ERGODIC THEORY
of
DIFFERENTIABLE DYNAMICAL SYSTEMS

Lai-Sang Young*

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90024
Email: lsy@math.ucla.edu

These notes are about the dynamics of systems with hyperbolic properties. The setting for the first half consists of a pair (f, μ), where f is a diffeomorphism of a Riemannian manifold and μ is an f-invariant Borel probability measure. After a brief review of abstract ergodic theory, Lyapunov exponents are introduced, and families of stable and unstable manifolds are constructed. Some relations between metric entropy, Lyapunov exponents and Hausdorff dimension are discussed. In the second half we address the following question: given a differentiable mapping, what are its natural invariant measures? We examine the relationship between the expanding properties of a map and its invariant measures in the Lebesgue measure class. These ideas are then applied to the construction of Sinai-Ruelle-Bowen measures for Axiom A attractors. The nonuniform case is discussed briefly, but its details are beyond the scope of these notes.

I have aimed these notes at readers who have a basic knowledge of dynamics but who are not experts in the ergodic theory of hyperbolic systems. To cover the material mentioned above in 40-50 pages, some choices had to be made. I wanted very much to give the reader some feeling for the flavor of the subject, even if that meant focusing on fewer ideas. I have not hesitated to include examples, informal discussions, and some of my favorite proofs. I did not try to mention all related results. For survey articles on similar topics see [ER] or [S5].

1. Review of Abstract Ergodic Theory

This section contains most of the ergodic theory background needed for these notes. A suitable reference for Sections 1.1 and 1.2 is [Wa]. I also like [S4]. For Section 1.3 see [Ro1] and [Ro2].

*This research is partially supported by NSF

1.1. Basic notions. Let $(X, \mathcal{B}, \mu)$ be a probability space, i.e. X is a set, $\mathcal{B}$ is a σ-algebra of subsets of X, and μ is a measure on $(X, \mathcal{B})$ s.t. $\mu(X) = 1$. If $(X_i, \mathcal{B}_i, \mu_i), i = 1, 2$, are probability spaces, a mapping $T : X_1 \to X_2$ is called *measurable* if $\forall\, A \in \mathcal{B}_2, T^{-1}A \in \mathcal{B}_1$. A measurable mapping T is said to be *measure preserving* if $\forall\, A \in \mathcal{B}_2, \mu_1(T^{-1}A) = \mu_2(A)$. We say that T is an *invertible* measure preserving transformation if T is bijective and both T and T^{-1} are measure preserving.

We use the notation $T : (X, \mathcal{B}, \mu) \circlearrowleft$ to denote a measure preserving transformation (henceforth abbreviated as *mpt*) of a probability space to itself. Mention of $\mathcal{B}$ is suppressed when the σ-algebra is understood. For instance, if X has a topological structure, then $\mathcal{B}$ is always the Borel σ-algebra (i.e. the σ-algebra generated by open sets) or the Borel σ-algebra completed with respect to μ. We also refer to μ as an *invariant measure* for T.

Two well known examples of *mpt*'s or flows are Hamiltonian systems and stationary stochastic processes. We mention also a few simple minded examples:

Example 1.1.1. Let $S^1 \cong \{z \in \mathbb{C} : |z| = 1\}$ and let μ be the Lebesgue measure on S^1. Then for fixed $\alpha \in \mathbb{R}, T : z \mapsto e^{2\pi i\alpha}z$, and $T : z \mapsto z^2$ are *mpt's.*

Example 1.1.2. Let $T : [-1, 1] \circlearrowleft$ be given by $Tx = 1 - 2x^2$. Then the measure with density $1/\pi cos(arcsinx)$ is an invariant probability measure. One way to see this is via the change of coordinates $x = h(\theta) = sin\frac{\pi}{2}\theta$. It suffices to verify that $g := h^{-1} \circ f \circ h$ satifies $|g'(\theta)| = 2$ and hence preserves Lebesgue measure.

Example 1.1.3. Let μ_0 be the probability measure on $X_0 := \{1, \ldots, k\}$ with $\mu_0\{i\} = p_i$, and let μ be the product measure of μ_0 on $X := \prod_{-\infty}^{\infty} X_0$. Let $T : X \circlearrowleft$ be the shift operator, i.e., if $(\underline{x})_i$ denotes the *ith* coordinate of $\underline{x} \in X$, then $(T\underline{x})_i = (\underline{x})_{i+1}$. This defines an *mpt* on X called the $(p_1, \ldots, p_k)$-*Bernoulli shift.*

As in every category, we do not distinguish between certain objects. We say that $T_1 : (X_1, \mathcal{B}_1, \mu_1) \circlearrowleft$ is *isomorphic* to $T_2 : (X_1, \mathcal{B}_2, \mu_2) \circlearrowleft$ if $\exists\, \tilde{X}_i \subset X_i$ with $\mu_i\tilde{X}_i = 1, i = 1, 2$, and an invertible *mpt* $S : \tilde{X}_1 \to \tilde{X}_2$ s.t. $S \circ T_1 = T_2 \circ S$.

Example 1.1.4. (baker's transformation) Let X be $[0, 1] \times [0, 1]$, and let μ be Lebesgue measure on X. Let $T(x, y) = (2x, \frac{1}{2}y)$ if $x < \frac{1}{2}, T(x, y) = (2x-1, \frac{1}{2}y+\frac{1}{2})$ if $x > \frac{1}{2}$. Then T is isomorphic to the $\left(\frac{1}{2}, \frac{1}{2}\right)$-Bernoulli shift.

Perhaps the first and most fundamental question one asks about an *mpt* is whether or not it can be decomposed into smaller non-interacting subsystems.

Definition 1.1.5. *$T : (X, \mathcal{B}, \mu) \circlearrowleft$ is called ergodic if $\forall\, A \in \mathcal{B}, T^{-1}A = A \Rightarrow \mu A =$ 0 or 1.*

It is straightforward to verify that ergodicity is equivalent to each of the following two conditions:

(1) $\forall A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0, \exists n > 0$ s.t. $\mu(T^{-n}A \cap B) > 0$.
(2) If $\varphi : X \to \mathbb{R}$ is measurable (or in L^p, any $0 < p \leq \infty$), and $\varphi \circ T = \varphi$ a.e., then φ is constant a.e. on X.

Next we state a version of the Ergodic Theorem:

Theorem 1.1.6. (Birkhoff Ergodic Theorem, 1932) *Consider $T : (X, \mathcal{B}, \mu) \circlearrowleft$, and let $\varphi \in L^1(\mu)$. Then $\exists\, \varphi^* \in L^1(\mu)$ s.t.*

$$\frac{1}{n}\sum_{0}^{n-1} \varphi \circ T^i \to \varphi^* \quad a.e.$$

Moreover, φ^* satisfies $\varphi^* \circ T = \varphi^*$ a.e. and $\int \varphi^* d\mu = \int \varphi d\mu$. It follows from this and (2) above that if (T, μ) is ergodic, then $\varphi^* = \int \varphi d\mu$ a.e.

Recall that in a probability space $(X, \mathcal{B}, \mu)$, two sets (or events) $A, B \in \mathcal{B}$ are called *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$.

Definition 1.1.7. *$T : (X, \mathcal{B}, \mu) \circlearrowleft$ is called mixing if $\forall A, B \in \mathcal{B}, \mu(T^{-n}A \cap B) \to \mu(A)\mu(B)$ as $n \to \infty$.*

A mixing transformation then is one for which every set, upon iteration of T, becomes asymptotically independent of every other set. This property is strictly stronger than ergodicity, which is equivalent to $\frac{1}{n}\sum\limits_{0}^{n-1} \mu(T^{-n}A \cap B) \to \mu A \cdot \mu B$ for every $A, B \in \mathcal{B}$ (an easy consequence of the Birkhoff Ergodic Theorem). An *mpt* is called *Bernoulli* if it is isomorphic to a Bernoulli shift. Approximating measurable sets by cylinder sets, one shows easily that Bernoulli transformations are mixing. To sum up: Bernoulli $\Rightarrow$ mixing $\Rightarrow$ ergodic.

We mention some topological counterparts to the notions we have introduced. If $T : X \circlearrowleft$ is a homeomorphism of a compact metric space, then a topological version of ergodicity is *topological transitivity*, which says that for every pair of open sets $A, B \subset X, \exists\, n$ s.t. $T^{-n}A \cap B \neq \phi$. This is also equivalent to every point in a residual subset of X having a dense orbit in X. The topological version of mixing is that for every open $A, B \subset X, \exists\, N$ s.t. $T^{-n}A \cap B \neq \phi\ \forall n \geq N$.

1.2. Metric entropy. Entropy measures the degree of randomness or complexity of a measure preserving transformation. Here is one way to motivate it:

Let us consider an experiment α with k possible mutually exclusive outcomes $A_1, \ldots, A_k$. Suppose that A_i occurs with probability p_i. We want to measure the amount of information gained by performing α, or equivalently, the amount of uncertainty one has when asked to predict the outcome of α in advance. What we seek is not really a function of α, but a function $H(p_1, \ldots, p_k)$ of probability vectors. (A k-tuple $(p_1, \ldots, p_k)$ is called a probability vector if $0 \leq p_i \leq 1$ and $\Sigma p_i = 1$.) Clearly, any reasonable H must have the following properties:

1) symmetry, i.e. $H(\cdots p_i \cdots p_j \cdots) = H(\cdots p_j \cdots p_i \cdots)$;
2) $H(1,0,\dots,0)=0$;
3) $H(0,p_1,\dots,p_k) = H(p_1,\dots,p_k)$;
4) $H\left(\frac{1}{k},\dots,\frac{1}{k}\right) \geq H(p_1,\dots,p_k)\ \forall\,(p_1,\dots,p_k)$, equality iff $p_i = \frac{1}{k}\ \forall\, i$;
5) $H \geq 0$ and is a continuous function of its coordinates for each fixed k.

Suppose now we are to perform two experiments α and β, the outcomes of α being $A_1,\dots,A_k$ with prob $(A_i) = p_i$, and the outcomes of β being $B_1,\dots,B_\ell$. Let $\pi_{ij} =$ prob $(A_i \cap B_j)$, or equivalently, prob$(B_j|A_i) = \frac{\pi_{ij}}{p_i}$. Then the amount of uncertainty involved in predicting the outcome of β given that A_i has occurred is $H\left(\frac{\pi_{i1}}{p_i},\dots,\frac{\pi_{ik}}{p_i}\right)$. Our intuition dictates that H must also satisfy

6) $H(\{\pi_{ij}\}) = H(\{p_i\}) + \sum\limits_i p_i H\left(\left\{\frac{\pi_{ij}}{p_i}\right\}\right)$.

Theorem 1.2.1. *The function*

$$H(p_1,\dots,p_k) = -\sum p_i \log p_i \qquad\qquad (0\log 0 = 0)$$

is, up to multiplication by a positive constant, the unique function satisfying 1) – 6).

Let us now transfer this intuition to *mpt's*. Let $(X,\mathcal{B},\mu)$ be a probability space, and let $\alpha = \{A_1,\dots,A_k\}$ and $\beta = \{B_1,\dots,B_\ell\}$ be measurable partitions of X, i.e. $A_i \in \mathcal{B}, X = \cup A_i, A_i \cap A_j = \phi\ \forall\, i \neq j$ etc. (All partitions in this subsection are finite.) The numbers

$$H(\alpha) := -\sum_i \mu A_i \log \mu A_i$$

and

$$H(\beta|\alpha) := \sum_i \mu A_i \left[-\sum_j \mu(B_j|A_i)\log\mu(B_j|A_i)\right]$$

have the following interpretations: We say that the α-address of $x \in X$ is A_i if $x \in A_i$. Then $H(\alpha)$, called the *entropy of the partition* α, measures the amount of uncertainty when one is asked to guess the α-address of a random point. Similarly, $H(\beta|\alpha)$, the *conditional entropy* of β *given* α, measures the average amount of uncertainty in guessing the β-address of a random point given its α-address.

We will use the following notations: If α and β are partitions, then the *join* of α and β, written $\alpha \vee \beta$, is defined to be $\{A \cap B : A \in \alpha, B \in \beta\}$. If $\alpha_1,\dots,\alpha_n$ are partitions, we also write $\bigvee\limits_1^n \alpha_i$. $T^{-1}\alpha$ is the partition $\{T^{-1}A, A \in \alpha\}$ and $\alpha < \beta$ means that β is finer than α, i.e. every $B \in \beta$ is contained in an element of α.

Let $T : (X, \mathcal{B}, \mu) \circlearrowleft$ be an *mpt*, and let α be a partition of X. We define

$$h(T;\alpha) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-1}\alpha\right)$$

and

$$h(T) := \sup_{\alpha} h(T;\alpha).$$

Continuing to interpret things as we have done, $H\left(\bigvee_{i=0}^{n-1} T^{-1}\alpha\right)$ measures the average information gain knowing the α-addresses of $x, Tx, \ldots, T^{n-1}x$ for a random x, and $h(T;\alpha)$ is the asymptotic information gain per iterate. It is easy to verify that $h(T;\alpha)$ can also be realized as $\lim_n H\left(\alpha|\bigvee_1^n T^{-i}\alpha\right)$.

In the definition of $h(T)$, one is supposed to take supremum over all partitions. This is neither practical nor necessary. To state the next lemma we need a couple of definitions. We will confuse a partition α with the σ-algebra generated by α, and if $\alpha_1, \alpha_2, \ldots$ are σ-algebras, then $\bigvee_{i=1}^{\infty} \alpha_i$ denotes the smallest σ-algebra containing all the α_i's. Two σ-algebras α and β are said to be *equal mod 0*, written $\alpha \doteq \beta$, if $\forall\, A \in \alpha$, $\exists\, B \in \beta$ s.t. $\mu(A \triangle B) = 0$ and vice versa.

Lemma 1.2.2. *Let $\{\alpha_n\}$ be a sequence of partitions s.t. $\alpha_1 < \alpha_2 < \cdots$ and $\bigvee_{n=1}^{\infty} \alpha_i \doteq \mathcal{B}$. Then $\lim_n h(T;\alpha_n) = h(T)$. Also, if α is a generator (i.e. $\bigvee_{-\infty}^{\infty} T^{-i}\alpha \doteq \mathcal{B}$ if T is invertible, $\bigvee_0^{\infty} T^{-i}\alpha \doteq \mathcal{B}$ if T is not), then $h(T;\alpha) = h(T)$.*

The concept of entropy goes back to Shannon in his work on information theory. Entropy as a measure of randomness for dynamical systems was introduced by Kolmogorov and Sinai in the late 50's. Let us look at some examples:

Example 1.2.3. Rotation of S^1. Let α be any partition of S^1 into intervals. One sees inductively that the cardinality of $\bigvee_0^{n-1} T^{-i}\alpha$ is $\leq n \cdot \text{card } \alpha$. So $h(T;\alpha) = 0$. The same argument tells us that interval exchange maps (maps that permute a finite number of intervals) also have entropy zero. For these maps, the growth of information is linear; entropy measures the *exponential rate* of information growth.

Example 1.2.4. Bernoulli shifts. Let T be the $(p_1, \ldots, p_k)$-Bernoulli shift, and let α be the partition of $X := \prod_{-\infty}^{\infty} \{1, \ldots, k\}$ by its 0th coordinate. Then α is a generator, and

$$h(T;\alpha) = \lim_n H\left(\alpha|\bigvee_1^n T^{-i}\alpha\right) = H(\alpha) = -\sum p_i \log p_i.$$

The following theorem gives some insight into the meaning of entropy. For a partition α, let $\alpha^n := \bigvee_0^{n-1} T^{-i}\alpha$, and let $\alpha^n(x)$ denote the element of α^n containing x.

Theorem 1.2.5. (Shannon-Breiman-McMillan, [Bi]) *Let $T : (X, \mathcal{B}, \mu) \circlearrowleft$ be an ergodic mpt, and let α be a finite partition. Then*

$$\lim_{n\to\infty} -\frac{1}{n}\log \mu\alpha^n(x) = h(T;\alpha) \quad \textit{for a.e. } x.$$

This theorem has the following interpretation, which can almost be taken to be the definition of entropy. It says that for every ergodic *mpt*, there is a number h s.t. the following holds: $\forall \varepsilon > 0$, if α is a sufficiently fine partition, then $\exists\ N$ s.t. $\forall\, n \geq N, \exists$ a set $X_n \subset X$ with $\mu X_n > 1 - \varepsilon$ s.t. X_n consists of $\sim e^{n(h\pm\varepsilon)}$ elements of α^n each having measure $\sim e^{-n(h\pm\varepsilon)}$.

For continuous maps of metric spaces, it is often convenient to estimate entropy using balls rather than partitions. Let $B(x,\varepsilon)$ denote the ball of radius ε about x. Given $\rho : X \to \mathbb{R}^+$, let $B(x,\rho;n) := \{y \in X : d(T^i x, T^i y) < \rho(T^i x)\,\forall\, 0 \leq i < n\}$, and write $B(x,\varepsilon;n)$ if $\rho \equiv \varepsilon$. The following is a version of Theorem 1.2.5.

Theorem 1.2.6. ([M1], [BrK]) *Assuming that $T : (X, \mathcal{B}, \mu) \circlearrowleft$ is ergodic, we have for a.e. x,*

$$\sup_{\varepsilon>0}\ \limsup_{n\to\infty} -\frac{1}{n}\log \mu B(x,\varepsilon;n) = h(T)\,.$$

One could also replace $B(x,\varepsilon;n)$ by $B(x,\rho_\varepsilon;n)$ where $\{\rho_\varepsilon\}$ is a family of functions satisfying $0 < \rho_\varepsilon \leq \varepsilon$ and $\int -\log\rho_\varepsilon d\mu < \infty$.

1.3. Measurable partitions and conditional measures. In this subsection we assume for technical reasons that $(X, \mathcal{B}, \mu)$ is a Lebesgue space, i.e., it is isomorphic to an interval endowed with Lebesgue measure together with at most a countable number of atoms. Completed Borel measures on compact metric spaces are always Lebesgue.

Sometimes it is necessary to deal with partitions that are not finite or even countable. Let us call η a measurable partition of $(X, \mathcal{B}, \mu)$ if there is a countable sequence of finite partitions $\alpha_1, \alpha_2, \ldots$ s.t. $\eta = \bigvee \alpha_i$. Given a measurable partition η of $(X, \mathcal{B}, \mu)$, there is the notion of a *canonical system of conditional probability measures* associated with η. By this we refer to a family $\{\mu_x^\eta, x \in X\}$ with the following properties:

1) for each $x \in X$, μ_x^η is a probability measure on $(X, \mathcal{B})$; it is supported on $\eta(x)$, the element of η containing x;
2) for every $E \in \mathcal{B}, x \mapsto \mu_x^\eta(E)$ is measurable;
3) for $E \in \mathcal{B}, \mu(E) = \int \mu_x^\eta(E) d\mu(x)$.

The family $\{\mu_x^\eta\}$ is essentially unique, in the sense that any two families satisfying 1) – 3) above agree for a.e. x. For a proper treatment see [Ro1].

