbation and denote its invariant splitting as $T_{\mathrm{p}}M^3 =$

$E^{\mathrm{u}}_{\mathrm{p}} \oplus E^{\mathrm{s}}_{\mathrm{p}} \oplus E^{\mathrm{c}}_{\mathrm{p}}$. These line fields are α-Hölder functions of $\mathrm{p} \in M^3$ and α is nearly equal to 1. The center unstable and center stable bundles are $E^{\mathrm{cu}} = E^{\mathrm{u}} \oplus E^{\mathrm{c}}$ and $E^{\mathrm{cs}} = E^{\mathrm{s}} \oplus E^{\mathrm{c}}$. Tangent to the bundles $E^{\mathrm{u}}, E^{\mathrm{s}}, E^{\mathrm{cu}}, E^{\mathrm{cs}}, E^{\mathrm{c}}$ are unique f-invariant laminations $\mathcal{W}^{\mathrm{u}}, \mathcal{W}^{\mathrm{s}}, \mathcal{W}^{\mathrm{cu}}, \mathcal{W}^{\mathrm{cs}}, \mathcal{W}^{\mathrm{c}}$. All of them have smooth leaves but they need not be smooth foliations (hence the word "lamination"). Of the five laminations, only $\mathcal{W}^{\mathrm{u}}, \mathcal{W}^{\mathrm{s}}$ are absolutely continuous in general.

On a neighborhood U of any point $\mathrm{p} \in M^3$ we introduce a smooth (x, y, z)-coordinate system such that

$$\mathrm{p} = (0,0,0), \quad \operatorname{span}\left(\frac{\partial}{\partial x}\right)_{\mathrm{p}} = E^{\mathrm{u}}_{\mathrm{p}}, \quad \operatorname{span}\left(\frac{\partial}{\partial y}\right)_{\mathrm{p}} = E^{\mathrm{s}}_{\mathrm{p}}, \quad \operatorname{span}\left(\frac{\partial}{\partial z}\right)_{\mathrm{p}} = E^{\mathrm{c}}_{\mathrm{p}},$$

the vectors $(\frac{\partial}{\partial x})_{\mathrm{p}}, (\frac{\partial}{\partial y})_{\mathrm{p}}, (\frac{\partial}{\partial z})_{\mathrm{p}}$ have unit length, and the orientations agree. For example we can use the exponential chart at p adapted to the splitting. Strictly speaking, we should write the coordinates of a point in U as $(x, y, z)_{\mathrm{p},f}$ to indicate that the coordinates depend on p and f. We denote by Z_{p} the set of points in U with zero z-coordinate and by π the projection $U \to Z_{\mathrm{p}}$, $\pi(x, y, z) = (x, y)$. The rectilinear square in Z_{p} with center p_0 and width 2w is

$$S = S(\mathrm{p}, \mathrm{p}_0, \mathrm{w}) = [\mathrm{x}_0 - \mathrm{w}, \mathrm{x}_0 + \mathrm{w}] \times [\mathrm{y}_0 - \mathrm{w}, \mathrm{y}_0 + \mathrm{w}],$$

while the rectilinear box with center $\mathrm{p}_0 = (\mathrm{x}_0, \mathrm{y}_0, 0)$, width 2w, and height 2h is

$$R(\mathrm{p}, \mathrm{p}_0, \mathrm{w}, \mathrm{h}) = S \times [-\mathrm{h}, \mathrm{h}].$$

See Figure 11.

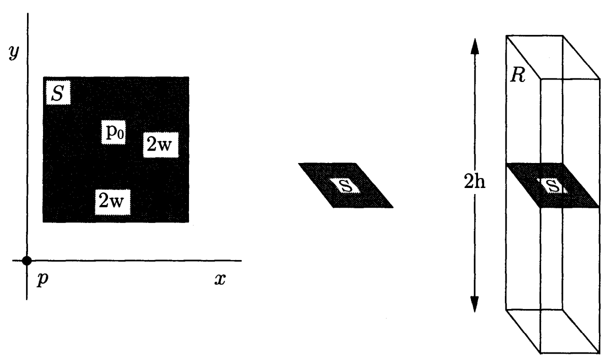

FIGURE 11. The square S and box R.

Since the laminations $\mathcal{W}^{\mathrm{cu}}, \mathcal{W}^{\mathrm{cs}}$ are transverse to Z_{p} and to each other they give transverse laminations

$$\mathcal{L}^{\mathrm{u}} = \mathcal{W}^{\mathrm{cu}} \cap Z_{\mathrm{p}} \quad \text{and} \quad \mathcal{L}^{\mathrm{s}} = \mathcal{W}^{\mathrm{cs}} \cap Z_{\mathrm{p}}$$

of Z_p. The leaves L^u, L^s of these laminations give a second coordinate grid on Z_p. It is Hölder, not smooth. Let $L^u_w(p_0)$ denote the arc of $L^u_{loc}(p_0)$ between the lines $x = x_0 \pm w$ on Z_p where $p_0 = (x_0, y_0, 0)$ and $w > 0$ is small. Similarly, let $L^s_w(p_0)$ denote the arc of $L^s_{loc}(p_0)$ between the lines $y = y_0 \pm w$ on Z_p. These two arcs through p_0 have length approximately equal to 2w and form the axes of a nonlinear square

$$\Sigma = \Sigma(p, p_0, w) = \bigcup L^u_{loc}(r) \cap L^s_{loc}(q)$$

where q ranges over $L^u_w(p_0)$ and r ranges over $L^s_w(p_0)$. See Figure 12. The

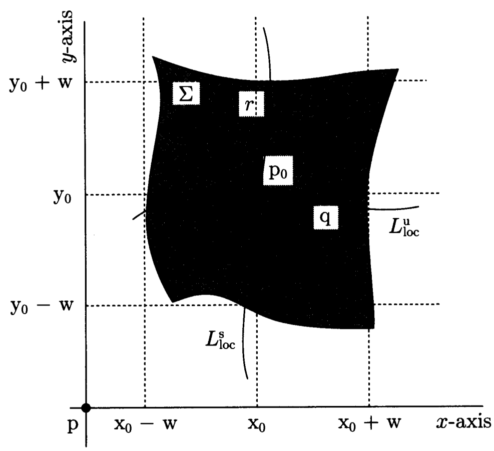

FIGURE 12. The base Σ of the julienne J in the plane Z_p.

