THE PESIN ENTROPY FORMULA

MARYAM CONTRACTOR

Abstract. We provide necessary background and results from hyperbolic dynamical systems and smooth ergodic theory to prove the Pesin entropy formula. We then briefly describe SRB measures.

Contents

1. Introduction

Dynamical systems theory studies ideas of structure-preserving operators. Thus, it is well-suited to applications in the physical sciences and beyond.

Arising in the late 19th century as the study of stable and unstable fixed points on a surface, dynamical systems theory evolved towards concentrating on *flows*, or structure-preserving maps. This study accelerated towards the mid-to-late 20th century and eventually resulted in ideas surrounding structural stability and hyperbolicity, which breaks down movement of a system into its stable and unstable parts [1].

In the 21st-century, work in information theory, coupled with preexisting dynamics study, led to proofs of the Margulis-Ruelle inequality and the Pesin entropy formula. These results describe that entropy of a dynamical system under certain conditions can be measured as the sum of the rates of divergence (or expansion) in that system. These ideas are discussed extensively in Section 7 and their applications in Section 8.

Date: August 23, 2023.

What follows are the required definitions and results necessary to rigorously prove Pesin's formula, alongside a few examples pertaining to measure and ergodic theory.

2. Measure Theory & Entropy

To understand the outcome(s) of a changing system, it is necessary to introduce the concept of a *probability measure*. Probability measures allow us to encode information about the outcome(s) of a given event. We use definitions from [2].

Definition 2.1. Let X be a set. A *σ-algebra* $\mathcal{A}$ is a collection of subsets of X that has the following properties:

(1) $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$
(2) if $A \in \mathcal{A}$ then $(\mathcal{A} \setminus A) \in \mathcal{A}$
(3) if $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

A pair $(X, \mathcal{A})$ is a *measurable space*, and elements of $\mathcal{A}$ are denoted as *measurable sets*.

Definition 2.2. Let X be a set and $\mathcal{A}$ be a σ-algebra on X. Then $\mathcal{A}$ is the *Borel σ-algebra* of X if $\mathcal{A}$ is comprised of all of the open sets of X, and a *Borel set* is an element of the Borel algebra $\mathcal{A}$.

We use σ-algebras to define which parts of an arbitrary set on which we can provide a well-defined measure.

Definition 2.3. Given a measureable space $(X, \mathcal{A})$, we define a *measure* as a function $\mu : \mathcal{A} \to [0, \infty]$ such that

(1) $\mu(\emptyset) = 0$
(2) The sum of the measure of pairwise disjoint sets is the measure of their sums, i.e. $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(B_i)$

One important measure is titled the Lebesgue measure. We will not provide the formal definition, but intuitively, Lebesgue measure provides a metric of volume in $\mathbb{R}^n$.

We call $(X, \mathcal{A}, \mu)$ a *measure space*. In the case where the measure of the whole space is 1, i.e. $\mu(X) = 1$, the triplet is called a *probability space*.

Example 2.4. To gain intuition for probability spaces, consider a coin flip. The two possible outcomes, heads and tails, are represented by X, so $X = \{\text{heads}, \text{tails}\}$. We may define $\mathcal{A}$ as $\{\text{heads}, \text{tails}, \emptyset, \{\text{heads}, \text{tails}\}\}$. Finally, μ is defined as follows: $\mu(\text{heads}) = 0.5$, $\mu(\text{tails}) = 0.5$, $\mu(\emptyset) = 0$, $\mu(\{\text{heads}, \text{tails}\}) = 1$. With $(X, \mathcal{A}, \mu)$ defined, we represent the probability of each outcome in the entire sample space X as a function of μ. Probability measures act similarly on more complex sample spaces; they are functions to represent all of the possible outcomes and their likelihood.

Definition 2.5. Given X and its σ-algebra $\mathcal{A}$, a measure μ is *absolutely continuous* with respect to another measure ν if $\mu(A) = 0$ when $\nu(A) = 0$ for $A \in \mathcal{A}$.

Absolute continuity lays the foundation for the first fundamental theorem of calculus to hold. Similarly, absolute continuity provides a foundation for the Radon-Nikodym Theorem.

Definition 2.6. Given X and its σ-algebra $\mathcal{A}$, a measure μ is *σ-finite* if X is the countable union of subsets $A_i \in \mathcal{A}$, where $\mu(A_i) < \infty$.

Theorem 2.7 (Radon-Nikodym)**.** *Let μ be a σ-finite positive measure on a measurable space $(X, \mathcal{A})$ and let ν be another measure on $(X, \mathcal{A})$ such that ν is absolutely continuous with respect to μ. Then there exists a μ-integrable non-negative function f which is measurable with respect to $\mathcal{A}$ s.t.*

$$\nu(A) = \int_A f d\mu$$

By the fundamental theorem of calculus and the Radon-Nikodym Theorem, there exists a function $\frac{d\mu}{d\nu}$ which can equate the two measures, called the *Radon Nikodym derivative*. The proof can be found in [2].

With measures defined, we are equipped to quantify change in a given system, which we represent using *mappings*.

Definition 2.8. If $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ are probability spaces, then a mapping $T : X_1 \to X_2$ is *measurable* if for all $A \in \mathcal{A}_2, T^{-1}A \in \mathcal{A}_1$. Measurable mappings are *measure preserving* if for all $A \in \mathcal{A}_2, \mu_1(T^{-1}A) = \mu_2(\mathcal{A}_1)$. If T is a bijective measure preserving transformation and T^{-1} is also measure preserving, then T is *invertible*.

Most often we will work with measure preserving transformations (or *mpt*s) which are operators, or send a probability space to itself.

