RC22837 (W0307-028) July 2, 2003
Mathematics

IBM Research Report

Stable Ergodicity

Charles Pugh
University of California at Berkeley
Berkeley, CA 94720

Michael Shub
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

STABLE ERGODICITY

CHARLES PUGH AND MICHAEL SHUB

June 30, 2003

1. Introduction

A dynamical system is **ergodic** if it preserves a measure and each measurable invariant set is a zero set or the complement of a zero set. No measurable invariant set has intermediate measure. See also Section 6. The classic real world example of ergodicity is how gas particles mix. At time zero, chambers of oxygen and nitrogen are separated by a wall. When the wall is removed the gasses mix thoroughly as time tends to infinity. In contrast think of the rotation of a sphere. All points move along latitudes, and ergodicity fails due to existence of invariant equatorial bands. Ergodicity is **stable** if it persists under perturbation of the dynamical system. In this paper we ask: "how common are ergodicity and stable ergodicity?" and we propose an answer along the lines of the Boltzmann hypothesis – "very."

There are two competing forces that govern ergodicity – hyperbolicity and the KAM phenomenon. The former promotes ergodicity and the latter impedes it. One of the striking applications of KAM theory and its more recent variants is the existence of open sets of volume preserving dynamical systems each of which possesses a positive measure set of invariant tori and hence fails to be ergodic. Stable ergodicity fails dramatically for these systems. But does the lack of ergodicity persist if the system is weakly coupled to another? That is, what happens if you have a KAM system or one of its perturbations that refuses to be ergodic, due to these positive measure sets of invariant tori, but somewhere in the universe there is a hyperbolic or partially hyperbolic system weakly coupled to it? – does the lack of egrodicity persist? The answer is "no," at least under reasonable conditions on the hyperbolic factor. See Section 13 for more details and the proof of

Theorem A. *If a volume preserving hyperbolic system with sufficiently strong hyperbolicity is weakly coupled to a KAM system then often the resulting dynamical system is not only ergodic, it is stably ergodic.*

In short,

Hyperbolicity trumps KAM, and
ergodicity often reigns à la Boltzmann.

Our theme has long been that "a little hyperbolicity goes a long way toward ergodicity," and we continue to hold hope for the

Main Conjecture. *Among the volume preserving partially hyperbolic dynamical systems, the stably ergodic ones form an open and dense set.*

The first author was supported in part by IBM. The second author was supported in part by NSF Grant #DMS-9988809.

Theorem A is a step in that direction. See also Section 18. In Section 7 we formulate Theorem E in which we give sufficient conditions for ergodicity and stable ergodicity of a partially hyperbolic system.

We also discuss at length some examples of stable ergodicity that arise in Lie group dynamics. We are interested in the following question: does stability of ergodicity in which the perturbations occur only in the Lie group context imply stable ergodicity? That is, is affine stable ergodicity **decisive** for general stable ergodicity? In this direction, we have

Theorem B. *In a large class of affine diffeomorphisms of homogeneous spaces, stable ergodicity within the class is decisive for general stable ergodicity.*

See Section 16 for details and the proof.

Notes. We assume for the most part that the phase space of our dynamical system is a smooth compact manifold M and that m is a smooth invariant volume form on M. Smooth means C^∞, although C^2, or in some cases $C^{1+\epsilon}$ with small $\epsilon > 0$, is usually good enough. Questions about ergodicity in the C^1 world seem to be fundamentally different.

In the case of a flow, for all $t \in \mathbb{R}$, $\varphi_t : M \to M$ is a diffeomorphism and $\varphi_{t+s} = \varphi_t \circ \varphi_s$. In the case of discrete time, t is restricted to $\mathbb{Z}$. Invariance of m means that $m(\varphi_t(S)) = m(S)$ for measurable sets S. We perturb a diffeomorphism in the function space of m-preserving C^2 diffeomorphisms, $\mathrm{Diff}^2_m(M)$, and we perturb a flow φ by perturbing its tangent vector field $\dot\varphi$ in the space of C^2 divergence-free vector fields, $\mathcal{X}^2_m(M)$.

For an easy example of ergodicity, think of the torus as the square with edges identified in the usual way, and think of the flow that translates all points along lines of slope α, where α is an irrational number. The flow preserves area on the torus and is ergodic. There are no measurable invariant sets of intermediate area. The proof is elementary, but not immediate. Although ergodic, the flow is not stably ergodic, for the slope α can be perturbed to become rational, and as in the case of the sphere, there are invariant bands that deny ergodicity.

The concept of ergodicity originated in statistical mechanics. See Gallavotti's book [Gall] for an extensive bibliography.

The KAM (Kolmogorov, Arnold, and Moser) phenomenon is built on an example of an area preserving diffeomorphism of the plane that has a fixed point at the origin. See de la LLave's article [de la Llave] for a thorough description of the theory. The origin is surrounded by a Cantor set of invariant closed curves on which the map is conjugate to irrational rotation. The curves bound invariant annuli, and this denies ergodicity. Enough annuli persist after perturbation in the area preserving C^4 topology to show that the diffeomorphism is persistently non-ergodic. It is also not partially hyperbolic, and hence does not contradict the Main Conjecture. The generalization to volume preserving dynamics in higher dimensions referred to above is due to Cheng and Sun [ChSu], Herman [Yoc], and others. Theorem E in Section 7 is the main result of our investigation of stable ergodicity that we made in collaboration with Keith Burns and Amie Wilkinson. The present paper borrows much from [BuPugShWi]. Other references appear in the relevant sections below.

2. Hyperbolicity

A flow is hyperbolic if its orbits look like those in Figure 1. Orbits that are

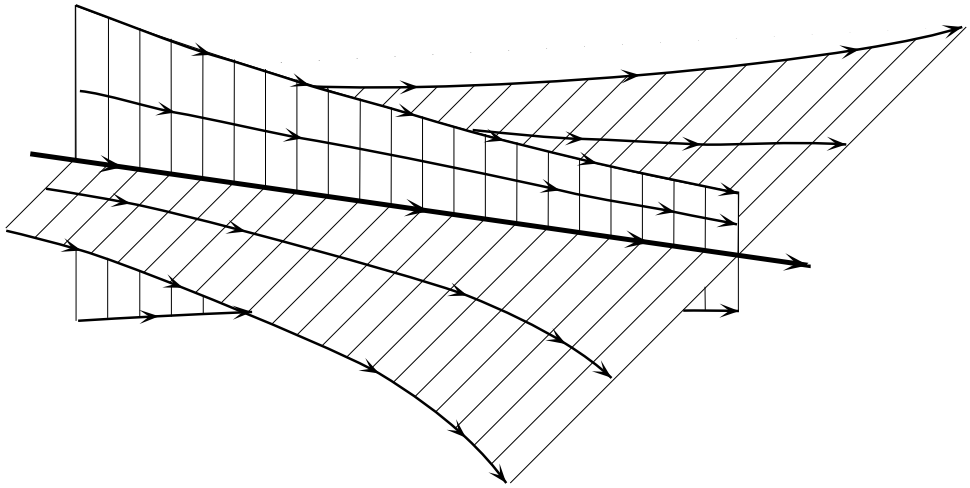

FIGURE 1. Local view of a hyperbolic orbit. The curves with arrowheads are orbits. The other lines are stable and unstable manifolds.

attracted toward each other asymptotically as time increases form a **stable manifold** and those that are repelled apart form an **unstable manifold**. We denote the unstable and stable manifolds systematically as W^u and W^s. A flow is hyperbolic if all its orbits are hyperbolic. The whole phase space of the flow is filled with stable and unstable manifolds. They foliate the phase space, and as discussed in Sections 4 and 9, the properties of these invariant foliations have a lot to do with ergodicity. We denote them as $\mathcal{W}^u$ and $\mathcal{W}^s$. Here is the formal definition.

Definition. A flow φ on M is **hyperbolic** if there is a splitting $TM = E^u \oplus E^o \oplus E^s$ such that

(a) E^u, E^o, E^s are continuous subbundles of TM that are invariant under $T\varphi$ in the sense that $T\varphi_t(E^u_p) = E^u_{\varphi_t p}$, $T\varphi_t(E^o_p) = E^o_{\varphi_t p}$, and $T\varphi_t(E^s_p) = E^s_{\varphi_t p}$, for all $p \in M$ and all $t \in \mathbb{R}$.

(b) The orbit bundle E^o is tangent to the orbits, $E^o_p = \text{span}(\dot\varphi(p))$ for all $p \in M$.

(c) $T\varphi$ exponentially expands the unstable bundle E^u in the sense that for some $\lambda > 1$ and a constant $c > 0$,

$$|T\varphi_t(v)| \geq c\lambda^t |v|$$

holds for all $v \in E^u$ and all $t \geq 0$.

(d) $T\varphi$ exponentially contracts the stable bundle E^s in the sense that for some constant C,

$$|T\varphi_t(v)| \leq C\lambda^{-t} |v|$$

holds for all $v \in E^s$ and all $t \geq 0$.

Obviously, φ is hyperbolic if and only if the time reversed flow $\psi_t(x) = \varphi_{-t}(x)$ is hyperbolic. Stability for one is unstability for the other. The leaves of $\mathcal{W}^u$ and $\mathcal{W}^s$ are the invariant manifolds, and they are everywhere tangent to E^u and E^s.

Since the bundle E^o is continuous, the orbits of φ are all nonsingular – a hyperbolic flow has no fixed points. The definition presupposes that TM carries a

norm since (c), (d) refer to the length of vectors in TM, but the particular choice of norm is irrelevant because we can choose the constants c, C at will.

One of the basic examples of ergodicity and hyperbolicity is the geodesic flow on the unit tangent bundle of a surface of constant negative curvature. It is the mother of all examples, so we will spend some time describing it in detail in the next section.

The definition of hyperbolicity for discrete time systems is the same, except that the orbit bundle E^o is zero. After all, the orbits of a diffeomorphism are zero-dimensional. The simplest example of a hyperbolic diffeomorphism is the **Thom map** (also called the cat map, see [ArAv]) of the 2-torus $f : \mathbb{T}^2 \to \mathbb{T}^2$ given by the matrix

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

The linear map defined by the preceding 2×2 matrix is an isomorphism $\mathbb{R}^2 \to \mathbb{R}^2$ that sends the integer lattice $\mathbb{Z}^2$ onto itself. This gives the diffeomorphism f on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. The eigenvalues of the matrix are $(\pm\sqrt{5} + 3)/2$ and the lines with slope $(\pm\sqrt{5} - 1)/2$ are the invariant manifolds.

Hyperbolic dynamical systems have a long history. How the eigenvalues at a fixed point determine the local phase portrait has been described by Hadamard, Poincaré, and many others. The global, systematic formulation of hyperbolicity appeared around 1960 and is due to Smale. See [Sm] for its best embodiment at the time. Globally hyperbolic systems are also called **Anosov systems** or, in Anosov's terms, systems that satisfy "Condition U." More than anything else, Anosov's thesis [An] has been the inspiration for our work on ergodicity. See Section 4.

3. The Geodesic Flow

A geodesic flow is defined geometrically as follows. Take a surface M equipped with a smooth Riemann structure. (A Riemann structure is a smooth choice of an inner product on each tangent space. If M happens to be a subset of $\mathbb{R}^3$ then one can choose the inner product on T_pM to be the one it inherits from $\mathbb{R}^3$.) Let v be a tangent vector, say at the point p. Through p and tangent to v there passes a unique geodesic, say $\gamma(t)$. Thus, γ locally minimizes arclength, $\gamma(0) = p$, and $\gamma'(0) = v$. The **geodesic flow** is defined by

$$\phi_t(v) = \gamma'(t).$$

It is a flow on the tangent bundle, not on the surface. Its orbits are curves of tangent vectors. In fact, $|\gamma'(t)| = |v|$, which is to say that the geodesic flow preserves the length of tangent vectors and it defines a flow on the unit tangent bundle, $T_1M = \{v \in TM : |v| = 1\}$. When speaking of the geodesic flow, it will be understood that it is the restriction of ϕ to the unit tangent bundle. If M has dimension n then T_1M has dimension $2n - 1$.

Remark. The fact that the geodesic flow occurs on the tangent bundle T_1M and not on the manifold M is a major stumbling block to overcome when first reading about the subject. Orbits of the geodesic flow are not geodesics. After all, geodesics cross each other and orbits never do. Although dimension constraints restrict us to drawing pictures of geodesics, the intent is to suggest tangent vectors in motion.

The unit tangent bundle supports a natural measure, Liouville measure. It is a smooth volume form. The geodesic flow leaves Liouville measure invariant.

The curvature of a surface is positive, zero, or negative in a neighborhood of a point p, according to whether two geodesics initially perpendicular to a short geodesic arc through p converge, stay parallel, or diverge. See Figure 2.

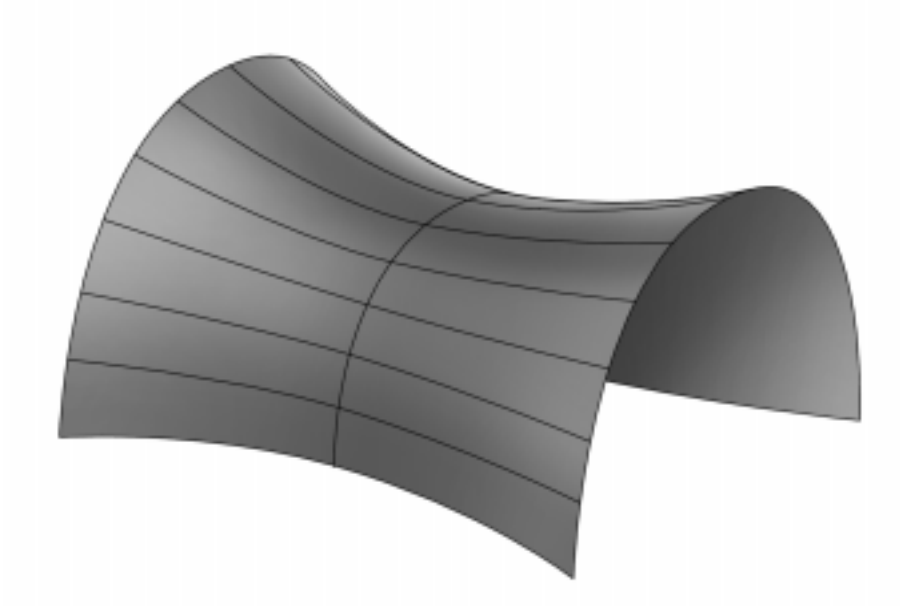

FIGURE 2. Divergence of initially parallel geodesics on a saddle.

Locally a surface of positive curvature resembles a sphere, a surface of zero curvature is a plane, and a surface of negative curvature resembles a saddle.

As will be clear in what follows, it is negative curvature that captures our interest, and in that case we give a geometric description of the geodesic flow as follows. We begin with the upper half plane $\mathbb{H} = \{x + iy = z \in \mathbb{C} : y > 0\}$, and equip it with the hyperbolic Riemann structure

$$\langle u, v \rangle_z = \frac{u \cdot v}{y^2}$$

where $u, v \in T_z\mathbb{H} \approx \mathbb{R}^2$ and $u \cdot v$ is the Euclidean dot product. With respect to this Riemann structure, the geodesics are the vertical half-lines and the semi-circles that meet the x-axis perpendicularly. (The geodesics all meet $\partial\mathbb{H}$ perpendicularly – the verticals do so at infinity.) With respect to the hyperbolic Riemann structure, $\mathbb{H}$ has constant curvature -1. See Figure 3.

Assume that M has constant negative curvature. Rescaling M lets us assume that the curvature is -1. The universal cover of M is $\mathbb{H}$, and the geodesic flow on T_1M is merely a quotient of the geodesic flow on $T_1\mathbb{H}$, so it is enough to study $\mathbb{H}$. (Equivalently, one could study the unit disc Δ equipped with the hyperbolic Riemann structure $\langle u, v \rangle_z = (4\, u \cdot v)/(1 - r^2)^2$ where $r = |z|$. For $\mathbb{H}$ is isometric to Δ. The factor 4 is chosen to get curvature -1.)