Example 1.3.1. Let $X = [0,1] \times [0,1]$, $\mu =$ any Borel probability measure on X, and $\eta(x,y) = \{x\} \times [0,1]$. Without worrying about technical details, let us try to understand what the measures $\mu^\eta_{(x,y)}$ look like. Let α_n be the partition of X into vertical strips of the form $\left[\frac{k}{2^n}, \ \frac{k+1}{2^n}\right) \times [0,1]$, and let $\mu^{(n)}_{(x,y)}$ be the usual conditional measures of μ wrt α_n, i.e. for $E \subset X$,

$$\mu^{(n)}_{(x,y)}(E) := \frac{\mu(E \cap \alpha_n(x,y))}{\mu\alpha_n(x,y)}$$

if $\mu\alpha_n(x,y) > 0$, anything you like otherwise. Since $\mu^{(n)}_{(x,y)}$ does not really depend on y, we will also call it $\mu^{(n)}_x$. Now $x \mapsto \mu^{(n)}_x(E), n = 1, 2, \ldots,$ is a martingale, so it converges a.e. as $n \to \infty$ to some function we call $\mu^{(\infty)}_x(E)$.

To define $\mu^\eta_{(x,y)}$ or μ^η_x on $\{x\} \times [0,1]$, it suffices to assign values to $\mu^\eta_x(F)$ for $F = F_{n,k} := \left[\frac{k}{2^n}, \ \frac{k+1}{2^n}\right), n = 1, 2, \ldots, 0 \le k < 2^n$, and to verify the usual compatibility conditions. This is easily done for a.e. x by letting $\mu^\eta_x(F_{n,k}) = \mu^\infty_x([0,1] \times F_{n,k})$.

Now it is possible to talk about entropy wrt measurable partitions that are not necessarily finite. The definition we gave for $h(T;\alpha)$ does not generalize, but an equivalent definition

$$h(T;\alpha) = \lim_n H\left(\alpha | \bigvee_1^n T^{-i}\alpha\right) = H\left(\alpha | \bigvee_1^\infty T^{-i}\alpha\right)$$

does if $H\left(\alpha | \bigvee_1^\infty T^{-i}\alpha\right)$ is interpreted to mean $\int -\log \mu^\eta_x(\alpha(x)) d\mu(x)$ where $\eta = \bigvee_1^\infty T^{-i}\alpha$. Indeed, $h(T)$ can be defined as $\sup h(T;\eta)$ with the supremum taken over all measurable partitions η. Much of our intuition for finite partitions carry over, but caution has to be exercised in a few places: for example, $\eta > \xi \not\Rightarrow h(T;\eta) \ge h(T;\xi)$, and $h(T;\eta)$ is not necessarily $h(T)$ when η is a generator (η could be the partition of X into points). See [Ro2] for a more detailed discussion.

2. Lyapunov Exponents

2.1. Oseledec's Theorem. Let $f : M \circlearrowleft$ be a differentiable map of a manifold M. The dynamics of f is determined at the linear level by $Df^n_x, n = 1, 2, \ldots$. To understand the geometry of Df^n_x, which we can think of as compositions of matrices, let us first understand the picture when a single linear map $A \in GL(m, \mathbb{R})$ is iterated. Let $v \in \mathbb{R}^m$. We are interested in the rate of growth of $|A^n v|$ as $n \to \infty$. It is easy to see, by writing A in its real Jordan form, for instance, that $\mathbb{R}^m$ can be decomposed into A-invariant subspaces $E_1 \oplus \cdots \oplus E_r$ in such a way that for each i, there is a number λ_i s.t. $\forall\, v \neq 0 \in E_i$,

$$\lim_{n\to\infty} \frac{1}{n} \log |A^{\pm n} v| = \pm \lambda_i \,.$$

The λ_i, of course, are the log of the moduli of the eigenvalues of A. Note that $\sum_i \lambda_i \cdot \dim E_i = \log |\det A|$.

Next we consider infinite sequences of matrices $\{A_i\}$ and ask about the growth rate of $|A_n \cdots A_0 v|$. It is easy to concoct sequences of $\{A_i\}$ for which these growth rates are not well defined. The theorem of Oseledec, which we will state shortly, is the matrix version of the Birkhoff Ergodic Theorem. It tells us that in the setting of ergodic theory, i.e. if f preserves a probability measure μ, then μ-a.e. the asymptotic behavior of Df^n_x is similar to that when a single linear map is iterated. We will state two versions of Oseledec's theorem, one for maps that are not necessarily invertible and the other one for maps that are. In fact, we will state Oseledec's theorem in its natural "cocycles" setting:

Let $T : (X, \mathcal{B}, \mu) \circlearrowleft$ be an *mpt* of a probability space, and let $A : X \to GL(m, \mathbb{R})$ be a measurable mapping. We assume that

$$\int \log^+ ||A|| d\mu < \infty \quad \text{and} \quad \int \log^+ ||A^{-1}|| d\mu < \infty \,.$$

(Here $\log^+ a = \max(\log a, 0)$.) For $n \geq 0$, we write $A^n(x) := A(T^{n-1}x) \cdots A(x)$ and $A^{-n}(x) = A(T^{-n}x)^{-1} \cdots A(T^{-1}x)^{-1}$. For $x \in X$ and $v \in \mathbb{R}^m$, we define

$$\overline{\lambda}_+(x, v) := \overline{\lim_{n\to\infty}} \frac{1}{n} \log |A^n(x) v| \,,$$
$$\underline{\lambda}_+(x, v) := \underline{\lim}_{n\to\infty} \frac{1}{n} \log |A^n(x) v| \,.$$

We write λ_+ when it is understood that $\overline{\lambda}_+ = \underline{\lambda}_+$. The numbers $\overline{\lambda}_-, \underline{\lambda}_-$ and λ_- are defined analogously when $A^n(x)$ is replaced by $A^{-n}(x)$.

Theorem 2.1.1. (Oseledec's Theorem: non-invertible version [O]) *Let $(T, \mu; A)$ be as above. Then at μ-a.e. x, $\exists$ a filtration of subspaces*

$$\{0\} = V_0(x) \subsetneqq V_1(x) \subsetneqq \cdots \subsetneqq V_{r(x)}(x) = \mathbb{R}^m$$

and numbers $\lambda_1(x) < \cdots < \lambda_{r(x)}(x)$ s.t.

(1) $\forall\, v \in V_i(x) - V_{i-1}(x), \quad \lambda_+(x, v) = \lambda_i(x)$*;*

(2) $\lim_{n\to\infty} \frac{1}{n} \log|\det A^n(x)| = \sum \lambda_i(x) \cdot [\dim V_i(x) - \dim V_{i-1}(x)]$.

The functions $x \mapsto r(x), \lambda_i(x)$ and $V_i(x)$ are measurable.

The functions $\lambda_i(x)$ together with their multiplicities $m_i(x) := \dim V_i(x) - \dim V_{i-1}(x)$ are called the *Lyapunov exponents* of $(T, \mu; A)$. It follows from our characterization of $V_i(x)$ that $A(x)V_i(x) = V_i(Tx)$, so that the functions $x \mapsto r(x), \lambda_i(x)$ and $\dim V_i(x)$ (but not $V_i(x)$!) are constant along orbits. Thus if (T, μ) is ergodic, then the Lyapunov spectrum of $(T, \mu; A)$ consists of a finite set of numbers $\lambda_1, \cdots, \lambda_r$ with multiplicities $m_1, \cdots, m_r$ respectively.

Oseledec's Theorem also gives information about growth in volume of higher dimensional parallelepipeds. For a linear subspace $W \subset \mathbb{R}^m$ define $\lambda_+(x, W) := \lim_{n\to\infty} \frac{1}{n} \log|\det(A^n(x)|W)|$. We claim that at a.e. x, for a "typical" k-dimensional subspace $W \subset \mathbb{R}^m$, $\lambda_+(x, W) = \sum_{i=m-k+1}^{m} \chi_i$ where $\chi_1 \le \chi_2 \le \cdots \le \chi_m$ is $\lambda_1(x) < \cdots < \lambda_{r(x)}(x)$ with λ_i counted m_i times. To see this, choose a basis $\{v_1, \ldots, v_m\}$ of $\mathbb{R}^m$ s.t. $\lambda_+(x, v_i) = \chi_i$ for all i and $v_{m-k+1}, \ldots, v_m \in W$. This is possible if W is in general position with respect to the $V_i(x)$. The second assertion in Theorem 2.1.1 then tells us that $\sum_{\text{all } i} \chi_i = \lambda_+(x, \mathbb{R}^m) \le \lambda_+(x, W) + \sum_{i=1}^{m-k} \chi_i$, so that $\lambda_+(x, W) \ge \sum_{i=m-k+1}^{m} \chi_i$. The reverse inequality is obvious.

If T is invertible, then as we iterate backwards, we obtain a filtration similar to that in the last theorem. By intersecting these two filtrations (and doing some work) one arrives at the following:

Theorem 2.1.2. (Oseledec's Theorem: invertible version [O]) *Let $(T, \mu; A)$ be as above, and assume that T is invertible. Then at μ-a.e. x, there exist numbers $\lambda_1(x) < \cdots < \lambda_{r(x)}(x)$ and a decomposition of $\mathbb{R}^m$ into*

$$\mathbb{R}^m = E_1(x) \oplus \cdots \oplus E_{r(x)}(x)$$

s.t.

(1) $\forall\, v \in E_i(x), \lambda_+(x, v) = -\lambda_-(x, v) = \lambda_i(x)$;
(2) *for* $j \ne k$, $\lim_{n\to\infty} \frac{1}{n} \log|\sin \sphericalangle(A^{\pm n} E_j(x), A^{\pm n} E_k(x))| = 0$.

The functions $x \mapsto r(x), \lambda_i(x)$ and $E_i(x)$ are measurable.

Letting $A(x) = Df_x$ Theorems 2.1.1 and 2.1.2 are easily adapted to apply to differentiable maps preserving Borel probability measures. We emphasize once again that in the case of diffeomorphisms, by iterating forwards alone one cannot pick out all the invariant subspaces $E_i(x)$. For example, if $\dim M = 2, \lambda_1(x) < 0$ and $\lambda_2(x) > 0$, then by iterating forwards alone one recognizes $E_1(x)$, but all the vectors in $T_xM - E_1(x)$ grow like $\sim e^{\lambda_2 n}$ and are indistinguishable from one another.

2.2. Examples.

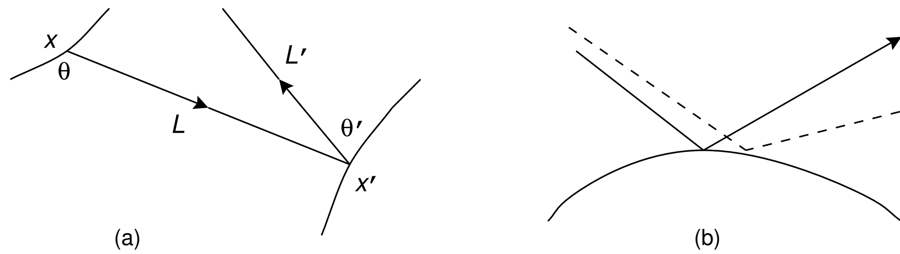
Example 2.2.1. Accepting that (T,μ) in Example 1.1.2 is ergodic, we have $\lambda = \int \log |DT| d\mu = \log 2$.

Example 2.2.2. (a) *Positive matrices.* Let $(T,\mu;A)$ be in the setting of Oseledec's Theorem, and suppose that for a.e. x, $A(x) \in SL(2,\mathbb{R})$ has strictly positive entries. We claim that $\lambda_1 < 0 < \lambda_2$. One way to see this is to consider $\begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and observe that if $a,b,c,d,v_1,v_2 > 0$ and $ad-bc=1$, then $v_1'v_2' > (1+2bc)v_1v_2$. Let $A(T^ix) = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$. Then for $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ with $v_1, v_2 > 0$, the product of the coordinates of $A^n(x)v$ is $> \left[\prod_{i=0}^{n-1} (1+2b_ic_i) \right] v_1v_2$. By the Birkhoff Ergodic Theorem, this number grows exponentially for a.e. x. Hence $|A^n(x)v|$ grows exponentially for a.e. x.

The classical theory of Perron and Frobenius tells us that when a positive matrix is iterated, every vector in the positive quadrant grows like its largest eigenvalue. Example 2.2.2(a), which in fact works in all dimensions, is a nonuniform or ergodic theory generalization of this result. It is due to Wojtkowski [W1], who had in mind applications to dynamical systems. We describe below some simple versions of these applications.

(b) *Invariant cones.* The setting is as in part (a), except that the positivity of $A(x)$ is replaced by the following condition. Let $\{u_1(x), u_2(x)\}$ be two measurable families of vectors so that $\int \log|\sin \sphericalangle(u_1,u_2)| d\mu > -\infty$. Without loss of generality, we may assume that u_1 and u_2 span a parallelogram of unit area. These vectors determine at each x a "cone" $C(x) := \{v \in \mathbb{R}^2 : v = a_1u_1 + a_2u_2, a_1a_2 \geq 0\}$. Our "invariant cones condition" here says that at a.e. x, $A(x)$ maps $C(x)$ into the interior of $C(Tx)$. Let $U(x)$ be the linear map that takes $\{u_1(x), u_2(x)\}$ to the usual basis of $\mathbb{R}^2$, and let $B(x) = U(Tx)A(x)U(x)^{-1}$. Since $\sphericalangle(u_1(T^nx), u_2(T^nx))$ does not tend to zero exponentially (this follows from the general fact that $\frac{1}{n}\varphi \circ T^n \to 0$ a.e. for $\varphi \in L^1$), we obtain that $(T,\mu;A)$ and $(T,\mu;B)$ have the same Lyapunov exponents.

(c) *Billiards.* Let $\Omega \subset \mathbf{R}^2$ be a compact region with piecewise smooth boundaries. Consider the flow generated by the uniform motion of a point mass in Ω with elastic reflections at $\partial\Omega$. We give $\partial\Omega$ an orientation, and let S be the set of unit vectors at $\partial\Omega$ pointing into the interior of Ω. Then S can be identified with $\partial\Omega \times [0,\pi]$. For $(x,\theta) \in S$, define $(x',\theta') = f(x,\theta)$, as follows: Let L be the directed line segment starting at $x \in \partial\Omega$, pointing into the interior of Ω and making an angle of θ with $\partial\Omega$. Then x' is the first point of contact of L with $\partial\Omega$, and if L' is the reflected line segment at x', then θ' is the angle L' makes with $\partial\Omega$. (See Fig. (a).) It is easy to see that f preserves the measure $\sin\theta dx d\theta$.



Billiards as dynamical systems were first studied by Sinai in [S2], where he proved the ergodicity of dispersing billiards, i.e. billiards in regions Ω whose boundary curves are concave (as in Fig. (b)). In the course of proving ergodicity, Sinai also proved that there is a positive Lyapunov exponent a.e.

This is not how it is done in [S2], but one way to study the Lyapunov exponents of billiards is to look for geometric conditions on $\partial\Omega$ that will guarantee natural families of invariant cones (see e.g. [W2]). It is easy to see intuitively what such a cone family might be for dispersing billiards. First observe that a tangent direction at (x, θ) can be represented by a curve in S through (x, θ), which in turn can be thought of a smooth family of rays $\{L(s), |s| < \epsilon\}$, with $L(0) = L_{(x,\theta)}$, the oriented line through x with direction θ. Let $C(x, \theta)$, the cone at (x, θ), be the set of all diverging rays. Since diverging rays diverge even more after bouncing off a concave piece of $\partial\Omega$ (Fig. (b)), this cone family is invariant and we conclude positive exponents.

For higher dimensional examples see e.g. [BB] and [W3]. See also [BuK].

We mention also that invariant cones play an important role in uniformly hyperbolic systems, conservative or non-conservative. We have taken our matrices in Example 2.2.2 to be in SL because the formulation in general is slightly different. In the absence of invariant cones, it is usually hard to estimate Lyapunov exponents. The following is a well known open problem:

Open problem. Consider the *standard map* $F_k : \mathbb{T}^2 \circlearrowleft$ defined by

$$F_k(x, y) = (-y + 2x + k \sin(2\pi x), x) \mod \mathbb{Z}^2$$

where k is a real parameter. Note that F_k is area preserving, and that for k large, F_k has an enormous amount of stretching – of order $\frac{k}{2\pi}$ – on all but a small part of the torus. The prevailing conjecture is that for a positive measure set of k, F_k has a positive exponent on a positive measure set. This has not been shown to be true (or false) for a single k.

The next example does not come from a diffeomorphism, but we hope that the action of these random matrices will give some insight into the behavior of $Df_x^n v$ in the absence of invariant cones.

Example 2.2.3. Suppose that we are equipped with a coin and a roulette wheel. We flip our coin, and write down a sequence of matrices as follows: If the i^{th} flip is a head(H), let $A_i = \begin{pmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{pmatrix}$ for some $\mu > 1$. If it is a tail(T), then we turn our roulette wheel, and if the marker stops at the angle θ, then we let $A_i = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Suppose that prob(H) $= p$, prob(T) $= 1-p$. For simplicity let us assume that the stopping places of the marker has uniform distribution on $[0, 2\pi]$ independent of where it started. We ask if the matrices so obtained have a positive Lyapunov exponent. (It is an exercise to verify that Oseledec's Theorem applies in this situation, and that λ_1 and λ_2 are nonrandom, i.e. they are the same two numbers for a.e. sequence of flips and turns.)

Let $A_0, A_1, \dots$ be the matrices corresponding to one trial. We consider $v_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and let $v_n = A^n v_0$. Let $n_1 < n_2 < \cdots$ be the times when T occurs and let θ_i be the angle v_{n_i+1} makes with the positive x-axis. (Remember that under our notation v_{n_i+1} is the result immediately after the rotation matrix A_{n_i} is applied.) Then for $0 \le n \le n_1, |v_n| = \mu^n$, and the component of v_{n_1+1} in the horizontal direction is $|v_{n_1}| \cdot |\cos\theta_1|$. This implies that for $n_1 < n \le n_2, |v_n| \ge \mu^{n-n_1-1}|\cos\theta_1|\,|v_{n_1}|$, and similarly, for $n_i < n \le n_{i+1}, |v_n| \ge \mu^{n-n_i-1}|\cos\theta_i|\,|v_{n_i}|$. So in general we have

$$\frac{1}{n}\log|v_n| \ge \frac{1}{n}\log(\mu^{n-\#T's \text{ in first } n \text{ flips}} \cdot \prod_{n_i<n}|\cos\theta_i|)\,.$$

Now $\theta_1, \theta_2, \dots$ are independent random variables, so the ergodic theorem or the law of large numbers tells us that for a typical sequence of flips and turns,

$$\lim_{n\to\infty}\frac{1}{n}\log|v_n| \ge p\log\mu + (1-p)E(\log|\cos\theta|)\,,$$

which is clearly > 0 for μ sufficiently large or p sufficiently near 1.