Now assume we have a well-defined mpt. We want to build some intuition for the outcome of the mpt under some given initial condition. The equivalent way to represent this question is that we want to know the quantity of information we lack prior to the event occurring. This concept is defined as *entropy*. For example, let's return to our coin flip example. Because there are only two outcomes (heads or tails), we can represent the outcome of n coin flips using only a binary code. Hence, the information gained per coin flip is 1 bit. We measure entropy using the following as a tool:

Definition 2.9. A *partition* α of a space X comprises of finitely many sets, each of which are pairwise disjoint, and whose union equals X. The *join* of two partitions α, β is $\alpha \vee \beta := \{A \cap B : A \in \alpha, B \in \beta\}$. We can also express the join of many partitions $\alpha_1, \dots, \alpha_n$ as $\bigvee_{i=1}^{n} \alpha_i$. Finally, $T^{-1}\alpha := \{T^{-1}A, A \in \alpha\}$.

Then we define entropy as the following.

Definition 2.10. Let $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ be an mpt and let α be a partition of X. Suppose for each $A_i \in \mathcal{A}$, $\mu(A_i) = p_i$. Then denote the *probability vector* of T as the n-tuple of probabilities for each event in $\mathcal{A}$, i.e. $(p_1, p_2, \dots, p_n)^T$. Define $H = -\sum p_i \log p_i$ (assume $0 \log 0 = 0$). Then the entropy h for α is defined as:

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

The entropy over all partitions $h(T)$ is the supremum over the entropies for each partition.

The impracticality involved in taking the supremum over all partitions makes the Shannon-Breiman-McMillan Theorem useful. Before introducing it, we discuss relevant properties of a dynamical system.

Definition 2.11. An mpt $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is *ergodic* if the following property holds: for any $A \in \mathcal{A}$ such that $T^{-1}A = A$, then $\mu(A) = 0$ or $\mu(A) = 1$. Equivalently, invariant functions on X are constant almost everywhere on X.

Remark 2.12. *Almost everywhere*, also denoted as *a.e.*, is defined to be everywhere except at sets with measure 0.

Now, let T be as defined in Definition 2.11.

Definition 2.13. T (or the measure μ under T) is *mixing* if for any $A_1, A_2 \in \mathcal{A}$, $\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$.

Mixing can intuitively summarize to: given two sets in space, the probability of their intersection after infinite iterations will approach their independent probabilities. Even more simply, the elements in a set A will be equidistributed–or "mixed"–in another set B after infinite iterations. Using this intuition, we then have that the only sets which remain exactly the same after just one iteration are trivial, either having a measure of zero or a measure of one. Thus, mixing is a stronger property than ergodicity. We also see that Bernoulli systems (those isomorphic to a Bernoulli shift, see Definition 3.1), are all mixing.

Theorem 2.14 (Shannon-Breiman-McMillan)**.** *Given an ergodic mpt* $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$*, almost everywhere we have*

$$\sup_{\epsilon>0} \limsup_{n\to\infty} \frac{-1}{n} \log \mu A(x, \rho_\epsilon; n) = h(T) \tag{2.15}$$

where $\{\rho_\epsilon\}$ *is comprised of functions such that* $0 < \rho_\epsilon \leq \epsilon$ *and* $\int -\log \rho_\epsilon d\mu < \infty$.

We summarize this to: the entropy of an ergodic mpt is the asymptotic rate of decrease of the mpt at individual points. The proof can be found in [3]. We also state a relevant theorem called the *Poincare recurrence theorem.*

Theorem 2.16 (Poincare Recurrence)**.** *Let* $(X, \mathcal{A}, \mu)$ *be a measure space and let* f *be an mpt on* $(X, \mathcal{A}, \mu)$*. Then for every* $A \in \mathcal{A}$*, the set*

$$\{x \in A \mid \text{there exists } N \text{ s.t. } f^n(x) \notin A, \text{for all } n > N\}$$

has measure zero.

The proof can be found in [4]. Poincare Recurrence states that an mpt will return to its starting point infinitely, except at measure zero. Maps which define specified points' return to their initial states are titled *Poincare maps*, and are useful to break down systems. In addition, Egorov's theorem will be helpful.

Theorem 2.17 (Egorov's)**.** *Let* $(X, \mathcal{A}, \mu)$ *be a measure space and let* $E \in \mathcal{A}$ *s.t.* $\mu(E) < \infty$*. Let* $\{f_n\}$ *be a sequence of measurable functions on* E *s.t.* f_n *is finite almost everywhere and* $\{f_n\}$ *is convergent in* E *to a finite limit (not necessarily uniform). Then for every* $\epsilon > 0$*, there exists some subset* $A \subset E$ *s.t.* $\mu(E - A) < \epsilon$ *s.t.* $\{f_n\}$ *converges uniformly on* A.

Egorov's theorem allows us to deduce a uniformly convergent sequence of functions from a sequence which is not necessarily uniformly convergent by taking a subset of the original set the sequence acts on. The proof can be found in [4]. Finally, we introduce the *product measure*, and cite an important result relating to it, called Fubini's theorem [2].

Definition 2.18. Let $(X, \mathcal{A}_1, \mu)$ and $(Y, \mathcal{A}_2, \nu)$ be two measure spaces where μ and ν are σ-finite. We call a *measurable rectangle* a set with a form $A_1 \times A_2$, where $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$. Let $\mathcal{C}_0$ be the collection of finite unions of disjoint measurable rectangles, such that for all $C \in C_0$,

(1) C is of the form $\bigcup_{i=1}^{n}(A_{1_i} \times A_{2_i})$ for $A_{1_i} \in \mathcal{A}_1, A_{2_i} \in \mathcal{A}_2$
(2) if $i \neq j$, $(A_{1_i} \times A_{2_i} \cap (A_{1_j} \times A_{2_j}) = \emptyset$.

We then have that $(A_1 \times A_2)^c = (A_1 \times A_2^c) \cup (A_1^c \times A_2)$, and the intersection of two measurable rectangles is a measurable rectangle. Thus, C_0 is an algebra. We can then define a *product σ-algebra*

$$\mathcal{A}_1 \times \mathcal{A}_2 = \sigma(C_0)$$

This definition gives rise to Fubini's theorem.