A **horocycle** is a circle in $\mathbb{H}$ that is tangent to $\partial\mathbb{H}$. This includes the case of a horizontal line ℓ. It is tangent to $\partial\mathbb{H}$ at infinity. Each vector $v \in T_1\mathbb{H}$ determines a horocycle H to which it and its negative are perpendicular. Let $\nu^+(H)$ be the

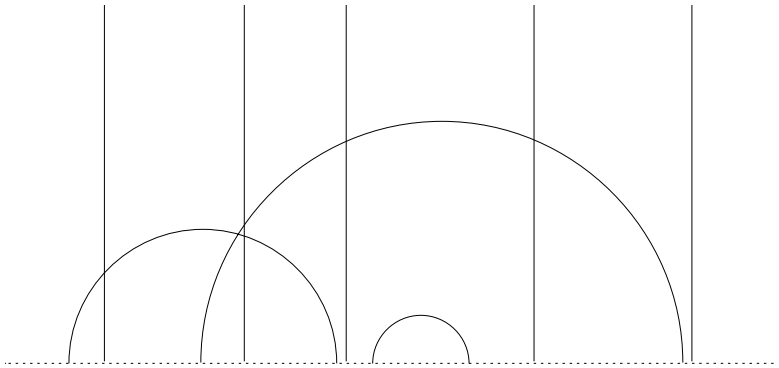

FIGURE 3. Geodesics in the upper half plane model of $\mathbb{H}$ are vertical lines and semi-circles perpendicular to the x-axis.

inward pointing unit normal bundle of H. When $H = \ell$, inward means upward because in both cases the vector aims its geodesic toward the point at which the horocycle is tangent to $\partial\mathbb{H}$. Under the geodesic flow Φ on $T_1\mathbb{H}$, vectors in $\nu^+(H)$ flow to vectors in $\nu^+(H_t)$ where H_t is inside H. The hyperbolic distance between vectors in $\nu^+(H_t)$ tends exponentially to zero as $t \to \infty$. To check this assertion, consider the horocycle ℓ through the imaginary number i. Vectors $v \in \nu^+(\ell)$ flow to upward unit vectors at points of the horizontal line ℓ_t through $e^t i$. The Euclidean length of $\Phi_t(v)$ is e^t but its hyperbolic length remains 1. The distance between vectors in $\nu^+(\ell_t)$ decreases exponentially because for large y, hyperbolic distance is much less than Euclidean distance. Isometries act transitively on $\mathbb{H}$, so if H is a general horocycle there is an isometry that carries ℓ to H and carries the upward flowing normal bundle $\nu^+(\ell_t)$ to the inward flowing normal bundle $\nu^+(H_t)$. See Figure 4.

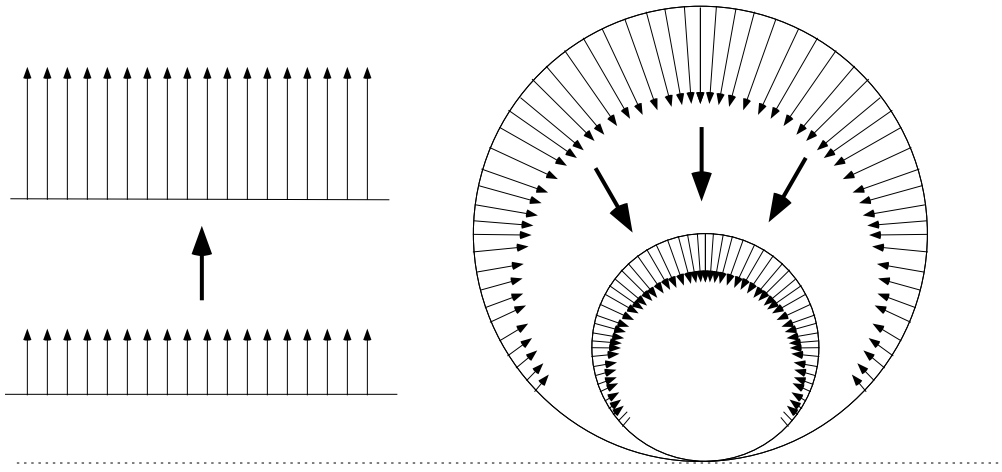

FIGURE 4. How vectors perpendicular to a horocycle flow in positive time.

This means that all vectors in $\nu^+(H)$ belong to the same stable manifold, and symmetrically, all vectors in the outward unit normal bundle $\nu^-(H)$ belong to the same unstable manifold:

$$W^s(v) = \nu^+(H) \qquad W^u(-v) = \nu^-(H)$$

for all $v \in \nu^+(H)$. In short, the stable and unstable manifolds are unit inward and outward normal bundles to horocycles. From this geometric description it follows that

Φ *and its quotient* ϕ *on* T_1M *are hyperbolic flows.*

See Section 14 for an algebraic description of all this.

4. Anosov's Thesis – the short version

The curvature concept makes sense also in higher dimensions, and has a similar description in terms of geodesic convergence/divergence. A central result in Anosov's thesis is

Anosov's Ergodicty Theorem. *The geodesic flow for a manifold of negative curvature is ergodic; in fact it is stably ergodic.*

See Section 8 for a sketch of the proof, and Section 12 for a proof that the time-t map itself is stably ergodic, $t \neq 0$. Anosov's proof proceeds in four four main steps.

First, by the Lobachevsky-Hadamard Theorem, negative curvature implies that the geodesic flow on the unit tangent bundle is hyperbolic: its tangent has a three-way splitting as above. (Keep in mind that the tangent to the geodesic flow lives on the second tangent bundle, or more exactly, it lives on $T(T_1M) \subset T^2M$.)

Second, according to the Hadamard-Perron Theorem, a hyperbolic flow has stable and unstable manifolds as in Figure 1, and they foliate the phase space of the flow.

Third, the Birkhoff Ergodic Theorem implies that a measurable invariant set consists essentially of whole unstable manifolds and essentially of whole stable manifolds. Here, "essentially" means "up to a zero set."

Fourth, a set consisting essentially of whole unstable and whole stable manifolds has measure zero or its complement does, and thus the geodesic flow is ergodic.

None of these steps was easy, in part because the classical ODE theorems needed to be re-proved globally, but the last two steps were hardest. The difficulty arises from the fact that although the unstable and stable manifolds themselves are smooth as individual manifolds, their assembly as foliations may not be smooth. To put it another way, the subbundles E^u, E^s integrate (in the sense of the Frobenius Theorem about integrating a distribution) to the invariant foliations $\mathcal{W}^u$, $\mathcal{W}^s$, but as subbundles they may not be smooth. They are only continuous.*

It turns out that in the two-dimensional case, a negatively curved surface's stable and unstable manifold foliations (they are merely general solutions of ODEs when

* If you have not thought about this before, you will find it instructive to devise a prototype of this situation, namely a continuous vector field in the plane having three properties: it has a unique integral curve through each point, the integral curves are smooth, but the vector field is nowhere Lipschitz. In terms of differential equations, you have smooth solutions of a non-smooth ODE. This is even possible with "smooth" replaced by "analytic," and by the same token, analyticity of the geodesic flow is no guarantee of smoothness of the invariant foliations.

M is two-dimensional) are better than continuous. They are of class C^1. This fact let Hopf give a proof of ergodicity for surfaces, [Ho]. In higher dimensions the foliations are not always C^1, but they have a weaker property that can be used in place of differentiability. It is absolute continuity and is defined as follows.

Consider a foliation $\mathcal{F}$ of an n-manifold N by leaves of dimension s. Assume that the leaves are smooth and that the tangent planes to the leaves vary continuously – as is the case with the stable manifold foliation. Embed smooth $(n-s)$-dimensional discs transverse to a leaf of $\mathcal{F}$, say at nearby points p, q in the leaf. See Figure 5. Call the discs D_p, D_q, and consider the natural **holonomy map** $h : D_p \to D_q$,

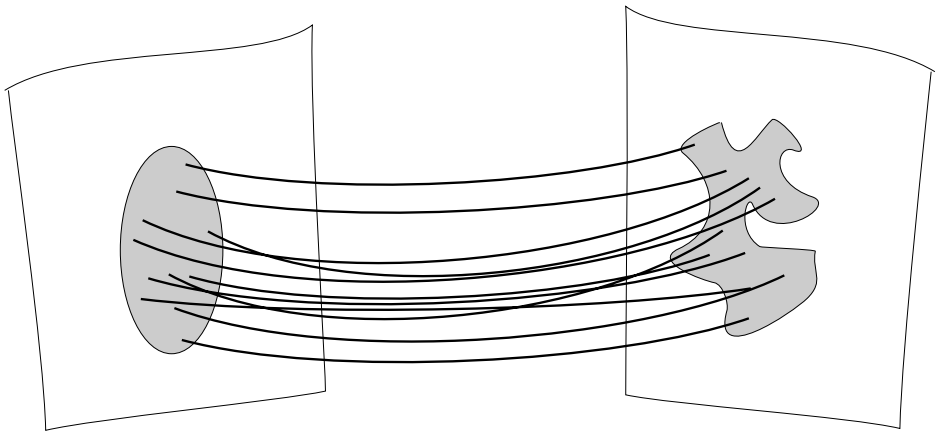

FIGURE 5. The holonomy map when $n = 3$ and $s = 1$.

which sends $y \in D_p$ to the unique point $h(y) \in D_q$ such that y and $h(y)$ lie in the same local leaf of $\mathcal{F}$. The holonomy map is a homeomorphism from a neighborhood of p in D_p to a neighborhood of q in D_q. A foliation is **absolutely continuous** if its holonomy maps transform zero sets of D_p to zero sets of D_q. The measure on the transversal discs can be any smooth $(n-s)$-dimensional volume, since all of them define the same collection of zero sets. Corresponding to a choice of such volumes, say m_p and m_q, it is a standard fact from measure theory that the holonomy map has an L^1 Radon-Nikodym derivative $\mathrm{RN}_x(h)$ satisfying

$$m_q(h(A)) = \int_A \mathrm{RN}_x(h)\, dm_p$$

for measurable subsets A of D_p.

The result that let Anosov proceed with his four step proof of ergodicity is that the stable and unstable foliations are absolutely continuous and the Radon-Nikodym derivatives of their holonomy maps are continuous. This is Theorem 10 in Anosov's thesis, and as he correctly remarks, it is the cornerstone to his analysis. He refers to it as a technical result because it addresses the technical issue of the degree of smoothness of a foliation. Previously, mathematicians had felt that showing the stable and unstable foliations were of class C^1 for manifolds of dimension ≥ 3 was mainly a matter of working harder. Anosov realized this was not so, and that the generalization lay in a different direction. What Hopf saw as C^1 foliations were actually the one-dimensional embodiment of absolutely continuous foliations whose holonomy maps have continuous Radon-Nikodym derivatives.

A key fact from elementary measure theory used by Anosov is the Lebesgue Density Theorem, which states that almost every point x of a Lebesgue measurable set A in Euclidean space is a **density point** of A in the sense that

$$\lim_{r\to 0} \frac{m(A \cap B_r(x))}{m(B_r(x))} = 1,$$

where $B_r(x)$ is the ball at x of radius r and m is Lebesgue measure. The fraction above is the concentration of A in the ball $B_r(x)$, and up to a zero set of points x, it approaches the characteristic function of A. The Lebesgue Density Theorem expresses, in our opinion, the single most useful feature of Lebesgue measure for dynamics, and it is a pity that it gets so little attention in standard measure theory courses. We will have more to say about density points in Section 8.

Remark. If the holonomy map h happens to be differentiable then its Radon-Nikodym derivative is the Jacobian determinant,

$$\mathrm{RN}_x(h) = \det(Dh)_x.$$

The derivative $(Dh)_x$ is an $(n-s)\times(n-s)$ matrix, so the existence of $\mathrm{RN}_x(h)$ without the existence of $(Dh)_x$ is rather like the grin of the Cheshire Cat. The determinant persists although the matrix does not.

5. Partial Hyperbolicity

In the late 60s and early 70s, together with Moe Hirsch in [HiPuSh1] and [HiPugSh2], we studied partially hyperbolic systems in the guise of normally hyperbolic foliations. Around the same time, Brin and Pesin [BriPe] formalized the idea of partially hyperbolic dynamics, and proved a version of our Theorem E, below. The whole idea is to relax hyperbolicity by permitting a center direction in addition to the stable and unstable directions. The simplest example is the time-t map of a hyperbolic flow. In that case the center direction is the tangent bundle to the orbits. Here is the formal definition.

Definition. A diffeomorphism $f : M \to M$ is **partially hyperbolic** if there is a continuous Tf-invariant splitting $TM = E^u \oplus E^c \oplus E^s$ such that Tf is hyperbolic on $E^u \oplus E^s$ and the hyperbolicity dominates Tf on E^c in the sense that for some τ, λ with $1 \le \tau < \lambda$ and positive constants c, C we have

(a) For all $v \in E^u$ and all $n \ge 0$, $c\lambda^n|v| \le |Tf^n(v)|$.
(b) For all $v \in E^s$ and all $n \ge 0$, $|Tf^n(v)| \le C\lambda^{-n}|v|$.
(c) For all $v \in E^c$ and all $n \ge 0$, $c\tau^{-n}|v| \le |Tf^n(v)| \le C\tau^n|v|$.
(d) The bundles E^u, E^s are non-zero.

Condition (d) is present to avoid triviality. Without it, every diffeomorphism would be partially hyperbolic, for we could take E^c as TM. Sometimes, one only requires $E^u \oplus E^s \neq 0$, but for simplicity we use the stronger assumption (d) in this paper.

Partial hyperbolicity means that under Tf^n, vectors in E^c grow or shrink more gradually than do vectors in E^u and E^s. The center vectors behave in a relatively neutral fashion. The definition can be recast in several different ways. For instance, expansion of E^u under positive iteration of Tf can be replaced by contraction under negative iteration. Also, non-symmetric rates can be used for expansion and contraction. More significantly, one could permit pointwise domination instead of

the absolute domination as above. This would permit the growth rates λ and τ of vectors in T_pM to depend (continuously) on the point $p \in M$. Although pointwise partial hyperbolicity would require $\sup \tau_p/\lambda_p < 1$, it would permit $\sup \tau_p \geq \inf \lambda_p$, contrary to the definition above. See [BuWi].

For dynamical systems in which time is modeled on $\mathbb{R}$ or a Lie group, the definition is similar. Necessarily the tangent bundle to the orbits is contained in the center bundle, $E^o \subset E^c$. For simplicity we concentrate on the discrete time case, and we use the symmetric, absolute definition of partial hyperbolicity given above.

The general question we have pursued for the past several years is:

How frequently is partial hyperbolicity the main reason that a dynamical system is ergodic or stably ergodic?

Anosov had shown that hyperbolic volume preserving diffeomorphisms and non-suspension, hyperbolic flows on compact manifolds are ergodic, and indeed are stably ergodic. See Section 8 for a proof in the discrete time case.

With Matt Grayson, we began by looking at the simplest non-hyperbolic case – the time-one map of the geodesic flow for a surface of constant negative curvature. (There is no important difference between the time-one map and the the time-t map, $t \neq 0$.) It is a volume preserving diffeomorphism of a compact 3-manifold, $f = \phi_1 : T_1M \to T_1M$. Ergodicity, and even stable ergodicity, of the geodesic flow ϕ *as a flow* was well known. But the question concerned the diffeomorphism on its own, not as part of a flow. That is, are all C^2 small volume preserving diffeomorphism perturbations of ϕ_1 ergodic? In [GrPugSh] we showed that the answer is "yes," thereby producing the first dynamical system that is stably ergodic but not hyperbolic. In her Berkeley thesis [Wi1],[Wi2] Amie Wilkinson showed that the same is true for surfaces of variable negative curvature. We took her result to higher dimensions as:

Theorem C. *If M is a compact Riemann manifold with negative curvature and ϕ is its geodesic flow then $f = \phi_t$ is stably ergodic, $t \neq 0$; i.e., if f' is volume preserving and the C^2 distance between f and f' is sufficiently small then f' is ergodic.*

This result and many others are corollaries of a general stable ergodicity theorem presented in Section 7. The proof of Theorem C appears in Section 12.

6. Ergodicity

The way that ergodicity was originally formulated in statistical mechanics involves averages. Given an L^1 function $g : M \to \mathbb{R}$, its average over the manifold is

$$\frac{1}{m(M)} \int_M g \, dm.$$

It is the **space average** of g. Without loss of generality, we assume $m(M) = 1$. The average of g along the forward orbit of p under a diffeomorphism f is the limit, if it exists,

$$B^+(g, p, f) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} g \circ f^k(p).$$

In the case of a flow φ we have

$$B^+(g,p,\varphi) = \lim_{T\to\infty} \frac{1}{T}\int_0^T g\circ\varphi_t(p)\,dt.$$

Using reverse orbits gives averages B^-. The averages $B^\pm$ are **time averages**. It is a well known fact that (a), (b), (c) are equivalent for volume preserving dynamical systems on M, and it is clear that (c) implies (d).