square is centered at p_0 and has width approximately equal to 2w. Transversality of the laminations implies that each of the intersections referred to in the definition is a single point. Consistent with this notation, we set $W^c_h(t)$ equal to the arc of $W^c_{loc}(t)$ between the planes $z = \pm h$. Then

$$J = J(p, p_0, w, h) = \bigcup_{t \in \Sigma} W^c_h(t)$$

is the julienne with center p_0, width 2w, and height 2h. See Figures 13, 14. The **vertical boundary** of J is the set $\partial^c J = \bigcup_{t \in \partial\Sigma} W^c_h(t)$.

Under reasonable conditions we will show that boxes and juliennes approximate each other well. To make this precise it will be useful to dilate coordinates and write

$$cR = R(p, p_0, cw, ch) \quad \text{and} \quad cJ = J(p, p_0, cw, ch)$$

where c is a real constant. We will assume that p_0, p_1, p_2 are nearby points of M and that p_0 lies on the horizontal plane in the (x, y, z) coordinate systems

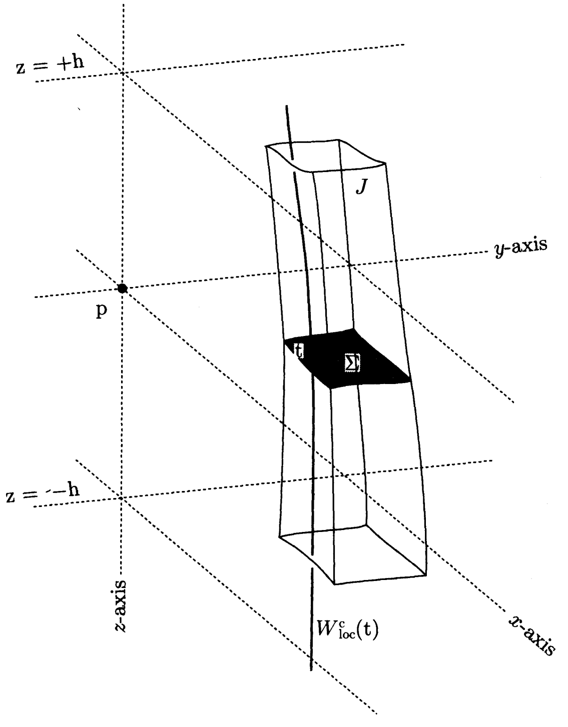

FIGURE 13. The julienne J consists of center manifold arcs between $z = -\mathrm{h}$ and $z = +\mathrm{h}$.

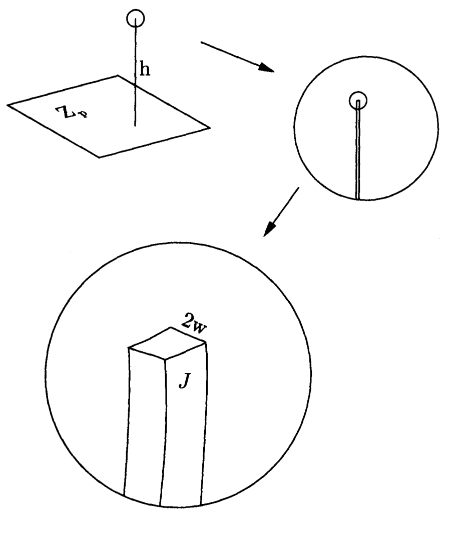

FIGURE 14. A julienne at three magnifications.

at both p_1 and p_2. That is, $p_0 \in Z_{p_1} \cap Z_{p_2}$. Then

$$J_1 = J(p_1, p_0, w, h), \quad R_1 = R(p_1, p_0, w, h),$$
$$J_2 = J(p_2, p_0, w, h), \quad R_2 = R(p_2, p_0, w, h)$$

are the appropriate sets to compare.

3.1 JULIENNE NESTING LEMMA. *Let* $c \in (0,1)$ *be fixed. If* $f \to \varphi_1$ *in the* C^2 *sense and* $w = h^{3/2} \to 0$, *while* $d(p_0,p_1)/h$ *and* $d(p_0,p_2)/h$ *stay bounded, then*

$$cR_1 \subset J_2 \subset c^{-1}R_1 \quad \textit{and} \quad cJ_1 \subset R_2 \subset c^{-1}J_1.$$

Remarks. If no relation between w and h is imposed then nesting fails. For example, if $h^{3/2} \gg w$ then the juliennes are very thin neighborhoods of the center manifold (an arc) through p_0 while the box is a very thin neighborhood of a vertical straight line segment. Generically these two arcs in 3-space meet only at p_0, so thin neighborhoods of them hardly intersect at all, much less nest. See Figure 15a. On the other hand, if $h \ll w$ then J_1, J_2 are virtually equal to the small nonlinear squares Σ_1, Σ_2 in Z_{p_1}, Z_{p_2}. Generically, these squares meet in a curve, so thin neighborhoods of them will not nest. See Figure 15b. As a special case, if $p_1 = p_2$, then the lemma asserts that a