Theorem 2.19 (Fubini's Theorem)**.** *Let $f : X \times Y \to \mathbb{R}$ be measurable with respect to $\mathcal{A}_1 \times \mathcal{A}_2$. Let μ and ν be σ-finite measures on X and Y (respectively). If either f is non-negative or $\int |f(x, y)| d(\mu \times \nu)(x, y) < \infty$, we have the following:*

(1) for each x, the function $y \mapsto f(x, y)$ is measurable with respect to $\mathcal{A}_2$
(2) for each y, the function $x \mapsto f(x, y)$ is measurable with respect to $\mathcal{A}_1$
(3) the function $g(x) = \int f(x, y)\nu dy$ is measurable with respect to $\mathcal{A}_1$
(4) the function $h(y) = \int f(x, y)\mu dx$ is measurable with respect to $\mathcal{A}_2$
(5) the following holds:

$$\begin{aligned}\int f(x, y) d(\mu \times \nu)(x, y) &= \int \int f(x, y) d\mu(x) d\nu(y) \\ &= \int \int f(x, y) d\nu(y) \mu(dx)\end{aligned}$$

Fubini's theorem states the existence of functions on product spaces and summarizes that on product spaces, the order of integration with respect to either measure is arbitrary.

3. Dynamical Systems

We've described a dynamical system as a structure-preserving map acting on a defined space, but here we make this description rigorous. These definitions are attributed to Young [5] and Teschl [6].

Definition 3.1. If μ_0 is a probability measure on $X_0 = \{1, \ldots, k\}$ where $\mu_0\{i\} = p_i$, and μ is the product measure of μ_0 on X defined as $\Pi_{-\infty}^{\infty} X_0$, then define a shift operator T as follows: if x_i is the ith coordinate of $x \in X$, then $T(x_i) = x_{i+1}$, defined as the $(p_1, \ldots, p_k)$ *Bernoulli shift*. Systems which are isomorphic to Bernoulli shifts are Bernoulli.

A Bernoulli shift involves shifting elements in a space one by one. Thus, one way to study complex flows is to reduce (or attempt to reduce) them to a Bernoulli shift.

To begin discussing flows we first describe a function f acting on an interval I to itself. Applying f, n number of times then leads to a *discrete dynamical system*. If f is invertible, then we can also apply the transform $-n$ times, or "go backwards".

Remark 3.2. Before continuing, we clarify that going forward, $C^k(X)$ will refer to the set of all functions that are k times continuously differentiable on a given space X.

Construction 3.3. Consider a differential equation $\frac{dx}{dt} = f(x)$ with an initial value of $x(0) = x_0$, where $f \in C^k(M, \mathbb{R}^n), k \geq 1$, and M is an open subset of $\mathbb{R}^n$. We can associate this ODE to a vector field on $\mathbb{R}^n$, representing values equal to $\frac{dx}{dt}$, or the "slopes" at given points. Integrating curves in this vector field with given initial conditions produces "integral curves", which represent specific solutions to the ODE. Based on existence and uniqueness theorems (which will not be proven here), there exists a unique maximal integral curve ϕ_x at each point x defined on an interval $I_x = (T_-(x), T_+(x))$. Define W as the following:

$$W = \bigcup_{x \in M} I_x \times \{x\} \subset \mathbb{R} \times M \tag{3.4}$$

Then the *flow* of the differential equation is defined as

$$\Phi : W \to M, (t, x) \to \phi(t, x) \tag{3.5}$$

where we define $\phi(t, x)$ as the maximal integral curve at each x. These flows establish *continuous dynamical systems*.

We can now treat f as an arbitrary flow (function) on X, a compact metric space. The *orbit* of a given point x is $\{f^n(x)\}_{n=-\infty}^{\infty}$, or the flow evaluated at a given point, iterated endlessly. If there exists some $n \in \mathbb{Z}$ such that $f^n(x) = x$, then we denote x as a *periodic point* and its orbit as periodic; we also call n its period. A *fixed point* x is a periodic point with period 1. A set $\Lambda \in X$ is *invariant* $f(\Lambda) = \Lambda$.

Now, we state the Birkhoff Ergodic Theorem [7]. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $f : X \to X$ be an mpt. Let ϕ be integrable with respect to μ; we may also denote this property as $\phi \in L^1(\mu)$. Define $\phi_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} \phi \circ f^j(x)$, and let the limit average $\hat{\phi}$ be $\lim_{n\to\infty} \phi_n(x)$.

Theorem 3.6 (Birkhoff Ergodic Theorem)**.** *Under the conditions stated above, $\hat{\phi}$ is defined a.e. and $\int \hat{\phi} d\mu = \int \phi d\mu$. In addition, if f is ergodic under μ, then $\hat{\phi}(x) = \int \phi d\mu$ a.e.*

Birkhoff's theorem asserts that a system's "time average" is equivalent to its "space average" almost everywhere for an mpt. Similarly to Poincare recurrence (Theorem 2.16), Birkhoff's theorem has a variety of applications when studying the asymptotic behavior of an mpt.

4. Manifolds

Often, flows are defined on some generalization of a surface to higher dimensions. These "higher-dimensional surfaces" are rigorously defined as *manifolds*. We assume the reader is familiar with the definition of a topological space and follow the definitions in [8].

Definition 4.1. Let M be a topological space. M is a *smooth topological manifold* if

(1) M is Hausdorff: for every $p, q \in M$, there exists disjoint open subsets $U, V \in M$ s.t. $p \in U$, $q \in V$.
(2) M is second countable: there exists a countable basis for the topology of M and
(3) M is locally Euclidean: for every $p \in M$, there exists an open neighborhood $U \subset M$ s.t. $U \ni p$ and U is diffeomorphic to an open subset in $\mathbb{R}^n$.
(4) there exists a maximal smooth atlas $\mathcal{A}$ associated to M, where a *maximal smooth atlas* denotes a family of charts on M which are pairwise smoothly compatible and which cannot be contained by any other smooth atlas on M

An intuitive summation of this definition is a surface which can be generalized to higher dimensions, and whose points are continuous or differentiable at all orders.

Definition 4.2. Let M be a smooth manifold and let $p \in M$. We call a linear map $X : C^\infty(M) \to \mathbb{R}$ a *derivation* at p if

$$X(fg) = f(p)Xg + g(p)Xf$$

for all $f, g \in C^\infty(M)$. The set of all derivations at p is called the *tangent space* at p, and we denote it as T^pM. We call an element $v \in T^pM$ a *tangent vector* at p.