(a) Ergodicity in the sense discussed in the introduction – only trivial measurable invariant sets.
(b) Only constant L^1 invariant functions.
(c) For each L^1 function g, the space average equals the time average for almost all orbits.
(d) For each continuous function g, the space average equals the time average for almost all orbits.

Remark. (c) is why ergodicity is good. Most orbits visit all parts of the phase space M in such a regular way that they sample the values of the function g fairly. It is said in statistical mechanics that the space average of a function may be sought, but hard to compute, while a trajectory that courses systematically through the phase space may be easier to find, and then calculating the average value of g along it will be simpler. Stable ergodicity is better yet. While ergodicity implies that for calculating the space average, the choice of sampling orbit is essentially irrelevant, stable ergodicity gives us latitude in choosing the dynamical system itself.

That (d) implies (c) is a consequence of the

Birkhoff Ergodic Theorem. *Let σ be a volume-reserving dynamical system on M. The maps $g\mapsto B^+g$, $g\mapsto B^-g$ are well defined continuous linear projections*

$$L^1(M)\to \mathrm{Inv}(\sigma)$$

where $\mathrm{Inv}(\sigma)$ *denotes the σ-invariant L^1 functions $M\to\mathbb{R}$. The two projections are equal.*

Proof that (d) $\Rightarrow$ (c). Let $\mathcal{C}$ denote the subset of continuous functions in L^1, and let K denote the constant functions. By (d), $B^\pm(\mathcal{C}) = K$. Since K is a one-dimensional subspace it is a closed set. Since $B^\pm$ is continuous and $\mathcal{C}$ is dense in L^1, $B^\pm(\mathcal{C})$ is dense in $B^\pm(L^1) = \mathrm{Inv}(\sigma)$ and hence $\mathrm{Inv}(\sigma) = K$, which is (c). □

In the proof of Theorem E we will use a simple fact about Birkhoff averages.

Lemma 6.1. *The forward Birkhoff average of a continuous function is constant on stable manifolds, and its reverse Birkhoff average is constant on unstable manifolds.*

Proof. Let f be the partially hyperbolic diffeomorphism and $g: M\to\mathbb{R}$ a continuous function. Let M^+ be the set of points at which B^+ is defined. If $p\in M^+$ and $q\in W^s(p)$, we claim that $B^+(q)$ exists and equals $B^+(p)$. Let $\epsilon>0$ be given. Since g is continuous there exists a large enough N so that if $k\geq N$ then

$|g(f^k(p)) - g(f^k(q))| < \epsilon/2$. Then, for $n \gg N$ we have

$$\begin{aligned}|B_n^+(p) - B_n^+(q)| &= \frac{1}{n+1}\Big|\sum_{k=0}^{n} g(f^k(p)) - g(f^k(q))\Big| \\ &\le \frac{1}{n+1}\Big(\sum_{k=0}^{N-1} + \sum_{k=N}^{n}\Big)|g(f^k(p)) - g(f^k(q))| \\ &\le \frac{2N}{n+1}\sup|g| + \Big(\frac{n-N+1}{n+1}\Big)\Big(\frac{\epsilon}{2}\Big) < \epsilon.\end{aligned}$$

Hence, for all $\epsilon > 0$ and all large n, $|B^+(p) - B_n^+(q)| \le \epsilon$ and hence $B^+(p) = B^+(q)$. This means that B^+ is constant on $W^s(p)$. Symmetrically, B^- is constant on $W^u(p)$ if $p \in M^-$ where M^- is the domain of definition of B^-. □

Remark. The same result holds for flows. Sums become integrals.

7. Sufficient Dynamical Conditions for Ergodicity

We have conjectured that ergodicity should be a consequence of partial hyperbolicity, an accessibility property, and a few technical conditions. A set $S \subset M$ is **u-saturated** if $p \in S$ implies $W^u(p) \subset S$; it is **s-saturated** if $p \in S$ implies $W^s(p) \subset S$, and it is **us-saturated** if it is saturated both ways.

Definition. An m-preserving diffeomorphism $f : M \to M$ belongs to the class $\mathcal{E}$ if

(a) f is partially hyperbolic.
(b) Each measurable us-saturated set is a zero set or its complement is a zero set. (This is the **e-accessibility** property.)
(c) The bundles E^u, $E^{cu} = E^u \oplus E^c$ E^c, $E^{cs} = E^c \oplus E^s$, and E^s are tangent to foliations, such that the E^u- and E^c-leaves subfoliate the E^{cu}-leaves, while the E^c- and E^s-leaves subfoliate the E^{cs}-leaves. (This is the **dynamical coherence** property.)
(d) The ratio $(\tau - 1)/(\lambda - 1)$ is sufficiently small. (This is the **center bunching** property.)

Theorem E. *Diffeomorphisms of class $\mathcal{E}$ are ergodic.*

Corollary. *Diffeomorphisms interior to $\mathcal{E}$ are stably ergodic.*

The corollary is of course a trivial consequence of the theorem. It says that to check stable ergodicity of f, it suffices that f has properties (a) - (d) above, and that they persist for perturbations. Properties (a) and (d), being essentially inequalities, always persist. The other two are more problematic. See Sections 11 and 12 for a further discussion of these hypotheses. They are valid for time-one maps of geodesic flows for manifolds of negative curvature and many other examples. Theorem E and its corollary imply Theorem C. See Section 12.

8. Absolute Continuity of the Invariant Foliations

Anosov's original proof that the stable and unstable foliations are absolutely continuous deals with the continuous differential forms that define them. Since these differential forms do not have exterior derivatives in the usual sense, this approach is necessarily difficult. In [PugSh1] we gave a proof of absolute continuity by somewhat different means. It bears a relation to [AnSi].

We say that a homeomorphism $h : X \to Y$ between topological measure spaces is **RN-regular** if its Radon-Nikodym derivative exists everywhere, is positive, and is continuous. As mentioned above, Theorem 10 of Anosov's thesis states that the holonomy maps are RN-regular. Our variation of the proof basically showed that the class of RN-regular maps is complete and that a naturally defined sequence of smooth maps h_n approximating the holonomy map h in the C^0 sense is Cauchy in the C^{RN} sense.

Key to the Cauchyness proof is that the dynamical system is better than class C^1. In fact, not only does the proof fail for non $C^{1+\epsilon}$ dynamical systems, the result is false. The stable and unstable foliations may fail to be absolutely continuous. This was first shown in an example of Bowen [Bow1], and later extended by Robinson and Young [RY]. See Section 21 for a discussion of the fascinating pathology that arises naturally and stably from some dynamical foliations that are not absolutely continuous.

One of the useful features of an absolutely continuous foliation $\mathcal{F}$ is a version of Fubini's theorem. See [PugSh1] for its natural proof.

Theorem 8.1. *If $Z \subset M$ is a zero set then the union of the leaves of $\mathcal{F}$ that meet Z in sets of positive leaf outer measure is a zero set. Conversely, if a set Z meets all leaves of $\mathcal{F}$ in leaf zero sets then it is a zero set.*

Leaf measure means a smooth s-dimensional volume on each leaf of $\mathcal{F}$.

Corollary 8.2. *The Birkhoff average of a continuous function is almost constant on almost every stable manifold, and almost constant on almost every unstable manifold.*

Proof. If $g : M \to \mathbb{R}$ is continuous then its Birkhoff averages $B^{\pm}(p)$ exist and are equal at all points of a set M_0 such that $Z = M \setminus M_0$ is a zero set. Let $B(p)$ be the common value of $B^{\pm}(p)$ for $p \in M_0$. The stable manifold foliation is absolutely continuous, so Theorem 8.1 implies that except for a zero set of stable manifolds, Z meets each W^s in a leaf zero set; i.e., M_0 meets almost every W^s in a set of full leaf measure. By Lemma 6.1, B^+ is constant on stable manifolds. Since $B = B^+$ on M_0, B is almost constant on almost every stable manifold. Symmetrically, the same is true for unstable manifolds. □

Note. The almost constant value of B on one stable manifold need not be the same as on another.

Here we see in the simplest case how to get ergodicity.

Theorem. *Hyperbolicity plus absolute continuity implies ergodicity.*

Proof. We assume $f : M \to M$ is hyperbolic and volume preserving. We claim f is ergodic, and so are its perturbations. As explained in Section 6, it is enough to check constancy of the Birkhoff averages of an arbitrary continuous function $g : M \to \mathbb{R}$. The Birkhoff average $B(p) = B(p, g, f)$ exists at all points of a set $M_0 \subset M$ of full measure. Theorem 8.1 implies that M_0 meets almost every stable and unstable manifold in sets of full leaf measure. Let Z_0 be the union of these invariant manifolds that fail to meet M_0 in sets of full leaf measure. It is a zero set, and every $p \in M_1 = M_0 \setminus Z_0$ has the property that $W^u(p)$ and $W^s(p)$ meet M_0 in sets of full leaf measure. Consider nearby points $p, q \in M_1$, and the local stable holonomy map $h : W^u_{\mathrm{loc}}(p) \to W^u_{\mathrm{loc}}(q)$. Since h is absolutely continuous it carries

the set $W^u_{\rm loc}(p) \cap M_0$ to a set of full leaf measure in $W^u_{\rm loc}(q)$. Sets of full measure intersect in sets of full measure, so $h(W^u_{\rm loc}(p)) \cap M_0$ intersects $W^u_{\rm loc}(q) \cap M_0$, say at the point y. This means that $h(x) = y$ for some $x \in W^u_{\rm loc}(p) \cap M_0$. That is, $y \in W^s(x)$. By Corollary 8.2, applied first in $W^u(p)$, then in $W^s(x)$, and last in $W^u(q)$, we have

$$B(p) = B(x) = B(y) = B(q).$$

For p, x, y, q all lie in M_0. Thus, B has the same (constant) value on $W^u_{\rm loc}(p) \cap M_0$ and $W^u_{\rm loc}(q) \cap M_0$, which shows that, except for a zero set, B is locally constant. Connectedness of M completes the proof. □

9. Juliennes

As discussed in Sections 4 and 8 the foliations $\mathcal{W}^u$ and $\mathcal{W}^s$ are RN-regular. From the measure theoretic point of view, this is as good as it gets, but to make full use of RN-regularity we need to understand the infinitesimal holonomy map h more geometrically. We need to understand how it affects density points. If h were differentiable this would be straightforward, but since h is only continuous we have a problem. Our solution is to redefine density points – instead of balls that shrink to a point we use sets that are more adapted to the dynamics. They are small, but highly eccentric in the sense that the ratio of their diameter to their inradius is large. (The inradius of a set is the radius of the largest ball it contains.) We refer to them as juliennes because they resemble slivered vegetables. See Figure 6.

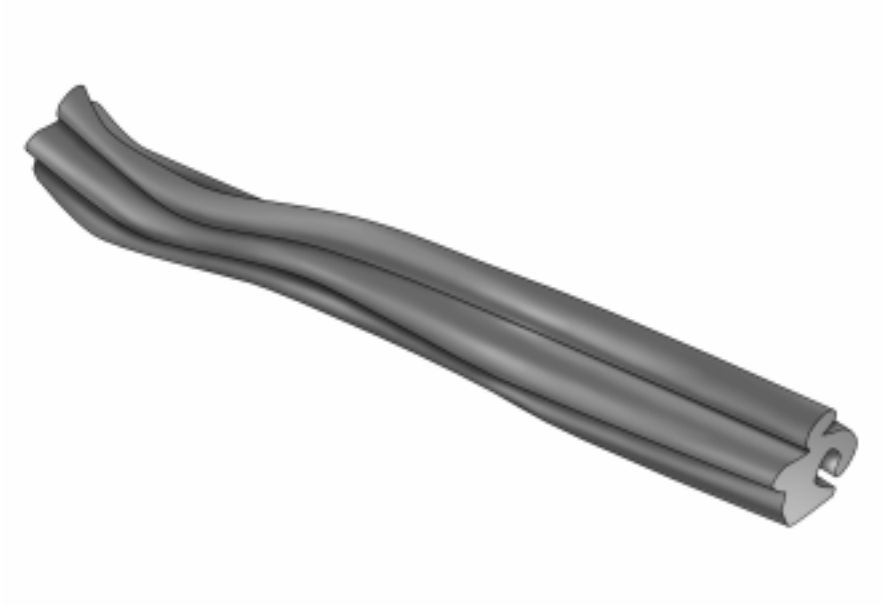

Figure 6. A three-dimensional center stable julienne with one-dimensional center.

If you have read Falconer's book about fractals [Fa], you will realize this is a risky business. For in [Nik], Nikodym constructs an example of a "paradoxical set"

N in the plane that has Lebesgue area zero, and yet for each $p \in N$ there is a sequence of rectangles R_n shrinking to p such that

$$\lim_{n\to\infty} \frac{m(N \cap R_n)}{m(R_n)} = 1.$$

m is Lebesgue area. The rectangles do not have axis-parallel edges, and their eccentricity tends to infinity. In terms of these rectangles, the zero set N acts like a set of full measure. See Figure 7.

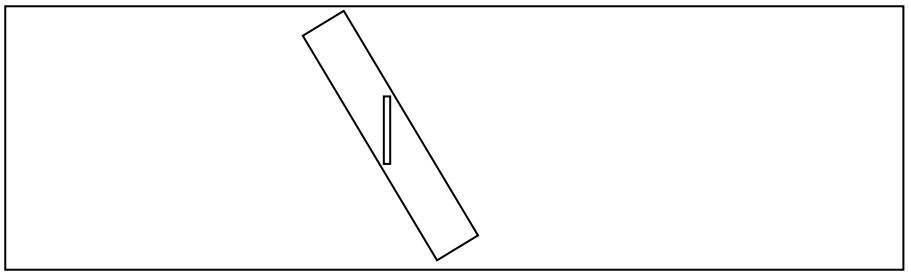

FIGURE 7. Nikodym rectangles shrinking to a point.

Our juliennes resemble Nikodym's rectangles in that the smaller they are, the thinner they are. Being non-linear, juliennes are potentially worse. The properties that distinguish juliennes from arbitrary eccentric rectangles are:

(a) The elongation axes of juliennes are Hölder controlled.
(b) The eccentricity of a julienne and its diameter are both exponential functions of a common number n, the number of dynamical iterations.
(c) The non-linearity, nesting, and shape-scaling properties of the juliennes are governed by a fixed, smooth dynamical system.

One of the facts we need about the juliennes is that they form a **density basis**. This means that they comprise a family $\mathcal{J} = \bigcup_{p\in M} \mathcal{J}(p)$ such that for each $p \in M$ there is a sequence of juliennes in $\mathcal{J}(p)$ that shrinks to p, and for every measurable set $A \subset M$

$$\lim_{J\downarrow p} \frac{m(A \cap J)}{m(J)} = \chi_A(p)$$

almost everywhere. (χ_A is the characteristic function of A, and $J \downarrow p$ means that $J \in \mathcal{J}(p)$ shrinks down to p.) The Lebesgue Density Theorem states that the family of balls in Euclidean space is a density basis. In particular, being a density basis rules out the Nikodym phenomenon.

Juliennes are constructed for a diffeomorphism of class $\mathcal{E}$ as follows. Using a fixed Riemann metric on M, write $W^u_{\text{loc}}(x)$ for the local unstable manifold of $x \in M$, and $W^u(x, r)$ for the disc of radius r in $W^u_{\text{loc}}(x)$. When r is small, this disc is quite round; i.e., its inradius approximates half its diameter. Define $W^c(x, r)$, $W^s(x, r)$ similarly. Fix an integer $n \geq 0$ and a point $p \in M$. Take a small disc D in the center manifold at p,

$$D = W^c(p, \sigma^n),$$

where $\sigma < 1$ depends on the hyperbolicity and center bunching constants. Let Y be the f^n-image of the disc $W^s(f^{-n}p, \tau^n)$ where $\tau < \sigma$ also depends on the hyperbolicity and bunching constants. When n is large, Y is much smaller than D,

but it may be quite eccentric. The **center stable julienne** is the local foliation product

$$J_n^{cs}(p) = \{z = W^c_{\rm loc}(q) \cap W^s_{\rm loc}(y) : q \in D \text{ and } y \in Y\}.$$

It is a neighborhood of p in $W^{cs}(p)$. A similar definition applies for the center unstable julienne, using in place of Y the set $X = f^{-n}W^u(f^n p, \tau^n)$. The **solid julienne** is the foliation product

$$J_n(p) = \{z \in W^u_{\rm loc}(x) \cap W^s_{\rm loc}(y) : x \in J^{cu}(p,n) \text{ and } y \in J^{cs}(p,n)\}.$$

See Figure 8.