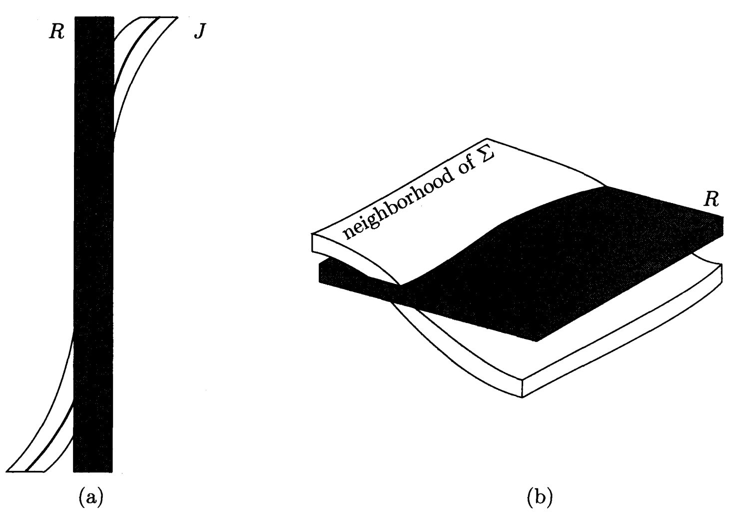

FIGURE 15. (a) h is too big, $h^{3/2} \gg w$. (b) h is too small, $h \ll w$.

julienne nests between two dilated boxes and a box nests between two dilated juliennes. Since c can be chosen nearly equal to 1, this provides the possibility of interchanging juliennes with boxes. By the way, the exponent 3/2 could be replaced by any fixed power between 1 and 2.

Proof. Let $\mathrm{p} = \mathrm{p}_1$ or p_2. The boundary of $\Sigma = \Sigma(\mathrm{p}, \mathrm{p}_0, \mathrm{w})$ consists of four arcs: two $\mathcal{L}^{\mathrm{u}}$-arcs through the endpoints of $L^{\mathrm{s}}_{\mathrm{w}}(\mathrm{p}_0)$ and two $\mathcal{L}^{\mathrm{s}}$-arcs through the endpoints of $L^{\mathrm{u}}_{\mathrm{w}}(\mathrm{p}_0)$. The tangents to these arcs approximate $E^{\mathrm{u}}_{\mathrm{p}}$ and $E^{\mathrm{s}}_{\mathrm{p}}$ uniformly as $\mathrm{w} \to 0$ and $f \to \varphi_1$. From this information alone, it follows that

$$\mathrm{c}S \subset \Sigma \subset \mathrm{c}^{-1}S.$$

See Figure 16. Similarly, the tangent vectors to the center manifolds $W^{\mathrm{c}}_{h}(\mathrm{t})$

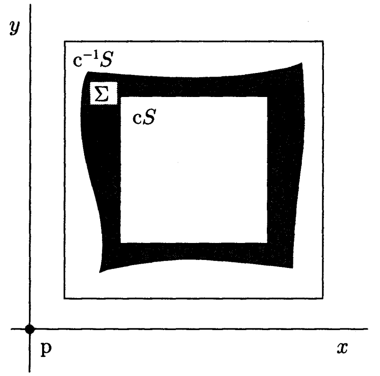

FIGURE 16. Nested squares in Z_{p}

comprising J differ from $E^{\mathrm{c}}_{\mathrm{p}}$ by at most Ch^{α}. For E^{c} is α-Hölder and the distance of any point of J to p is no more than $\mathrm{h} + d(\mathrm{p}, \mathrm{p}_0) = 2\mathrm{h}$. It follows that $W^{\mathrm{c}}_{\mathrm{h}}(\mathrm{t})$ differs from the vertical segment $\mathrm{t} \times [-\mathrm{h}, \mathrm{h}]$ by at most $\mathrm{Ch}^{\alpha+1}$. As $f \to \varphi_1$, $\alpha \to 1$. When $\alpha > 1/2$, $\mathrm{h}^{\alpha+1}$ is on a smaller order than $\mathrm{w} = \mathrm{h}^{3/2}$. That is, the linear projection $\pi(W^{\mathrm{c}}_{\mathrm{h}}(\mathrm{t}))$ has diameter $\ll \mathrm{w}$ so the projection of the vertical boundary of J fits between $\mathrm{c}S$ and $\mathrm{c}^{-1}S$. This implies that

$$\mathrm{c}R \subset J \subset \mathrm{c}^{-1}R \quad \text{and} \quad \mathrm{c}J \subset R \subset \mathrm{c}^{-1}J,$$

which takes care of the case $\mathrm{p}_1 = \mathrm{p}_2$.

Next, suppose that $\mathrm{p}_1 \neq \mathrm{p}_2$ and consider the box R_2 as it appears to an observer in the coordinate systems at p_1. It is not rectilinear. The change of variables between the $(x, y, z)_{\mathrm{p}_1}$ coordinate system and the $(x, y, z)_{\mathrm{p}_2}$ coordinate system uniformly approaches the identity map in the C^1 sense as $\mathrm{w} \to 0$ and $f \to \varphi_1$. For $d(\mathrm{p}_1, \mathrm{p}_2) \leq d(\mathrm{p}_1, \mathrm{p}_0) + d(\mathrm{p}_0, \mathrm{p}_2)$ which is dominated by h, and $\mathrm{h} \to 0$. Thus, at scale w, $S_2 \to S_1$. Similarly, at scale h, the planes $z = \pm\mathrm{h}$ in the p_2-coordinate system converge to the planes $z = \pm\mathrm{h}$ in the p_1-coordinate system. The trajectories of φ comprising R_2 are intrinsically defined; they are independent of which coordinate system is used. It follows that

$$\mathrm{c}R_1 \subset R_2 \subset \mathrm{c}^{-1}R_1.$$

To get the corresponding assertion with R_2 replaced by J_2, we let $e = \sqrt{c}$, which is merely another constant in $(0,1)$. We apply what we have just proved to the boxes eR_2 and $e^{-1}R_2$. As $\mathrm{w} \to 0$ and $f \to \varphi_1$, we can bracket the former between $eeR-1$ and $e^{-1}eR_1$, while we can bracket the latter between $ee^{-1}R_1$ and $e^{-1}e^{-1}R_1$. Since they themselves bracket J_2, concatenation gives

$$\mathrm{c}R_1 = e^2R_1 \subset eR_2 \subset J_2 \subset e^{-1}R_2 \subset e^{-2}R_1 = \mathrm{c}^{-1}R_1.$$

The same manipulation with $e = \sqrt{c}$ proves that $\mathrm{c}J_1 \subset R_2 \subset \mathrm{c}^{-1}J_1$. Q.E.D.