Using manifolds, we can concretely apply the theory of ordinary differential equations to *integral curves* on M, but the specifics will not be required here, so we omit this. However, we will describe specific functions and properties of functions on manifolds. For example, we define an *attractor*:

Definition 4.3. Let $f : M \to M$ be smooth on M. We call a compact invariant set Λ *attracting* if there is a neighborhood $U \subset \Lambda$ such that for all $x \in U$, $\lim_{n\to\infty} d(f^n(x), \Lambda) = 0$. If Λ is topologically transitive we call it an attractor. We define topologically transitive as follows: for any two open sets $U, V \subseteq \Lambda$, there exists some $t \in M$ s.t. $f(t, U) \cap V \neq \emptyset$.

Finally, we give a definition of foliations, which will be useful in the future to "break down" a manifold into subspaces.

Definition 4.4. A *foliation* $\mathcal{F}$ on M is a partition of M into disjoint subsets $A_1, \ldots A_n$ such that

(1) For all $p \in M$, there exists some neighborhood U such that $A_i \cap U = \bigcup A_j$, where each A_j are pairwise disjoint and $A_j \in \mathcal{F}$
(2) Each leaf A_i is a smooth submanifold in M
(3) The tangent spaces of the leaves at each point p on M span a subspace of the tangent space T_pM
(4) For all $p \in M$, $p \in A_i$ for some $A_i \in \mathcal{F}$, and $\bigcup_{i=1}^n A_i = M$

We will also define a Riemannian manifold.

Definition 4.5. A smooth manifold M is *Riemannian* if there exists a defined positive-definite inner product g_p on the tangent space T_pM at all $p \in M$. The family g_p of inner products is titled the Riemannian metric.

5. Lyapunov Exponents & Hyperbolicity

Lyapunov exponents allow us to quantify the asymptotic rates of contraction or expansion in a dynamical system.

Let $v \in \mathbb{R}^m$ and let A be an invertible $m \times m$ matrix, where $a \in \mathbb{R}$ for all $a \in A$. If we write A in its Jordan canonical form, we see that A is decomposed into invariant subspaces $E_1 \oplus \cdots \oplus E_n$, such that for each $i \leq n$, there exists some λ_i such that for all $v \in E_i$ $(v \neq 0)$, $\lim_{n\to\infty} \frac{1}{n} \log |A^n v| = \lambda_i$. Here, λ_i are the eigenvalues and encode information about the rate of change of A after n repetitions at the given vector v.

Theorem 5.1 (Oseledet's, or the Multiplicative Ergodic Thm.)**.** *Assume $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ is an mpt of a probability space, and A is an invertible $m \times m$ matrix which represents T. Then almost everywhere, there exists a filtration of subspaces:*

$$\{0\} = V_0(x) \subsetneq V_1(x) \subsetneq \cdots \subsetneq V_r(x) = \mathbb{R}^m$$

and associated $\lambda_1(x) < \cdots < \lambda_r(x)$ such that for all $v \in V_i(x) - V_{i-1}(x), \lambda_+(x, v) = \lambda_i(x)$ and $\lim_{n\to\infty} \frac{1}{n} \log |\det A^n(x)| = \sum \lambda_i(x)[\dim V_i(x) - \dim V_{i-1}(x)]$.

We call each λ_i with its associated $\dim V_i(x) - \dim V_{i-1}(x)$ *Lyapunov exponents* and their multiplicities (or fractal dimensions), respectively. The proof can be found in [9].

As a consequence of the Multiplicative Ergodic Theorem we can break a manifold down into "stable" and "unstable" foliations as follows [14]:

Definition 5.2. Let $\lambda_1 > \cdots > \lambda_n$ be the positive Lyapunov exponents of (f, μ), where f is an ergodic mpt acting on a compact C^2 Riemannian manifold and μ is the associated Borel invariant probability measure. We then define n nested invariant foliations

$$W^1 \subset \cdots \subset W^n, W^i = \oplus_{j \leq i} E_j$$

Then if $\lambda_{r-s+1} > \cdots > \lambda_r$ are the negative Lyapunov exponents,

$$W^1 \subset \cdots \subset W^s, W^i = \oplus_{j > r-i} E_j$$

The concept of a system being partially stable and partially unstable can be helpful when analyzing particular systems. This property is titled *hyperbolicity.*

Definition 5.3. Let M be a compact $\mathbb{R}^\infty$ Riemannian manifold, and $f : M \to M$ be a diffeomorphism (a differentiable map on a manifold with a differentiable inverse). Let $T_x M$ be the tangent space at x on M. $\Lambda \subset M$ is *uniformly hyperbolic* set if:

(1) for all $x \in \Lambda, T_x M = E^s(x) \oplus E^u(x)$
(2) $Tf(E^s(x)) = E^s(f(x)), Tf(E^u(x)) = E^u(f(x))$, and
(3) for some $C \geq 1$ and $0 < \lambda < 1$, $|Tf^n(v)| \leq C\lambda^n |v|$ for all $x \in A, v \in E^s(x), n \geq 0$ and similarly
(4) $|Tf^{-n}(v)| \leq C\lambda^n |v|$ for all $x \in A, v \in E^u(x), n \geq 0$

Foliations as defined in Definition 5.2 will prove helpful when defining unstable and stable foliations on a manifold on which we wish to deduce entropy, relating to an important result titled the Stable Manifold Theorem.