FIGURE 8. The solid julienne is a local foliation product.

The julienne bases are

$$\mathcal{J}^{cs} = \bigcup_p \{J_n^{cs}(p) : n \in \mathbb{N}\} \quad \mathcal{J}^{cu} = \bigcup_p \{J_n^{cu}(p) : n \in \mathbb{N}\} \quad \mathcal{J} = \bigcup_p \{J_n(p) : n \in \mathbb{N}\}.$$

It is clear that as $n \to \infty$, $J_n^{cu}(p)$, $J_n^{cs}(p)$, and $J_n(p)$ shrink down to p. In [PugSh3] we show that the families of these juliennes are density bases. The center unstable juliennes are a density basis on $W^{cu}(p)$ with respect to the smooth leaf measure m_p^{cu} on $W^{cu}(p)$, the center stable juliennes are a density basis on $W^{cs}(p)$ with respect to the smooth leaf measure m_p^{cs} on $W^{cs}(p)$, and the solid juliennes are a density basis with respect to the smooth measure m on M. The proof is based on two properties, which we state for the solid juliennes.

(a) Scaling: for any fixed $k \geq 0$,

$$\frac{m(J_n(p))}{m(J_{n+k}(p))}$$

is uniformly bounded as $n \to \infty$.

(b) Engulfing: there is a uniform L such that

$$J_{n+L}(p) \cap J_{n+L}(q) \neq \emptyset \quad \Rightarrow \quad (J_{n+L}(p) \cup J_{n+L}(q)) \subset J_n(p).$$

These are properties possessed by the family of round balls in Euclidean space and they underlie the proof of the Lebesgue Density Theorem.

Denote the solid julienne density points of a measurable set $A \subset M$ as

$$D_J(A) = \{p \in M : \lim_{n\to\infty} \frac{m(A \cap J_n(p))}{m(J_n(p))} = 1\}.$$

The fact that the juliennes form a density basis implies that $A = D_J(A)$ modulo a zero set. Although $D_J(A)$ and the set of Lebesgue density points differ by a zero set, it is a crucial zero set.

The final property of juliennes we use is **julienne quasi-conformality**. It describes how juliennes behave under holonomy, and is what we mean when we say that juliennes are adapted to the dynamics. There is a uniform $k \geq 0$ such that if $p, q \in M$ are connected by an arc on an unstable manifold that has length ≤ 1 then the unstable holonomy map $h : W^{cs}_{\text{loc}}(p) \to W^{cs}_{\text{loc}}(q)$ satisfies

$$J^{cs}_{n+k}(q) \subset h(J^{cs}_n(p)) \subset J^{cs}_{n-k}(q).$$

As $n \to \infty$ the juliennes have progressively more elongated shapes but they nest in a way similar to balls. Furthermore, h does not disrupt this nesting much. It is as though, judged in a world where juliennes, not balls, are the norm, the holonomy map preserves shape.

An intuitive description of differentiability of a map $h : X \to Y$ is often given in visual terms. Magnifying glasses of increasing power are placed at $p \in X$ and $q = h(p) \in Y$, and thereby h is viewed more and more locally. Differentiability means that shapes of small sets near p are affected nearly linearly under increasing magnification. Quasi-conformality means that the shapes are not grossly distorted. Julienne quasi-conformality requires new magnifying glasses. They should magnify eccentrically elongated neighborhoods of p and q, the eccentricity increasing with the magnification, and then shapes of small sets near p should not appear grossly distorted. With the right choice of such magnifying glasses, the holonomy maps do not grossly distort shape.

Some manipulation with the foregoing julienne properties shows that h preserves center stable julienne density points. That is, if p is a density point of a set $W \subset W^{cs}(p)$ with respect to $\mathcal{J}^{cs}(p)$ and $q = h(p)$ then q is a density point of $h(W)$ with respect to $\mathcal{J}^{cs}(q)$. Further manipulation lets us pass from leaf juliennes to solid juliennes and obtain the following theorem. Corresponding to saturation as explained in Section 7 we say that $A \subset M$ is **almost u-saturated** if, except for a zero set, it consists of almost whole unstable manifolds. That is, $A = A_0 \cup Z$ such that Z is a zero set and if $p \in A_0$ then $W^u(p) \setminus A$ has leaf measure zero. The same applies to stable leaves. We say that A is **almost us-saturated** if it is almost u-saturated and almost s-saturated.

Theorem J. [PugSh3] *If $A \subset M$ is measurable and almost u-saturated then its set of solid julienne density points is u-saturated. The same is true for s-saturation.*

Corollary. *If $A \subset M$ is measurable and almost us-saturated then its set of solid julienne density points is us-saturated and differs from A by a zero set.*

Proof. Let A_0 be a set that equals A modulo a zero set Z and consists of almost whole unstable manifolds. Because the juliennes are a density basis,

(a) $D_J(A) = D_J(A_0)$.
(b) $A = D_J(A_0)$ modulo a zero set.

Note that (a) is an exact equality, not an equality modulo a zero set, because $D_J(Z) = \emptyset$. Theorem J asserts that $D_J(A_0)$ consists of whole unstable manifolds, so (a) implies that $D_J(A)$ consists of whole unstable manifolds. Symmetrically, $D_J(A)$ consists of whole stable manifolds, so it is us-saturated. (a) and (b) imply that $D_J(A)$ equals A modulo a zero set. □

10. The Proof of Theorem E

Let $f \in \mathcal{E}$ be given. We must prove it is ergodic. By the Birkhoff Ergodic Theorem and its consequences discussed in Section 6, it suffices to check that the Birkhoff average of an arbitrary continuous function $g : M \to \mathbb{R}$ is constant as an element of L^1. That is, B is essentially constant, where $B(p) = B(g, p, f)$ is the common value of $B^{\pm}(g, p, f)$ at points in the set M_0 where both averages exist and are equal. Fix any $c \in \mathbb{R}$ and consider the set

$$A = B^{-1}(0, c) = \{p \in M_0 : B(p) < c\}.$$

A is measurable because B is measurable. By Corollary 8.2, B is almost constant on almost every stable manifold and almost every unstable manifold. Therefore A is almost us-saturated. By the corollary to Theorem J in Section 9, the set $D_J(A)$ of julienne density points of A equals A modulo a zero set and is us-saturated. e-accessibility implies that $D_J(A)$ or its complement is a zero set. Therefore A or its complement is a zero set. This is true for each c, and shows that B is constant. For non-constancy of B would imply that for some c, both $B^{-1}(-\infty, c)$ and its complement $B^{-1}[c, \infty)$ have positive measure.

11. Center Bunching and Dynamical Coherence

The center bunching condition quantifies how neutral the center is. We regard center bunching as largely technical but have been unable to get around it. Writing $Tf = T^u f \oplus T^c f \oplus T^s f$, the center growth rate is

$$\tau = \limsup_{|n|\to\infty} \|T^c f^n\|^{1/|n|},$$

and the overall growth rate is

$$\kappa = \limsup_{|n|\to\infty} \|T f^n\|^{1/|n|}.$$

The unstable and stable growth rates are

$$\lambda_u = \limsup_{n\to-\infty} \|T^u f^n\|^{1/n} \qquad \lambda_s = \limsup_{n\to\infty} \|T^s f^n\|^{1/n},$$

and the hyperbolic growth rate is $\lambda = \max\{\lambda_u, 1/\lambda_s\}$. It is automatic that $\lambda \le \kappa$. Partial hyperbolicity requires $1 \le \tau < \lambda$. As shown in [An] and [HiPugSh2], the number

$$\theta_0 = \frac{\log \lambda - \log \tau}{\log \kappa}$$

has the property that for all $\theta < \theta_0$, the partially hyperbolic splitting is θ-Hölder. In [PugSh3], we defined center bunching as the requirement that

$$\tau^{2+2/\theta_0} < \lambda.$$

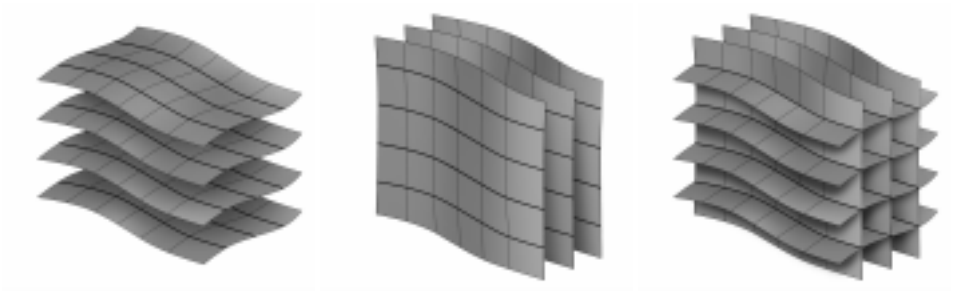

FIGURE 9. Dynamical Coherence.

Clearly this means that $(\tau - 1)/(\lambda - 1)$ is small. In [BuWi2], Burns and Wilkinson have shown that a weaker form of center bunching suffices for Theorem E, namely

$$\tau^{1/\theta_0} < \lambda,$$

which is equivalent to $S + 1 > W$, where

$$S = \frac{\log \lambda}{\log \tau} \qquad W = \frac{\log \kappa}{\log \lambda}.$$

S is the logarithmic separation between the hyperbolic and center bands of the spectrum of Tf, while W is the logarithmic width of the hyperbolic bands.

We adopt the Burns-Wilkinson definition of center bunching. Since it is an inequality, center bunching is an open condition. It is stable under perturbation. Burns and Wilkinson point out two interesting facts:

(a) If $\theta_0 > 1/2$ then center bunching is automatic.
(b) If the center bunching condition is violated, the center bundle need not be integrable. See [Wi1].

(b) means that center bunching affects dynamical coherence.

Dynamical coherence requires that the various foliations exist and fit together correctly. See Figure 9. By [HiPugSh2],

(a) The invariant foliations $\mathcal{W}^u$ and $\mathcal{W}^s$ always exist.
(b) Existence of the center foliation $\mathcal{W}^c$ implies existence of a center unstable and a center stable manifold, W^{cu} and W^{cs}, through each center manifold. The families of these center unstable and center stable manifolds are invariant under f, but do not a priori foliate M.
(c) W^{cu} and W^{cs} are sub-foliated by $\mathcal{W}^u$ and $\mathcal{W}^s$ respectively.
(d) If the center bundle is uniquely integrable then the families of center unstable and center stable manifolds do foliate M, and we have dynamical coherence.

Integrability, let alone unique integrability, of E^c is an open question. It is not hard to prove the existence of a semi-invariant family of "center plaques," but in general, as stated above, Wilkinson showed that without center bunching, they may not cohere to make a foliation. It is unknown whether center bunching implies the existence of a center foliation, and even if a center foliation does exist, it is not known to persist under dynamical perturbation without some extra hypothesis such as "plaque expansiveness," see [HiPugSh2].

Brin [Bri3] proves dynamical coherence from an assumption of quasi-isometricness of the stable and unstable foliations. Moreover, Brin, Burago and Ivanov [BriBuIv] establish dynamical coherence for all partially hyperbolic diffeomorphisms of the 3-torus.

We suspect that dynamical coherence is always stable under C^1 perturbations. Apart from the cases mentioned above the main tool we have in this direction is:

Proposition 11.1. [PugSh2, Proposition 2.3] *If the center foliation $\mathcal{W}^c$ exists and is of class C^1, then f is dynamically coherent, as is any f' close enough to f in the C^1 topology.*

For hyperbolic flows, or any of the affine diffeomorphism examples we consider below, or the direct products in Theorem A, the center foliations are C^1 so stable dynamical coherence is not an issue.

12. Accessibility and the Proof of Theorem C

Accessibility is a concept in control theory. Given smooth vector fields $X_1, \ldots, X_k$ and a point $p \in M$, one says that q is **accessible** from p if there is a path from p to q that consists of finitely many arcs, each of which is an X_i-trajectory arc, $1 \leq i \leq k$. The points p might represent the configurations of a multi-jointed robot arm, and the vector fields might generate motions of the arm due to flexing the joints. It is natural to ask what the set of accessible points looks like. Lie brackets of the vector fields play a large role in finding the answer.

We have adapted the accessibility notion to our situation. Points of a us-saturated set can be joined by a **us-path**, i.e., a path consisting of finitely many smooth arcs alternately in stable and unstable manifolds. See Figure 10. A key

Figure 10. A us-path alternates between unstable and stable arcs. It is potentially like a corkscrew.

difference between this and control theory accessibility is that lack of smoothness of E^u and E^s precludes the use of Lie brackets.

There are three types of accessibility to consider:

(a) e-accessibility as defined above – us-saturated sets are zero sets or complements of zero sets.
(b) us-accessibility – M itself is us-saturated.
(c) h-accessibility – us-accessibility plus homotopy nontriviality. See below.

h-accessibility requires that for each p and q in M there is a us-path γ from p to q that is **homotopically nontrivial** in the following sense. There should exist a continuous n-parameter family $\Gamma_\mu : [0,1] \to M$ of us-paths from p to points in a ball neighborhood U of q, where n is the dimension of M, such that

(i) $\Gamma_0 = \gamma$.
(ii) For some $(n-1)$-sphere S around $\mu = 0$ in μ-space, the map $\mu \mapsto \Gamma_\mu(1)$ sends S into $U \setminus q$ and has non-zero degree on the $(n-1)^{\text{st}}$ homology groups.

It is clear that:

$$\text{h-accessibility} \Rightarrow \text{us-accessibility} \Rightarrow \text{e-accessibility}.$$

Since the unstable and stable bundles depend continuously on f, h-accessibility is stable under perturbation. It is unknown whether this is true for us-accessibility, although in [PugSh3, Section 11] there is a non-dynamical example indicating that it may not be. Brin [Bri1] has dynamical examples in which e-accessibility is not stable.

As a consequence we have

Proof of Theorem C. Given a Riemann structure on M with negative curvature, we must show that the time-one map f of its geodesic flow is stably ergodic. Before perturbation, f is partially hyperbolic, dynamically coherent, and center bunched because these are properties of the geodesic flow. The center constant of f is $\tau = 1$.

Partial hyperbolicity and center bunching are stable properties.

The center foliation is smooth because it is the orbit foliation of the geodesic flow. By Proposition 11.1, f is stably dynamically coherent.

From Katok and Konnenko [KaKo] the bundle $E^u \oplus E^s$ is differentiable and given by a contact form. The everywhere non-integrability of this bundle implies that we can generate the flow direction by us-paths and verify h-accessibility. Therefore, f is stably e-accessible. The upshot is that f lies in the interior of $\mathcal{E}$ and so by Theorem E, it is stably ergodic. □

The previous proof amounts to checkable conditions for stable ergodicity, namely

Theorem 12.1. *If $f \in \mathcal{E}$, E^c is C^1, and we have h-accessibility then f is stably ergodic.*

An added condition that makes us-accessibility equivalent to h-accessibility is smoothness.

Theorem 12.2. [PugSh2] *If E^u and E^s are C^1 then us-accessibility implies h-accessibility.*

Remark. One may view this as a middle stage in the evolution of stably ergodic diffeomorphisms, at least from the point of view of accessibility considerations. The initial stage is due to Anosov who used the fact that all hyperbolic volume-preserving diffeomorphisms have us-accessibility. Persistence under perturbations is automatic in the Anosov case. Our results use h-accessibility to get stability, while in the latest stage Rodriguez Hertz uses only e-accessibility, [RH]. See Sections 15 and 16. He has found the first example of a stably e-accessible diffeomorphism that is not h-accessible.

Note. In our other papers on stable ergodicity we called e-accessibility essential accessibility. The "e" could also indicate "ergodic accessibility." It is an unfortunate

coincidence that in topology, "essential" refers to the property of not being null homotopic used in the definition of h-accessibility. At times we referred to h-accessibility as engulfing accessibility.

13. Proof of Theorem A

Let $f : M \to M$ and $g : N \to N$ be volume preserving diffeomorphisms. We have in mind the case that f is hyperbolic and g is KAM. The product diffeomorphism $f \times g : M \times N \to M \times N$ represents the dynamics of the uncoupled system. The dynamics on M and N evolve independently. If either of the diffeomorphisms f or g is not ergodic then the diffeomorphism $f \times g$ is also not ergodic. If $h : M \times N \to M \times N$ is an approximation of $f \times g$ then h is a **weak coupling** of f and g.