We remark that in Example 2.2.3 it is in fact true that the larger Lyapunov exponent is positive for any $\mu > 1$ and $p > 0$. This follows from a theorem of Furstenberg [Fu] which says that for *iid* matrices in $SL(2,\mathbb{R})$, it is very hard for both exponents to be zero. For conditions guaranteeing distinct exponents in n-dimension, see [GR] and [GM].

2.3. Proof of Oseledec's Theorem in dimension 2. There now exist more than one proof of Oseledec's Theorem. See e.g. [M3] and [Ru4]. Our exposition here follows [Ru4]. See also [L2]. There are two ingredients in this proof: one is the subadditive ergodic theorem, and the other is the geometry of matrices.

Theorem 2.3.1. (Subadditive ergodic theorem [Ki]) *Let $T : (X, \mathcal{B}, \mu) \circlearrowleft$ be an mpt. For $n = 1, 2, \dots$ let $\varphi_n : X \to \mathbb{R}$ be s.t.*

1. $\int \varphi_1^+ d\mu < \infty$,
2. $\varphi_{m+n} \le \varphi_m + \varphi_n \circ T^m$ *a.e. for* $m, n \ge 1$.

Then there exists $\varphi^* : X \to \mathbb{R} \cup \{-\infty\}$ *with* $\varphi^* \circ T = \varphi^*$ *a.e. and* $\int (\varphi^*)^+ d\mu < \infty$ *s.t.*

$$\frac{1}{n}\varphi_n \to \varphi^* \qquad a.e.$$

Clearly, if "$\le$" is hypothesis 2 is replaced by "$=$", then Theorem 2.3.1 is simply the Birkhoff Ergodic Theorem with $\varphi_n = \sum_{i=0}^{n-1} \psi \circ T^i$ for some given $\psi : X \to \mathbb{R}$. To prove Oseledec's Theorem, we apply Theorem 2.3.1 to

$$\varphi_n^{(k)}(x) = \log \left[\sup_{\substack{W \subset \mathbb{R}^m \\ k-\text{dim subsp}}} |\det(A^n(x)|W)| \right], n = 1, 2, \ldots .$$

It is easy to verify that $\varphi_n^{(k)}$ is subadditive for $k < m$, additive for $k = m$.

We will not get into the proof of the subadditive ergodic theorem here, but will discuss the geometric part of the proof in dimension 2 in some detail. To pinpoint the issues, we divide the argument into the following paragraphs:

(1) *Preliminaries from linear algebra.* Every $A \in GL(2, \mathbb{R})$ can be written as $A = O_2 D O_1$, where O_1, O_2 are orthogonal and D is diagonal. We will assume that D is always in the form $D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$ where $0 < d_1 \le d_2$. The numbers d_1 and d_2 are called the *singular values* of A. If $d_1 < d_2$, then up to multiplication by ± 1, A has two well defined unit vectors u_1 and u_2 corresponding to its directions of minimum and maximum stretch. More precisely, $|Au_1| = d_1, |Au_2| = d_2$. Note that $u_1 \perp u_2$ and $Au_1 \perp Au_2$.

The goal of the next 3 paragraphs is to find V_1, the first nontrivial invariant subspace in the filtration in Theorem 2.1.1. In dimension 2, if the two Lyapunov exponents are equal, then $V_1 = \mathbb{R}^2$ and there is nothing to do. If not, then V_1 consists exactly of those vectors that grow like the smaller of the two exponents.

(2) *Growth of singular values of* A^n. For a typical x, let $d_1^{(n)}(x) \le d_2^{(n)}(x)$ denote the singular values of $A^n(x) := A(T^{n-1}x) \cdots A(x)$. We claim that

* $(d_1^{(n)})^{\frac{1}{n}} \to$ some d_1,
* $(d_2^{(n)})^{\frac{1}{n}} \to$ some d_2.

This is true because $\frac{1}{n} \log d_2^{(n)} = \frac{1}{n} \log ||A^n||$ converges by the subadditive ergodic theorem, and $\frac{1}{n} \log d_1^{(n)} d_2^{(n)} = \frac{1}{n} \log |\det A^n|$ converges by the usual ergodic theorem.

If $d_1 = d_2$, then we are done, so we assume from here on that $d_1 < d_2$.

(3) *Convergence of directions of minimum stretch.* Again fix a typical x, and let $u_1^{(n)}$ and $u_2^{(n)}$ be respectively the directions that are stretched the least and the most by $A^n(x)$. We will show that

* $u_1^{(n)} \to$ some u_1 as $n \to \infty$;
* in fact, $\sphericalangle(u_1^{(n)}, u_1) \leq$ const $\cdot(1+\epsilon)^n \left(\frac{d_1}{d_2}\right)^n$.

Proof: Write $u_1^{(n)} = v_1 + v_2$ wrt the splitting $u_1^{(n+1)} \oplus u_2^{(n+1)}$, and let $\theta_n = \sphericalangle(u_1^{(n)}, u_1^{(n+1)})$. On the one hand we have

$$|A^{n+1}v_2| \approx |\sin\theta_n| \cdot d_2^n \approx |\theta_n| \cdot d_2^n;$$

on the other we have

$$|A^{n+1}v_2| \leq |A(T^n x)A^n(x)u_1^{(n)}| \lesssim ||A(T^n x)|| \cdot d_1^n \, .$$

So

$$|\theta_n| \lesssim ||A(T^n x)|| \cdot \left(\frac{d_1}{d_2}\right)^n \leq (1+\epsilon)^n \cdot \left(\frac{d_1}{d_2}\right)^n$$

(again because $\frac{1}{n}\varphi \circ T^n \to 0$ a.e. for every $\varphi \in L^1$). Hence $u_1^{(n)} \to$ some u_1 and $\sphericalangle(u_1^{(n)}, u_1) \leq \sum\limits_{i \geq n} |\theta_i| \leq$ const $\cdot(1+\varepsilon)^n \left(\frac{d_1}{d_2}\right)^n$.

(4) *Invariant subspaces.* Let $V_1(x)$ be the subspace generated by u_1, where u_1 is the limit vector in the last paragraph. We need to check that

* $\frac{1}{n}\log|A^n u_1| \to \log d_1$,
* $\forall\, v \notin V_1$, $\frac{1}{n}\log|A^n v| \to \log d_2$.

Write $u_1 = w_1 + w_2$ wrt $u_1^{(n)} \oplus u_2^{(n)}$. Then $|A^n w_1| \approx d_1^n$ and $|A^n w_2| \lesssim \left(\frac{d_1}{d_2}\right)^n \cdot d_2^n \approx d_1^n$. Since $A^n w_1 \perp A^n w_2$, we have $|A^n u_1| \approx d_1^n$, proving the first assertion. For $v \notin V_1$ write $v = v_1 + v_2$ wrt $u_1^{(n)} \oplus u_2^{(n)}$. Then $|v_2| \geq$ some $c > 0$ for all sufficiently large n. With $|A^n v_1| \leq d_1^n$ while $|A^n v_2| \geq c d_2^n$, our conclusion follows.

(5) *The invertible case.* Let $\tilde{d}_1, \tilde{d}_2, \tilde{u}_1^{(n)}, \tilde{u}_2^{(n)}$ etc. denote the corresponding objects when T is iterated backwards. First observe that $d_1 = \tilde{d}_2^{-1}, d_2 = \tilde{d}_1^{-1}$. To see this, let $\Lambda \subset X$ be a set of nearly full measure s.t. on Λ the singular values of A^n and A^{-n} have essentially reached their correct limits. For $x \in \Lambda \cap T^{-n}\Lambda$, compare $A^n(x)$ and $A^{-n}(T^n x)$.

Now let $E_1(x) = V_1(x)$ be as in the last paragraph, and let $E_2(x) = \tilde{V}_1(x)$. If $E_1(x) = E_2(x)$, then we would have $E_1(T^n x) = E_2(T^n x)$ as well. From this we would deduce that $\forall\, v \in E_1(x)$,

$$|v| = |A^{-n}(T^n x)A^n(x)v| \sim \left(\frac{d_1}{d_2}\right)^n |v|\,,$$

which is absurd.

3. Nonuniformly Hyperbolic Systems: Local Theory

Let $f : M \circlearrowleft$ be a diffeomorphism of a Riemannian manifold preserving a Borel probability measure μ. Let $\{\lambda_i\}$ denote the Lyapunov exponents of (f, μ), and let $\oplus E_i$ denote the corresponding splitting in the tangent spaces. The goals of this section are:

(1) to introduce coordinate changes so that at μ-a.e. x, Df_x reflects the Lyapunov exponents at x in *one* iterate, i.e. $||Df_x v|| \sim e^{\lambda_i(x)}||v|| \ \forall\, v \in E_i(x)$;
(2) to transfer the Oseledec splitting at the linear level to stable and unstable manifolds of the nonlinear map in neighborhoods of typical trajectories.

Both of these objectives will be accomplished, but as we shall see, it is at the expense of something else. Nevertheless, coordinate systems of this type – sometimes called *Lyapunov charts* – are very useful to work in, because they allow us to think of a complicated nonlinear map as locally very simple. These charts are particularly useful for proving relations involving Lyapunov exponents, for through them we see f as essentially a linear map with eigenvalues e^{λ_i}.

The material in Section 3.1 and most of Section 3.3 is due to Pesin. It is part of his work ([P1], [P2]) on the ergodic theory of diffeomorphisms preserving a smooth measure – although for this part the smoothness of μ is irrelevant. Ruelle first published these results for arbitrary Borel probability measures [Ru4]. Our exposition follows [P1], but not completely. We will confine our discussion to diffeomorphisms, mentioning only briefly in the last subsection what one could do for endomorphisms and maps with singularities. For the reader primarily interested in maps in these categories, the reference is [KaS].

3.1. Lyapunov charts. Throughout this subsection, let $f : M \circlearrowleft$ be a C^2 diffeomorphism of a compact Riemannian manifold, and let μ be an f-invariant Borel probability measure. For simplicity let us deal only with the case where (f, μ) is ergodic. Let $\lambda_1 > \cdots > \lambda_r$ be the Lyapunov exponents of (f, μ), and let

$T_xM = \oplus E_i(x)$ be the corresponding splitting. (Notice that we have written our Lyapunov exponents in descending order. This is a switch from Section 2 but will be more convenient for future purposes.) Since we will be making rather large changes of coordinates, let us be a little pedantic with notation: let $\langle\langle\cdot,\cdot\rangle\rangle_x$ denote the inner product on T_xM associated with the given Riemannian metric, while $\langle\cdot,\cdot\rangle$ is reserved for the usual inner product on Euclidian space. The norms corresponding to $\langle\langle\cdot,\cdot\rangle\rangle$ and $\langle\cdot,\cdot\rangle$ are denoted by $||\cdot||$ and $|\cdot|$ respectively. Distances between points on M given by the Riemannian metric are denoted by $d(\cdot,\cdot)$.

Let $\Lambda \subset M$ be the set of points satisfying the conclusions of Theorem 2.1.2. We fix $\varepsilon > 0$ s.t. $\varepsilon \ll \min_{i\neq j}|\lambda_i - \lambda_j|$ and introduce a new inner product $\langle\langle\cdot,\cdot\rangle\rangle'_x$ on T_xM by specifying it for the following pairs of vectors:

* for $u, v \in E_i(x)$, let

$$\langle\langle u,v\rangle\rangle'_x := \sum_{-\infty}^{\infty} \frac{\langle\langle Df_x^n u, Df_x^n v\rangle\rangle_{f^n x}}{e^{2n\lambda_i + 2|n|\varepsilon}};$$

* for $u \in E_i(x), v \in E_j(x), i \neq j$,

$$\langle\langle u,v\rangle\rangle'_x = 0\,.$$

It is straightforward to verify that the infinite sum above converges, and that for $v \in E_i(x)$, we have

$$e^{\lambda_i - \varepsilon}||v||'_x \leq ||Df_x v||'_{fx} \leq e^{\lambda_i+\varepsilon}||v||'_x\,.$$

We have thus achieved the goal of making Lyapunov exponents show up in one iterate, at least at the tangent space level.

Next we introduce a family of charts $\{\Phi_x, x \in \Lambda\}$, where each Φ_x is a diffeomorphism between a small neighborhood of 0 in $\mathbb{R}^m$ and a neighborhood of x in M. As usual, let $m_i = \dim E_i$. We write $\mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_r}$ and let $L_x : T_xM \to \mathbb{R}^m$ be a linear map taking $E_i(x)$ to $\{0\} \times \cdots \times \mathbb{R}^{m_i} \times \cdots \times \{0\}$ in such a way that L_x carries $\langle\langle\cdot,\cdot\rangle\rangle'_x$ to $\langle\cdot,\cdot\rangle$. Our chart Φ_x is then defined to be the restriction of $\exp_x \circ L_x^{-1}$ to a small neighborhood of 0, where $\exp_x := T_xM \to M$ is the exponential map. The size of this neighborhood will be specified later.

Let us now compare distances in our chart to actual distances on the manifold. We may assume that the ranges of Φ_x are uniformly small for all x, so that $\exp_x$ in the relevant domains are uniformly near the identity. The problem then boils down to comparing $||v||_x$ and $|L_x v|$, which is $= ||v||'_x$. It follows from our definition that $||v||_x \leq ||v||'_x \,\forall\, v \in E_i(x)$. For arbitrary $v \in T_xM$, writing $v = \oplus v_i \in \oplus E_i(x)$, we have $||v||_x \leq \sum ||v_i||_x \leq r||v||'_x$. To prove the opposite inequality, let us introduce the functions

$$\alpha(x) := \max_{i=1,\ldots,r} \sup_{\substack{v\in E_i(x)\\ ||v||_x=1}} \sup_{n\in\mathbb{Z}} \frac{||Df_x^n v||_{f^n x}}{e^{\lambda_i n + \frac{\varepsilon}{2}|n|}}$$

and

$$\beta(x) := \sup_{\substack{(v_1,\dots,v_r) \\ v_i \in E_i(x)}} \frac{||v_1||_x \cdots ||v_r||_x}{||v_1 \wedge \cdots \wedge v_r||_x}$$

where $||v_1 \wedge \cdots \wedge v_r||_x$ denotes the r-dimensional volume of the parallelepiped whose sides are $v_1, \dots, v_r$. It follows from our definition of $\langle\langle \cdot, \cdot \rangle\rangle'_x$ that $\forall\, v \in E_i(x), ||v||'_x \leq C\alpha(x)||v||_x$ for some constant C. Let $v = \oplus v_i \in \oplus E_i(x)$. Then $||v||'_x \leq \Sigma ||v_i||'_x \leq rC\alpha(x)||v_j||_x$ for some j. Now $||v||_x \geq ||v_j||_x \cdot |\sin\theta|$ where θ is the angle between v_j and the subspace spanned by $v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_r$. Hence $||v||'_x \leq rC\alpha(x)\beta(x)||v||_x$.

Let $\tilde{f}_x$ be the connecting map between the chart at x and the chart at fx, i.e. $\tilde{f}_x = \Phi_{fx}^{-1} \circ f \circ \Phi_x = L_{fx} \circ \exp_{fx}^{-1} \circ f \circ \exp_x \circ L_x^{-1}$, defined wherever it makes sense. Since the second derivatives of $\exp, \exp^{-1}$ and f are uniformly bounded in x, the Lipschitz constant of $D\tilde{f}_x$, which we denote by $Lip(D\tilde{f}_x)$, is essentially determined by the Lipshitz constant of L_{fx}. Let $\ell(x) := C_0\alpha(x)\beta(x)$ where C_0 is chosen large enough that $||v||'_x \leq \ell(x)||v||_x$ and $Lip(D\tilde{f}_x) \leq \ell(x)$. Then if we choose the domain of Φ_x to be $R(\varepsilon\ell(x)^{-1})$, the ball of radius $\varepsilon\ell(x)^{-1}$ about 0, we would be guaranteed that $\forall\, z \in R(\varepsilon\ell(x)^{-1}), |D\tilde{f}_x(z) - D\tilde{f}_x(0)| < \varepsilon$.

Let us summarize the results of our discussion in the following thoerem:

Theorem 3.1.1. *Let Λ be the set of points satisfying the conclusions of Oseledec's Theorem, and let $\varepsilon \ll \min\limits_{i \neq j} |\lambda_i - \lambda_j|$ be fixed. Then there is a measurable function $\ell : \Lambda \to [1, \infty)$ and a family of charts $\{\Phi_x : R(\varepsilon\ell(x)^{-1}) \to M\}_{x \in \Lambda}$ with the following properties:*

i) * $\Phi_x(0) = x$;
   * $D\Phi_x(\{0\} \times \cdots \times \mathbb{R}^{m_i} \times \cdots \times \{0\}) = E_i(x)$;
   * $\forall\, z, z' \in R(\varepsilon\ell(x)^{-1})$,

$$K^{-1} d(\Phi_x z, \Phi_x z') \leq |z - z'| \leq \ell(x) d(\Phi_x z, \Phi_x z')$$

   where K is a constant that depends only on the dimension of M.

ii) *Let $\tilde{f}_x = \Phi_{fx}^{-1} \circ f \circ \Phi_x$, defined wherever it makes sense. Then*
   * $e^{\lambda_i - \varepsilon}|v| \leq |D\tilde{f}_x(0)v| \leq e^{\lambda_i + \varepsilon}|v| \;\forall\, v \in \{0\} \times \cdots \times \mathbb{R}^{m_i} \times \cdots \times \{0\}$;
   * $Lip(\tilde{f}_x - D\tilde{f}_x(0)), Lip(\tilde{f}_x^{-1} - D\tilde{f}_x^{-1}(0)) < \varepsilon$;
   * $Lip(D\tilde{f}_x), Lip(D\tilde{f}_x^{-1}) \leq \ell(x)$.

For purposes of proving the existence of stable and unstable manifolds, however, our charts have a defect: the sizes of their domains vary too irregularly along orbits. Let us postpone explaining what exactly we mean by this or why it may cause a problem till after Section 3.2, but accept for now that it would be useful to have the following property:

Addendum to Theorem 3.1.1. *The function ℓ may be chosen to satisfy $\ell(f^{\pm}x)/\ell(x) < e^{2\varepsilon}$.*

Here is a trick that can be used to turn α and β into functions that vary slowly along orbits. Let

$$\alpha_1(x) := \sup_{n\in\mathbb{Z}} \alpha(f^n x)e^{-\varepsilon|n|}\,,$$
$$\beta_1(x) := \sup_{n\in\mathbb{Z}} \beta(f^n x)e^{-\varepsilon|n|}\,.$$

Then we automatically have $\alpha_1(f^{\pm}x)/\alpha_1(x) < e^{\varepsilon}$ and $\beta_1(f^{\pm}x)/\beta_1(x) < e^{\varepsilon}$ *provided that* $\alpha_1, \beta_1 < \infty$ a.e. Assuming this last proviso checks out, replacing α and β by α_1 and β_1 in the definition of ℓ gives the property in the Addendum.