Theorem 5.4 (Stable Manifold)**.** *Let $U \subset \mathbb{R}^n$ be open and $\phi_t : U \to U$ the flow of a C^1 vector field $f : U \to \mathbb{R}^n$. Assume that x_0 is a fixed point of f and let $E^s \to \mathbb{R}^n$ be the stable subspace for $Df(x_0)$, or the span of all eigenvectors for eigenvalues with negative real part. Let $E^{cu} = E^c \oplus E^u \subset \mathbb{R}^n$ be the center-unstable subspace for $Df(x_0)$, corresponding to eigenvectors of eigenvalues with positive or zero real part. Then there exists some $r > 0$ and some C^1 function $\psi : B(x_0, r) \cap E^s \to E^{cu}$ such that the set*

$$W^s := \{x + \psi(x) \mid x \in B(x_0, r) \cap E^s\}$$

obeys the following:

(1) W^s is positively invariant and
(2) Given an initial condition $x \in W^s$,

$$\lim_{t\to\infty} \phi_t(x) = x_0$$

Thus we have that near a fixed point on a manifold, there exists some local stable manifold in which points asymptotically approach the fixed point. The proof can be found in [10].

6. The Lorenz Attractor

As an application of the past four sections, we provide a brief introduction to the Lorenz attractor. The general proof and results not shown here can be found in [11]. Lorenz's system of differential equations is known for highlighting a phenomenon called sensitivity to initial conditions.

Definition 6.1. Let f be a diffeomorphism on a compact manifold M. A dynamical system displays *sensitivity to initial conditions* if there exists some $\delta > 0$ such that for all $\epsilon > 0, x \in M$, there exists some $y \in M, n \in \mathbb{N}$ such that $|x - y| < \epsilon$ and $|f^n(x) - f^n(y)| > \delta$.

Such a property, which could be summarized as, minute changes in initial conditions can lead to vastly different outcomes in the system, lead to the phenomenon commonly called the "Butterfly Effect". The Lorenz equations are a system which exemplify this effect:

$$\frac{dx}{dt} = 10(y - x)$$
$$\frac{dy}{dt} = 28x - y - xz$$
$$\frac{dz}{dt} = xy - \frac{8}{3}z$$

It has previously been proven by Tucker that under certain initial conditions and parameters, the Lorenz equations form a robust attractor A which contains an SRB measure ν with a positive Lyapunov exponent. SRB measures are defined concretely in Section 8, but the specifics are not requried for the following proof. We prove here that this measure is mixing.

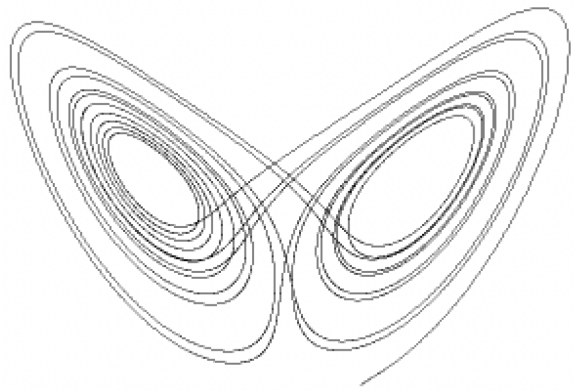

FIGURE 1. Image of the Lorenz attractor under the given parameters [13].

Firstly, we provide some background for the Lorenz equations. Their origin is an equilibrium point with two negative eigenvalues and one positive eigenvalue, denoted as $\lambda_{ss} < \lambda_s < 0 < \lambda_u$, and $\lambda_u > |\lambda_s|$.

To simplify our proof, we would like to construct a one-dimensional map that acts as a representative for the Lorenz attractor.

From the Poincare Recurrence Theorem (Theorem 2.16), we have that points about the origin will eventually return there, and thus we may define a Poincare map around a locally linearized origin.

Construction 6.2. Begin by linearizing the cube $[-1, 1]^3$ about the origin. Let

$$\Sigma = \{x_1, x_2, x_3) : |x_1|, |x_2| \leq 1, x_3 = 1\}$$

$$\tilde{\Sigma} = \{(x_1, x_2, x_3) : x_1 = \pm 1, |x_2|, |x_3| \leq 1\}$$

Then define the first hit map $P_0 : \Sigma \to \tilde{\Sigma}$ as follows:

$$P_0(x_1, x_2, 1) = (e^{\lambda_u r_0} x_1, e^{\lambda_{ss} r_0} x_2, e^{\lambda_s r_0}) = (\text{sgn}(x_1, \tilde{x_2}, \tilde{x_3})$$

where r_0 represents the initial time of flight of the map.

Let $e^{\lambda_u r_0} x_1 = \text{sgn} x_1$. We then have that $r_0 = \frac{-(\ln |x_1|)}{\lambda_u}$. Substituting, we have

$$P_0(x_1, x_2, 1) = (\text{sgn} x_1, |x_1|^{\beta} x_2, |x_1|^{\alpha})$$

where $\alpha = \frac{|\lambda_s|}{\lambda_u} \in (0, 1)$ and $\beta = \frac{|\lambda_{ss}|}{\lambda_u} > 0$.

We call the compact region $N \subset \Sigma$ such that the Poincare first return map $P : N \setminus W^s(0) \to N$ is well-defined the "trapping region", and its existence was proven by Tucker.

We then decompose $P = P_1 \circ P_0$, where P_0 is the "forward" map as defined above and P_1 is the "returning" map. Since P is well-defined, we then also have that P_1 is well-defined and thus $r_1 \leq \infty$. Thus the total time of flight $r = r_0 + r_1$ is smooth except for the logarithmic singularity at $x_1 = 0$.

From this lemma, we can acquire a hyperbolicity estimate about the origin. Before, however, we define a *cone* [12].

Definition 6.3. Let $E \subset \mathbb{R}^n$ s.t. $0 < \dim E < n$ and let F be such that $E \oplus F = \mathbb{R}^n$. The *standard unit cone* determined by E and F is the set

$$K(E, F) = \{v = (v_1, v_2) : v_1 \in E, v_2 \in F, |v_2| \leq |v_1|\}$$

A *cone* in $\mathbb{R}^n$ with core E is the image $T(K(E, F))$ where $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear automorphism s.t. $T(E) = E$.

A *cone field* is the collection of cones at points.