Suppose that f is hyperbolic with splitting $TM = E^u \oplus E^s$. Then $f \times g$ has an invariant splitting $T(M \times N) = E^u \oplus TN \oplus E^s$. If the hyperbolicity of f is sufficiently large then $f \times g$ is partially hyperbolic and center bunched.

The following result is what we meant in Theorem A by saying that a weak coupling of $f \times g$ is often stably ergodic.

Theorem 13.1 (**Theorem A).** [ShWi1], [BuPugShWi] *Let f and g be volume preserving diffeomorphisms of M and N as above. If $f \times g$ is partially hyperbolic and center bunched then it can be arbitrarily well approximated by a volume preserving stably ergodic diffeomorphism h.*

Proof. Since the center foliation are the fibers $p \times N$, $p \in M$, the center foliation is smooth and hence all diffeomorphisms in a neighborhood of $f \times g$ are dynamically coherent. It remains to find a diffeomorphism h in the neighborhood with h-accessibility. But this is accomplished using the Brin quadrilateral construction as in [ShWi1],[BuPugShWi]. □

Remark. Let f and g be volume preserving diffeomorphisms of M and N, with f hyperbolic.

(a) There is an $n \geq 1$ such that $f^n \times g$ is arbitrarily well approximated by stably ergodic diffeomorphisms. For the hyperbolicity of f^n tends to ∞ as $n \to \infty$.
(b) Revisiting the proofs in [ShWi1] and [BuPugShWi], it can be shown that the theorem holds whenever f is the time-t map of a geodesic flow for a manifold of constant negative curvature, and $|t|$ is large.
(c) From our Main Conjecture it would follow that not only is $f \times g$ approximable by stably ergodic diffeomorphisms, but that stable ergodicity is locally generic.
(d) Volume preserving may be replaced by symplectic in the theorem.

The situation for product diffeomorphisms is even better when f and g are both partially hyperbolic. The proof of the following theorem is straightforward.

Theorem 13.2. *Let f and g be volume preserving diffeomorphisms of M and N as above. Suppose that f and g are partially hyperbolic, stably dynamically coherent, h-accessible, and are center bunched with the same bunching conditions. Then $f \times g$ is stably ergodic.*

Corollary 13.3. *The Cartesian product of time-t maps of geodesic flows for manifolds of negative curvature is stably ergodic.*

Proof. Let ϕ_i be the time-t_i map of the geodesic flow for M_i where M_i has negative curvature, $t_i \neq 0$, and $i = 1, \ldots, k$. The center constants are all $\tau_i = 1$, so the maps ϕ_i have the same center bunching conditions. As in the proof of Theorem C in Section 12, each ϕ_i is stably dynamically coherent and has the h-accessibility property. So the same is true for f. By Theorem 13.2, $f = \phi_1 \times \cdots \times \phi_k$ is stably ergodic. □

14. The Algebraic Description of the Geodesic Flow

In the next section, we will define a class of affine diffeomorphisms of homogeneous spaces to which our theory applies, but first we discuss the special case of the time-t map of the diagonal flow on $\mathrm{SL}(2,\mathbb{R})$ and quotients, namely geodesic flows for compact surfaces of constant negative curvature.

Real 2×2 matrices $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $ad - bc = 1$ comprise the special linear Lie group $G = \mathrm{SL}(2,\mathbb{R})$. Positive diagonal matrices have $a, d > 0$ and $b = c = 0$; they form a subgroup D, with a 1-parameter representation

$$D_t = \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix}.$$

The **diagonal flow** on G is left multiplication by D_t, its orbits being $D_t A$. Fix $t > 0$ and let f be the time-t map of the flow. It is left multiplication by $g = D_t$,

$$f = L_g : G \to G.$$

We claim that the affine diffeomorphism f is partially hyperbolic with respect to any right invariant Riemann structure. (To get such a Riemann structure, choose any inner product on $T_{\mathrm{id}}G$ and define the inner product on the general tangent space T_xG so that $TR_x : T_{\mathrm{id}}G \to T_xG$ is an isometry, where R_x is right multiplication by x.) Since f is linear, the effect of Tf on $T_{\mathrm{id}}G$ is then isometric to the adjoint

$$\mathrm{Ad}(T_{\mathrm{id}}f) = T_gR_{g^{-1}} \circ T_{\mathrm{id}}L_g : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a & e^t b \\ e^{-t}c & d \end{bmatrix}.$$

The three relevant subgroups of G are U, D, L, where U and L consist of upper and lower triangular matrices

$$\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}.$$

The left cosets of U, D, and L foliate G, and the map f preserves the foliations. According to the preceding adjoint formula, the effect of Tf on a vector $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ tangent to U at the identity is isometric to expansion by e^t. By right invariance of the Riemann structure the same is true at all $x \in G$. Thus f expands along the U-coset foliation, and symmetrically it contracts along the L-coset foliation, while it sends each D-coset isometrically onto itself.

This verifies partial hyperbolicity of f; the left U-cosets form the unstable foliation, the left D-cosets form the center foliation, and the left L-cosets form the stable foliation. Clearly f is center bunched. Its center constant is $\tau = 1$. It is dynamically coherent because, as sets of matrix products, $UD = DU$ and $LD = DL$ are groups. It has the us-accessibility property because the smallest subgroup containing U and L is the whole group $\mathrm{SL}(2,\mathbb{R})$. Since its stable and unstable bundles are smooth, Theorem 12.2 gives h-accessibility.

There is a close and well known relation between the diagonal flow and the geodesic flow for a compact surface M of constant negative curvature – the geodesic flow is a quotient of the diagonal flow. This one sees as follows.

As in Section 4, rescale M so that its curvature is -1, and consider the universal covering space of M. It is the upper half plane $\mathbb{H} = \{x + iy = z \in \mathbb{C} : y > 0\}$, which we equip with the hyperbolic Riemann structure

$$\langle u, v\rangle_z = \frac{u \cdot v}{y^2}$$

where $u, v \in T_z\mathbb{H} \approx \mathbb{R}^2$ and $u \cdot v$ is the Euclidean dot product. The curvature of $\mathbb{H}$ is -1, the covering map $\pi : \mathbb{H} \to M$ is a local isometry, and M is just $\mathbb{H}$ modulo a uniform discrete subgroup Γ of isometries of $\mathbb{H}$. We claim that the geodesic flow for M is smoothly conjugate to a quotient of the the diagonal flow. This means that the time-t map of the geodesic flow can equally well be considered in a purely algebraic context.

The isometries $\mathbb{H} \to \mathbb{H}$ are Möbius transformations

$$\mu(z) = \frac{az + b}{cz + d}$$

where a, b, c, d are real numbers and $ad - bc = 1$. That is, they are elements of $\mathrm{SL}(2, \mathbb{R})$ which act on $\mathbb{H}$ according to the previous formula for μ. The map

$$h : \mathrm{SL}(2, \mathbb{R}) \to \mathrm{Isom}(\mathbb{H})$$
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \frac{az + b}{cz + d}$$

is a $2 : 1$ epimorphism whose kernel is $\{\mathrm{id}, -\mathrm{id}\}$.

Above, we discussed the left diagonal flow on $\mathrm{SL}(2, \mathbb{R})$, but there is also the right diagonal flow whose orbits are AD_t. Under inversion $\mathrm{inv} : \mathrm{SL}(2, \mathbb{R}) \to \mathrm{SL}(2, \mathbb{R})$, it is smoothly conjugate to the time-reversed left diagonal flow as

$$\mathrm{inv}(AD_t) = D_{-t}\,\mathrm{inv}(A).$$

We claim that the orbits of the right diagonal flow correspond to geodesics in $\mathbb{H}$, i.e., that $\mu(D_t(i))$ is a geodesic. If μ is the identity, this is obvious because

$$D_t(i) = \frac{e^{t/2}i + 0}{0i + e^{-t/2}} = e^t i$$

is the unit speed geodesic that is tangent to the upward vector $e_2 \in T_i\mathbb{H}$ at $t = 0$. Since μ acts on $\mathbb{H}$ as an isometry, the curve $\mu(D_t(i))$ is also a geodesic.

Define $k : \mathrm{SL}(2, \mathbb{R}) \to T_1\mathbb{H}$ by

$$k : A \mapsto T_i A(e_2).$$

(In this expression, $T_i A$ is the tangent at i to the Möbius transformation defined by A.) We claim that k is a $2 : 1$ semi-conjugacy from the right diagonal flow to the geodesic flow Φ on $T_1\mathbb{H}$. By the chain rule

$$\frac{d}{dt}\Big|_{t=0} AD_t(i) = T_i A(e_2) \in T_{A(i)}\mathbb{H},$$

which implies that k sends orbits of the right diagonal flow to orbits of Φ. As is shown in complex variables courses, Möbius transformations have the the following

transitivity property: given unit vectors $u \in T_z\mathbb{H}$, $u' \in T_{z'}\mathbb{H}$, there exists a unique μ such that

$$T_z\mu(u) = u'.$$

Thus k is onto. It is $2:1$ because the only ambiguity between Möbius transformations and matrices in $\mathrm{SL}(2,\mathbb{R})$ is that A and $-A$ correspond to the same μ.

Now we return to the geodesic flow ϕ on T_1M. Since $M = \mathbb{H}/\Gamma$ where Γ is a uniform discrete subgroup of isometries, ϕ is the quotient of the geodesic flow Φ on $T_1\mathbb{H}$, which is semi-conjugate to the left diagonal flow. Pulling Γ back to a subgroup $h^{-1}\Gamma \subset \mathrm{SL}(2,\mathbb{R})$, we observe that ϕ is conjugate to the left diagonal flow on $\mathrm{SL}(2,\mathbb{R})/h^{-1}\Gamma$. For $h^{-1}\Gamma$ automatically includes the normal subgroup $\{\mathrm{id}, -\mathrm{id}\}$, and this removes the $\pm A$ ambiguity.

Remark. In the definition of affine diffeomorphisms below, we prefer group multiplication on the left (it is consistent with function composition) and division by subgroups such as Γ on the right. Other left/right choices make little mathematical difference. However, one should note that the orbits of the left diagonal flow do not correspond to geodesics in the same fashion as do orbits of the right diagonal flow: $D_tA(i)$ is not a geodesic.

In summary, we have:

The geodesic flow for a surface of constant negative curvature is conjugate to the diagonal flow on a quotient of $\mathrm{SL}(2,\mathbb{R})$*, and its stable and unstable manifolds correspond to quotients of the cosets of the U and L subgroups. All facts about time-t maps of the diagonal flow project to facts about time-t maps of the geodesic flow.*

15. Affine Diffeomorphisms

We now place linear hyperbolic diffeomorphisms and hyperbolic geodesic flows in the unifying context of affine diffeomorphisms of homogeneous spaces.

Suppose that G is a connected Lie group, $A : G \to G$ is an automorphism, B is a closed subgroup of G with $A(B) = B$, $g \in G$ is given, and the **affine diffeomorphism**

$$f : G/B \to G/B$$

is defined as $f(xB) = gA(x)B$. It is covered by the diffeomorphism

$$\bar{f} = L_g \circ A : G \to G,$$

where $L_g : G \to G$ is left multiplication by g.

We have already seen some examples of affine diffeomorphisms. For the Thom diffeomorphism of Section 2, $G = \mathbb{R}^2$, $B = \mathbb{Z}^2$ the automorphism $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and $g = 0$. For the algebraic version of the geodesic flow, $G = \mathrm{SL}(2,\mathbb{R})$, $B = \Gamma$, and the affine diffeomorphism is given by left translation, $f = L_g$ where $g = \begin{bmatrix} e^{\frac{1}{2}} & 0 \\ 0 & e^{-\frac{1}{2}} \end{bmatrix}$. The automorphism A is the identity.

An affine diffeomorphism $\bar{f}$ induces an automorphism of the Lie algebra $\mathfrak{g} = T_eG$, $\mathfrak{a}(\bar{f}) = \mathrm{Ad}_g \circ T_eA$, where Ad_g is the adjoint action of g, and $\mathfrak{g}$ splits into generalized eigenspaces,

$$\mathfrak{g} = \mathfrak{g}^u \oplus \mathfrak{g}^c \oplus \mathfrak{g}^s,$$

such that the eigenvalues of $\mathfrak{a}(\bar{f})$ are respectively outside, on, or inside the unit circle. These eigenspaces and the direct sums $\mathfrak{g}^{cu} = \mathfrak{g}^u \oplus \mathfrak{g}^c$, $\mathfrak{g}^{cs} = \mathfrak{g}^c \oplus \mathfrak{g}^s$ are Lie subalgebras and hence tangent to connected subgroups G^u, G^c, G^s, G^{cu}, G^{cs}.

Theorem 15.1. [PugShSt] *Let $f : G/B \to G/B$ be an affine diffeomorphism as above such that G/B is compact and supports a smooth G-invariant volume. Let G^* be any of the groups $G^u, G^c, G^s, G^{cu}, G^{cs}$. Then the orbits of the left G^*-action on G/B foliate G/B. Moreover, f exponentially expands the G^u-leaves, exponentially contracts the G^s-leaves, and affects the G^c-leaves subexponentially.*

Remark. If G/B is compact and the subgroup B is discrete then G/B always supports a smooth G-invariant volume, so the hypothesis on the existence of such a volume is redundant. (Invariance refers to left multiplication by elements of G.) On the other hand, if B is not discrete the theorem is false without the volume hypothesis. Consider $G = \mathrm{GL}(n, \mathbb{R})$ and B the subgroup of upper triangular matrices. If $A \in \mathrm{GL}(n, \mathbb{R})$ has real eigenvalues whose moduli are distinct, left multiplication by A on G/B is a Morse-Smale diffeomorphism with $n!$ fixed points. The orbits of G^u are the unstable manifolds of these points and vary in dimension from 0 to $n(n-1)/2$. They do not foliate. See [ShVa], and see [St1] as a general reference for dynamical systems on homogeneous spaces as well as for many illuminating examples.

Now we characterize partial hyperbolicity, bunching, dynamical coherence, and accessibility in the context of affine diffeomorphisms. Let $\mathfrak{h}$ denote the smallest Lie subalgebra of $\mathfrak{g}$ containing $\mathfrak{g}^u \cup \mathfrak{g}^s$. It is not hard to see that $\mathfrak{h}$ is an ideal in $\mathfrak{g}$. We call it the **hyperbolic Lie subalgebra** of $\bar{f}$, and we denote by H the connected subgroup of G tangent to $\mathfrak{h}$, calling it the **hyperbolic subroup** of $\bar{f}$. Finally, let $\mathfrak{b}$ denote the Lie algebra of B, $\mathfrak{b} \subset \mathfrak{g}$.

Theorem 15.2. [PugSh3], [BreSh] *Let $f : G/B \to G/B$ be an affine diffeomorphism as above such that G/B is compact and supports a smooth G-invariant volume. Then*

(a) *f is partially hyperbolic if and only if the hyperbolic Lie subalgebra of $\bar{f}$ is not contained in the Lie algebra of B, $\mathfrak{h} \not\subset \mathfrak{b}$.*
(b) *If f is partially hyperbolic then it is center bunched and dynamically coherent.*
(c) *f has the us-accessibility property if and only if $\mathfrak{g} = \mathfrak{b} + \mathfrak{h}$.*
(d) *f has the e-accessibility property if and only if $\overline{HB} = G$.*

When the stable and unstable foliations are smooth, as in Theorems 15.1 and 15.2, then by Theorem 12.2, us-accessibility is stable. Thus we have:

Theorem 15.3. [PugSh3] *Let $f : G/B \to G/B$ be an affine diffeomorphism as above such that G/B is compact and supports a smooth G-invariant volume. Then f is stably ergodic among C^2 volume preserving diffeomorphisms of G/B if (merely) the hyperbolic Lie subalgebra $\mathfrak{h}$ is large enough that $\mathfrak{g} = \mathfrak{b} + \mathfrak{h}$.*

If G is simple then any nontrivial $\mathfrak{h}$ is large enough since it is an ideal. Thus,

Corollary 15.4. *Let $f : G/B \to G/B$ be as above, with G is simple. Then f is stably ergodic among C^2 volume preserving diffeomorphisms of G/B if (merely) the hyperbolic Lie subalgebra $\mathfrak{h}$ is nonzero, $\mathfrak{h} \neq 0$.*

This gives a generalization of the ergodicity of our geodesic flow example in all dimensions. Suppose that $A \in \mathrm{SL}(n,\mathbb{R})$ has some eigenvalues that are not of modulus one, and suppose that Γ is a uniform discrete A-invariant subgroup of $\mathrm{SL}(n,\mathbb{R})$. Set $M = \mathrm{SL}(n,\mathbb{R})/\Gamma$. Then left multiplication by A, $L_A : M \to M$, is stably ergodic in $\mathrm{Diff}^2_m(M)$. The case where n is large and all but two eigenvalues have modulus one is interesting, in that the dimension of G^u and G^s is $n-1$ while the dimension of G^c is $(n-1)^2$, so the dimension of G^c is much larger than that of G^u and G^s.