That $\beta_1 < \infty$ a.e. follows from Theorem 2.1.2(2), which implies in particular that $\lim\limits_{n\to\pm\infty} \frac{1}{|n|}\log\beta(f^n x) = 0$. To prove the a.e. finiteness of α_1, it is convenient to first modify α by letting $\alpha(x)$ be the maximum of the old $\alpha(x)$ and

$$\alpha'(x) := \max_{i=1,\dots,r}\ \sup_{\substack{v\in E_i(x)\\ ||v||_x=1}}\ \sup_{n\in\mathbb{Z}}\ \frac{e^{n\lambda_i - \frac{\varepsilon}{2}|n|}}{||Df_x^n v||_{f^n x}}\,.$$

Using this new definition $\alpha(x)$, we have $\forall\, v \in E_i(x)$ and $\forall\, n$,

$$(\alpha(x)e^{\frac{\varepsilon}{2}|n|})^{-1}e^{\lambda_i n}||v||_x \le ||Df_x^n v||_{f^n x} \le \alpha(x)e^{\frac{\varepsilon}{2}|n|}e^{\lambda_i n}||v||_x\,,$$

from which one verifies easily that $\alpha(f^m x) \le \alpha(x)^2 e^{\varepsilon|m|}\ \ \forall\, m \in \mathbb{Z}$. This proves $\alpha_1(x) < \infty$.

To summarize again: we have constructed a family of point dependent coordinate changes $\{\Phi_x\}$. These coordinate changes are chosen so that if $\tilde{f}_x = \Phi_{fx}^{-1}\circ f\circ\Phi_x$ and $\Phi_x(0) = x$, then $D\tilde{f}_x(0)$ reflects the Lyapunov exponents of f. We pay for this, however, by allowing Φ_x to distort the Riemannian metric on M by arbitrarily large amounts. This arbitrarily large distortion also causes $D^2\tilde{f}_x$ to be large, and to keep $\tilde{f}_x$ on the entire chart to be C^1 near $D\tilde{f}_x(0)$ we are forced to restrict our chart to an arbitrarily small neighborhood of x. We have arranged for all these "arbitrarily large" and "arbitrarily small" quantities to be controlled by a single function ℓ, which can in fact be arranged to fluctuate slowly along orbits. Later on we will see that this slow variation of ℓ is hardly noticeable in arguments involving exponential estimates.

For each $n \in \mathbb{Z}^+$, let $\Lambda_n := \{x \in \Lambda : \ell(x) \le n\}$. One way to view a nonuniformly hyperbolic system is to see M as, up to a set of measure zero, the increasing union $\bigcup\limits_{n=1}^{\infty}\Lambda_n$, where the Λ_n are sets on which we have uniform estimates. Indeed, on each $\Lambda_n, x \mapsto E_i(x)$ varies continuously, and the splitting $\oplus E_i$ can be extended to the closure of Λ_n with the same uniform estimates. (To prove the continuity of $x \mapsto E_i(x)$, prove the continuity of $x \mapsto \bigoplus_{j \le i} E_j(x)$ and $x \mapsto \bigoplus_{j \ge i} E_j(x)$.) If for instance $\lambda_i \neq 0 \, \forall \, i$, then for large n, $\bar{\Lambda}_n$ can be seen as a sort of approximation to our nonuniformly hyperbolic system by uniformly hyperbolic sets. The sets $\bar{\Lambda}_n$, however, are usually not f-invariant. For approximations of nonuniformly hyperbolic systems by uniformly hyperbolic f-invariant sets, see [Ka].

3.2. Stable manifold theorem for a fixed point. The purpose of this subsection is to recall the graph transform proof of unstable manifolds for hyperbolic fixed points. We give only the basic geometric ideas leading to the existence of local unstable manifolds of Lipschitz class. For complete proofs see [HP] or [Sh].

The setting and notations of this subsection are as follows. Let $\mathbb{R}^m = R^u \oplus R^s$, where $R^u = \mathbb{R}^k \times \{0\}$ and $R^s = \{0\} \times \mathbb{R}^{m-k}$ for some $0 < k < m$. The coordinates of $x \in \mathbb{R}^m$ with respect to the splitting $R^u \oplus R^s$ are denoted by (x^u, x^s). On $\mathbb{R}^m$, it will be convenient to use the norm $|x| := \max\{|x^u|, |x^s|\}$, when $|x^u|$ and $|x^s|$ are the usual Euclidean norms. Given a linear map T, we define its *norm* $||T||$ and *minimum norm* $m(T)$ to be

$$||T|| := \sup_{x \neq 0} \frac{|Tx|}{|x|}, \quad m(T) := \inf_{x \neq 0} \frac{|Tx|}{|x|} \, .$$

For $r > 0, R(r)$ refers to $R^u(r) \times R^s(r)$, where $R^u(r) := \{x \in R^u := |x| \le r\}$.

Standing hypothesis. We assume throughout Section 3.2 that $f : \mathbb{R}^m \mapsto \mathbb{R}^m$ is a C^1 diffeomorphism with $f(0) = 0$ such that when restricted to some $R(r)$, it is a Lipschitz small perturbation of a hyperbolic linear map T. More precisely, we assume that

(a) $T : \mathbb{R}^m \mapsto \mathbb{R}^m$ is a linear map leaving invariant R^u and R^s; and that if $T^u = T|R^u, T^s = T|R^s$, then $m(T^u) > 1, ||T^s|| < 1$;
(b) when restricted to $R(r), Lip(f - T) < \epsilon$.

Estimate #1. (Effects of nonlinearity) Let $x, y \in R(r)$. Then
(i) $|(fy)^u - (fx)^u| \ge m(T^u)|y^u - x^u| - \epsilon|y - x|$,
(ii) $|(fy)^s - (fx)^s| \le ||T^s|| \; |y^s - x^s| + \epsilon|y - x|$.

Proof. Since $|((f-T)y)^u - ((f-T)x)^u| \le |(f-T)y - (f-T)x| \le \epsilon|y-x|$, it follows that $|(fy)^u - (fx)^u| \ge |(Ty)^u - (Tx)^u| - \epsilon|y-x| \ge m(T^u)|y^u - x^u| - \epsilon|y-x|$. ∎

Estimate #2. (Invariant cones). $\forall x, y \in E(r)$ with $|y^u - x^u| \ge |y^s - x^s|$, we have
(i) $|(fy)^u - (fx)^u| \ge (m(T^u) - \epsilon)|y^u - x^u|$,

(ii) $|(fy)^u - (fx)^u| \geq \frac{m(T^u)-\varepsilon}{||T^s||+\varepsilon} \cdot |(fy)^s - (fx)^s|$.

Proof. Use Estimate #1 and the assumption that $|y-x| = |y^u - x^u| \geq |y^s - x^s|$. ∎

We are now ready to introduce the graph transform $\Gamma = \Gamma_f$. Let $\mathcal{G}$ denote the set of all continuous functions $g : R^u(r) \mapsto R^s(r)$ s.t. $g(0) = 0$ and $Lip(g) \leq 1$. We make $\mathcal{G}$ into a metric space by taking the sup norm on functions. For $g \in \mathcal{G}$, the *graph transform of g by f*, if well defined, is the function $\Gamma g : R^u(r) \mapsto R^s(r)$ satisfying

$$\text{graph } \Gamma g = f(\text{graph } g) \cap R(r)\,.$$

Without some conditions on f, there is no reason to assume that $f(\text{graph } g) \cap R(r)$ is the graph of any function at all.

Lemma 3.2.1. *In addition to the standing hypothesis, we assume that*

(c) $m(T^u) - \epsilon > 1$ *and*

(d) $(||T^s|| + \epsilon)/(m(T^u) - \epsilon) < 1$.

Then Γ is a well defined mapping of $\mathcal{G}$ into itself with $Lip(\Gamma) \leq ||T^s|| + 2\epsilon$.

Proof. Let $g \in \mathcal{G}$. To show that Γg is defined, it suffices to show that the mapping $t \mapsto (f(t, g(t)))^u$ maps $R^u(r)$ homeomorphically onto a set containing itself. This is guaranteed by Estimate #2 (i) and assumption (c) above. Once Γg is defined, the same reasoning as in Estimate #2 (ii) shows that $\text{Lip}(\Gamma g) \leq (||T^s|| + \epsilon)/(m(T^u) - \epsilon)$.

To estimate the contraction constant of Γ, let $g_1, g_2 \in \mathcal{G}$, and let $x_1, x_2 \in R(r)$ be s.t. $x_i \in$ graph g_i and $x_1^u = x_2^u$. We will argue that at $(fx_1)^u$, the values of Γg_1 and Γg_2 differ by $< (||T^s|| + 2\epsilon) \cdot |x_2 - x_1| \leq (||T^s|| + 2\epsilon) \cdot |g_2 - g_1|$. Since $(fx_i)^s = (\Gamma g_i)(fx_i)^u$, we have

$$|\Gamma g_1(fx_1)^u - \Gamma g_2(fx_1)^u| \leq |(fx_1)^s - (fx_2)^s| + |(fx_2)^s - \Gamma g_2(fx_1)^u|\,.$$

The first term above is $\leq (||T^s|| + \epsilon)|x_1 - x_2|$ by Estimate #1. The second term is $= |\Gamma g_2(fx_2)^u - \Gamma g_2(fx_1)^u|$, which is $\leq |(fx_1)^u - (fx_2)^u|$ since $Lip(\Gamma g_2) \leq 1$, and this is $\leq \epsilon|x_2 - x_1|$ by a variant of Estimate #1. (Draw a picture!) ∎

Theorem 3.2.2. (Lipschitz version of Local Unstable Manifold Theorem) *Let f be as above. Then $\exists!$ $g \in \mathcal{G}$ s.t. $f^{-1}(graph(g)) \subset graph(g)$.*

It is in fact true that if f is C^r (any $r \geq 1$) then so is g. For a proof see [HP]. The graph of g is called the *local unstable manifold* of f at 0, written $W^u_{loc}(0)$ or $W^u_r(0)$. If f^{-1} satisfies the same hypotheses as f with respect to T^{-1} (instead of T), then $W^u_r(0)$ can be characterized as follows: $x \in W^u_r(0) \Rightarrow f^{-n}x \to 0$ exponentially fast as $n \to \infty$; $x \notin W^u_r(0) \Rightarrow f^{-n}x \notin R(r)$ for some $n \geq 0$. (The statement about $x \notin W^u_r(0)$ is proved by comparing x to $y \in W^u_r(0)$ with $x^u = y^u$ and applying f^{-n} to both.)

3.3. Stable and unstable manifolds for nonuniformly hyperbolic systems. In this subsection we describe the stable and unstable "foliations" on M and discuss how the ideas in Section 3.2 can be used to prove these more general results. For complete proofs we refer the reader to [P1], [FHY] or [Ru4].

For simplicity consider the situation where (f,μ) is ergodic and $\lambda_i \neq 0\,\forall\, i$. Let $E^u(x) = \bigoplus_{\lambda_i>0} E_i(x), E^s(x) = \bigoplus_{\lambda_i<0} E_i(x); \lambda^+ = \min_{\lambda_i>0} \lambda_i$, and $\lambda^- = \max_{\lambda_i<0} \lambda_i$. Let $\{\Phi_x\}$ be a family of Lyapunov charts. (See Section 3.1). We write the domain of Φ_x as R_x, i.e. $R_x = R(\varepsilon\ell(x)^{-1}) \subset \mathbb{R}^m$, and write $R_x = R^u_x \times R^s_x$, where R^u_x and R^s_x are disks in the subspaces corresponding to $E^u(x)$ and $E^s(x)$. As before we let $\tilde{f}_x : R_x \to R_{fx}$ denote the map corresponding to f.

Let $\mathcal{G}^u_x = \{g : R^u_x \to R^s_x | g(0) = 0 \text{ and } Lip(g) \leq 1\}$, and consider the graph transform $\Gamma_x : \mathcal{G}^u_x \to \mathcal{G}^u_{fx}$ defined using $\tilde{f}_x$. Since $\ell(f(x))^{-1} > \ell(x)^{-1}e^{-\varepsilon}$, we are assured that $\tilde{f}_x(\text{graph } g)$ stretches all the way across R_{fx} for every $g \in \mathcal{G}^u_x$. Indeed, except for the fact that the domain and range of Γ_x are different function spaces, there is no difference between the situation here and that in Section 3.2. The same proofs therefore show that Γ_x is well defined and is a contraction in the C^0 norm. It follows that if $0 : R^u_{f^{-n}x} \to R^s_{f^{-n}x}$ is the function that is identically zero, then $\Gamma_{f^{-1}x} \circ \cdots \circ \Gamma_{f^{-n}x} 0$ converges as $n \to \infty$ to a function $g^u_x : R^u_x \to R^s_x$.

We have so far followed our argument in Section 3.2 almost verbatim, and have obtained a Lipshitz function g^u_x at a.e. x. To further study the smoothness properties of g^u_x, we will have to be more careful, because $Lip(D\tilde{f}_x)$ is not uniform in x. Nevertheless, the arguments in the fixed point case go through. We summarize the results:

Proposition 3.3.1. *For a.e. x, there is a function $g^u_x : R^u_x \to R^s_x$ with the following properties:*

(1) $g^u_x(0) = 0, Dg^u_x(0) = 0, Lip(g^u_x) \leq 1,$ *and* $Lip(Dg^u_x) \leq C\ell(x)$*;*
(2) $(\tilde{f}_{f^{-1}x})^{-1}(\text{graph } g^u_x) \subset \text{graph } (g^u_{f^{-1}x})$, *and* $\forall\, z_1, z_2 \in \text{graph } (g^u_x)$,

$$|(\tilde{f}_{f^{-1}x}\circ\cdots\circ\tilde{f}_{f^{-n}x})^{-1}(z_1) - (\tilde{f}_{f^{-1}x}\circ\cdots\circ\tilde{f}_{f^{-n}x})^{-1}(z_2)| < e^{(-\lambda^+ + 2\varepsilon)n}|z_1 - z_2|\,;$$

(3) *graph*(g^u_x) *consists exactly of those points* $z \in R_x$ *s.t.* $f^{-n}(\Phi_x z) \in \Phi_{f^{-n}x}(R_{f^{-n}x})\,\forall\, n \geq 0$ *.*

If f is replaced by f^{-1}, we obtain at a.e. x a function $g^s_x : R^s_x \to R^u_x$ with analogous properties.

Bringing these results back to the manifold, we obtain the following:

Theorem 3.3.2. *Given a family of Lyapunov charts $\{\Phi_x\}$, we have defined a.e. two measurable families of embedded disks called* local stable *and* unstable manifolds. *The local unstable manifold at x, written $W^u_{\ell oc}(x)$, is given by $\Phi_x(\text{graph } g^u_x)$.*

It is a disk through x, having the same dimension as $E^u(x)$ and tangent at x to $E^u(x)$. For $y \in W^u_{loc}(x), d(f^{-n}x, f^{-n}y) \to 0$ exponentially fast as $n \to \infty$. Local unstable manifolds are invariant in the sense that $f^{-1}W^u_{\ell oc}(x) \subset W^u_{\ell oc}(f^{-1}x)$. Local stable manifolds, written $W^s_{\ell oc}(\cdot)$, have similar properties: they are tangent to $E^s(\cdot)$ and satisfy $fW^s_{\ell oc}(x) \subset W^s_{\ell oc}(fx)$; forward orbits of points in each leaf converge exponentially.

We remark that the exponential decrease in $d(f^{-n}x, f^{-n}y)$ for $y \in W^u_{\ell oc}(x)$ follows from Proposition 3.1.1(2) and the fact that along orbits the distortion of metric in charts varies more slowly than $e^{(\lambda^+ - 2\varepsilon)n}$.

The *(global) unstable manifold* at x is defined to be

$$W^u(x) := \{y \in M : \overline{\lim_{n\to\infty}} \frac{1}{n} \log d(f^{-n}x, f^{-n}y) < 0\}\,.$$

Similarly, the *(global) stable manifold* at x is

$$W^s(x) := \{y \in M : \overline{\lim_{n\to\infty}} \frac{1}{n} \log d(f^{n}x, f^{n}y) < 0\}\,.$$

We claim that at a.e. $x, W^u(x) = \bigcup_{n\geq 0} f^n W^u_{\ell oc}(f^{-n}x)$, where $\{W^u_{\ell oc}(\cdot)\}$ is the family of local $W^u_{\ell oc}$-leaves defined using any system of Lyapunov charts. This description of $W^u(x)$ tells us that $W^u(x)$ is an immersed submanifold. To prove the "$\subset$" part of the claim, suppose $y \in W^u(x)$ satisfies $d(f^{-n}x, f^{-n}y) \lesssim e^{-\varepsilon' n}$ for some $\varepsilon' > 0$. Choose some $\varepsilon'' \ll \varepsilon'$, and consider a system of charts defined using ε''. Because chart sizes vary so slowly, $f^{-n}y$ is in the chart at $f^{-n}x$ for all large n, and we conclude that in fact $d(f^{-n}x, f^{-n}y) \lesssim e^{-\lambda^+ n}$. So for any given system of charts $\{\Phi_x\}, \exists\, N$ s.t. $\forall\, n \geq N, f^{-n}y \in \Phi_{f^{-n}x} R_{f^{-n}x}$. From this it follows that $f^{-N}y \in W^u_{\ell oc}(f^{-N}x)$, proving our claim. Needless to repeat, dual statements hold for $W^s(x)$.

From the definition of W^u (resp. W^s), it is clear that for $x, y \in M$, either $W^u(x) \cap W^u(y) = \phi$ or $W^u(x) = W^u(y)$. Also, the union of W^u-leaves is a set of full measure. For this reason, we sometimes refer to W^u as the "unstable foliation" – even though it is not a foliation in the usual sense of the word.

What we have described so far are stable and unstable manifolds corresponding to E^s and E^u for systems with no zero exponents. Observe that in Section 3.2, what is needed for the graph transform method to work is that $m(T^u) < 1$ and that $||T^s|| < m(T^u)$ – whether T^s is contracting or not is irrelevant. This suggests that we can obtain invariant manifolds corresponding to certain other splittings of the Lyapunov spectrum.

Continuing to assume that (f, μ) is ergodic, we let $\lambda_1 > \cdots > \lambda_r$ be the Lyapunov exponents of (f, μ). Suppose that $\lambda_i > 0$. We define the *ith unstable manifold* at x to be

$$W^i(x) := \{y \in M : \overline{\lim_{n\to\infty}} \frac{1}{n} \log d(f^{-n}x, f^{-n}y) \leq \lambda_i\}\,.$$

It can be shown that for a.e. $x, W^i(x)$ is a C^2 immersed submanifold through x, that it has dimension $\sum_{j\le i} \dim E_j$ and is tangent to $\bigoplus_{j\le i} E_j(x)$ at x. These W^i-manifolds are obtained from $W^i_{\ell oc}$-manifolds as in our previous discussion, and the existence of $W^i_{\ell oc}(\cdot)$ is proved using graph transform methods as before. Clearly, the W^i-manifolds are f-invariant, and if $\lambda_{i+1} > 0$, then $W^i(x) \subset W^{i+1}(x)$. (One might wish that there was an unstable manifold tangent to $E_i(x)$ for each i with $\lambda_i > 0$, but there is in general no such nonlinear structure that is f-invariant.)