Lemma 6.4. *The return map P admits a forward invariant cone field, or there exists some cone $\mathcal{C}(u) \subset \Sigma$, for all $u \in N \setminus W^s(0)$, such that $(dP)_u\mathcal{C}(u) \subset \mathcal{C}(Pu)$. In addition, there exists the hyperbolic constants $c > 0, \pi > 1$, such that for each $u \in N \setminus W^s(0)$,*

$$||(dP^n)_u v|| \geq c\pi^n ||v||$$

for $v \in \mathcal{C}(u), n \geq 1$.

As an immediate consequence, we can make assumptions regarding a foliation on P with $C^{1+\epsilon}$ leaves, $\epsilon > 0$. If $I = [-1, 1]$, define a singular one-dimensional map $f : I \to I$ by dividing along stable leaves. Then $r : I \to \mathbb{R}^+$ becomes a singular map. If $J = I \setminus \{0\}$, then on J,

(1) f is $C^{1+\epsilon}$
(2) $|f^{(n)}| \geq c\pi^n$ for all $n \geq 1$
(3) $C^{-1}x^{\alpha-1} \leq f'(x) \leq C|x|^{\alpha-1}$
(4) $r(x) \to \infty$ as $x \to 0$ and
(5) $|r(x) - r(y)| \leq C|\ln|x| - \ln|y||$, for all $x, y \in \mathbb{R} \setminus \{0\}$

Using this construction, we have reduced the three-dimensional initial Lorenz equations (or attractor) to a Poincare map and then to a one-dimensional continuous expanding map. We can also work backwards, by taking the natural extension of f and then taking the suspension of P by the roof function r. We could also recover the Lorenz flow as the natural extension of the suspension semiflow of the map f by r, which is the geometric Lorenz flow construction. We omit the precise definitions for such suspension and extensions. We begin our analysis of this one-dimensional function, which will allow us to prove that the Lorenz attractor is mixing.

Definition 6.5. Let $f : I \to I$ such that $f(0)$ be undefined. Let $\lim_{x\to 0^+} f(x) = -1$ and $\lim_{x\to 0^-} f(x) = 1$ for $x \in I$. If f satisfies conditions (1), (2), and (3) as stated above, then we call f a *Lorenz-like expanding map*. In addition, a semiflow f_t is called a *geometric Lorenz semiflow* if it is the suspension of a Lorenz-like expanding map by a roof function r which satisfy conditions (4) and (5). We call a flow f_t a *geometric Lorenz flow* if it is the natural extension of a geometric Lorenz semiflow.

Definition 6.6. We define the map $f : I \to I$ as locally eventually onto ($l.e.o$) if for all open $U \subset J$, there exists some $k \geq 0$ such that $f^k U \ni (0, 1)$.

If at some iteration of f the map on any open set eventually contains $(0, 1)$, it is locally eventually onto. Lorenz-like expanding maps which are $l.e.o$ admit a unique ergodic probability measure μ, which when suspended, $\nu = u^r$, defines an SRB

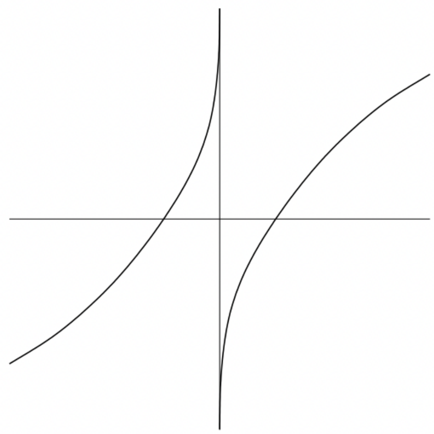

FIGURE 2. The reduced Lorenz-like expanding map about the origin [11].

measure for the geometric Lorenz flow. This would prove that the Lorenz attractor, which is defined as a geometric Lorenz flow, is mixing.

It was previously known that semiflows and their suspended flows are Bernoulli, and thus also mixing. Thus, we wish to show that Lorenz semiflows are weak mixing, from which the general result would follow.

Theorem 6.7. *Let f_t be a geometric Lorenz flow. Let the associated one-dimensional map f be l.e.o. Then f_t is Bernoulli and mixing.*

Proof. Let $f : I \to I$ be a Lorenz-like expanding map. We have that f admits an induced map F which has the following properties: there exists an open interval $Y \subset I$ s.t. $Y \ni 0$ and a.e. there exists a partition $\mathcal{P} = \{\omega\}$ of Y where each $\omega \in \mathcal{P}$ is an interval. The return function $R : Y \to \mathbb{N}$ is constant on each ω such that $F(x) = f^{R(x)}(x)$ restricts to a diffeomorphism $F|_\omega : \omega \to Y$ for each $\{\omega\}$ and such that the following holds.

1. there exists $\lambda > 1$ such that $|DF_\omega| \geq \lambda$, for all ω
2. for all $\omega, \log g_\omega$ is Holder (uniformly in ω, where g is the differential matrix of the inverse of $F|_\omega : \omega \to Y$),
3. R is Lebesgue integrable, and finally
4. $0 \notin f^{\bar{k}}\omega, 0 \leq k < R(\omega)$, for all ω

From prior results, we have if these conditions are satisfied then there exists a unique invariant ergodic measure μ that is absolutely continuous with respect to Lebesgue and whose support includes Y. These conditions allow for the following lemma to hold:

Lemma 6.8. *If $f : I \to I$ admits an induced map $F : Y \to Y$ which satisfies the above conditions with absolutely continuous ergodic measure μ, then let $r : I \to \mathbb{R}^+$ be a Holder roof function with a logarithmic singularity at 0. Let $\phi : I \to S^1$ be a μ-measurable function such that, almost everywhere,*

$$e^{ir} = (\phi \circ T)\phi^{-1}$$

Then ϕ has a version that is Holder on Y.

Due to some technicalities we omit the proof. Let $r : I \to \mathbb{R}^+$ be a Holder roof function with a logarithmic singularity at 0 and let f_t be the corresponding geometric Lorenz semiflow with ergodic measure μ^r.