To understand the degree to which a set A saturates a julienne J we would like a nonlinear Fubini estimate that relates the measure of A in J to its measure in the $\mathcal{W}^{\mathrm{cs}}$-slices of J. This is not quite straightforward due to the lack of absolute continuity of $\mathcal{W}^{\mathrm{cs}}$. The **density** of one measurable set X in another measurable set Y is the conditional measure of $X \cap Y$ in Y, namely

$$m(X\colon Y) = \frac{m(X \cap Y)}{m(Y)}$$

where m is the measure and we assume $0 < m(Y) < \infty$. Given measurable sets $X_i, \ldots, X_n$, the minimum of $m(X_i\colon X_j)$ over $1 \le i,\ j \le n$ is their **mutual density**. By taking c near 1, the Julienne Nesting Lemma lets us assume that

the mutual density of R_1, R_2, J_1, J_2 is near 1.

For it is clear that $m(\mathrm{c}R\colon \mathrm{c}^{-1}R) = \mathrm{c}^6 \doteq 1$ when $\mathrm{c} \doteq 1$.

The plane $x = \xi$ meets the box $R = R(\mathrm{p}, \mathrm{p}_0, \mathrm{w}, \mathrm{h})$ in a **center stable slice**

$$R_\xi^{\mathrm{cs}} = \{(x,y,z) \in R\colon x = \xi\}.$$

The center stable **midslice** of R is $R_{x_0}^{\mathrm{cs}}$, which we also denote as R^{cs}. See Figure 17. (As usual, p_0 has coordinates $(\mathrm{x}_0, \mathrm{y}_0, 0)$.) The foliation of R by these slices is smooth and Fubini's Theorem applies to it. As a further abuse of notation, we use the same letter "m" to indicate M-volume, slice area, and density respecting either of these measures.

3.2 LEMMA. *Let $\theta < d < 1$ and $.1 \le \rho \le .9$ be given constants. If f C^2-approximates φ_1 well enough and $\mathrm{w} = \mathrm{h}^{3/2}$ is small enough then for any essentially $\mathcal{W}^{\mathrm{u}}$-saturated set A,*

(5) $m(A\colon R) \ge \rho \Longrightarrow m(A^{\mathrm{u}}\colon R^{\mathrm{cs}}) \ge d\rho$

(6) $m(A^{\mathrm{u}}\colon R^{\mathrm{cs}}) \ge \rho \Longrightarrow m(A\colon R) \ge d\rho$,

where $R = R(\mathrm{p,p,w,h})$ and A^{u} is the $\mathcal{W}^{\mathrm{u}}$-saturate of A. See Figure 18.

Proof. First note that since A is essentially $\mathcal{W}^{\mathrm{u}}$-saturated, A^{u} differs from A by a zero set, so $m(A\colon R) = m(A^{\mathrm{u}}\colon R)$ in (5), (6). The letter "d"

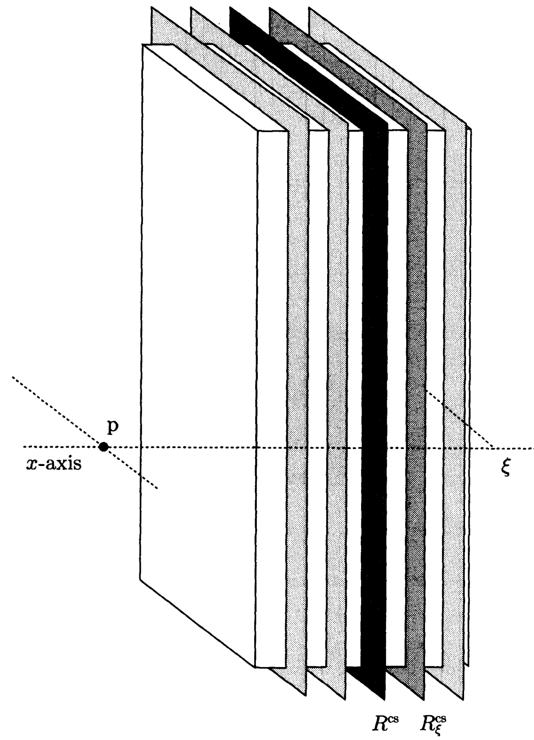

FIGURE 17. Center stable slices of R, extended beyond R.

stands for "deteriorates": the density of A in R deteriorates to density $d\rho$ in the slice. Choose constants $\mathrm{c}, \mathrm{r} \in (0, 1)$ such that

$$\mathrm{r}(1 - 10(1 - \mathrm{c}^3)) \geq d.$$

This is possible since for $\mathrm{c} = \mathrm{r} = 1$, we have $1 > d$. Let $A^{\mathrm{u}}_{\mathrm{c}}$ denote the set of arcs $W^{\mathrm{u}}_{\mathrm{loc}}(\mathrm{q}) \cap R$ where $\mathrm{q} \in A^{\mathrm{u}} \cap \mathrm{c}R$. Clearly

$$A^{\mathrm{u}}_{\mathrm{c}} \cap \mathrm{c}R = A^{\mathrm{u}} \cap \mathrm{c}R.$$

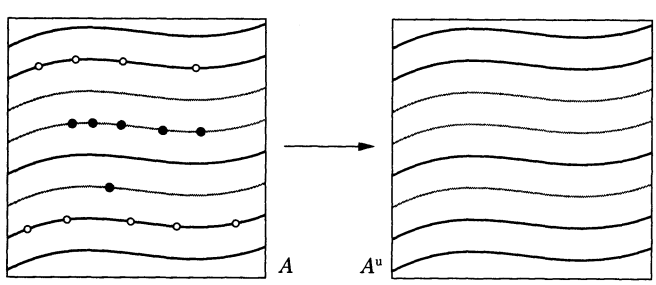

FIGURE 18. $\mathcal{W}^{\mathrm{u}}$-leaves in A are black while those in its complement are gray. A^{u} discards leaves which meet A in sets of leaf-measure zero, while it fills in leaves which meet A in sets of full leaf measure.