At the other extreme are abelian groups. If $G = \mathbb{R}^n$ and $B = \mathbb{Z}^n$ then translations on the torus, $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ may be ergodic if the entries of the element defining the translation are rationally independent, but they are never stably ergodic. For an automorphism A, however, the hyperbolic Lie subalgebra equals $\mathbb{R}^n$ if and only if A is hyperbolic.

It may be instructive here to recall that an automorphism A of $\mathbb{T}^n$ is ergodic if and only if A has no eigenvalues that are roots of unity, or equivalently if the orbits of all nonzero lattice points under A (or under its transpose A^*) are infinite. The classical proof of this fact is simple and illustrative of the different techniques used to study affine diffeomorphisms and nonlinear dynamics in general. Here it is.

Suppose that A has no eigenvalues which are roots of unity. Let $B \subset \mathbb{T}^n$ be an A-invariant set of positive measure, and let χ_B be its characteristic function. Then $\chi_B \circ A = \chi_B$. Writing Fourier series we have

$$\chi_B(x) = \sum a_z e^{<z,x>} = \sum a_z e^{<A^* z,x>} = \chi_B \circ A.$$

The sums are taken over the lattice points $z \in \mathbb{Z}^n$. Thus $a_z = a_{A^* z}$ and as the orbits are infinite, $a_z = 0$ for all $z \neq 0$. Thus $\chi_B(x)$ is almost everywhere constant, equal to 1, and B has measure 1.

A little bit of algebra quickly shows that the hypothesis that A has no eigenvalues which are roots of unity is equivalent to the hypothesis that $\overline{H\mathbb{Z}^n} = \mathbb{R}^n$ where H is the hyperbolically generated subgroup of $\mathbb{R}^n$. We sketch one direction. If $\overline{H\mathbb{Z}^n} \neq \mathbb{R}^n$ then H is an invariant proper rational subspace of $\mathbb{R}^n$ and the characteristic polynomial of A splits over the rationals and hence the integers into a product of two monic polynomials over the integers, one with all roots off the unit circle and one with all roots on the unit circle. But the roots of a monic polynomial over the integers with all roots on the unit circle are all roots of unity.

We have concentrated on the accessibility condition because accessibility is a topological property and as such it is not difficult to stipulate easily verifiable conditions which guarantee that it persists under small perturbations.

In a recent remarkable paper, Federico Rodriguez Hertz gives the first examples of a stably e-accessible diffeomorphisms that are not us-accessible, [RH]. They are ergodic, non-hyperbolic diffeomorphisms of tori. The first such occurs in dimension four, and was written down by Peter Walters [Wa] as

$$\begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -6 \\ 0 & 0 & 1 & 8 \end{bmatrix}.$$

Another example is due to Doug Lind [Li].

Rodriguez Hertz sometimes uses a technical assumption on the automorphism A, namely

($*$) The characteristic polynomial of A is irreducible over the integers and is not a polynomial in t^n for all $n > 1$.

This assumption is true for the examples of Walters and Lind.

Theorem 15.5. [RH] *Let A be an ergodic toral automorphism of $\mathbb{T}^n$.*

(a) *If $n \leq 5$ then A is stably ergodic in $\mathrm{Diff}^{22}_m(\mathbb{T}^n)$.*
(b) *If $n \geq 6$, E^c is two-dimensional, and A satisfies ($*$) then A is stably ergodic in $\mathrm{Diff}^5(\mathbb{T}^n)$.*

The differentiability degrees 22 and 5 are not misprints.

Part of Rodriguez Hertz' proof involves an alternative. Either the perturbation is us-accessible or the stable and unstable manifold foliations are differentiably conjugate to the foliations of the linear example and hence the perturbation has the e-accessibility property.

Problem. Is every ergodic toral automorphism stably ergodic in the C^r topology for some r?
In Section 16 we will discuss a more general version of this problem.

The next result shows that at least they often lie in the closure of the stably ergodic diffeomorphisms.

Theorem 15.6. [ShWi1] *Every ergodic toral automorphism of $\mathbb{T}^n$ that is an isometry on the center bundle E^c can be approximated arbitrarily well in $\mathrm{Diff}^\infty_m(\mathbb{T}^n)$ by a stably us-accessible, stably ergodic diffeomorphism.*

At this point it is natural to ask if every e-accessible diffeomorphism can be approximated by a stably us-accessible diffeomorphism. In fact we have conjectured in [PugSh2, Conjecture 4] and [PugSh3, Conjecture 2] that partially hyperbolic diffeomorphisms which are us-accessible are open and dense in the partially hyperbolic diffeomorphisms, volume preserving or not. See also Conjecture 18.4 in Section 18.

There are connections between ergodic automorphisms of the torus and Salem polynomials. These are integral, monic, reciprocal polynomials of even degree $d \geq 4$ having one positive real root λ outside the unit circle, one inside the unit circle, and the remaining roots on the circle. Reciprocal means that $a_{d-i} = a_i$ for all $0 \leq i \leq d/2$. If the characteristic polynomial of the automorphism A of $\mathbb{T}^d$ is a Salem polynomial then A is ergodic. Such automorphisms are at the opposite extreme from the class covered by Theorem 15.5 (except in dimension 4). David Boyd lists Salem polynomials with small λ in [Boy]. The corresponding automorphisms of the torus are poorly bunched. For degree four polynomials Boyd has told us how to get all Salem polynomials. Namely, let $P(x) = x^4 - ax^3 + bx^2 - ax + 1$ with a positive. Then P is Salem if and only if $-2a - 2 < b < 2a - 2$. A general reference for Salem numbers is [D-G,G-H,Be].

Further examples of partially hyperbolic stably ergodic diffeomorphisms are considered in [BuPugShWi]. These include skew products, frame flows, and Anosov-like diffeomorphisms. We discuss skew products below.

16. Decisiveness and the Proof of Theorem B

In this section we make use of the idea that stability with respect to a small class of perturbations is **decisive** for stability with respect to a larger class. For

example, structural stability of a linear map $\mathbb{R}^n \to \mathbb{R}^n$ where only diagonal linear perturbations are permitted is equivalent to structural stability where all C^1 small perturbations are permitted. The usage is made clear in what follows.

Theorem 16.1 (Theorem B). *Suppose that $f : G/B \to G/B$ is an affine diffeomorphism such that $M = G/B$ is compact and supports a smooth G-invariant volume. Assume that G is simple. Stable ergodicity of f with respect to perturbation by left translations is decisive for stable ergodicity of f in* $\mathrm{Diff}^2_m(M)$.

The proof is based on the theorem of Starkov which appears in the appendix to this paper.

Theorem 16.2. [St2] *With hypotheses as in Theorem 16.1, except that G need not be simple, the following are equivalent.*

(a) *f is stably ergodic under perturbation by left translations.*
(b) *$\overline{HB} = G$ where H is the hyperbolically generated subgroup of G.*

Proof of Theorem B. We may assume that $B \neq G$ and that f is stably ergodic with respect to perturbation by left translations. Theorem 16.2 implies that $\overline{HB} = G$ so H is non-trivial. As H is normal and G is simple $H = G$. Now by Theorem 15.3, f is stably ergodic in $\mathrm{Diff}^2_m(M)$. □

A second class where decisiveness occurs is skew products. Let G be a compact Lie group and let $f : N \to N$ be a hyperbolic diffeomorphism of a compact manifold N. Each smooth map $\varphi : N \to G$ defines a **skew product** diffeomorphism $f_\varphi : M \to M$ where $M = N \times G$ and

$$f_\varphi(p, g) = (f(p), \varphi(p)g).$$

Left translations are isometries of G in the bi-invariant metric, which implies that f_φ is partially hyperbolic and center bunched. If f preserves a smooth volume m_N, then f_φ preserves the smooth volume $m = m_N \times \nu$, where ν is normalized Haar measure on G. Skew products over f are also called G-extensions of f. The set of all C^2 skew products over f,

$$\mathrm{Skew}(f, G) = \{f_\varphi : \varphi \in C^2(N, G)\},$$

is a closed subset of $\mathrm{Diff}^2_m(M)$.

The ergodic properties of such skew products were studied by Brin [Bri2]. He proved that ergodic diffeomorphisms form an open and dense subset of $\mathrm{Skew}(f, G)$. Burns and Wilkinson have shown:

Theorem 16.3. [BuWi] *With f, G, N, M as above*

(a) *Stable ergodicity is open and dense in Skew(f, G).*
(b) *If $f : N \to N$ is infranil (see Section 20) then stable ergodicity of f_φ in* $\mathrm{Skew}(f, G)$ *is is decisive for stable ergodicity in* $\mathrm{Diff}^2_m(M)$.

The proof is based on a classification of those skew products that are not stably ergodic. Stable ergodicity of skew products has been studied in more generality, and there is now a considerable literature on the subject. Some of the references are [ParPo], [FiPa], [FiMeTo], [FiNi1], [FiNi2], and [Wal].

We end with a question from [BuPugShWi] of a very different nature. We have used both the strong unstable and strong stable foliations in our proof of ergodicity, but we don't know an example where this is strictly necessary.

Question. For a partially hyperbolic C^2 ergodic diffeomorphism f with the e-accessibility property, are the unstable and stable foliations already uniquely ergodic?

Unique ergodicity of $\mathcal{W}^u$ and $\mathcal{W}^s$ was proved by Bowen and Marcus [BowMar] in the case where f is the time-one map of a hyperbolic flow. Rodriguez Hertz' result adds more cases in which the invariant foliations are uniquely ergodic, namely those in which they are differentiably conjugate to the invariant foliations of a linear ergodic toral automorphism.

In the topological category Bonnatti, Díaz, and Ures [BonDíPujUr] prove the minimality of the stable and unstable foliations for an open and dense set of robustly transitive diffeomorphisms.

17. The Mautner Phenomenon

Theorem B and Rodriguez Hertz' results make the decisiveness situation very suggestive: *Is perturbation by left translations decisive for stable ergodicity of affine diffeomorphisms in* $\mathrm{Diff}^r_m(M)$ *where* $M = G/B$ *and* r *is large enough?*

The problem revolves around the question of what happens to the accessibility classes under perturbations when $\overline{HB} = G$. Starkov's proof that $\overline{HB} = G$ implies the stable ergodicity of the affine diffeomorphism f also shows that the group H can only get larger under perturbation. Thus the perturbed accessibility classes contain the old ones. Since the H-orbits are smooth it follows by the arguments of [PugSh2] that under C^2 perturbations, accessibility classes can't get smaller (at least in the homological sense) but they do move about. On the other hand they might get larger even to the extent of becoming the whole manifold. It is conceivable that the only way the accessibility classes do not get larger is if they stay differentiable. Then a kind of rigidity argument along the lines of KAM theory may show that e-accessibility persists. This is Rodriguez Hertz' argument in the cases he has so far accomplished.

Here is how a slight variant of our proof of Theorem E goes for the special case of an affine diffeomorphism f. (Smoothness of the invariant foliations obviates juliennes and absolute continuity.) Let $k \in L^1(M)$ be an f-invariant function, where $M = G/B$ as above. The set of group elements $g \in G$ that leave k invariant is closed. By Lemma 6.1 and the smoothness of the invariant foliations, k must be almost constant on unstable and stable manifolds; i.e., it is almost constant on the G^u and G^s orbits. Hence for *every* element $g \in G^u \cup G^s$, $k(gx) = k(x)$. (This is the analogue of our Theorem J.) Thus k is invariant for every $g \in H$, the hyperbolically generated subgroup. (This is the analogue of the corollary to Theorem J.) Now since k is defined on G/B, k is constant on HB orbits. As $\overline{HB} = G$, k is constant.

The proof we presented here is an instance of what is called the **Mautner phenomenon** in the ergodic theory of flows on homogeneous spaces or group representation theory. Our Theorem J, its corollary, and Theorem E may be thought of as a non-linear generalization of the Mautner phenomenon.

Gelfand and Fomin [GeFo] view the geodesic flow as a flow on a homogeneous space, and the Mautner phenomenon [Mau1],[Mau2] was initially applied in that context. It has taken on a much wider scope in the theory of flows on homogeneous spaces and representation theory. The hyperbolically generated group H was considered by Auslander and Green in [AuGr] and a generalization by Moore [Mo]. See

[St1] for a general discussion. There are elegant proofs of the Mautner phenomenon in the literature, different from the one outlined above, and sometimes more general; see for example the one by Parry in [Par].

18. The Main Conjecture

For $r \geq 1$, let us denote by $\mathrm{PH}^r(M)$ the partially hyperbolic C^r diffeomorphisms of M, and by $\mathrm{PH}^r_m(M)$ those that are volume preserving. We equip these spaces with the C^r topology. We denote by SE the set of stably ergodic diffeomorphisms. In Section 1 we stated our Main Conjecture that for $r \geq 2$, SE is an open and dense subset of $\mathrm{PH}^r_m(M)$. Besides the results discussed above, most of the progress on our Main Conjecture is limited to the case where the dimension of the center bundle E^c is one. Here are three results in that case.

Theorem 18.1. [Do2] *If the dimension of M is three then $\mathrm{SE} \cap \mathrm{PH}^2_m(M)$ is open and dense in $\mathrm{PH}^2_m(M)$.*

Theorem 18.2. [BuPugWi] *If the dimension of M is three, then for almost all hyperbolic flows on M (i.e., for those that are not suspensions) the time-t maps are stably ergodic for all $t \neq 0$.*

Theorem 18.3. [BonMaViWi] *Among partially hyperbolic C^2 volume preserving diffeomorphisms with one-dimensional center bundle, the stably ergodic ones are C^1 dense.*

Remark. Theorem 18.1 essentially verifies our Main Conjecture in dimension three. Theorem 18.2 is not a direct corollary of Theorem 18.1 since time-t maps of hyperbolic flows do not form an open subset of $\mathrm{Diff}^2_m(M)$. Theorem 18.3 has mixed topologies: each $f \in \mathrm{PH}^2_m$ with one-dimensional center can be C^1 approximated by an $f' \in \mathrm{SE} \cap \mathrm{PH}^2_m$.

In general, the Main Conjecture can be split into two parts. The first part concerns the prevalence of stable accessibility.

Conjecture 18.4. [PugSh2, Conjecture 4] and [PugSh3, Conjecture 2] For $r \geq 1$, e-accessibility holds for open and dense subsets of PH^r and PH^r_m.

In this direction there are three results.

Theorem 18.5. [NiTö] *Conjecture 18.4 is valid for all $r \geq 1$ when the center is one-dimensional and there exist two nearby compact center leaves.*

Theorem 18.6. [Do2] *Conjecture 18.4 is valid when the manifold has dimension three and $r = 2$.*

Theorem 18.7. [DoWi] *For all $r \geq 1$, among partially hyperbolic C^r diffeomorphisms with one-dimensional center bundle, the stably us-accessible (and therefore stably e-accessible) ones are C^1 dense.*

The second part of the Main Conjecture concerns the bunching and dynamical coherence hypotheses in Theorem E.

Conjecture 18.8. [PugSh3, Conjecture 3] A partially hyperbolic C^2 volume preserving diffeomorphism with the e-accessibility property is ergodic.

Not much progress has been made on Conjecture 18.8 except that Burns and Wilkinson [BuWi2] have improved the center bunching conditions.

19. Dominated Splitting

We have seen above that partial hyperbolicity is necessary for stable ergodicity in the context of affine diffeomorphisms. As Tahzibi shows in [Ta], this is not generally the case. Diffeomorphisms of tori introduced by Bonatti and Viana in [BonVi] are stably ergodic but not partially hyperbolic. They do however satisfy a similar but more general notion introduced by Mañé [Mañ1], [Mañ2], [Mañ3] and independently by Liao [Liao] and Pliss [Pli] of having a dominated splitting.

A continuous Tf-invariant splitting $E \oplus F$ is **dominated** if there is a norm on TM such that point by point, Tf affects all vectors in E_p more expansively than any vectors in F_p. Domination requires neither $\|Tf|F_p\| < 1$ nor $\|(Tf|E_p)^{-1}\| < 1$. Rather, for all unit vectors $u \in E_p$, $v \in F_p$, one should have

$$|Tf(u)| > |Tf(v)|.$$

For a partially hyperbolic diffeomorphism f, there are two dominated splittings: $E \oplus F = (E^u \oplus E^c) \oplus E^s$ and $E \oplus F = E^u \oplus (E^c \oplus E^s)$.