To sum up then:

Theorem 3.3.3. *Let $\lambda_1 > \cdots > \lambda_u$ be the positive Lyapunov exponents of (f, μ). Then defined a.e. on M are u nested invariant "foliations"*

$$W^1 \subset \cdots \subset W^u\,,$$

where W^i is the unstable foliation corresponding to $\bigoplus_{j\le i} E_j$. Similarly, if $\lambda_{r-s+1} > \cdots > \lambda_r$ are the negative exponents, then we have

$$W^{(1)} \subset \cdots \subset W^{(s)}\,,$$

where $W^{(i)}$ is the stable foliation corresponding to $\bigoplus_{j>r-i} E_j$.

3.4. Noninvertible maps and maps with singularities. Let $S \subset M$ be a singularity set in the sense that $f : (M - S) \to M$ is a C^2 diffeomorphism onto its image, and let μ be an invariant measure with $\mu\{x \in M : f^n x \notin S\ \forall n \in \mathbb{Z}\} = 1$. We discuss briefly how the material in Section 3 can be adapted to handle this slightly more complicated situation.

First, since the domain of f is no longer compact, we must verify the integrability of $\log ||Df^{\pm}||$ to make sure that Lyapunov exponents are defined. Second, in our construction of Lyapunov charts, the second derivative of f is assumed to be bounded. Integrability of $\log ||D^2 f^{\pm}||$ will suffice, but $||D^2 f||$ must play a role in our definition of ℓ.

Assuming that the derivatives are under control, there is still the possibility that the orbits of "typical" points may approach S too fast. If this happens, stable and unstable manifolds may not exist. Let us explain what may go wrong. Suppose that at x there is a piece of local unstable manifold of size $\delta > 0$. For the sake of argument, let us assume that each iterate of f^{-1} shrinks $W^u_\delta(x)$ by a factor of $\frac{1}{2}$, and that $f^{-n} W^u_\delta(x)$ is always roughly perpendicular to S. Since $f^{-n} W^u_\delta(x) \cap S = \phi$, we must have $d(f^{-n}x, S) > \frac{\delta}{2^n}\ \forall n \ge 0$. In other words, in order for $W^u_{loc}(x)$ to exist, the backward orbit must not approach S faster than the rate of $\frac{1}{2^n}$.

This problem is usually handled by imposing a condition on μ. Let $U(S, \epsilon)$ denote the ϵ-neighborhood of S in M. Suppose we require that $\forall\, \epsilon > 0, \mu U(S, \epsilon) <$

$C\epsilon^\alpha$ for some $C, \alpha > 0$. Then $\sum_{n\geq 0} \mu f^n U(S, e^{-\lambda^+ n}) = \sum \mu U(S, e^{-\lambda^+ n}) < \infty$ and we conclude by the Borel-Cantelli Lemma that for a.e. x, $\exists\, N(x)$ s.t. $f^{-n}x \notin U(S, e^{-\lambda^+ n})\, \forall\, n \geq N(x)$. This guarantees the existence of $W^u_{\delta(x)}(x)$ for a.e. x. For more details see [KaS].

For noninvertible maps, one possibility is to consider the inverse limit of f, i.e., to construct Lyapunov charts etc. not for a.e. x, but for a.e. *history* of a.e. x. If in addition $\det(Df) = 0$ some places, then the set $\{\det(Df) = 0\}$ will have to be treated as a singularity set as described above.

4. Entropy, Lyapunov exponents and dimension

We have seen that Lyapunov exponents, in particular positive Lyapunov exponents, give the rates at which nearby orbits diverge. They provide us with a *geometric* way of measuring the complexity of a map. Metric entropy, on the other hand, is a purely *probabilistic* notion. It measures randomness in the sense of information and predictability. In this section we compare these two invariants. As we do so, a third invariant, the *dimension* of the *invariant measure*, will present itself.

Unless otherwise stated, we assume throughout Section 4 that $f : M \circlearrowleft$ is a C^2 diffeomorphism of a compact Riemannian manifold M preserving a Borel probability measure μ. Additional conditions on f or μ will be stated in each theorem. We let $\{\lambda_i\}$ denote the Lyapunov exponents of (f, μ), and let m_i denote the multiplicity of λ_i. We write $a^+ = \max(a, 0)$, so that $\{\lambda_i^+\}$ refers to the positive exponents of (f, μ). The metric entropy of f with respect to μ is written $h_\mu(f)$. (See Section 1.2.)

4.1. Entropy and Lyapunov exponents: discussion of results.

The two most basic results in this direction are:

Theorem 4.1.1. (Pesin formula) *If μ is equivalent to the Riemannian measure on M, then*

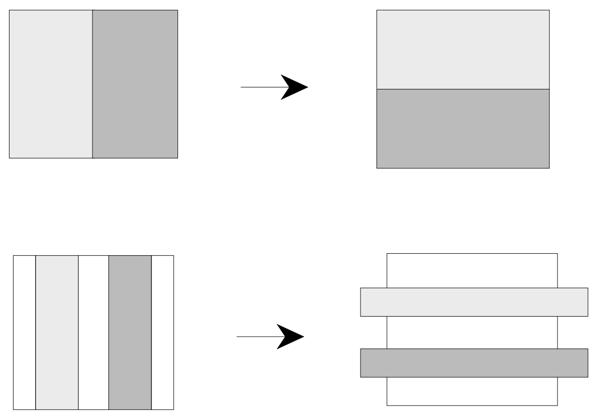
$$h_\mu(f) = \int \sum \lambda_i^+ m_i d\mu \qquad (*)$$

Theorem 4.1.2. (Ruelle's inequality) *For C^1 mappings (that are not necessarily invertible) we have*

$$h_\mu(f) \leq \int \sum \lambda_i^+ m_i d\mu\,.$$

Theorem 4.1.1 is first proved in [P2]; see also [M1]. Theorem 4.1.2 is proved in [Ru3]; this result is also attributed to Margulis (unpublished). Rough sketches of their proofs are given in Section 4.2. The following examples are no doubt overly simplistic, but in some ways they do illustrate what is going on:

Example 4.1.3. (baker v.s. horseshoe)



The first map is the baker transformation (see Example 1.1.4 in these notes). The second is easily extended to Smale's horseshoe. For simplicity, we assume that both maps are affine on the shaded regions, and in both cases, we take μ to be the probability measure that makes the system isomorphic to the $\left(\frac{1}{2}, \frac{1}{2}\right)$-Bernoulli shift. For the baker map, μ is Lebesgue measure, and $h_\mu(f) = \log 2 = \lambda_1$, the positive exponent. For the horseshoe, μ is supported on a Cantor set of Lebesgue measure zero, and $h_\mu(f) = \log 2 < \lambda_1$.

Here is one way to interpret these results. Entropy is created by the exponential divergence of nearby orbits (see e.g. Theorem 1.2.6). In a conservative system, all the expansion goes back into the system to make entropy, hence Pesin's formula. A strict inequality, on the other hand, corresponds to the situation where some of the expansion is "wasted", and that can happen only if there is some "leakage" from the system.

Before Theorems 4.1.1 and 4.1.2 were proved in their present generality, they had been known for some time in the context of Anosov systems. It was also proved for Anosov systems ([S1], [S3]) and later for Axiom A attractors ([Ru1]) that measures satisfying (*) have a special geometric property, namely that their conditional measures on unstable manifolds are equivalent to the Riemannian measure on W^u. This geometric characterization of measures satisfying (*) has since been extended to the nonuniform setting by Ledrappier and others.

Theorem 4.1.4. ([LS], [L1], [LY]) *Assume that (f, μ) has a positive Lyapunov exponent a.e. Then Pesin's formula holds iff μ has absolutely continuous conditional measures on unstable manifolds.*

We define what it means to have absolutely continuous conditional measures on W^u. A measurable partition ξ of M is said to be *subordinate to* W^u if for μ-a.e. $x, \xi(x) \subset W^u(x)$ and contains an open neighborhood of x in $W^u(x)$. Let $\{\mu_x^\xi\}$ denote a canonical system of conditional measures of μ with respect to ξ (see Section 1.3), and let m_x^u denote the Riemannian measure induced on $W^u(x)$.

Definition 4.1.5. *We say that μ has absolutely continuous conditional measures on unstable manifolds if for every measurable partition ξ that is subordinate to W^u, μ_x^ξ is absolutely continuous wrt m_x^u for μ-a.e. x.*

Measures with the properties in Theorem 4.1.4 are sometimes called *SRB measures*, because they were first constructed for Anosov systems and Axiom A attractors by Sinai, Ruelle and Bowen ([S1],[S3], [BR],[Ru1]; see also [Bo]). The existence of these measures is the subject of Section 6.

4.2. Sketches of proofs. Main ideas in the proofs of Ruelle's inequality and Pesin's formula are given below. We will not attempt to outline a proof of Theorem 4.1.4 here, except to remark that it involves working with entropy with respect to partitions that are subordinate to W^u (see Section 1.3), and that the "only if" part uses Jensen's inequality.

Proof of Ruelle's inequality. For simplicity we assume that (f, μ) is ergodic. For $\epsilon > 0$, let α_ϵ be a partition of M into approximate ϵ-boxes, and let δ_1, δ_2, and δ_3 be prescribed small numbers.

First we choose N s.t. $\forall x$ in a good set G with $\mu G > 1 - \delta_1, Df^N$ looks like what the Lyapunov exponents say it should.

Next we choose $\epsilon > 0$ small enough that

- in the ϵ-neighborhood of every $x \in G, Df^N$ is a good approximation of f^N; we assume in fact that if $\alpha_\epsilon(x) \cap G \neq \phi$, then $Df^N\alpha_\epsilon(x)$ is contained in an $\epsilon e^{(\lambda_1+\delta_2)N} \times \cdots \times \epsilon e^{(\lambda_r+\delta_2)N}-$ box (λ_i counted with multiplicity), and
- $h(f^N) \leq h(f^N; \alpha_\epsilon) + \delta_3$.

Now $h(f) = \frac{1}{N}h(f^N)$ and $h(f^N; \alpha_\epsilon) \leq H(f^{-N}\alpha_\epsilon|\alpha_\epsilon)$. We estimate this latter quantity by

$$H(f^{-N}\alpha_\epsilon|\alpha_\epsilon) \leq \sum_{A\in\alpha_\epsilon} \mu A \cdot \log r_{N,\epsilon}(A)$$

where $r_{N,\epsilon}(A)$ is the number of elements of $f^{-N}\alpha_\epsilon$ that meet A, or, equivalently, the number of elements of α_ϵ that intersect $f^N A$. If $A \cap G \neq \phi$, then we have control on the size and shape of $f^N A$, obtaining $r_{N,\epsilon}(A) \lesssim e^{N\Sigma(\lambda_i^+ + \delta_2)m_i}$. If $A \cap G = \phi$, then $r_{N,\epsilon}(A) \leq e^{C_0 N}$ where C_0 is a constant depending only on $||Df||$. We have thus proved

$$\frac{1}{N}H(f^{-N}\alpha_\epsilon|\alpha_\epsilon) \leq (1-\delta_1)\sum(\lambda_i^+ + \delta_2)m_i + \delta_1 C_0\,,$$

which gives the desired result. ■

Proof of Pesin's formula. We follow the proof in [M1], and use the notations in Section 3. Let us first give the argument assuming that the function ℓ in Theorem 3.1.1 is uniformly bounded. Let $U(x) = \Phi_x R_x = \Phi_x R(\epsilon \ell(x)^{-1})$ and define $U_n(x) = \{y \in M : f^i y \in U(f^i x) \; \forall \, 0 \le i \le n\}$. Then by Theorem 1.2.6,

$$\mu U_n(x) \sim e^{-nh_\mu(f)}$$

if the charts are sufficiently small. We also estimate $\mu U_n(x)$ as follows. Let P_0 be a "plane" in R_x having the same dimension as R^u_x and parallel to R^u_x. Then $\tilde{f}_x P_0 \cap R_{fx}$ is the graph of a function from $R^u_{fx} \to R^s_{fx}$ with small slope, and $P_1 := \tilde{f}_x^{-1}(\tilde{f}_x P_0 \cap R_{fx})$ has area $\sim e^{-\Sigma \lambda_i^+ m_i}$. Continuing to iterate, we see inductively using arguments similar to those in (3.2) that for each n, $\tilde{f}^n_x P_{n-1} \cap R_{f^n x}$ is a manifold roughly parallel to $R^u_{f^n x}$ and $P_n := \tilde{f}_x^{-n}(\tilde{f}^n_x P_{n-1} \cap R_{f^n x})$ has area $\sim e^{-n\Sigma\lambda_i^+ m_i}$. Fill R_x with planes like P_0. It follows that $\mu U_n(x) \approx \text{Leb } \Phi_x^{-1} U_n(x)$, which has the desired estimate.

The problem is that in general, ℓ is unbounded and $U(x)$ can be arbitrarily small. One might try to use the ρ-version of Theorem 1.2.6 with $\rho(x) = \text{diam } U(x)$, but there is no information on the integrability of $\log \ell$. Mañé used the following trick (in addition to proving the ρ-version of Theorem 1.2.6 to make the trick work): Fix some n_0, and consider returns to Λ_{n_0}. Let $t(x)$ be the first return time, and define

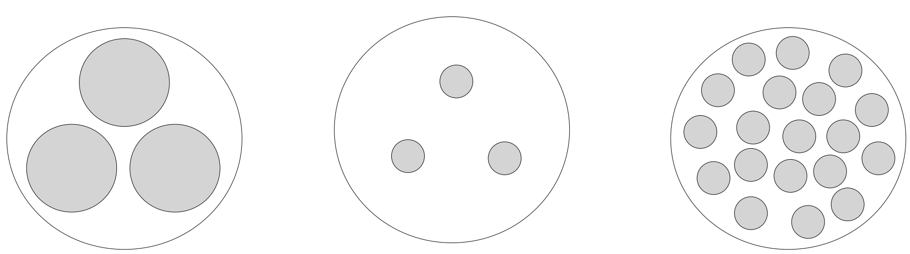
$$\rho(x) = \begin{cases} e^{-t(x) \cdot \max\limits_i \lambda_i(x)} & x \in \Lambda_{n_0} \\ 1 & x \notin \Lambda_{n_0} \end{cases} .$$

Then $B(x, \rho; n) \subset U_n(x)$, and some version of the estimates above continue to hold. Also, $\int -\log \rho d\mu < \infty$ because $\int_{\Lambda_{n_0}} t d\mu < \infty$, so Theorem 1.2.6 applies.

■

4.3. Relation to dimension. The theorems in Section 4.1 suggest that the size of the "gap" in Ruelle's inequality might be related to certain geometric properties of the invariant measure. The following example further illustrates this point:

Example 4.3.1. (Affinely constructed Cantor sets)



(a) (b) (c)

Each picture represents a fractal Λ defined as follows: let B be the larger region, and let $f : B \to \mathbb{R}^2$ be s.t. $f^{-1}B \cap B$ is the union of the smaller regions. Then $\Lambda = \bigcap\limits_{n \geq 0} f^{-n}B$. The following seems intuitively clear: with entropy fixed, Hausdorff dimension decreases as Lyapunov exponents are increased (compare pictures (a) and (b)); and with Lyapunov exponents fixed, dimension goes up with entropy (compare (b) and (c)).

For the definition and properties of Hausdorff dimension, see e.g. [Fa]. Since entropy and Lyapunov exponents reflect a.e. behavior with respect to an invariant measure, it is reasonable to expect that if Hausdorff dimension is involved, it will see only the set of "typical" points. Let X be a metric space, and let m be a Borel probability measure on X. Recall that for $x \in X$ and $\epsilon > 0, B(x, \epsilon)$ denotes the ball of radius ϵ centered at x.

Definition 4.3.2. *We say that the dimension of* m*, written* $\dim(m)$*, is well defined and is equal to* α *if for* m*-a.e.* x*,*

$$\lim_{\epsilon \to 0} \frac{\log mB(x, \epsilon)}{\log \epsilon} = \alpha \, .$$

The Hausdorff dimension of m *is defined to be*

$$HD(m) = \operatorname*{Inf}_{\substack{Y \subset X \\ mY = 1}} HD(Y)$$

where $HD(Y)$ *denotes the Hausdorff dimension of the set* Y*.*

The notion $\dim(m)$ is not always well defined. It is easy to construct examples of measures for which the limits as $\epsilon \to 0$ do not exist. On the other hand, if m is an ergodic invariant measure for a locally bi-Lipschitz map, then once these limits exist, they are constant a.e. It is also easy to see that if $\dim(m)$ is well defined, then it is equal to $HD(m)$ (see e.g. [Y]).

We begin with a "conformal" version of the result in [LY].

Theorem 4.3.3. *Let* f *be a* C^2 *mapping. Assume that* (f, μ) *is ergodic, and that* $\lambda_i = \lambda \, \forall i$*. Then* $\dim(\mu)$ *is well defined and*

$$h_\mu(f) = \lambda \cdot \dim(\mu) \, .$$

The situation in general is slightly more complicated. Let $\lambda_1 > \cdots > \lambda_u$ be the positive exponents arranged in descending order, and as in (3.4), let $W^i(x)$ be the unstable manifold tangent to $\bigoplus\limits_{j \leq i} E_j(x)$. For each i, let ξ_i be a partition subordinate to W^i, and let $\{\mu_x^i\}$ be a family of conditional measures of μ wrt ξ_i. In the next theorem, the numbers σ_i have the interpretation of being "partial dimensions" of μ in the directions of the subspaces corresponding to λ_i.

Theorem 4.3.4. (Part II of [LY]) *Let f be a C^2 diffeomorphism and assume that (f,μ) is ergodic. Then corresponding to each λ_i, there is a number $\sigma_i, 0 \le \sigma_i \le m_i$, such that*

(1) *for every ξ_i as above, $\dim(\mu_x^i)$ is well defined for a.e. x, and their common value, written $\dim(\mu|W^i)$, is equal to $\sum_{j\le i}\sigma_j$;*

(2) $h_\mu(f) = \sum \lambda_i^+ \sigma_i$.

We will indicate the idea of proof in the next subsection. Suffice it to observe here that since $0 \le \sigma_i \le m_i$, (2) agrees with Ruelle's inequality; and that if μ is smooth on W^u, then $\sigma_i = m_i \, \forall i$ and (2) agrees with the entropy formula (*) in Section 4.1.

We have expressed entropy in terms of Lyapunov exponents and dimension. One might ask if $\dim(\mu)$ or $HD(\mu)$ can also be expressed in terms of the other two. This has some practical use, since the Hausdorff dimension of a set tells us how many variables are needed to describe it completely. (See [M2] for a precise statement.) For each $\lambda_i > 0$, let σ_i be as above. For $\lambda_i < 0$, we define σ_i by considering f^{-1}.