Since the vectors tangent to these arcs $W^{\mathrm{u}}_{\mathrm{loc}}(\mathrm{q}) \cap R$ approximate $E^{\mathrm{u}}_{\mathrm{p}}$ and since the constant c is fixed, these arcs stretch all the way across R when w is small and f C^2-approximates φ_1 well. See Figure 19. That is, the $\mathcal{W}^{\mathrm{u}}$-holonomy map is well defined:

$$\pi^{\mathrm{u}}\colon\ A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}} \to A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}}_x, \quad -\mathrm{w} \le x \le \mathrm{w}.$$

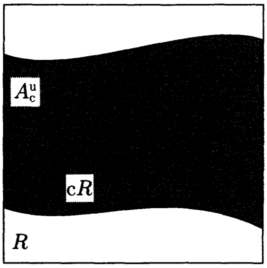

FIGURE 19. The arcs $W^{\mathrm{u}}_{\mathrm{loc}}(\mathrm{q})$ stretch across R without hitting the top or bottom.

According to Lemma 2.2, its Radon-Nikodym derivative is nearly 1. This lets us assume

$$\mathrm{r} \le \text{Radon-Nikodym derivative of } \pi^{\mathrm{u}} \le \mathrm{r}^{-1}$$

on R^{cs}. Consequently, the area of the slices $A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}}_x$ are all about the same, from which (5), (6) follow easily. More precisely, assume that $m(A\colon R) \ge \rho$. Then

$$\begin{aligned} m(A^{\mathrm{u}}\colon R^{\mathrm{cs}}) \ge m(A^{\mathrm{u}}_{\mathrm{c}}\colon R^{\mathrm{cs}}) &\ge \frac{\frac{\mathrm{r}}{2\mathrm{w}}\int_{-\mathrm{w}}^{\mathrm{w}} m(A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}}_x)dx}{m(R^{\mathrm{cs}})} = \frac{\mathrm{r}}{m(R)} m(A^{\mathrm{u}}_{\mathrm{c}} \cap R) \\ &= \frac{\mathrm{r}}{m(R)}(m(A^{\mathrm{u}} \cap R) - (m(A^{\mathrm{u}} \cap R) - m(A^{\mathrm{u}}_{\mathrm{c}} \cap R)) \\ &\ge \mathrm{r}\left(\rho - \frac{m(R \setminus \mathrm{c}R)}{m(R)}\right) \\ &= \mathrm{r}(\rho - (1-\mathrm{c}^3)) = \mathrm{r}\left(1 - \frac{1-\mathrm{c}^3}{\rho}\right)\rho \ge \mathrm{r}(1 - 10(1-\mathrm{c}^3))\rho \ge d\rho, \end{aligned}$$

which verifies (5). The proof of (6) is similar; assume that $m(A^{\mathrm{u}}\colon R^{\mathrm{cs}})\rho \ge \rho$. Then

$$m(A\colon R) \ge \frac{\int_{-\mathrm{w}}^{\mathrm{w}} m(A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}}_x)dx}{m(R^{\mathrm{cs}})} \ge \mathrm{r}\frac{m(A^{\mathrm{u}}_{\mathrm{c}} \cap R^{\mathrm{cs}})}{m(R^{\mathrm{cs}})} \ge \mathrm{r}(1 - 10(1-\mathrm{c}^2))\rho > d\rho.$$

Q.E.D.

The nonlinear **center stable slice** of a julienne $J = J(\mathrm{p}, \mathrm{p}_0, \mathrm{w}, \mathrm{h})$ is defined similarly, using the center stable foliation $\mathcal{W}^{\mathrm{cs}}$ instead of the linear planes $x=$ constant. Namely, we set

$$J_x^{\mathrm{cs}} = W_{\mathrm{loc}}^{\mathrm{cs}}(x) \cap J$$

for points x in the $\mathcal{L}^{\mathrm{u}}$-arc $L_{\mathrm{w}}^{\mathrm{u}}(\mathrm{p}_0)$. (Recall that this arc is the intersection of the $z=0$ plane with the center unstable manifold through p_0.) The **midslice** of J is the center stable slice through p_0; it is denoted as J^{cs}. See Figure 20.

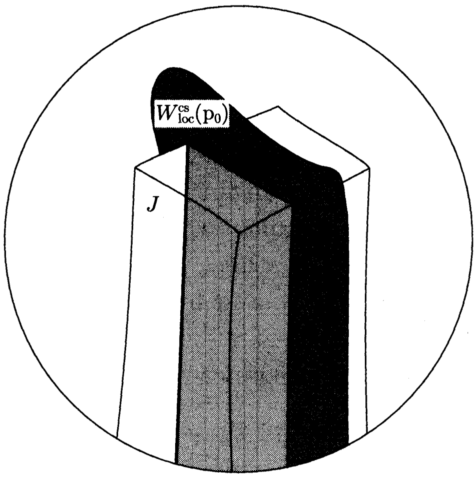

FIGURE 20. The center stable midslice of the julienne, viewed at the top.

3.3 LEMMA. *Under the same hypotheses as in Lemma* 3.2,

(7) $m(A\colon\, J) \geq \rho \Rightarrow m(A^{\mathrm{u}}\colon\, J^{\mathrm{cs}}) \geq d\rho$
(8) $m(A^{\mathrm{u}}\colon\, J^{\mathrm{cs}}) \geq \rho \Rightarrow m(A\colon\, J) \geq d\rho$,

where $J = J(\mathrm{p,p,w,h})$ *and* A^{u} *is the* $\mathcal{W}^{\mathrm{u}}$*-saturate of* A.

Proof. This follows immediately from Lemma 3.2 and the consequence of Lemma 3.1 which states that the mutual densities of R and J, and of R^{cs} and J^{cs}, are nearly equal to 1. Q.E.D.