There is a considerable literature on dominated splitting and its relationship to stable or robust transitivity. In fact the literature has gotten so large that we do not attempt to survey it here. See instead the articles by Díaz [Dí], Pujals [Puj], and the references therein. It would be interesting to extend more of the results on partial hyperbolicity and stable ergodicity in this direction. From the literature and the examples currently known it is natural to speculate as Tahzibi does that a dominated splitting is necessary for stable ergodicity.

In fact there is the following recent result of Bochi, Fayad, and Pujals. Let $\mathrm{Diff}^{1+}_m(M)$ be the set of m-preserving diffeomorphisms whose derivative is Hölder continuous, equipped with the C^1 topology, and let SE be the subset of stably ergodic ones. (As usual, m is a smooth volume on the compact manifold M.)

Theorem 19.1. [BoFaPu] *There is an open and dense subset of* SE *each of whose elements f is non-uniformly hyperbolic, admits dominated splitting, and is Bernoulli.*

A similar result is proved concerning dominated splittings and robust transitivity in [ArMa].

20. Existence

Which manifolds support partially hyperbolic diffeomorphisms? Already the question as to which manifolds support hyperbolic diffeomorphisms or flows is not well understood, so the question will be difficult.[†] There may be some hope in understanding the problem in low dimensions. In dimension three, Pujals has suggested that all manifolds that support partially hyperbolic diffeomorphisms either support hyperbolic flows or are circle bundles over the torus. Bonatti and Wilkinson [BonWi] take a step in this direction, and Brin, Burago and Ivanov [BriBuIv] prove that there are no partially hyperbolic diffeomorphisms on S^3, or on any other compact 3-manifold with a finite fundamental group.

Here are two off-the-top-of-the-head questions. Let $f : M \to M$ be a partially hyperbolic diffeomorphism.

[†] From the definition of partial hyperbolicity the tangent bundle must admit a non-trivial direct sum decomposition and this is already a restriction. But the class seems much more restricted.

(1) We have made examples of partially hyperbolic diffeomorphisms on homogeneous spaces, fiber bundles, and direct products. Does there exist a manifold N of dimension less the dimension of M, other than a point, and a locally trivial fibration $\pi : M \to N$?
(2) Do the strong stable and unstable manifolds represent non-trivial homology classes in the homology of M? (See Ruelle-Sullivan [RueSu] for the Anosov case.)

Relevant to these questions is a beautiful conjecture dating back to the 60s about the classification of hyperbolic (i.e., Anosov) diffeomorphisms. Let N be a simply connected nilpotent Lie group, Γ a uniform discrete subgroup of N, and A an automorphism of N that preserves Γ. Then $A : N/\Gamma \to N/\Gamma$ is an affine diffeomorphism. If the tangent $T_eA : T_eN \to T_eN$ is a hyperbolic linear map then $A : N/\Gamma \to N/\Gamma$ is hyperbolic. The manifold N/Γ is called a nilmanifold and the affine diffeomorphism is a nilmanifold Anosov diffeomorphism. Examples of these diffeomorphisms were constructed by Smale [Sm], and he suggested that they may comprise all the Anosov diffeomorphisms up to topological conjugacy. Shub [Sh1] realized that there may be an additional finite group of symmetries, and he constructed an example of an Anosov diffeomorphism on a manifold that is not itself a nilmanifold but is finitely covered by one.

$$\begin{array}{ccc} N & \xrightarrow{\quad A \quad} & N \\ \downarrow & & \downarrow \\ N/\Gamma & \xrightarrow{\quad A \quad} & N/\Gamma \\ \pi\downarrow & & \downarrow\pi \\ M & \xrightarrow{\quad f \quad} & M \end{array}$$

π is finite-to-one. These examples are called infranil Anosov diffeomorphisms, and no new examples of Anosov diffeomorphisms have been added in thirty-six years. At the time Smale was cautious about calling his suggestion a conjecture, but by now it is certainly reasonable.

Conjecture. Every Anosov diffeomorphism is topologically conjugate to an infranil example.

There are a lot of partial results but we do not survey them here. Credence is lent to the conjecture by the fact that the corresponding conjecture is true for expanding maps, [Gro].

21. Fubini's Nightmare

In a product measure space $X \times Y$, Fubini detects a zero set Z slice by slice. Almost every slice Z_x should be a zero set in Y. Slices look like local leaves (plaques) of a foliation, and one would expect a Fubini type of theorem. In fact, quite the opposite is true for many dynamically defined foliations – each slice of a zero set can have full slice measure – and one must face the fact that this anti-Fubini phenomenon, dubbed "Fubini's Nightmare" by Flaminio, is natural and perhaps even typical.

It all comes down to center subbundle E^c, and often in the measure theoretic sense E^c can be almost eradicated. One begins with a linear hyperbolic diffeomorphism $A : \mathbb{T}^2 \to \mathbb{T}^2$ having a splitting $T(\mathbb{T}^2) = E^u_0 \oplus E^s_0$. The center leaves of the partially hyperbolic skew product $f_0 = A \times \mathrm{id} : \mathbb{T}^3 \to \mathbb{T}^3$ are circles $z \times S^1$ tangent to the bundle E^c_0. These circles are fibers of a normally hyperbolic invariant fibration,

$$\begin{array}{ccc} \mathbb{T}^3 & \xrightarrow{\quad f \quad} & \mathbb{T}^3 \\ \pi\downarrow & & \downarrow\pi \\ \mathbb{T}^2 & \xrightarrow{\quad A \quad} & \mathbb{T}^2, \end{array}$$

the stability of which was one of the starting points for our study of normally hyperbolic foliations and laminations in [HiPugSh2]. We showed that corresponding to each small perturbation f of f_0 there is a unique equivariant fibration

$$\begin{array}{ccc} \mathbb{T}^3 & \xrightarrow{\quad f \quad} & \mathbb{T}^3 \\ \pi_f\downarrow & & \downarrow\pi_f \\ \mathbb{T}^2 & \xrightarrow{\quad A \quad} & \mathbb{T}^2. \end{array}$$

The fibers of π_f are smooth simple closed curves, but π_f may only be Hölder.

By the results of Burns and Wilkinson (Theorem 16.3) f_0 can first be approximated by a stably ergodic skew product f_1, and then by the construction of Shub and Wilkinson [ShWi2] f_1 can be perturbed to a volume preserving diffeomorphism $f : \mathbb{T}^3 \to \mathbb{T}^3$ for which the Lyapunov exponent along its center leaves is almost everywhere positive. The Shub-Wilkinson diffeomorphism f is partially hyperbolic. It has a unique invariant splitting

$$T(\mathbb{T}^3) = E^u \oplus E^s \oplus E^c,$$

which is continuous and defined everywhere, but it also has a unique Lyapunov splitting

$$T(\mathbb{T}^3) = L^u \oplus L^s$$

such that vectors in L^u have positive Lyapunov exponent while vectors in L^s have negative Lyapunov exponent. The subbundles L^u, L^s are defined almost everywhere and are measurable, instead of continuous. Likewise their expansion/contraction rates are only measurable. Thus,

$$L^u = E^u \oplus E^c \ \text{ and } \ L^s = E^s \ \text{ almost everywhere,}$$

which is what we mean by almost eradicating of the center bundle. On a set $P \subset M$ of full measure, the center bundle is part of the Lyapunov unstable bundle. A diffeomorphism like f whose Lyapunov exponents are nonzero almost everywhere is called **non-uniformly hyperbolic**. Such dynamical systems were first considered by Pesin [Pe] who did much of the fundamental work on them.

But what does this mean for the center leaves? They are simple closed curves whose lengths are bounded above and below. Thus, arclength in a center leaf cannot grow exponentially under f^n, and it follows that P meets each center leaf in a set of arclenth zero, so the center foliation $\mathcal{W}^c$ of f, although unique and natural, is not absolutely continuous. Rather, it is absolutely singular. The pathology persists under perturbation, which is why we think it may be the rule, not the exception.

It is a nice exercise to picture the juliennes for these examples. Are they small hourglasses with exponentially thin multiple waists?

Previously, Katok had constructed a dynamically invariant foliation that is not absolutely continuous; it was presented by Milnor [Mi] under the title *Fubini foiled*. Subsequently, Dolgopyat [Do2] proved a similar theorem for the time-one map f of the geodesic flow for surfaces of negative curvature. He shows f has a neighborhood $\mathcal{N} \subset \mathrm{Diff}^2_m(T_1M)$ such that for all f' in an open-dense subset $\mathcal{N}' \subset \mathcal{N}$, f' has nonzero Lyapunov exponent in the center direction. It then follows from arguments communicated to us long ago by Mañé [Mañ4] that these maps have non-absolutely continuous center foliations.

Worse than non-absolute continuity of a foliation $\mathcal{F}$ is the property that for some sets P of full measure,

$$k = \sup\{\#(F \cap P) : \ F \text{ is a leaf of } \mathcal{F} \ \} < \infty.$$

P meets every leaf in at most k points. Katok's example in [Mi] has this property with $k = 1$. Ruelle and Wilkinson [RueWi] and Katok have shown that the Shub-Wilkinson examples have a similar property: for each such example there is a finite number k such that the set

$$P = \{p \in M : \ \text{the Lyapunov exponent in the center direction is positive}\}$$

meets almost every center leaf in exactly k points. Katok has pointed out that it is possible to have $k > 1$; this is achieved by versions of the Shub-Wilkinson construction that commute with a finite group of symmetries. It would be interesting to prove that there are such examples with $k = 1$. Then it would follow from entropy considerations that the Hölder continuous fibration π_f is a Lebesgue isomorphism from $\mathbb{T}^3$ to $\mathbb{T}^2$ whose fibers are C^1 circles. Of course, it is well known that $\mathbb{T}^3$ and $\mathbb{T}^2$ are Lebesgue isomorphic, but it would be somewhat surprising to have an explicitly constructible example of such an isomorphism arising naturally via considerations of dynamics. The fibers π_f are easy to calculate numerically since they are given by the contraction mapping theorem.

There is a dramatic difference between manifolds that support hyperbolic diffeomorphisms, or even partially hyperbolic diffeomorphisms, and those that support non-uniformly hyperbolic diffeomorphisms. In [DoPe], Dolgopyat and Pesin prove the latter exist on all manifolds of dimension ≥ 2. Their examples are isotopic to the identity, while the Shub-Wilkinson non-uniformly hyperbolic examples are not. This leads one to the question: *Are there non-uniformly hyperbolic diffeomorphisms of manifolds of dimension* ≥ 2 *in all isotopy classes?*

22. Other measures

Even though a dynamical system refuses to preserve a smooth volume m, it may preserve other measures that capture the behavior of almost all (with respect to m) orbits. Such SRB measures were constructed by Sinai, Ruelle and Bowen in the 70s for Axiom A dynamical systems. See [Si], [Rue], [BowRu], [Bow2]. Axiom A systems include hyperbolic diffeomorphisms and flows, and also some systems with chaotic attractors. The SRB measure is supported on the attractor, so the space average of a function is its average over the attractor. Ergodicity means that for almost all (with respect to m) orbits, the forward time average equals the space average. Stable ergodicity in the SRB context presents a tantalizing avenue of research.

At a recent 65th birthday celebration for Yasha Sinai and David Ruelle, Lai-Sang Young [You] surveyed the work on these measures. We recommend this article to anyone interested in pursuing the subject, and mention a few references closely tied to stable ergodicity or partial hyperbolicity, [AlBonVi],[BonVi], [BuDoPe], [Do1], [Do3], [PeSi].

Rufus Bowen died suddenly of a brain aneurism in 1978. He was born in 1947. So he would anyway have been too young for a 65th birthday party.

References

[AlBonVi] Alves, J.F., C. Bonatti and M. Viana, *SRB measures for partially hyperbolic systems whose central direction is mostly expanding*, Invent. Math. **140** (2000), 351-398.

[An] Anosov, D. V., *Geodesic flows on closed Riemannian manifolds of negative curvature*, Proc. Steklov. Inst. Math. **90** (1967).

[AnSi] Anosov, D. V. and Ya. G. Sinai, *Some smooth ergodic systems*, Russian Math. Surveys **22 No.5** (1967), London Math. Soc., 103-167.

[ArMa] Arbieto, A. and C. Mateus, C^1 *Robust Transitivity*, preprint.

[ArAv] Arnold, V.I. and A.Avez, *Théorie Ergodique des Systèmes Dynamiques*, Gautier-Villars, Paris, 1967

[AuGr] Auslander,L. and L. Green, *G-induced flows*, Amer. J. Math. **88** (1966), 43-60.

[BoFaPu] Bochi, J., B. Fayad and E. Pujals, *Stable ergodicity implies Bernoulli*, preprint.

[BonDíPujUr] Bonatti, C., L. Díaz and R.Ures, *Minimality of strong stable and unstable foliations for partially hyperbolic diffeomorphisms*, Preprint (2001).

[BonMaViWi] Bonatti, C., C. Matheus, M. Viana and A. Wilkinson, *Abundance of stable ergodicity*, preprint

[BonVi] Bonatti, C. and M. Viana, *SRB measures for partially hyperbolic systems whose central direction is mostly contracting*, Israel J. Math. **115** (2000), 157-194.

[BonWi] Bonatti, C. and A.Wilkinson, *Transitive partially hyperbolic diffeomorphisms on 3-manifolds*, preprint.

[Bow1] Bowen, R. *A horseshoe with positive measure*, Invent. Math. **29** (1975), 203-204.

[Bow2] Bowen, R., *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics **470**, Springer-Verlag, Berlin-New York, 1975.

[BowMar] Bowen, R. and B. Marcus, *Unique ergodicity for horocycle foliations*, Israel J. Math. **26** (1977), 43-67.

[BowRu] Bowen, R. and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. **29** (1975), 181-202.

[Boy] Boyd, D.W., *Small Salem Numbers*, Duke Math. J., **44**(1977), 315-328.

[BreSh] Brezin, J. and M. Shub, *Stable ergodicity in homogeneous spaces*. Bol. Soc. Brasil. Mat. (N.S.) **28** (1997), no. 2, 197-210.

[Bri1] Brin, M., *Topological transitivity of one class of dynamical systems and flows of frames on manifolds of negative curvature*, Func. Anal. Appl. **9** (1975), 9-19.

[Bri2] Brin, M., *The topology of group extensions of C systems*, Mat. Zametki **18** (1975), 453-465.

[Bri3] Brin, M., *On dynamical coherence*, Ergodic Th. and Dynamical Sys.**23** (2003), 395-401.

[BriBuIv] Brin, M., D.Burago and S.Ivanov *Partial hyperbolicity in dimension 3*, in preparation.

[BriPe] Brin, M. and Ja. Pesin, *Partially hyperbolic dynamical systems*, Math. USSR Izvestija **8** (1974), 177-218.

[BuDoPe] Burns, K, D. Dolgopyat and Ya. Pesin *Partial hyperbolicity, Lyapunov exponents and stable ergodicity*, preprint

[BuPugShWi] Burns, K., C. Pugh, M.Shub and A. Wilkinson, *Recent Results About Stable Ergodicity* in Proceedings of Symposia in Pure Mathematics Vol. 69, "Smooth Ergodic Theory and Its Applications" (Katok, A., R. de la Llave, Y. Pesin, H. Weiss, Eds.), AMS, Providence, R.I., 2001, 327-366.

[BuPugWi] Burns, K., C. Pugh and A. Wilkinson, *Stable ergodicity and Anosov flows*, Topology **39** (2000), 149-159.

[BuWi] Burns, K. and A. Wilkinson, *Stable ergodicity of skew products*, Ann. Sci. École Norm. Sup. **32** (1999), 859-889.

[BuWi2] Burns, K. and A. Wilkinson, *Better center bunching*, in preparation.

[ChSu] Cheng, C.-Q., and Y.-S. Sun, *Existence of invariant tori in three dimensional measure-preserving mappings*, Celestial Mech. Dynam. Astronom. **47** (1989/90), 275-292.

[D-G,G-H,Be] Descomps-Guilloux,A., M. Grandet-Hugot and M.J.Bertin, *Pisot and Salem Numbers*, Springer, 1993

[Dí] Díaz, L.,*The end of domination*, Abstracts, New Directions in Dynamical Systems, Kyoto, Japan, August 5-15, 2002.