Question. *Suppose (f,μ) is ergodic and $\lambda_i \neq 0 \, \forall i$. Is it true that*

$$HD(\mu) = \dim(\mu) = \sum_{\text{all } i} \sigma_i \quad ?$$

Since we know that $\dim(\mu|W^u)$ and $\dim(\mu|W^s)$ are well defined, it remains only to show that they add properly. A technical difficulty is that the leaves of W^u and W^s are in general not packed together very nicely; they form a coordinate system that is usually no better than Hölder continuous. This is the case even for uniformly hyperbolic systems (see e.g. [HP]). We list some situations in which the question above has been answered in the affirmative:

(a) either W^u or W^s is C^1 (e.g. algebraic toral automorphisms) [L3];

(b) μ is an SRB measure [L3] (this uses the absolute continuity of the W^s foliation – see Section 6.2); and

(c) the "conformal" case, referring here to the situation where all of the positive exponents are equal and all of the negative exponents are equal. The proof is identical to that in [Y], where M is assumed to have dimension 2.

4.4. Ideas behind the dimension formula. *The "conformal" case: noninvertible maps.* Consider the situation in Theorem 4.3.3. Let $B(x,\epsilon\,;n) = \{y \in M : d(f^k x, f^k y) < \epsilon \ \forall \ 0 \le k \le n\}$. Pretending that locally f is a dilation by e^λ, we have

$$B(x,\epsilon\,;n) \sim B(x,\epsilon e^{-\lambda n}).$$

Assuming that Df is nonsingular, estimates of this type are easily made precise using Lyapunov charts. (See Sections 3 and 4.2.) By Theorem 1.2.6, we have

$$\mu B(x, \epsilon\ ; n) \sim e^{-nh}$$

where $h = h_\mu(f)$. Comparing these two expressions and letting $\tau = e^{-\lambda n}$, we have

$$\mu B(x, \tau) \sim \tau^{\frac{h}{\lambda}},$$

and so $\dim(\mu)$ exists and $= \frac{h}{\lambda}$.

∎

The "conformal" case: invertible maps. Consider situation (c) at the end of Section 4.3. That is, f is a diffeomorphism with no zero exponents, all of its positive exponents are $= \lambda$ and all of its negative exponents are $= \lambda'$. Let $B(x, \epsilon; n, n') := \{y \in M : d(f^k x, f^k y) < \epsilon\ \forall - n' \le k \le n\}$. Choosing $n, n' \in \mathbb{Z}^+$ s.t. $n\lambda \approx -n'\lambda'$, we have $B(x, \epsilon; n, n') \sim B(x, \epsilon e^{-\lambda n}) \sim B(x, \epsilon e^{\lambda' n'})$. The same argument as above gives

$$e^{-(n+n')h} \sim \mu B(x, \epsilon; n, n') \sim (e^{-n\lambda})^{\dim(\mu)}\ .$$

Equating exponents, we obtain

$$\dim(\mu) = h\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\ .$$

∎

Note that if more than one positive exponent is present, then the "eccentricity" of $B(x, \epsilon; n, n')$ approaches ∞ as $n, n' \to \infty$ for any choice of n, n'. $\operatorname{Dim}(\mu)$, as with Hausdorff dimension, cannot be estimated using only ellipsoids that are exponentially eccentric, and so the argument in the last paragraph fails.

The general picture. The strategy here is to work with one exponent at a time and to work our way up the entire hierarchy of unstable manifolds $W^1 \subset W^2 \subset \cdots W^u$. For each i, we introduce a notion of entropy along W^i, written h_i, measuring the randomness of f along the leaves of W^i and ignoring what happens in the transverse directions. We also prove that the dimensions of the conditional measures are well defined. For brevity write $\delta_i = \dim(\mu|W^i)$. The proof consists of the following steps:

(i) $h_1 = \delta_1 \lambda_1$;
(ii) $h_i - h_{i-1} = (\delta_i - \delta_{i-1})\lambda_i$ for $i = 2, \ldots, u$;
(iii) $h_u = h_\mu(f)$.

The proof of (i) is the same as before since it only involves one exponent. To give an idea of why (ii) is true, consider the action of f on the leaves of W^i, and pretend somehow that a quotient dynamical system can be defined by collapsing the leaves of W^{i-1} inside W^i. This "quotient" dynamical system has exactly one Lyapunov exponent, namely λ_i. It behaves as though it leaves invariant a measure with dimension $\delta_i - \delta_{i-1}$ and has entropy $h_i - h_{i-1}$. A fair amount of technical work is needed to make this precise, but once properly done, it is again the single exponent principle at work. Letting $\sigma_1 = \delta_1$, $\sigma_i = \delta_i - \delta_{i-1}$ for $i = 2, \dots, u$, and summing the equations in (ii) over i, we obtain $h_u = \sum_{i=1}^{u} \sigma_i \lambda_i$. Step (iii) says that zero and negative exponents do not contribute to entropy. The influence of negative exponents is easily ruled out, and an argument similar to that in step (ii) tells us that entropy does not increase as we go from the unstable foliation to the "center unstable foliation". This completes the outline to the proof in [LY].

5. Expanding Maps and the Lebesgue Measure Class

5.1. The space of invariant measures. Most of the material in this subsection is contained in [Wa].

Let X be a compact metric space, and let $\mathcal{M}(X)$ be the set of all Borel probability measures on X. Let $C(X)$ denote the Banach space of continuous real-valued functions on X with the sup norm, and let $C(X)^*$ denote the space of continuous linear functionals on $C(X)$. By the Riesz Representation Theorem we know that there is a one-to-one correspondence between $\mathcal{M}(X)$ and $\{\alpha \in C(X)^* : \alpha(\varphi) \geq 0 \,\forall\, \varphi \geq 0 \text{ and } \alpha(1) = 1\}$. This bijection assigns to each $\mu \in \mathcal{M}(X)$ the functional $\alpha_\mu \in C(X)^*$ defined by $\alpha_\mu(\varphi) = \int \varphi d\mu$. We assume that $\mathcal{M}(X)$ is given the weak* topology it inherits from $C(X)^*$, i.e. $\mu_n \to \mu$ iff $\int \varphi d\mu_n \to \int \varphi d\mu$ for every $\varphi \in C(X)$. With this topology, it is straightforward to check that $\mathcal{M}(X)$ is a nonempty, compact, convex, metrizable space (e.g. the compactness of $M(X)$ follows from the compactness of the unit ball in $C(X)^*$).

Let T be a continuous map of X to itself, and let $\mathcal{M}_T(X)$ denote the set of T-invariant Borel probability measures on X.

Lemma 5.1.1. *$\mathcal{M}_T(X) \neq \phi$.*

Proof. Pick an arbitrary x in X, and let $\mu_n := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x}$, where δ_z denotes the Dirac measure at z (i.e. $\delta_z(E) = 1$ iff $z \in E$ for every Borel set E). Since $\mathcal{M}(X)$ is compact, a subsequence of μ_n converges to some $\mu \in \mathcal{M}(X)$. To verify that μ is

T-invariant is equivalent to showing that $\int \varphi \circ T d\mu = \int \varphi d\mu \, \forall \, \varphi \in C(X)$. This is true because $|\int \varphi d\mu_n - \int \varphi \circ T d\mu_n| \leq \frac{2}{n}|\varphi|$.

■

$\mathcal{M}_T(X)$ is in fact a closed convex subset of $\mathcal{M}(X)$. Being a compact convex subset in a locally convex topological vector space, $\mathcal{M}_T(X)$ is the convex hull of its extreme points (the Krein-Milman Theorem). Recall that μ is an extreme point of $\mathcal{M}_T(X)$ if $\mu \neq c\mu_1 + (1-c)\mu_2$ for some $\mu_1, \mu_2 \in \mathcal{M}_T(X)$ and $0 < c < 1$. Now it is not hard to convince ourselves that $\mu \in \mathcal{M}_T(X)$ is an extreme point iff (T, μ) is ergodic. As a consequence of the Choquet Representation Theorem then, we have

Proposition 5.1.2. (Ergodic decomposition of invariant measures) *Let $\mathcal{E}_T(X)$ denote the set of ergodic invariant measures of T. Then corresponding to every $\mu \in \mathcal{M}_T(X)$, there is a Borel probability measure τ_μ on $\mathcal{E}_T(X)$ s.t.*

$$\mu = \int \nu d\tau_\mu(\nu)\,,$$

i.e. $\int \varphi d\mu = \int (\int \varphi d\nu) d\tau_\mu(\nu)$ for every $\varphi \in C(X)$.

Theorem 5.1.2 gives some justification to our making assumptions on ergodicity, such as what we did in Section 3. When μ is supported on a finite or countable number of atoms, i.e. when $\mu = \Sigma c_i \mu_i$ for $\mu_i \in \mathcal{E}_T(X), c_i \in \mathbb{R}^+$, then the μ_i are often referred to as the "ergodic components" of μ. In general, μ disintegrates into an uncountable family of invariant measures, and it only makes sense to talk about a.e. ergodic component.

Suppose now we wish to use ergodic theory techniques to study the almost everywhere behavior of a mapping or a differential equation on a compact domain of $\mathbb{R}^n$. To do ergodic theory we need an invariant measure. We saw earlier on that $\mathcal{M}_T(X)$ is never empty, so existence is not a problem. It is often the case, however, that $\mathcal{M}_T(X)$ is a rather large set, containing infinitely many ergodic measures. (For example, if T has an invariant subset that is topologically conjugate to a horseshoe, then there are infinitely many periodic orbits, each of which supports an ergodic measure.) Moreover, the dynamics of (T, μ) depends seriously on μ; typical orbits with respect to mutually singular invariant measures may have little to do with each other. *Which invariant measure, then, should we pick*?

The answer to this question depends on the purpose of the investigation. If one is interested in capturing the maximum amount of randomness, then the measure of maximal entropy (if it exists) would be most relevant. Other circumstances may warrant the consideration of measures satisfying certain constraints or reflecting certain geometric properties. In the remainder of this section and in the next, we will adopt the viewpoint that the Lebesgue measure class is of natural interest, because sets of positive Lebesgue measure are in some sense the only sets that can be observed physically. We are particularly interested in the role played by Lebesgue measure in "dissipative" systems, i.e. systems that are not *a priori* volume preserving.

5.2. Expanding maps and absolutely continuous invariant measures. We say that $f : M \circlearrowleft$ is *expanding* or *uniformly expanding* if $\exists\, \lambda > 1$ s.t. $\forall\, x \in M$ and $\forall\, v \in T_xM, ||Df_xv|| \geq \lambda||v||$. Let m denote Lebesgue or Riemannian measure on M. Invariant probability measures that are absolutely continuous with respect to m will be referred to as *acim's*.

Theorem 5.2.1. [KrS] *Let $f : M \circlearrowleft$ be a C^2 uniformly expanding map of a compact Riemannian manifold. Then f admits an acim, with a density bounded above and below by two positive constants.*

Before giving a proof, let us give a naïve explanation for why the expanding property is condusive to having smooth invariant measures. Let φ_0 be a density on M, i.e. $\varphi_0 : M \to \mathbb{R}$ is a function satisfying $\varphi_0 \geq 0$ and $\int \varphi_0 dm = 1$. Let us suppose that φ_0 is supported on a small disk D on which f is $1-1$. We transport the probability measure $\varphi_0 dm$ forward by f, obtaining a new probability measure with density $\varphi_1(x) = \frac{\varphi_0(f^{-1}x)}{|\det Df(f^{-1}x)|}$. Notice that if f is expanding, then $\varphi_1 dm$ is in some sense more evenly distributed than before, whereas if f is contracting, then the new measure will lump up. As this process is iterated, one could imagine that an expanding map would spread the measure $\varphi_0 dm$ around more and more evenly, giving rise eventually to a smooth invariant measure. (This is not a proof, because f^n is not $1-1$ on $D\,\forall\, n \geq 1$.

Proof of Theorem 5.2.1. We begin with a few easy facts about f:

(1) $\exists \epsilon_0 > 0$ and $\lambda_0 > 1$ s.t. $d(x, y) < \epsilon_0 \Rightarrow d(fx, fy) \geq \lambda_0 d(x, y)$.
(2) If $deg(f) = k$, then every $x \in M$ has exactly k inverse images.
(3) If ϵ_1 is sufficiently small, then restricted to any ϵ_1-disk D in M, f^{-n} has exactly k^n well defined branches $\forall\, n > 0$. (For $n = 1$, this is an immediate consequence of (2). For $n > 1$, use (1) and $\epsilon_1 < \epsilon_0$.)

Let $\nu_0 = m$ normalized. For $n = 1, 2, \ldots$, define $\nu_n := f^n_* \nu_0$, i.e. $\nu_n(E) = \nu_0(f^{-n}E)$ for every Borel set E, and let $\varphi_n = \frac{d\nu_n}{d\nu_0}$.

We claim that

$$(*) \qquad \exists\, \alpha, \beta > 0 \ \text{ s.t. } \ \alpha \leq \varphi_n \leq \beta\ \forall\, n\,.$$

To prove this claim, we need the following distortion estimate:

Lemma 5.2.2. *Let f be as above. Then $\exists\, C_0$ (independent of n) s.t. $\forall\, x, y \in X$, if $d(f^ix, f^iy) < \epsilon_0 \ \ \forall\, i \leq n$, then*

$$\frac{\det Df^n(x)}{\det Df^n(y)} \leq e^{C_0 d(f^n x, f^n y)}\,.$$

Proof of Lemma 5.2.2.

$$\begin{aligned}\log \frac{\det Df^n(x)}{\det Df^n(y)} &\le \sum_{i=0}^{n-1} |\log\det Df(f^i x) - \log\det Df(f^i y)| \\ &\le \sum_{i=1}^{n-1} C_1 d(f^i x, f^i y) \quad \text{for some } C_1 \\ &\le \sum_{i=1}^{n-1} C_1 \lambda_0^{-(n-i)} d(f^n x, f^n y)\,.\end{aligned}$$

We remark that this estimate relies on the fact that f is C^2.

■

Continuing with our proof of the theorem, consider x, y in some ϵ_1-disk D. Then $\varphi_n = \sum_i \varphi_n^i$, where each φ_n^i is the contribution to the density of ν_n by pushing along the ith branch of $f^{-n}|D$. Assuming that $\epsilon_1 < \epsilon_0$, Lemma 5.2.2 tells us that

$$\frac{\varphi_n^i(x)}{\varphi_n^i(y)} \le e^{C_0 d(x,y)}\,.$$

Summing over i, we obtain

$$\frac{\varphi_n(x)}{\varphi_n(y)} \le e^{C_0 d(x,y)}\,.$$

This together with $\int \varphi_n d\nu_0 = 1$ and the compactness of M proves the claim(*).

Let μ be an accumulation point of $\{\frac{1}{n}\sum\limits_{i=0}^{n-1}\nu_i\}_{n=1,2,\dots}$. Then clearly μ is invariant and has a density with the same upper and lower bounds as the φ_n's. This completes our proof.

■

With a little more work, one can show that (f^n, μ) is ergodic $\forall\, n \ge 1$. It is in fact true that the natural extension of (f, μ) is isomorphic to a Bernoulli shift.

We mention also the *Perron-Frobenius* or *transfer operator* approach, which often gives more precise information on the invariant density and the manner in which it is approximated. Let $\mathcal{F}$ be a suitable space of real valued functions on M, and let $\mathcal{L} : \mathcal{F} \to \mathcal{F}$ be the operator defined by

$$\mathcal{L}(\varphi)(x) := \sum_{y \in f^{-1}x} \frac{\varphi(y)}{|\det Df(y)|}\,.$$

Assuming that everything is well defined, we have that if $\nu = \varphi dm$, then $f_*\nu = \mathcal{L}(\varphi)dm$, so that fixed points of $\mathcal{L}$ correspond exactly to invariant measures of f. It has been shown, for instance, that if f is C^r and $\mathcal{F}$ is the space of C^{r-1} functions on M, then $\mathcal{L}$ is a quasi-compact operator with a simple eigenvalue at 1 [Ru5]. For more information on transfer operators see the article by V. Baladi in this volume.

5.3. Non-uniformly expanding maps. Theorem 5.2.1 contains a basic principle that has been generalized in a number of ways. In Section 6 a similar construction will be carried out on unstable manifolds. Another generalization involves relaxing the uniform assumption on the expanding property. We mention some results for 1-dimensional maps.

Consider $f_a(x) = 1 - ax^2, x \in [-1,1], a \in [0,2]$. Clearly, f_a is not uniformly expanding for any a. In fact, near the critical point, f transforms any bounded density in one step to a density with inverse square-root singularity. One of the first results on the existence of *acim's* for maps of this kind is the following.

Theorem 5.3.1. [Ru2] *Let f be a quadratic map as above, with its critical point at c. Suppose that for some $k \in \mathbb{Z}^+, f^k c = x_0$, where x_0 is an expanding periodic point, i.e. $\exists\, p \in \mathbb{Z}^+$ s.t. $f^p x_0 = x_0$ and $|(f^p)'x_0| > 1$. Then f admits an acim.*

Idea of Proof. Away from c, f is essentially expanding, so we need only be concerned with x near c. For x near c, we have $|f'(x)| \sim |x-c|$ and $|f(x) - f(c)| \sim |x-c|^2$. Suppose for simplicity that $k = 1$ and $|(f^p)'x_0| = \lambda^p$ for some $\lambda > 1$. Then the orbit of $f(x)$ will stay near the orbit of x_0 for n times, where n is determined by $\lambda^n|x-c|^2 \sim 1$. Hence $|(f^{n+1})'x| \sim |x-c|\lambda^n \sim \frac{1}{|x-c|} \sim \lambda^{\frac{1}{2}n}$, so f^n is expanding in a neighborhood of c afterall.

The result in Theorem 5.3.1 has been extended a number of times. One intermediate step is the "Misiurewicz case", where $\{f^n c, n \geq 1\}$ is bounded away from c. Under suitable conditions this implies that the forward orbit of c is trapped in an expanding invariant set. (See [Mi].) Around 1980, Jakobson proved a theorem that tells us that in the quadratic family, it is in fact "normal" to admit an *acim.*

Theorem 5.3.3. [J] *There is a positive Lebesgue measure set Δ in parameter space s.t. $\forall\, a \in \Delta, f_a$ admits an acim.*

On a very basic level, the reasoning we used in Theorem 5.3.1 is valid when the critical orbit does not approach the critical point *too closely too soon*, and the parameters in Δ are those for which this approach is sufficiently controlled. To say all this precisely involves a nontrivial amount of bookkeeping which we will omit due to limited space.

There are now several proofs of this important theorem; see e.g. [BC1] and [Ry]. The list of related results is too long to enumerate, so let me just mention a small (random) sample: [CE], which first used a positive exponent along the critical orbit as a criterion for the existence of *acim*; [NS], which gives perhaps the weakest known condition for the existence of *acim*; [T] and [TTY], which formulate some of these results in the context of generic 1-parameter families; and [Re], which contains a complex version of Jakobson's thoerem.

The fact that Δ does not fill up all of parameter space is not a weakness of Jakobson's Theorem. It is well known that for a in many parameter intervals, f_a

has an attractive periodic orbit. (It has been announced recently by G. Swiatek and independently M. Lyubich that the set of parameters with this property is dense.) The point we wish to make here is that by casually inspecting a given map, it is usually impossible to determine whether or not it admits an *acim* – except when the map is uniformly expanding. In general, there is a delicate interplay between the expanding and contracting parts, in a way that is reminiscent of (and in fact intimately connected to) the problems one has when trying to decide on the positivity of Lyapunov exponents. (See Section 2.3.)