3.4 LEMMA. *As* $W = \mathrm{h}^{3/2} \to 0$ *and* $f \to \varphi_1$ *in the* C^2 *sense,*

(9)
$$\frac{m(J^{\mathrm{cs}})}{4\mathrm{wh}} \rightrightarrows 1$$

where $J = J(\mathrm{p,p,w,h})$ *and* m *denotes area on the leaf* $W_{\mathrm{loc}}^{\mathrm{cs}}(\mathrm{p})$. *Also, if* J_+^{cs} *denotes the part of* J^{cs} *lying to the right of* $W_{\mathrm{h}}^{\mathrm{c}}(\mathrm{p})$, *then*

(10)
$$\frac{m(J_+^{cs})}{2\mathrm{wh}} \rightrightarrows 1.$$

(*See Figure* 21.)

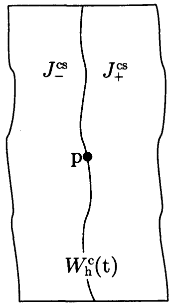

FIGURE 21. Although the edges of J^{cs} are bumpy, they stay away from each other.

Proof. (9) is an immediate consequence of Lemma 3.1 and the fact that $W^{cs}_{loc}(p)$ converges uniformly to E^{cs}_p in the C^1 sense. The proof of (10) amounts to re-inspecting the proof of Lemma 3.1. If q denotes the right endpoint of the base $L^s_w(p)$ of J^{cs} then the distance between $W^c_h(p)$ and $W^c_h(q)$ cannot differ from w by more than $Ch^{1+\alpha}$, and $h^{1+\alpha} \ll w$. At scale w the two center manifolds are nearly parallel lines at a distance of w apart, and of length 2h. Thus, the area of J^{cs}_+ is well approximated by 2wh. Q.E.D.

It goes without saying that all these geometric assertions about center stable slices hold equally well for center unstable slices.

4. Juliennes and holonomy

Next we discuss $\mathcal{W}^u$-holonomy at unit displacement. Consider J^{cs}, the center stable midslice of the julienne $J = J(p, p, w, h)$. If $p' \in W^u_1(p)$ then the $\mathcal{W}^u$-holonomy π^u carries J^{cs} homeomorphically into $W^{cs}_{loc}(p')$. According to Lemma 2.4, π^u is α-Hölder and α is near 1. Moreover, its Radon-Nikodym derivative nearly equals 1. Although π^u is locally quite different from an isometry, we claim that its overall effect on J^{cs} is nearly isometric in the following sense.

4.1 JULIENNE HOLONOMY LEMMA. *Let a constant* $c \in (0,1)$ *be given. If* p, p', J *are as above, if* $h = w^{2/3}$ *is small enough, and if* f C^2*-approximates* φ_1 *well enough then*

$$cJ'^{cs} \subset \pi^u(J^{cs}) \subset c^{-1} J'^{cs}$$

where $J' = J(p', p', w, h)$.

Proof. According to Lemma 2.3, π^u sends center manifolds to center manifolds, and restricted to each center manifold it is Lipschitz with Lipschitz constant approximately equal to 1. Thus, the image $\pi^u(J^{cs})$ is the union of center manifolds, all of whose heights are on the order of $(1 + o(1))h$. Since

π^u is α-Hölder, the top and bottom horizontal arcs of J^{cs} map to arcs of diameter $\leq \mathrm{C}\mathrm{w}^{\alpha}$ for some uniform constant C. Q.E.D.

The point of the whole construction: $\mathrm{C}\mathrm{w}^{\alpha}$ **is very small compared to** h**, even though it is huge compared to** w**.**

Set $e = \sqrt{c}$. Since $\mathrm{C}\mathrm{w}^{\alpha} \ll \mathrm{h}$, the π^u-image of the top boundary arc of J^{cs} lies between the planes $z = e\mathrm{h}$ and $z = e^{-1}\mathrm{h}$ in the (x, y, z) coordinate system at p′, and the bottom arc has image between the planes $z = -e^{-1}\mathrm{h}$ and $z = -e\mathrm{h}$. See Figure 22.

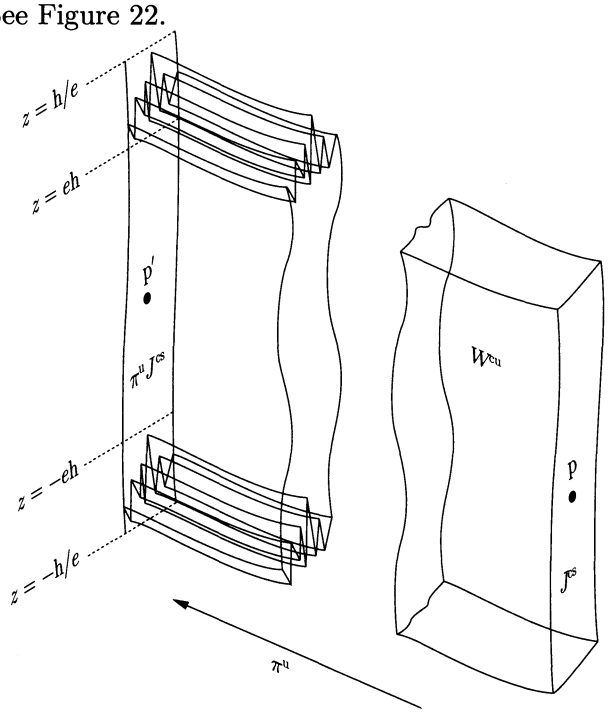

FIGURE 22. The julienne J^{cs} and its holonomy image $\pi^u J^{cs}$ in J'^{cs}.

The same reasoning does *not* imply that the width of $\pi^u(J^{cs})$ is $\leq 2\mathrm{c}^{-1}\mathrm{w}$. Instead, we will use the fact that the Radon-Nikodym derivative of π^u is nearly equal to 1. Let J^{cs}_+ denote the right half of J^{cs}; it is $\bigcup W^c_{\mathrm{h}}(\mathrm{t})$ where t ranges over the right half of the arc $L^s_{\mathrm{w}}(\mathrm{p})$. (These points t have positive y coordinate. Recall that $L^s_{\mathrm{loc}}(\mathrm{p})$ is the intersection of $W^{cs}_{\mathrm{loc}}(\mathrm{p})$ with the plane $z = 0$.) Let $W^c_{\mathrm{h}}(\mathrm{q})$ be the right edge of J^{cs}_+. Consider also the right half J'^{cs}_+ of J'^{cs}, and especially consider its right edge, $W^c_{\mathrm{h}}(\mathrm{q}')$. The holonomy map π^u preserves orientation and carries $W^c_{\mathrm{h}}(\mathrm{q})$ into a local center manifold $W^c_{\mathrm{loc}}(\mathrm{r})$. See Figure 23.