[de la Llave] de la Llave, R. *A Tutorial on KAM Theory,* in Proceedings of Symposia in Pure Mathematics Vol. 69, "Smooth Ergodic Theory and Its Applications" (Katok, A., R de la Llave, Y. Pesin, H. Weiss, Eds.), AMS, Providence, R.I., 2001, 175-292.

[Do1] Dolgopyat, D., *On dynamics of mostly contracting diffeomorphisms*, Comm. in Math. Physics , **213** (2000) 181-201.

[Do2] Dolgopyat, D., *On dynamics of partially hyperbolic systems on three manifolds*, preprint (1999).

[Do3] Dolgopyat, D., *On differentiability of SRB states for partially hyperbolic systems*, preprint.

[DoPe] Dolgopyat, D. and Ya.Pesin, *Every compact manifold carries a completely hyperbolic diffeomorphism,* , Ergod. Theory and Dynam. Systems , **22** (2002), 409-437.

[DoWi] Dolgopyat, D. and A.Wilkinson *Stable accessibility is C^1 dense*, preprint.

[Fa] Falconer, K. *The Geometry of Fractal sets*, Cambridge University Press, London, 1985.

[FiNi1] Field, M. and M. Nicol, *Ergodic theory of equivariant diffeomorphisms I: Markov partitions*, preprint (1999).

[FiNi2] Field, M. and M. Nicol, *Ergodic theory of equivariant diffeomorphisms II: stable ergodicity*, preprint (1999).

[FiPa] Field, M. and W. Parry, *Stable ergodicity of skew extensions by compact Lie groups*, Topology **38** (1999), 167-187.

[FiMeTo] Field, M., I. Melbourne and A. Torok *Stable ergodicity for smooth compact Lie group extensions of hyperbolic basic sets, and stable mixing for hyperbolic flows*, preprint.

[Gall] Gallavotti, G. *Statistical Mechanics: a Short Treatise*, Springer-Verlag, NY, 1999.

[GeFo] Gelfand, I.M. and S.V. Fomin, *Geodesic flows on manifolds of constant negative curvature*, Uspeki Mat. Nauk. **7** (1952) 118-137; English translation., Amer. Math. Soc. Trans. (2) **1** (1955), 49-65.

[GrPugSh] Grayson, M., C. Pugh and M. Shub, *Stably ergodic diffeomorphisms*, Ann. Math. **140** (1994), 295-329.

[Gro] Gromov, M., *Groups of polynomial growth and expanding maps*, Inst. des Hautes Études Sci. Publ. Math. **53**, (1981), 53-73.

[HiPuSh1] Hirsch, M., C. Pugh and M.Shub, *Invariant Manifolds* , Bull. AMS, **76**(1970), 1015-1019.

[HiPugSh2] Hirsch, M., C. Pugh and M. Shub, *Invariant manifolds*, Lecture Notes in Mathematics, **583**, Springer-Verlag, 1977.

[Ho] Hopf, E., *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261-304.

[KaKo] Katok, A., and A. Kononenko, *Cocycles' stability for partially hyperbolic systems*, Math. Res. Lett. **3** (1996), 191-210.

[Liao] Liao, S.T., *On the stability conjecture*, Chinese Ann. Math. **1** (1980), 9-30.

[Li] Lind, D., *Dynamical properties of quasihyperbolic toral automorphisms*, Ergodic Th. and Dyn. Syst. **2** (1982), 49-68.

[Mañ1] Mañé, R., *Persistent manifolds are normally hyperbolic*, Bull. Amer. Math. Soc. 80 (1974), 90-91.

[Mañ2] Mañé, R., *Oseledec's theorem from the generic viewpoint*, Proc. of Int. Congress of Math. (1983) Warszawa 1269-1276.

[Mañ3] Mañé, R., *A proof of the C^1 stability conjecture*. Inst. des Hautes Études Sci. Publ. Math. No. 66, (1988), 161-210.

[Mañ4] Mañé, R., private communication.

[Mau1] Mautner, F.I., *Geodesic flows and unitary representations*, Proc. Nat. Acad. Sci. USA **40** (1954), 33-36

[Mau2] Mautner, F.I., *Geodesic flows on symmetric Riemann spaces*, Ann. of Math. (2) **65** (1957), 416-431.

[Mi] Milnor, J. *Fubini foiled: Katok's paradoxical example in measure theory*, Math. Intelligencer, **19** (1997), 30-32.

[Mo] Moore, C.C. *Ergodicity of flows on homogeneous spaces*, Amer. J. Math. **88** (1966), 154-178.

[Nik] Nikodym, O. *Sur la mesure des ensembles plans dont tous les points sont rectilinearment accessibles.* Fund. Math., **10**, 1927, 116-168.

[NiTö] Niţică, V. and A. Török, *An open dense set of stably ergodic diffeomorphisms in a neighborhood of a non-ergodic one*, preprint (1998).

[Par] Parry W., *Dynamical systems on nilmanifolds.* Bull. London Math. Soc. **2** (1970), 37-40.

[ParPo] Parry W., and M. Pollicott, *Stability of mixing for toral extensions of hyperbolic systems*, Tr. Mat. Inst. Steklova **216** (1997), Din. Sist. i Smezhnye Vopr., 354-363.

[Pe] Pesin, Ya., *Characteristic Lyapunov exponents and smooth ergodic theory*, Russian Math. Surveys **32** (1977), no. 4(196), 55-112, 287.

[PeSi] Pesin, Ya.B. and Ya. G. Sinai, *Gibbs measures for partially hyperbolic attractors*, Ergod. Th. and Dynam. Sys., **2** (1982), 417-438.

[Pli] Pliss, V. A., *On a conjecture of Smale*, (Russian) Differencialć nye Uravnenija **8** (1972), 268-282.

[PugSh1] Pugh, C. and M. Shub, *Ergodicity of Anosov actions*, Invent. Math. **15** (1972), 1-23.

[PugSh2] Pugh, C. and M. Shub, *Stably ergodic dynamical systems and partial hyperbolicity*, J. of Complexity **13** (1997), 125 - 179.

[PugSh3] Pugh, C. and M. Shub, *Stable ergodicity and julienne quasiconformality*, J. Eur. Math. Soc. **2** (2000), 1-52.

[PugShSt] Pugh, C., M. Shub and A. Starkov, Erratum to Stable ergodicity and julienne quasiconformality, J. Eur. Math. Soc. 2, 1-52 preprint.

[Puj] Pujals, E., *Tangent bundles dynamics and its consequences* in Li, Ta Tsien (ed.) et al., Proceedings of the international congress of mathematicians, ICM 2002, Beijing, China, August 20-28, 2002. Vol. III: Invited lectures. Beijing: Higher Education Press. 327-338 (2002). .

[RY] Robinson, C., and Young, L-S., *Nonabsolutely continuous foliations for an Anosov diffeomorphism*, Invent. Math. **61** (1980), 159-176.

[RH] Rodriguez Hertz, F., *Stable Ergodicity of Certain Linear Automorphisms of The Torus*, to appear.

[Rue] Ruelle, D., *A measure associated with Axiom-A attractors*, Amer. J. Math., **98** (1976), 619-654.

[RueSu] Ruelle, D. and D.Sullivan, *Currents, flows and diffeomorphisms*, Topology **14** (1975), 319-327.

[RueWi] Ruelle, D. and A. Wilkinson, *Absolutely singular dynamical foliations*, Comm. Math. Phys. **219** (2001), 481-487.

[Sh1] Shub, M. *Endomorphisms of Compact Differentiable Manifolds*, Amer. J. Math. **XCI**(1969), 175-199.

[ShVa] Shub, M. and A. Vasquez, *Some Linearly Induced Morse-Smale Systems, the QR Algorithm and the Toda Lattice*, Contemporary Mathematics, Vol. 64, The Legacy of Sonya Kovalevskaya, Linda Keen ed., AMS, Providence 1987, 181-194.

[ShWi1] Shub, M. and A. Wilkinson, *Stably ergodic approximation: two examples*, Ergod. Th. and Dyam. Syst. **20**, (2000), 875-894.

[ShWi2] Shub, M. and A. Wilkinson, *Pathological foliations and removable zero exponents*, Invent. Math. **139**, (2000), 495-508.

[Si] Sinai, Ya., *Gibbs measures in ergodic theory*, Russian Math. Surveys **27** (1972), 21-69.

[Sm] Smale, S., *Differentiable dynamical systems*, Bull. AMS, **73** (1967), 747-817.

[St1] Starkov, A.N., *Dynamical Systems on Homogeneous Spaces*, Translations of Mathematical Monographs, (190) AMS, Providence, RI., 2000.

[St2] Starkov, A.N., *Stable ergodicity among left translations*, Appendix to this article.

[Ta] Tahzibi, A., *Stably ergodic diffeomorphisms which are not partially hyperbolic*, to appear.

[Wal] Walkden, C., *Stable ergodic properties of cocycles over hyperbolic attractors*, Comm. Math. Phys. **205** (1999), 263-281.

[Wa] Walters, P., *Affine transformations of tori*, Trans. Amer. Math. Soc., **131** (1968), 40-50.

[Wi1] Wilkinson, A., *Stable ergodicity of the time one map of a geodesic flow*, Ph.D. Thesis, University of California at Berkeley (1995).

[Wi2] Wilkinson, A., *Stable ergodicity of the time one map of a geodesic flow*, Ergod. Th. and Dynam. Syst. **18** (1998), 1545-1588.

[Yoc] Yoccoz, J.-C., *Travaux de Herman sur les tores invariants*, Séminaire Bourbaki, Vol. 1991/92, Astérisque No. 206 (1992), Exp. No. 754, **4**, 311-344.

[You] Young, L.-S., *What are SRB measures, and which dynamical sysytems have them?* preprint, to appear in JSP

Appendix: Stable Ergodicity among Left Translations

Alexander Starkov[‡]

Let G be a connected Lie group and $B \subset G$ be a closed subgroup. The homogeneous space G/B is said to be of **finite volume** if G/B admits a finite G-invariant measure. An **affine map** $f : G/B \to G/B$ is a composition $f = L_g \circ A$ where $L_g : G/B \to G/B$ is left translation by an element $g \in G$ and $A : G/B \to G/B$ is the map induced by an automorphism $\bar{A} : G \to G$ such that $\bar{A}(B) = B$. As explained in Section 15 above, f induces an automorphism $\mathfrak{a}(f) = \mathrm{Ad}_g \circ T_e A$ of the Lie algebra $\mathfrak{g} = T_e G$ where Ad_g is the adjoint action of g. The $\mathfrak{a}(f)$-invariant subspace of generalized eigenvectors with eigenvalues off the unit circle is a Lie algebra $\mathfrak{h}$ tangent to the **hyperbolic subgroup** H of G.

Theorem 1. *Let $f : G/B \to G/B$ be an affine map of the finite volume homogeneous space G/B with hyperbolic subgroup H. Then the following are equivalent:*

(a) *f is stably ergodic among left translations, i.e. there exists a neighborhood $O(g) \subset G$ such that the affine map $f_x = L_x \circ A$ is ergodic for each $x \in O(g)$.*

(b) *f is stably a K-automorphism among left translations, i.e. there exists a neighborhood $O(g) \subset G$ such that the affine map $f_x = L_x \circ A$ is a K-automorphism for each $x \in O(g)$.*

(c) *G equals the closure $\overline{HB}$.*

Proof. We need a simple property that the hyperbolic subgroup $H = H_g$ can only enlarge when we perturb g.

Proposition 1. There exists a neighborhood $O(g) \subset G$ such that H is contained in the hyperbolic subgroup H_x of $f_x = L_x \circ A$ for every $x \in O(g)$.

Proof. Recall that the Lie algebra $\mathfrak{h} \subset \mathfrak{g}$ of H is the smallest ideal in $\mathfrak{g}$ invariant under $\mathrm{Ad}_g \circ d\bar{A}$ such that the operator $\mathrm{Ad}_g \circ d\bar{A}$ on $\mathfrak{g}/\mathfrak{h}$ has all its eigenvalues of absolute value 1. It follows that $\mathfrak{h}$ is invariant under $d\bar{A}$ as well as under every operator $\mathrm{Ad}_x,\ x \in G$.

Let $\mathfrak{h}_x$ be the Lie algebra of H_x. Notice that $\mathfrak{h}_x$ is invariant under $\mathrm{Ad}_g \circ d\bar{A}$.

Suppose that H is not contained in H_x. Then the operator $\mathrm{Ad}_g \circ d\bar{A}$ on $\mathfrak{g}/\mathfrak{h}_x$ has an eigenvalue λ with $|\lambda| \neq 1$. Taking a quotient eliminates eigenvalues but

[‡] The author was supported by the Leading Scientific School Grant, No 457.2003.1.

does not change the ones that remain. Thus λ does not depend on x. As $x \to g$, $\mathrm{Ad}_x \circ d\bar{A} \to \mathrm{Ad}_g \circ d\bar{A}$, contrary to the fact that all eigenvalues of $\mathrm{Ad}_x \circ d\bar{A}$ on $\mathfrak{g}/\mathfrak{h}_x$ are of modulus 1. □

According to Dani [Da], [Da1], the affine map f is a K-automorphism on G/B if and only if $G = \overline{HB}$. It follows from Proposition 1 that (c) implies (b), and (b) evidently implies (a). It remains to show that (a) implies (c), i.e., that $\overline{HB} \neq G$ implies that f is not stably ergodic.

If $\overline{HB} \neq G$ we take $G' = G/H, B' = \overline{HB}/H$, and let f' be the affine map of $G'/B' = G/\overline{HB}$ induced by f. Notice that the hyperbolic subgroup of G' for f' is trivial (so f' has zero entropy). It suffices to prove that f' is not stably ergodic. So, replacing f by f' if needed, we can assume without loss of generality that H is trivial and G/B is not trivial. We do so.

Recall that the cases when G is either semisimple or solvable (as well as a large part of the general situation) were fully considered by Brezin and Shub [BS].

Remark. In short, the idea used in [BS] is as follows. If G is solvable and H is trivial, then G/B admits a toral quotient on which f acts as a pure translation; hence f is not stably ergodic. The case when G is semisimple reduces to the situation when $f = L_g$ and G has finite center. Since H is trivial, g is a quasi-unipotent element of G. Then by the generalized Jacobson-Morozov Lemma, g can be approximated by elements of finite order; hence again f is not stably ergodic.

Now, to prove that f is not stably ergodic if H is trivial, we only need to reduce the general situation to these two basic cases. To do this, we apply the following result of Starkov [St] and Witte [Wi] which solves certain problems related to the structure of finite volume homogeneous spaces.

Proposition 2. Let G/B be a finite volume homogeneous space, and let $f : G/B \to G/B$ be an ergodic affine map. Let R be the solvable radical of G. Then the product group RB is closed.

Now there are two cases. Let $G = SR$ be a Levi decomposition of G, R being the solvable radical and S being a maximal semisimple connected subgroup of G. If RB is a proper subgroup of G, then the space $G/RB \simeq S/RB \cap S$ is isomorphic to a nontrivial homogeneous space of S. We fall into the semisimple case, and it follows from Proposition 3.3 of [BS] that f is not stably ergodic.

If $G = RB$ then $G/B \simeq R/B \cap R$, and hence we are in the solvable case. It follows from Proposition 3.4 of [BS] that f is not stably ergodic. □

References

[BS] Brezin, J. and M. Shub, *Stable ergodicity in homogeneous spaces*, Bol. Soc. Bras. Mat. **28** (1997), No. 2, 197-210.

[Da] Dani, S. *Kolmogorov automorphisms on homogeneous spaces*, Amer. J. Math. **98** (1976), 119-163.

[Da1] Dani, S. *Spectrum of an affine transformation*, Duke Math. J. **44** (1977), 129-155.

[St] Starkov, A.N., *On a criterion for the ergodicity of G-induced flows*, Uspekhi Mat. Nauk, **42**(1987), No. 3, 197-198; English transl., Russian Math. Surveys **42** (1987), No. 3, 233-234.

[Wi] Witte, D. *Zero-entropy affine maps on homogeneous spaces*, Amer. J. Math. **109** (1987), 927-961.

Mathematics Department, University of California, Berkeley California, 94720
E-mail address: pugh@math.berkeley.edu

Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario, M5S 3G3, Canada, and, IBM T.J. Watson Research Center, Yorktown Heights, NY, 10598-0218, USA
E-mail address: shub@math.toronto.edu and mshub@us.ibm.com

All-Russian Institute of Electrotechnics, Istra, Moscow region, Russia, 143500
E-mail address: RDIEalex@istra.ru