Lemma 6.9. *Assume the geometric Lorenz semiflow is not weak mixing. Then there exists a constant $a > 0$ and a measurable eigenfunction $\psi : X \to S^1$ continuous on $\bigcup_{k \geq 0} f^k Y$ such that, almost everywhere,*

$$e^{iar} = (\psi \circ f)\psi^{-1} \tag{6.10}$$

Proof. By the definition of weak mixing, $\phi \circ f_t = e^{iat}\phi$ has no measurable solutions $\phi : X^r \to S^1$, for $a > 0$. Let ϕ be a measurable solution. By Fubini's Theorem (Theorem 2.19), there exists $\epsilon > 0, r > \epsilon$ almost everywhere such that $\phi \circ f_t(x, \epsilon) = e^{iat}\phi(x, \epsilon)$ for almost every $x \in X$.

Let $t = r(x), \psi(x) = \phi(x, \epsilon)$. If $f_{r(x)}(x, \epsilon) = (fx, \epsilon)$, then $\psi \circ f = e^{iar}\psi$, so ψ is a measurable solution to (6.10). We then have from Lemma 6.8 that there exists some solution ϕ that is Holder continuous on Y.

We wish to show that ψ is continuous on $\bigcup_{k \geq 0} f^k Y$. Let $z = f^k y, y \in Y$. Since $f(0)$ is undefined, we have that $f^j y \neq 0$, for $j \in [k-1]$. Thus there exists an open $U \subset Y$ such that $U \ni y$ and $0 \notin f^j U$ for $0 \leq j \leq k-1$. Then there exists $\gamma \in (0, 1)$ such that ψ is C^γ on Y and e^{iar_k} is C^1 on U. Let $z_i = f^k y_i$, $y_i \in U$, $i = 1, 2$. Applying (6.10),

$$\psi(z_1)\psi(z_2)^{-1} = \psi(y_1)\psi(y_2)^{-1} e^{iar_k(y_1)} e^{-iar_k(y_2)}$$

We then have by statement (2) above that

$$\psi(z_1)\psi(z_2)^{-1} \leq D|y_1 - y_2|^\gamma \leq D(c\lambda^k)^{-\lambda}|z_1 - z_2|^\gamma$$

So we have that ψ is Holder continuous on $f^k U$ and thus at z. □

We now have the necessary tools to prove Theorem 6.7. Assume for the sake of contradiction that f_t is not weak mixing. By Lemma 6.9, there exists some $a > 0$ and a measurable eigenfunction $\psi : I \to S^1$ such that (6.10) is satisfied and ψ is continuous on $\bigcup_{k \geq 0} f^k Y$. Then since f is *l.e.o*, ψ is continuous on $(-1, 1)$, so ψ is continuous at $f(\pm 1)$. Iterating (6.10) we have

$$e^{iar} e^{iar \circ f} = (\psi \circ f^2)\psi^{-1}$$

Evaluate along a sequence $x_n > 0$ as $x_n \to 0$. We have $\psi(x_n) \to \psi(0), \psi(f^2 x_n) \to \psi(f(-1))$, and $r(f x_n) \to r(-1)$. Then $(\psi \circ f^2)\psi^{-1}$ and $e^{iar \circ f}$ both converge as $n \to \infty$. But for any n sufficiently large, we have $b_n > r\frac{\epsilon}{2^n}$ such that $e^{iar(x_n)}$ is divergent.

We have that r is continuous on $I\backslash\{0\}$ and $r(x)\to\infty$ as $x\to 0$, so there exists some $x_n>0$ such that $r(x_n)=b_n$. Thus, $e^{iar(x_n)}$ diverges, which is a contradiction. $\square$

7. The Pesin Entropy Formula

In the late 1970s, Ruelle proved the following inequality connecting the metric entropy of a diffeomorphism f with invariant probability measure μ on a compact manifold M as the following:

$$h_\mu(f)\le\int_M\sum\lambda_i^+m_i d\mu \tag{7.1}$$

In 1978, Pesin published the conditions under which equality holds; when μ is absolutely continuous relative to the Lebesgue measure and f' is Holder continuous. Here we closely follow the proof by Mane of this formula, which does not rely on the use of Lyapunov charts or stable manifold theory [15].

Notation 7.2. For clarity we will let $\chi : M\to R$ be defined as the sum of the products of Lyapunov exponents and their multiplicities, i.e. $\chi=\sum\lambda_i^+m_i$. Thus the formula will henceforth be reduced to $h_\mu(f)=\int_M\chi d\mu$.

For the duration of this proof we assume $g : M\to M$ is an mpt on a manifold M and let $\rho : M\to(0,1)$ be a function such that $\log\rho$ is μ-integrable. For $x\in M, n\ge 0$, define S_n as the following:

$$S_n(g,\rho,x)=\{y\mid d(g^j(x),g^j(y))\le\rho(g^j(x)),0\le j\le n\} \tag{7.3}$$

To summarize, we choose S_n at x to be the set of all points whose distance from x is less than or equal to $\rho(x)$ on any iteration of g up until the given n. Now, let g and ρ be measurable with respect to a measure μ, and let g be invariant with respect to μ. Define the following:

$$h_\mu(g,\rho,x)=\limsup_{n\to\infty}\frac{1}{n}(-\log\mu(S_n(g,\rho,x))) \tag{7.4}$$

Here we define entropy of the system with respect to μ,g,ρ,x as the measure of the greatest asymptotic average of the points bounded by ρ. We begin the proof with the following lemmas:

Lemma 7.5. *Let $\sum_1^\infty x_n$ be a series, such that $0<x_n<1$ for all n, and $\sum_1^\infty nx_n$ converges. Then $\sum_1^\infty x_n\log(\frac{1}{x_n})$ converges.*

Proof. Define $N:=\{n\in\mathbb{N}\mid\log(\frac{1}{x_n})<n\}$. If for some $n\in\mathbb{N}, n\notin N$, we have

$$\log(\frac{1}{x_n})\ge n$$

so $e^n\ge\frac{1}{x_n}$, and thus $e^{-n}\ge x_n$. Then we have

$$\begin{aligned}&\sum_1^\infty x_n\log(\frac{1}{x_n})\\ &\le\sum_{n\in N}nx_n+\sum_{n\notin N}x_n\log\frac{1}{x_n}\end{aligned}$$

We also have that since $t\log(1/t) \leq \frac{2}{e}$ for all $t \in \mathbb{N}$,

$$\sum_1^\infty x_n \log \frac{1}{x_n}$$
$$\leq \sum_{n\in N} nx_n + \frac{2}{e}\sum_{n\notin N}\sqrt{x_n}$$
$$\leq \sum_1^\infty nx_n + \frac{2}{e}\sum_1^\infty e^{-\frac{1}{2}n}$$

And since these two sums converge, the result follows. □

We now prove that we can find some countable partition on our manifold such that the diameter of each atom is bounded by ρ almost everywhere.