We claim that r approximates q′ at scale w. Suppose first that $\mathrm{r} \in L^s_{\mathrm{cw}}(\mathrm{p}')$. That is, r is too close to p′. Two dimensionality of J'^{cs} implies that $W^c_{\mathrm{loc}}(\mathrm{r})$ lies

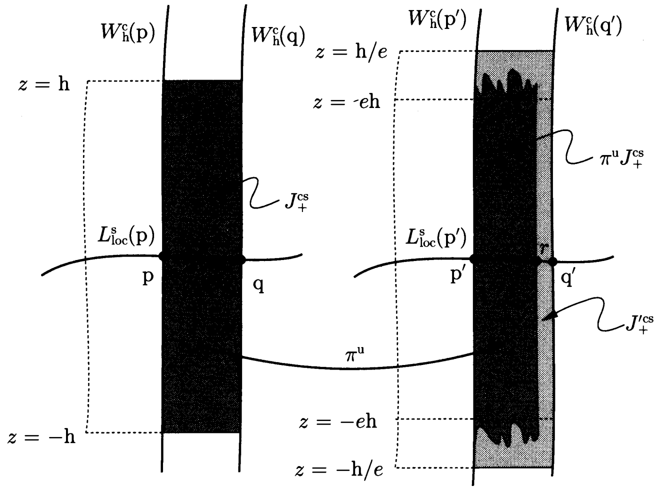

FIGURE 23. J^{cs}_+ and its π^u holonomy image.

wholly between $W^c_{loc}(p')$ and $W^c_{loc}(cq')$ where cq' denotes the right endpoint of $L^s_{cs}(p')$. See Figure 24a. According to Lemma 3.4, the area of the part of $W^{cs}_{loc}(p')$ bounded by $W^c_{loc}(p')$, $W^c_{loc}(cq')$, and the arcs where $W^{cs}_{loc}(p')$ meets the planes $z = \pm e^{-1}h$ is well approximated by $(cw)(2e^{-1}h) = 2ewh$. Moreover, this region contains $\pi^u(J^{cs}_+)$ which has area well approximated by 2wh. Since $e < 1$, this is a contradiction.

On the other hand if r is too far from p′ in the sense $c^{-1}q'$ lies to the its left, then $\pi^u(J^{cs}_+)$ contains the region bounded by $W^c_{loc}(p'), W^c_{loc}(c^{-1}q')$, and the arcs where the planes $z = \pm eh$ meet $W^{cs}_{loc}(p')$. See Figure 24b. The area of this region is well approximated by $(c^{-1}w)(2eh) = 2e^{-1}wh$, while the larger region $\pi^u(J^{cs}_+)$ has area well approximated by 2wh. Since $e^{-1} > 1$, this is a contradiction. Thus, r lies somewhere between cq′ and $c^{-1}q'$, so that

$$cJ'^{cs}_+ \subset \pi^u(J^{cs}_+) \subset c^{-1}J'^{cs}_+.$$

The same holds for the left-hand half of the juliennes, and the lemma is proved. Q.E.D.

4.2 JULIENNE HOLONOMY DENSITY LEMMA. *Let* $c,\rho \in (0,1)$ *be constants such that* $.1 \le c\rho < \rho \le .9$, *and let* f,φ,J,p,p',w,h *have the same meanings as in Lemma* 4.1. *If* $h = w^{2/3}$ *is small enough, if* f C^2 *approximates* φ_1 *well enough, and if* A *is an essentially* $\mathcal{W}^u$*-saturated set which has density* $\ge \rho$ *in* J, *then* A *has density* $\ge c\rho$ *in* J'.

Proof. The effect here is to let us "turn the corner". Let $d = c^{1/4}$. By Lemma 3.3, if $w = h^{2/3}$ is small enough and f C^2 approximates φ_1 well enough,

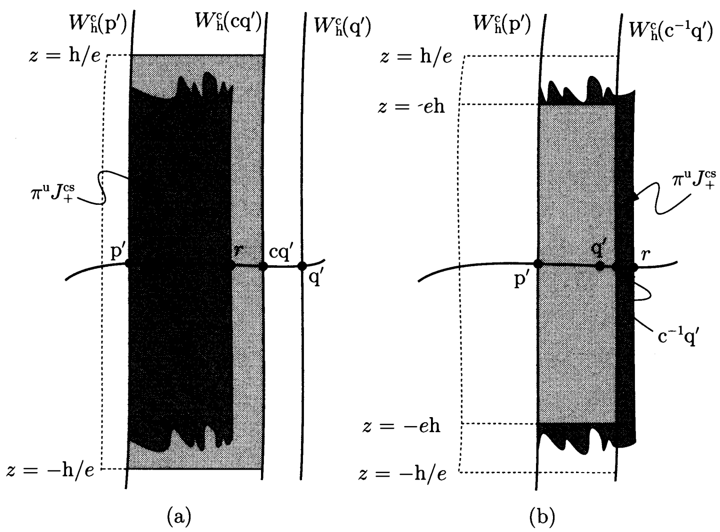

FIGURE 24. (a) r is too near p$'$. (b) r is too far from p$'$.

the density of $A^{\rm u}$ in $J^{\rm cs}$ deteriorates by no worse than d,

$$m(A^{\rm u}\colon\ J^{\rm cs}) \geq d\rho.$$

Since $A^{\rm u}$ is $\mathcal{W}^{\rm u}$-saturated, $\pi^{\rm u}$ sends $A^{\rm u}\cap J^{\rm cs}$ onto $A^{\rm u}\cap\pi^{\rm u}(J^{\rm cs})$. The Radon-Nikodym derivative of $\pi^{\rm u}$ is nearly equal to 1, and so

$$m(A^{\rm u}\colon\ \pi^{\rm u}(J^{\rm cs})) \geq d^2\rho.$$

By Lemma 4.1, the density of $J'^{\rm cs}$ and $\pi^{\rm u}(J^{\rm cs})$ in each other is nearly equal to 1 and so

$$m(A^{\rm u}\colon\ J'^{\rm cs}) \geq d^3\rho.$$

By Lemma 3.3, it follows that $m(A\colon\ J') \geq d^4\rho = {\rm c}\rho$. Q.E.D.