6. Sinai-Ruelle-Bowen Measures

6.1. Natural invariant measures for attractors. The setting in this section is as follows: U is an open set in a manifold or Euclidian space, $f : U \circlearrowleft$ is a diffeomorphism that takes U into itself, and $\Lambda \subset U$ is a compact f-invariant set such that $\forall\, x \in U, f^n x \to \Lambda$ as $n \to \infty$. We call Λ an *attractor* and U its *basin of attraction*. We have in mind a genuinely dissipative situation, and think of the dynamics of $f|\Lambda$ as "chaotic" – although there is no need to assume either one of these conditions.

The following are common practices:

(1) An experimenter assumes that the time evolution of some physical process leads to a strange attractor Λ. To find the average value of an observable φ on Λ, he or she starts off the process at some initial condition x_0, runs the experiment up to time T, measures $\varphi(f^t x_0)$, and concludes that $\frac{1}{T}\int_0^T \varphi(f^t x_0)dt$ is approximately the "average" of φ on Λ.

(2) To make a computer picture of an attractor for a map f, one often picks some convenient starting point x_0 in U, plots $x_i = f^i x_0$ for $i <$ some reasonably large n, and calls the resulting plot "the picture" of Λ.

In both of these scenarios an underlying assumption seems to be that *for almost every initial condition, the orbit has a well defined distribution as* $n \to \infty$*, and this distribution is independent of the choice of initial condition.* More precisely, it seems to be taken for granted that there is a probability measure μ such that for *a.e.* $x \in U$, $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i x} \to \mu$ as $n \to \infty$. In (1), the experimenter in effect measures the average of φ with respect to μ. In (2), the "picture" of Λ is the picture of the measure $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i x_0}$, which is an approximation of μ. We said "*a.e.* x"

because clearly not every orbit has the same asymptotic behavior (e.g. some orbits are periodic and others are not.) In the lack of any other reference measure, "*a.e.*" presumably refers to Lebesgue. Also, μ is automatically f-invariant (see Proposition 5.1.1).

If f is volume preserving and ergodic, then the assumption above is valid on account of the Birkhoff Ergodic Theorem. The situation is much more delicate in the dissipative case: what is at issue here is the existence of a singular invariant measure that is somehow able to influence the statistical behavior of Lebesgue-a.e. point in a much larger set. Mathematically there is no *a priori* reason why such an invariant measure should exist; but when it does, its significance from the physical standpoint is obvious.

For Anosov systems and Axiom A attractors, this very special invariant measure is known to exist. It is called a *Sinai-Ruelle-Bowen* or *SRB measure*. I will describe in detail the construction of this measure for Axiom A attractors and report on some of the progress since.

6.2. Axiom A attractors: geometric properties. For the geometric theory of uniformly hyperbolic sets see [Sm], [Sh] or the article by J.-C. Yoccoz in this volume.

Throughout Sections 6.2 and 6.3, let f be a C^2 diffeomorphism of a Riemannian manifold M, and let $\Lambda \subset M$ be a compact f-invariant set. We say that f is *uniformly hyperbolic* on Λ if there is a continuous splitting of the tangent bundle over Λ into a direct sum of two Df-invariant subbundles, written

$$T\Lambda = E^u \oplus E^s \,,$$

so that for all $x \in \Lambda$ and $n > 0$, the following hold:

$$v \in E^u(x) \Rightarrow |Df_x^{-n} v| \leq C\lambda^n |v|$$

and

$$v \in E^s(x) \Rightarrow |Df_x^{n} v| \leq C\lambda^n |v|$$

where $\lambda < 1$ and $C > 0$ are constants independent of x.

Definition 6.2.1. *We say that Λ is a uniformly hyperbolic attractor or an Axiom A attractor if*

1) *$f|\Lambda$ is uniformly hyperbolic;*
2) *$\exists$ a compact neighborhood U of f s.t. $fU \subset U$ and $\Lambda = \bigcap_{n \geq 0} f^n U$;*
3) *$f|\Lambda$ is topologically transitive.*

Let us assume throughout that Λ is not just a periodic sink, i.e. E^u is nontrivial. We do not rule out the Anosov case where $\Lambda = U = M$. The simplest example of an Axiom A attractor that is not Anosov is the “solenoid”. (The article by B. Kitchens in this volume has more details.) We mention also that either $f|\Lambda$ is topologically mixing (see Section 1.1), or by taking a power of f, Λ decomposes into N components on each one of which f^N is topologically mixing.

Recall the definitions of local stable and unstable manifolds from Section 3. Because we have a continuous splitting of $T\Lambda$ into $E^u \oplus E^s$ with uniform expansion and contraction, we are guaranteed continuous families of local stable and unstable manifolds on Λ. More precisely, let D be the unit disk having the same dimension as E^u, and let $Emb^r(D, M)$ denote the space of C^r embeddings of D into M. Let d^u denote distance along W^u-manifolds.

Theorem 6.2.2. *Let f be uniformly hyperbolic on Λ. Then*

(a) *$\exists \delta > 0$ and a continuous map $\Psi : \Lambda \to Emb^2(D, M)$ s.t. $\Psi(x) = W^u_\delta(x)$ where $W^u_\delta(x)$ is the local unstable manifold at x of radius δ;*

(b) *$\exists C' > 0$ and $\lambda' < 1$ s.t. $\forall\, x \in \Lambda$ and $\forall\, y \in W^u_\delta(x)$,*

$$d^u(f^{-n}x, f^{-n}y) < C'(\lambda')^n \;\forall\, n \geq 0.$$

Global stable and unstable manifolds are defined as in Section 3.3. The next two propositions describe first the local then the global structure of an Axiom A attractor in terms of its stable and unstable manifolds.

Proposition 6.2.3. (Local product structure) *Let $\delta > 0$ be as above. Then*

(a) *$\exists \varepsilon, 0 < \varepsilon < \delta$ s.t. $\forall\, x, y \in \Lambda$ with $d(x, y) < \varepsilon, W^u_\delta(x)$ meets $W^s_\delta(y)$ in exactly one point, which we call $[x, y]$;*

(b) *for every $z \in \Lambda$, the set $\{[x, y] : x \in W^s_\delta(z), y \in W^u_\delta(z), d(x, y) < \varepsilon\}$ contains an open neighborhood of z in Λ.*

Proposition 6.2.4.

(a) $\Lambda = \bigcup\limits_{x\in\Lambda} W^u(x)$.

(b) $U \subset \bigcup\limits_{x\in\Lambda} W^s(x)$.

Proof of Proposition 6.2.4. To prove (a), let $y \in W^u(x)$. If $y \notin \Lambda$, then $y \notin f^n U$ for all large n, or, equivalently, $f^{-n}y \notin U$ for all large n. This contradicts $d^u(f^{-n}x, f^{-n}y) \to 0$. (b) follows from the fact that $\bigcup\limits_{x\in\Lambda} W^s(x)$ contains an open neighborhood of Λ. That in turn is a consequence of $W^u_\delta(x) \subset \Lambda$ and the continuity of $x \mapsto W^s_\delta(x)$.

■

Except for the case when f is Anosov, Λ is a Lebesgue measure zero set (see e.g. [Bo]). Part (a) of the last proposition tells us that Λ is made up of W^u-leaves. The W^s-foliation, on the other hand, is a genuine foliation defined on the open set U. In the next subsection it will be important for us to know how smooth this foliation is. Now there are two different issues concerning the smoothness of a foliation: one is the smoothness of each leaf, and the other is how smoothly these leaves are packed together. For a uniformly hyperbolic set, the leaves are C^r if the map is C^r. These leaves, however, are usually not packed together very nicely.

The following notation will be used. Given a foliation $\mathcal{F}$ and and a transversal Σ, let m_Σ denote the Lebesgue measure induced on Σ, and let d_Σ denote the induced distance on Σ. For $x \in \Sigma$, let $B_{\Sigma_1}(x,\delta) := \{y \in \Sigma : d_\Sigma(x,y) < \delta\}$, and let $\mathcal{F}(x)$ denote the leaf of $\mathcal{F}$ through x.

Definition 6.2.5. *Let $\mathcal{F}$ be a continuous foliation with smooth leaves, and let Σ_1 and Σ_2 be two transversals to $\mathcal{F}$. A "Poincaré map" from Σ_1 to Σ_2 (if it can be defined) is a continuous map $\theta : \Sigma_1 \to \Sigma_2$ s.t. $\theta(x) \in \mathcal{F}(x) \cap \Sigma_2$. We say that $\mathcal{F}$ is absolutely continuous if for every choice of $(\Sigma_1, \Sigma_2; \theta)$, $m_{\Sigma_1}(A) = 0$ iff $m_{\Sigma_2}(\theta A) = 0$ for all $A \subset \Sigma_1$.*

Theorem 6.2.6. *The W^s-foliation of an Axiom A attractor for a C^2 diffeomorphism is absolutely continuous.*

Sketch of proof. In this proof, "$\approx$" means "equal up to a multiplicative constant", and "$\lesssim$" means "less a constant times". These constants are understood to be independent of point or iterate, although they do depend on the two transversals Σ_1 and Σ_2.

Let Σ_1 and Σ_2 be two C^2 transversals to W^s, and let θ be a Poincaré map from Σ_1 to Σ_2. Let $A \subset \Sigma_1$. We will assume that A is compact, and prove that $m_{\Sigma_2}(\theta A) \lesssim m_{\Sigma_1}(A)$. Let $\mathcal{O}$ be a small neighborhood of A in Σ_1 s.t. $m_{\Sigma_1}(\mathcal{O}) \le 2m_{\Sigma_1}(A)$, and let $\delta > 0$ be a small number. As n increases, $d(f^n x, f^n(\theta x)) \to 0$, and $B_{f^n\Sigma_1}(f^n x, \delta)$, and $\theta(B_{f^n\Sigma_1}(f^n x, \delta))$ become more and more parallel to each other. (We will use θ to denote the Poincaré maps from $f^n\Sigma_1$ to $f^n\Sigma_2$ as well as that from Σ_1 to Σ_2.) Let N be large enough that for every $x \in A$,

- $f^{-N} B_{f^N\Sigma_1}(f^N x, \delta) \subset \mathcal{O}$;
- $m_{f^N\Sigma_1} B_{f^N\Sigma_1}(f^N x, \delta) \approx m_{f^N\Sigma_2} \theta(B_{f^N\Sigma_1}(f^N x, \delta))$.

Let $\{B_1, \ldots, B_k\}$ be a covering of $f^N A$ by δ-balls centered at points in $f^N A$. By the Besicovitch Covering Lemma [G], we may assume that no point in $f^N\Sigma_1$ lies in more than C of the B_i's, where C is a number that depends only on the dimension of Σ_1 (and the extent to which $f^n\Sigma_1$ locally differs from Euclidean space).

$$Claim: \quad m_{\Sigma_1}(f^{-N}B_i) \approx m_{\Sigma_2}(\theta(f^{-N}B_i))\,\forall\, i$$

To prove the the claim, observe that

- $m_{f^N\Sigma_1}(B_i) \approx m_{f^N\Sigma_2}(\theta B_i)$,
- $\forall\, x \in f^{-N}B_i, |\det D(f^N|\Sigma_1)(x)| \approx |\det D(f^N|\Sigma_2)(\theta x)|$ (this is because $d(f^j x, f^j\theta x) \lesssim \lambda^j$, and $f^j\Sigma_1$ and $f^j\Sigma_2$ both become more and more parallel to W^u-manifolds);

- $\forall\, x_1, x_2 \in f^{-N}B_i, |\det D(f^N|\Sigma_1)(x)| \approx |\det D(f^N|\Sigma_2)(x_2)|$, and the same is true for pairs of points in $\theta(f^{-N}B_i)$ (this is the same distortion estimate as in (5.4.1); in addition we are using implicitly the uniform Lipschitzness of $x \mapsto D(f|f^j\Sigma_1)(x)) \;\forall\; j < N$.

Putting the last few lines together, we have

$$\begin{aligned} m_{\Sigma_1}(f^{-N}B_i) &\approx |\det D(f^{-N}|B_i)(x_i)| \cdot m_{f^N\Sigma_1}(B_i) \quad \text{for some } x_i \in B_i \\ &\approx |\det D(f^{-N}|\theta B_i)(\theta x_i)| \cdot m_{f^N\Sigma_2}(\theta B_i) \\ &\approx m_{\Sigma_2}\theta(f^{-N}B_i)\,. \end{aligned}$$

From here one concludes easily that

$$m_{\Sigma_2}(\theta A) \leq \sum_i m_{\Sigma_2}\theta(f^{-N}B_i) \approx \sum_i m_{\Sigma_1}(f^{-N}B_i) \leq 2C m_{\Sigma_1}(A)\,.$$

■

For a complete proof see e.g. [PS1].

We remark that a C^2 (or $C^{1+\alpha}$) assumption on f is needed to guarantee the absolute continuity of W^s. (See e.g. [RY].) On the other hand, increasing the smoothness of f will not improve the smoothness of the W^s-foliation.

6.3. Construction of SRB measures for Axiom A attractors. Throughout this section let m denote the Lebesgue or Riemannian measure on the manifold M. As before, if W is a submanifold of M, then m_W denotes the induced Riemannian measure on W.

Theorem 6.3.1. *Let $\Lambda \subset U$ be an Axiom A attractor of a C^2 diffeomorphism f. Then there exists a unique f-invariant Borel probability measure μ on Λ such that for every continuous function φ on U,*

$$\frac{1}{n}\sum_{i=0}^{n-1}\varphi(f^i(x)) \to \int \varphi d\mu$$

for m-a.e. x in U. An equivalent characterization of μ is that it has absolutely continuous conditional measures on unstable manifolds (see Section 4.1). Moreover, (f,μ) is ergodic, and if $f|\Lambda$ is topologically mixing then (f,μ) is isomorphic to a Bernoulli shift.

The results in Theorem 6.3.1 were first proved for Anosov diffeomorphisms by Sinai ([S1], [S3]). They were later generalized to Axiom A attractors by Ruelle and Bowen (see e.g. [Ru1], [BR] or [Bo]). In these proofs, Markov partitions were used to connect the dynamics of f to certain 1-dimensional lattice systems in statistical mechanics, and μ was realized as a Gibbs state or equilibirum state. We will not use Markov partitions in our proofs, both because Markov partitions are not necessary for our purposes, and because they are not available for systems more general than Axiom A. Other than that, our construction is quite similar to that of Sinai. See also [PS]. The Bernoulli property is proved in [A]. See also [OW].

We now present a proof of part of Theorem 6.3.1. We will prove everything up to the ergodicity of (f, μ). For the last assertion, see [Bo] or [L2].

Proof of Theorem 6.3.1

Step I. Construction of μ with absolutely continuous conditional measures on W^u-manifolds. Let Σ be a W^s_ε-disk, and let $\mathcal{N} = \mathcal{N}(\Sigma, \varepsilon) := \cup\{W^u_\varepsilon(x) : x \in \Lambda \cap \Sigma\}$. For small enough $\varepsilon, \mathcal{N}$ is topologically a product. In this proof we will refer to $\mathcal{N}$ as a "canonical neighborhood" in Λ and let $\mathcal{N} = \cup D_\alpha$ denote the partition of $\mathcal{N}$ into local unstable disks. The aim of Step I is to construct an invariant measure μ with the property that for every $\mathcal{N}$, if $\{\mu_\alpha\}$ is a canonical family of conditional measures on $\{D_\alpha\}$, then μ_α is equivalent to m_{D_α}.

Pick an arbitrary $x_0 \in \Lambda$ and let $L = W^u_{\ell oc}(x_0)$. Let μ_0 be m_L normalized. We transport μ_0 forward by f^i, and call the resulting measure μ_i. That is, μ_i lives on f^iL, and $\mu_i(E) = \mu_0(f^{-i}E)$ for all Borel sets E. Consider a canonical neighborhood $\mathcal{N}$. Let $\mathcal{N}_i$ be the union of D_α-disks in $\mathcal{N}$ that are completely contained in f^iL, and let $\hat{\mu}_i := I_{\mathcal{N}_i} \cdot \mu_i, I_E$ being the indicator function of the set E. We claim that $\mu_i(\mathcal{N}) - \hat{\mu}_i(\mathcal{N}) \to 0$ as $i \to \infty$. This is true because for $x \in L$ with $d^u(x, \partial L) > \delta C' {\lambda'}^i, f^ix \in D_\alpha \Rightarrow D_\alpha \subset f^iL$. (We are assuming of course that ε is smaller than the δ in Theorem 6.2.3.)

To ensure that we obtain an invariant measure, consider the averages $\{\frac{1}{n}\sum\limits_{i=0}^{n-1} \mu_i\}$ and let μ be an accumulation point. Then for every $\mathcal{N}$ the corresponding subsequence of $\frac{1}{n}\Sigma\hat{\mu}_i$ will also converge to μ. We argue that μ has the desired conditional measures on D_α. Recall the proof of the existence of smooth invariant measures for expanding maps in Section 5.2. A virtually identical argument can be carried out for $f|W^u$, because restricted to W^u-manifolds f is uniformly expanding and has uniformly bounded C^2 derivatives. The distortion lemma (Lemma 5.2.2) guarantees that $\exists \alpha, \beta > 0$ (independent of i) s.t. on each $D_\alpha \subset \mathcal{N}_i$, we have

$$\alpha \leq \frac{d\hat{\mu}_i}{dm_{D_\alpha}} \leq \beta \,.$$

These bounds are also valid for $\frac{1}{n}\Sigma\hat{\mu}_i$, and so are passed on to all accumulation points. (To see more details for this last step, do as in Example 1.3.1.)

We make some general definitions before going on to Step II. Let ν be a probability measure on M. We say that $x \in M$ is *future generic with respect to ν*

if for every continuous function $\varphi : M \to \mathbb{R}$, $\frac{1}{n}\sum\limits_{0}^{n-1}\varphi(f^i x) \to \int \varphi d\nu$ as $n \to \infty$. We say that x is *past generic wrt* ν if $f^i x$ above is replaced by $f^{-i}x$. The word *generic* is used to mean both future and past generic.

From the Ergodic Theorem, we know that if ν is an ergodic measure, then ν-a.e. x is generic wrt ν. Also, using the Ergodic Decomposition Theorem (Theorem 5.1.2), we know that for every invariant measure ν, ν-a.e. x is generic wrt some ergodic measure, which we call $\nu_{(x)}$.

We now return to the setting of Theorem 6.3.1. Recall that μ is the measure constructed in Step I, while m is Lebesgue measure on the manifold M.

Step II. "Local ergodicity" of (f, μ). Let x be a density point of μ. The aim of this step is to prove that there is a neighborhod V of x in M and an ergodic measure ν (depending on x) s.t.

(i) μ-a.e. $z \in V$ is generic wrt ν;
(ii) m-a.e. $z \in V$ is future generic wrt ν.

We remark that the ergodicity of (f, μ) is equivalent to μ-a.e. z being generic wrt the same measure; hence (i) can be interpreted as a statement of "local ergodicity". Assertion (ii) is of a different nature; it is about the relation between m and ν.

The following is a slight modification of a very standard argument due to Hopf. Let $\mathcal{N}$ be a canonical neighborhood centered at x. We disintegrate μ on $\mathcal{N} = \cup D_\alpha$ and let W be one of the D_α-disks with the following two properties: W is near enough to x that $x \in \bigcup\limits_{y\in W} W^s_\delta(y)$; and m_W-a.e. $y \in W$ is generic wrt some $\mu_{(y)}$. (We are using here the fact that the conditional measures of μ on $\{D_\alpha\}$ are equivalent to m_{D_α}.) For $y, y' \in W$, since $d(f^{-n}y, f^{-n}y') \to 0$ as $n \to \infty$, we must have $\mu_{(y)} = \mu_{(y')}$. Call this common ergodic measure ν. Similar reasoning tells us that for m_W-a.e. y, if $z \in W^s_\delta(y)$ then z is future generic wrt ν.

We now use the absolute continuity of W^s (Theorem 6.2.6) to conclude that for every $D_\alpha \subset \mathcal{N}$, m_{D_α}-a.e. $z \in (D_\alpha \cap \bigcup\limits_{y\in W} W^s_\delta(y))$ is future generic wrt ν. Passing from m_{D_α} back to the conditional measures of μ on D_α completes the proof of (i).

The proof of (ii) is similar, except that we should first foliate an entire neighborhood of x in M by disks that are roughly parallel to W^u. No invariance properties are required of these disks; we only ask that they form a smooth foliation. The absolute continuity of W^s is then used between W and every one of these disks.

Step III. (Global) ergodicity of (f, μ) *and the orbit distribution of* m*-a.e.* $z \in U$. As in Step II, we need to show

(i) μ-a.e. z is generic wrt μ;
(ii) m-a.e. $z \in U$ is future generic wrt μ.

The argument is essentially local ergodicity combined with topological transitivity. More precisely, let V and V' be two open sets on which (i) and (ii) in Step II hold for the ergodic measures ν and ν' respectively. Using the topological transitivity of $f|\Lambda$, one can choose n s.t. $f^nV \cap V' \neq \phi$. Since points in $f^nV \cap V'$ are future generic to both ν and ν', we have $\nu = \nu'$.

The proof of (ii) is similar, except that here we take V to be an open set as in the last paragraph and V' to be an arbitrary open set intersecting Λ. It is convenient to take V' to be of the form $\bigcup_{y\in W} W^s_\delta(y)$. Choose n s.t. f^nV crosses V' completely in the W^u-direction, meaning that f^nV meets every W^s-leaf in V'. This forces m-a.e. z in V' to be future generic wrt μ. Once we have this property in a neighborhood of Λ, it is easily passed on to all of U.

■

6.4. The nonuniform case. The purpose of this subsection is to discuss some possible generalizations of Theorem 6.3.1 to attractors that are not Axiom A.

A. The role of nonuniform hyperbolic theory (as described in Sections 2,3 and 4)

We saw in Sections 2, 3 and 4 that if f is a diffeomorphism preserving a probability measure μ with no zero exponents, then stable and unstable manifolds are defined $a.e.$ and it makes sense to talk about whether or not μ has absolutely continuous conditional measures on W^u. We did not mention this in Section 3, but W^s is in fact absolutely continuous in the nonuniform setting, in the sense that restricted to a fixed Λ_n (see Section 3.1), local stable manifolds of a fixed size are guaranteed and Poincaré maps between transversals to these $W^s_{\ell oc}$-manifolds are absolutely continuous. (See e.g. [PS2].)

It follows from general theory then that if f admits an ergodic measure μ with no zero exponents *and* with absolutely continuous conditional measures on W^u, then using the absolute continuity of W^s we have that the set of points that are future generic wrt μ is a positive Lebesgue measure set. *What nonunifom hyperbolic theory does not tell us is under what conditions f will admit an invariant measure with these properties.* In other words, without the Axiom A assumption, there is no guarantee that Step I in the proof of Theorem 6.3.1 will work.

For purposes of the present discussion let us agree to call a measure with the properties in the last paragraph an *SRB* measure (even though the definition here is more restrictive than the one we used in Section 4.) The question of existence of SRB measures is a very delicate one. It is analogous to the problem of existence of absolutely continous invariant measures for nonuniformly expanding maps (see Section 5.3) – only that it is more complicated because one has to consider not only rates of expansion but their directions as well.

B. Against the existence of SRB measures ....

First let us understand why Step I in the proof of Theorem 6.3.1 may fail without the assumption of uniform hyperbolicity. Suppose we pick an arbitrary

$W^u_{\ell oc}$-leaf L and push forward Lebesgue measure on L. The definition of a local unstable manifold says that $x, y \in L \Rightarrow d(f^{-n}x, f^{-n}y) \to 0$ as $n \to \infty$, but there is no stipulation on what happens to points in L under forward iterates. It is easy to imagine, for instance, that parts of L may, under forward iterations of f, fall into the basins of periodic sinks.

The concern we have just expressed is not unfounded, for it is well known that periodic sinks are easily created near homoclinic tangencies. In fact, according to the theory of Newhouse [N], under fairly mild conditions and assuming that the map is area contracting, a generic diffeomorphism has infinitely many sinks.

Another scenario which may cause Step I in the proof of Theorem 6.3.1 to fail is the presence of nonhyperbolic fixed points. In [HY] the authors consider surface diffeomorphisms that are hyperbolic everywhere except at one fixed point, where the larger eigenvalue is equal to 1 and the smaller one is < 1. Pushing forward Lebesgue measure on a piece of $W^u_{\ell oc}$-leaf results in the Dirac measure at this point.

C. Some results

The Hénon mappings are a 2-parameter family of maps of the plane given by

$$T_{a,b} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix} .$$

Numerical studies suggest that for parameter values near $a = 1.4$ and $b = 0.3$, $T_{a,b}$ appears to have an attractor with an SRB measure. I know of no rigorous results in this parameter range. Near $a = 2$ and $b = 0$ Benedicks and Carleson developed an elaborate machinery for analyzing the dynamics of $T_{a,b}$ for a positive measure set of parameters, and Benedicks and Young subsequently constructed SRB measures for these attractors. We summarize their results:

Theorem 6.4.1. ([BC], [BY]) *Let $T_{a,b}$ be as above. Then*

(1) *there is a rectangle $\Delta = (a_0, a_1) \times (0, b_1)$ in parameter space s.t. $\forall (a, b) \in \Delta, T = T_{a,b}$ has an attractor Λ;*
(2) *for all sufficiently small $b > 0, \exists$ a positive measure set Δ_b s.t. $\forall a \in \Delta_b, T = T_{a,b}$ admits an SRB measure μ; moreover, μ is unique, it is supported on the entire attractor, and (T, μ) is isomorphic to a Bernoulli shift.*

An immediate consequence of (2) is that if (a, b) is a “good” parameter, then a positive Lebesgue measure set of points in $\mathbb{R}^2$ is future generic wrt μ.

The first assertion is quite easy, and it is just as easy to convince oneself that even though T has invariant cones away from some neighborhood of the y-axis, the attractor cannot possibly be Axiom A. A lot of the delicate analysis goes into proving that restricted to the unstable manifold W of a hyperbolic fixed point, T behaves very much like 1-dimensional maps with quadratic singularities. This is only proved for certain “good” parameters for which a certain inductive argument can be carried through. (Note that for $b = 0, T_{a,b}$ is essentially f_a, the 1-dimensional maps discussed in Section 5.3.) SRB measures are constructed by pushing forward Lebesgue measure on some small piece of W – let us call it L. What is different from the Axiom A proof is that here uniform expansion is proved only for a positive percentage of the iterates, and even then only for a positive measure subset of points in L. While the results in [BC] and [BY] are stated only for the Henon family, they clearly hold for families with similar qualitative properties.

Hénon-type attractors have been shown to appear in certain homoclinic bifurcations [MV]. (See the article by M. Viana in this volume.) These attractors too have SRB measures for the same reasons as above. Newhouse and Yakobson have recently announced that they are able to duplicate most of the results in Theorem 6.4.1.

D. Outlook

Very little is known about how often nonuniformly hyperbolic attractors admit SRB measures; the results in Section C are the first and essentially only ones. In some sense, this is the dissipative version of a question we asked in Section 2: how prevalent are positive Lyapunov exponents for, say, area preserving maps? Both problems require that one prove sustained exponential growth in the face of possible cancellations, and techniques are not well developed for handling either.

A generic theorem in the future might look something like this: "Consider a typical k-parameter family of maps with attractors. Suppose that these maps are hyperbolic on most (though not all) of phase space. Then for a positive measure set of parameters the attractor has an SRB measure." "Typical" might mean satisfying certain transversality conditions, and the stronger the hyperbolicity observed, the bigger the set of good parameters. I can also imagine for area preserving maps similar statements about positive Lyapunov exponents on positive measure sets.

That leads to another point I would like to make. The uniform case excepted, SRB measures and positive exponents are, in general, not likely to be open conditions. For 1-dimensional maps we have already seen how different types of dynamical behaviors are delicately intertwined. Even if we accept the picture in the last paragraph, positive exponents for area preserving maps are likely to co-exist side by side with elliptic islands, and SRB measures are likely to share the stage with infinitely many sinks. In the absence of invariant cones, it seems nearly impossible to know for certain if a given dynamical system admits an SRB measure or has positive Lyapunov exponents.

A more realistic goal might be some sort of approximate knowledge, the kind of theorems that would enable one to say – based on an understanding of qualitative properties rather than complete knowledge of fine details – something like: "this diffeomorphism has a 99% chance of having a Lyapunov exponent of order $\log x$ on at least 99% of this region of its phase space ..... ", or something along these lines.

Acknowledgement. I am grateful to L. Fahlberg-Stojanovska and F. Ledrappier for reading over this manuscript. Their comments are much appreciated. I would like to thank M. Wojtkowski for helpful conversations, S. Sutton for her excellent typing, and P. Fan for general assistance. I wish also to acknowledge the hospitality of the IHES, where part of this manuscript was written.

References

[A] Azencott, R., Diffeomorphismes d'Anosov et schemas de Bernoulli, CRAS Paris **A270** (1970) 1105-1107.

[BB] Ballman, W. and Brin, M., On the ergodicity of geodesic flows, Ergod. Th. and Dyn. Sys. **2** (1982) 311-315. (Note: gap in proof of ergodicity).

[BC1] Benedicks, M. and Carleson, L., On iterations of 1-ax^2 on (-1,1), Annals of Math. **122** (1985) 1-25.

[BC2] _______ and _______, The dynamics of the Henon map, Ann. Math. **133** (1991) 73-169.

[BY] _______ and Young, L.-S., SBR measures for certain Henon maps, Inventiones Math. **112** (1993) 541-576.

[Bi] Billingsley, P., Ergodic theory and information, John Wiley and Sons, Inc., New York-London-Sydney (1965).

[Bo] Bowen, R., *Equilibrium states and the ergodic theory of Anosov diffeomorphisms,* Springer Lectures Notes in Math. **470** (1975).

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

[BrK] Brin, M. and Katok, A., On local entropy, *Geometric Dynamics,* Springer Lecture Notes **1007** (1983) 30-38.

[BuK] Burns, K. and Katok, A., Infinitesimal Lyapunov functions, invariant cone families, and stochastic properties of smooth dynamical systems, to appear in Erg. Th. & Dynam. Sys.

[CE] P. Collet and J. P. Eckmann, Positive Lyapunov exponents and absolutely continuity, *Ergod. Th. and Dynam. Syst.* **3** (1983), 13-46.

[ER] Eckmann, J.-P. and Ruelle, D., Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57** (1985) 617-656.

[FHY] Fathi, A., Herman, M. and Yoccoz, J.-C., A proof of Pesin's stable manifold theorem, Springer Lecture Notes in Math. **1007** (1983) 117-215.

[Fa] Falconer, K., *Fractal Geometry, mathematical foundations and applications*, Wiley (1990).

[Fu] Furstenberg, H., Noncommuting random products, Trans. AMS **108** (1963) 377-428.

[G] M. de Guzmán, *Differentiation of Integrals in R^n*, Springer Lecture Notes in Math. **481** (1975), 2.

[GM] Goldsheid, I. and Margulis, G., Lyapunov indices of a product of random matrices, Russ. Math. Surveys **44**:5 (1989) 11-71.

[GR] Guivarc'h, Y. and Raugi, A., Frontière de Furstenberg, propriétés de contraction et theorèmes de convergence, Z. Wahrscheinlichkeitstheorie verw. Geb. **69** (1985), 187-242.

[HP] M. W. Hirsch and C. Pugh, Stable manifolds and hyperbolic sets, Proc. Sym. in Pure Math. **14**, A.M.S., Providence, RI (1970).

[HY] Hu, H. and Young, L.-S., Nonexistence of SBR measures for some systems that are "almost Anosov", to appear in Erg. Th. & Dyn. Sys.

[J] Jakobson, M., Absolutely continuous invariant measures for one-parameter families of one-dimensional maps, Commun. Math. Phys. **81** (1981) 39-88.

[Ka] Katok, A., Lyapunov exponents, entropy and periodic orbits for diffeomorphisms, Publ. Math. IHES **51** (1980) 137-174.

[KaS] _______ and Strelcyn, J.M., *Invariant manifolds, entropy and billiards; smooth maps with singularities,* Springer Lecture Notes in Math. 1222 (1986).

[KrS] Krzyzewski, K. and Szlenk, W., On invariant measures for expanding differentiable mappings, Studia Math. **33** (1969), 83-92.

[Ki] Kingman, C., Subadditive processes, Springer Lecture Notes in Math. **539** (1976).

[L1] Ledrappier, F., Preprietes ergodiques des mesures de Sinai, Publ. Math. IHES **59** (1984) 163-188.

[L2] _______, Quelques propriétés des exposants caractéristiques, Springer Lecture Notes in Math **1097** (1984), 305-396.

[L3] _______, Dimension of invariant measures, Teubner-Texte zur Math. **94** (1987) 116-124.

[LS] _______ and Strelcyn, J.-M., A proof of the estimation from below in Pesin entropy formula, Ergod. Th. & Dynam. Sys. **2** (1982) 203-219.

[LY] _______ and Young, L.-S., The metric entropy of diffeomorphisms, Annals of Math. **122** (1985) 509-574.

[M1] Mañé, R., A proof of Pesin's formula, Ergod. Th. & Dynam. Sys. **1** (1981) 95-102.

[M2] _______, On the dimension of compact invariant sets of certain non-linear maps, Springer Lecture Notes in Math. **898** (1981) 230-241 (see erratum).

[M3] _______, *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag (1987).

[Mi] M. Misiurewicz, Absolutely continuous measures for certain maps of the interval, *Publ. Math. IHES* **53** (1981), 17-51.

[MV] Mora, L. and Viana, M., Abundance of strange attractors, to appear in Acta Math.

[N] Newhouse, S., Lectures on Dynamical Systems, Progress in Math. **8**, Birkhauser (1980) 1-114.

[NS] Nowicki, T. and van Strein, S., Absolutely continuous invariant measures under a summability condition, Invent. Math **105** (1991) 123-136.

[O] Oseledec, V. I., A multiplicative ergodic theorem: Liapunov characteristic numbers for dynamical systems, Trans. Moscow Math. Soc. **19** (1968) 197-231.

[OW] Ornstein, D. and Weiss, B., Geodesic flows are Bernoullian, Israel J. Math **14** (1973), 184-198.

[P1] Pesin, Ya. B., Families if invariant manifolds corresponding to non-zero characteristic exponents, Math. of the USSR, Izvestjia **10** (1978) 1261-1305.

[P2] _______, Characteristic Lyapunov exponents and smooth ergodic theory, Russ. Math. Surveys **32** (1977) 55-114.

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

[PS1] Pugh, C. and Shub, M., Ergodicity of Anosov actions, Inventiones Math. **15** (1972) 1-23.

[PS2] _______ and _______, Ergodic attractors, Trans. AMS, Vol. 312, No. 1 (1989) 1-54.

[Re] Rees, M., Positive measure sets of ergodic rational maps, *Ann. Scient. Ec. Norm. Sup.* 4^0 série **19** (1986) 93-109.

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

[Ro1] V. A. Rohlin, On the fundamental ideas of measure theory, A.M.S. Transl. (1) **10** (1962) 1-52.

[Ro2] _______, Lectures on the theory of entropy of transformations with invariant measures, Russ. Math. Surveys **22**:5 (1967) 1-54.

[Ru1] Ruelle, D., A measure associated with *Axiom A* attractors, Amer. J. Math. **98** (1976) 619-654.

[Ru2] _______, Applications conservant une measure absolument continue par raport a dx sur [0, 1], Commun. Math. Phys. **55** (1987) (47-52).

[Ru3] _______, An inequality of the entropy of differentiable maps, Bol. Sc. Bra. Mat. **9** (1978) 83-87.

[Ru4] _______, Ergodic theory of differentiable systems, Publ. Math. IHES **50** (1979) 27-58.

[Ru5] _______, The thermodynamics formalism for expanding maps, Commun. Math. Phys., Vol. 125 (1989) 239-262.

[Ry] Rychlik, M., A proof of Jakobson's theorem, Ergod. Th. and Dyn. Sys. **8** (1988) 93-109.

[Sh] Shub, M., *Global Stability of Dynamical Systems*, Springer (1987).

[S1] Sinai, Ya. G., Markov partitions and C-diffeomorphisms, Func. Anal. and its Appl. **2** (1968) 64-89.

[S2] _______, Dynamical systems with elastic reflections: ergodic properties of dispersing billiards, Russ. Math. Surveys **25**, No. 2 (1970) 137-189.

[S3] _______, Gibbs measures in ergodic theory, Russ. Math. Surveys **27** No. 4 (1972) 21-69.

[S4] _______, *Introduction to Ergodic Theory*, Princeton Univ. Press (1976).

[S5] _______ (ed.), *Dynamicals systems II,* Encyclopaedia of Math. Sc. Vol. **2**, Springer-Verlag (1989).

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

[T] Tsujii, M., Positive Lyapunov exponents in families of 1-dimensional systems, Invent. Math. **11** (1993) 113-117.

[TTY] Thieullen, P., Tresser, C. and Young, L.-S., Positive Lyapunov exponent for generic 1-parameter families of unimodal maps, C.R. Acad. Sci. Paris, t. 315 Série I(1992) 69-72; longer version to appear in J. d'Analyse.

[Wa] Walters, P., *An introduction to Ergodic Theory*, Grad. Texts in Math., Springer-Verlag (1981).

[W1] Wojtkowski, M., Invariant families of cones and Lyapunov exponents, Erg. Th. & Dyn. Sys. **5** (1985) 145-161.

[W2] _______, Principles for the design of billiards with nonvanishing Lyapunov exponents, Commun. Math. Phys. **105** (1986) 391-414.

[W3] _______, Systems of classical interacting particles with nonvanishing Lyapunov exponents, Springer Lecture Notes in Math **1486** (1991).

[Y] Young, L.-S., Dimension, entropy and Lyapunov exponents, Ergod. Th. & Dynam. Sys. **2** (1982) 109-129.