5. Ergodicity

Now we pull together the estimates in the preceding sections and prove our main theorem—"the small volume-preserving perturbation f of the time-one map φ_1 of the geodesic flow on a surface of constant negative curvature is ergodic". In the introduction we showed that if f is not ergodic then there exist measurable sets $A, B \subset M^3$ which are disjoint, have positive volume, and are composed of essentially complete $\mathcal{W}^{\rm u}$-leaves and essentially complete $\mathcal{W}^{\rm s}$-leaves. (To do this we used Birkhoff averages, which is where the volume-preserving hypothesis entered the picture.)

Let $a \in A$ and $b \in B$ be density points of A and B. For all sufficiently small h, the density in a small cube with side 2h about a or b is $\geq .9$. Divide the cubes into rectilinear boxes of width 2w and height 2h where $w = h^{3/2}$. (This requires $h^{-1/2}$ to be an integer.) It follows that the density of A in one of these boxes $R(a, a_0, w, h)$ near a is $\geq .9$, as is the density of B in one of the boxes $R(b, b_0, w, h)$ near b. According to Lemma 3.1, these boxes are highly concentrated in the corresponding juliennes $J_{a_0} = J(a_0, a_0, w, h)$ and $J_{b_0} = J(b_0, b_0, w, h)$, so we may assume that the density of A in J_{a_0} is $\geq .8$, as is the density of B in J_{b_0}. (Note that in this application of Lemma 3.1, $p_1 = a$, and $p_2 = a_0 = p_0$. The fact that a_0 is not necessarily equal to a is the reason that we dealt with $p_0 \neq p_1$ in Lemma 3.1. Note also that the hypotheses $d(p_1, p_0)/h$ and $d(p_2, p_0)/h$ being bounded are met because a_0 lies in the cube of edge 2h around a.)

Let N be the number supplied by Lemma 1.1. Any two points of M can be connected by a $\mathcal{W}^{u,s}$ path with N edges. Fix the constant $c = (.5)^{1/N}$. If necessary decrease h so that the hypothesis of Lemma 4.2 is satisfied for this choice of c. There is a $\mathcal{W}^{u,s}$ path $a_0 = p_0 \to p_1 \to p_2 \to \cdots \to p_N = b_0$. Lemma 4.2 asserts that at each step in the path the density of A in the julienne decays at worst by the factor c. Therefore, the density of A in the N^{th} julienne, $J(p_N, w, h) = J_{b_0}$, is at least half of its original density of .8 in J_{a_0}. That is, $m(A\colon J_{b_0}) \geq .4$, contrary to the fact that the disjoint set B has density $\geq .8$ in the same julienne. Q.E.D.

IBM T.J. WATSON RESEARCH CENTER, YORKTOWN HEIGHTS, NY, first and third authors
UNIVERSITY OF CALIFORNIA, BERKELEY

REFERENCES

[1] D.V. ANOSOV, Geodesic flows on closed Riemannian manifolds with negative curvature, Proc. Steklov Inst. Math. **90**. English translation, A.M.S., 1967.

[2] D.V. ANOSOV and YA.G. SINAI, Some smooth ergodic systems, Upsekhi Mat. Nauk **22** (1967), 107–172 (= Russian Math Surveys, **22**(1967), 103–167).

[3] M.I. BRIN and JA.B. PESIN, Partially hyperbolic dynamical systems. Izv. Akad. Nauk SSR, **38**(1974) (= Math. USSR Izvestija **8**(1974), 177–218).

[4] Z. BUCZOLICH, Density points and bi-Lipschitz functions in $\mathbb{R}^m$, Proc. A.M.S. **116**(1992), 53–59.

[5] G.A. HEDLUND, Metric transitivity of the geodesics on closed surfaces of constant negative curvature, Ann. of Math. **35**(1934), 787–808.

[6] M.W. HIRSCH, C.C. PUGH, and M. SHUB, *Invariant Manifolds*, Springer Lecture Notes **583**, Heidelberg, 1977.

[7] E. HOPF, Statistik der geodatischen Linien in Mannigfaltigkeiten negativer Krummung, Berichten der Mathematisch-Physischen Klasse der Sashsichen Akademie der Wissenschaften-Leipzig, **11**(1939), 261–304.

[8] R. MAÑÉ, *Ergodic Theory and Differentiable Dynamics*, Ergebnisse der Mathematik, Springer-Verlag, Berlin, 1987.

[9] W. PARRY, Dynamical systems on nil-manifolds. Bull. London Math. Soc. **2**(1970), 37–40.

[10] C.C. Pugh and M. Shub, Ergodicity of Anosov actions, Invent. Math. **15**(1972), 1–23.
[11] ______, Ergodic attractors. Trans. A.M.S. **312**(1989), 1–54.
[12] M. Shub, Instability, Symp. on diff. eq. and dyn. systems, Springer Lecture Notes **206**, 28–29, 1969.
[13] ______, Topologically transitive diffeomorphisms of T^4, Symp. on diff. eq. and dyn. systems, Springer Lecture Notes **206**, 39–40, 1969.
[14] ______, *Global Stability of Dynamical Systems.* Springer Verlag, New York, 1986.
[15] P. Walters, Affine transformations of tori, Trans. A.M.S. **131**(1968), 40–50.

(Received October 26, 1992)