Lemma 7.6. *There exists a countable partition $\mathcal{P}$ of M such that $H(\mathcal{P}) < \infty$ and if $\mathcal{P}(x)$ denotes the atom $P \in \mathcal{P}$ s.t. $P \ni x$, $\operatorname{diam}\mathcal{P}(x) \leq \rho(x)$ almost everywhere.*

Proof. Choose constants $C > 0, r_0 > 0$ such that for all $0 < r \leq r_0$, there exists a partition $\mathcal{P}'_r$ whose $\mathcal{P}'(x) \leq r$ and $|\mathcal{P}'_r| \leq C(\frac{1}{r})^{\dim M}$. Define U_n as follows:

$$U_n := \{x \mid \frac{1}{e^{n+1}} \leq \rho(x) \leq \frac{1}{e^n}, n \geq 0\}$$

Since $\log\rho$ is integrable, we have that $\sum_1^\infty n\mu(U_n)$ is finite, so applying Lemma 7.5 to $\mu(U_n)$, we have that $\sum_1^\infty \mu(U_n)\log\frac{1}{\mu\cdot U_n}$ also converges.

Let $r_n = \frac{1}{e^{n+1}}$, $n \in \mathbb{N}$, and let $Q \in \mathcal{P}'_{r_n}$. Define a new partition $\mathcal{P}$ as follows:

$$\mathcal{P} := Q \cap U_n, \mu(Q \cap U_n) \neq 0$$

We know that $H(\mathcal{P}) = \sum_n^\infty(-\sum_{P\in\mathcal{P}}\mu(P)\log\mu(P))$. Since $-\sum_{i=1}^m x_i\log x_i \leq (\sum_1^m x_i)(\log m - \log\sum_1^m x_i), 0 < x_i \leq 1, i = \{1, 2, \ldots, m\}$, we then have that

$$H(\mathcal{P}) \leq \sum_0^\infty \mu(U_n)[\log|P_{r_n} - \log\mu(U_n)]$$
$$\leq \sum_0^\infty \mu(U_n)[\log C + \dim M(n+1) - \log\mu(U_n)]$$

Since we defined $|P_{r_n}| \leq (\frac{C}{r})^{\dim M}$, the result follows. □

Using these lemmas, we have that for some g-invariant measure μ and some ν which is absolutely continuous with respect to μ, the following holds. The metric entropy with respect to μ is greater than or equal to entropy as defined in (7.4) with respect to ν.

Proposition 7.7. *Let μ be a g-invariant probability measure on M and let ν be absolutely continuous with respect to μ (not necessarily g-invariant). Then we have that a.e.*

$$h_\mu(g) \geq \int_M h_\nu(g, \rho, x)d\mu$$

Proof. Take $\mathcal{P}$ as the same partition from Lemma 7.6. By the Shannon-Breiman-McMillan Theorem (Theorem 2.14) and our definition of metric entropy (Definition 2.10), we have

$$h_\mu(g) \geq h_\mu(\mathcal{P}, g) = \int_M \lim_{n\to\infty} \frac{1}{n}[-\log \mu(\mathcal{P}_n(x))]d\mu$$

Thus it is left to show that $\lim_{n\to\infty} \frac{1}{n}[-\log \mu(\mathcal{P}_n(x))] \geq h_\nu(g, \rho, x)$ a.e.

Let $\mathcal{P}_\infty$ be a σ-algebra generated by $\mathcal{P}_n, n \geq 0$ and let $k : M \to \mathbb{R}$ be ν-integrable, measurable with respect to $\mathcal{P}_\infty$, and such that $\int_A k d\nu = \mu(A), A \in \mathcal{P}_\infty$. By the Radon-Nikodym Theorem (Theorem 2.7), we have that a.e.

$$\lim_{n\to\infty} \frac{\mu(\mathcal{P}_n(x))}{\nu(\mathcal{P}_n(x))} = k$$

Thus we have that

$$\begin{aligned}
&\frac{1}{n}\log \mu(\mathcal{P}_n(x)) \\
&= \frac{1}{n}\log \nu(\mathcal{P}_n(x)) + \frac{1}{n}\log \frac{\mu(\mathcal{P}_n(x))}{\nu(\mathcal{P}_n(x))} \\
&\implies \lim_{n\to\infty} \frac{1}{n}[-\log_\mu(\mathcal{P}_n(x))] = \lim_{n\to\infty} \frac{1}{n}[-\log_\nu(\mathcal{P}_n(x))]
\end{aligned}$$

Since $\mathcal{P}_n(x) \subset S_n(g, \rho, x)$, we have that

$$\lim_{n\to\infty} \frac{1}{n}[-\log \mu(\mathcal{P}_n(x))] \geq h_\nu(g, \rho, x), a.e.$$

and the result follows.

$\square$

Our ultimate argument will now be that $\int_M h_\nu(g, \rho, x)d\mu \geq \int_M \chi d\mu$, but we must introduce some concepts and lemmas first.

Definition 7.8. If E is a normed space and $E_1 \oplus E_2$ is a splitting of E, we define $\gamma(E_1, E_2)$ as the supremum of the norms of the projections $\pi_1 : E \to E_1, \pi_2 : E \to E_2$. We also define an (E_1, E_2) *graph* G if there exists an open set $U \subset E_2$, and a C^1 map $\Psi : U \to E$ such that $G = \{x + \Psi(x) \mid x \in U\}$. We also call

$$\sup\{\frac{||\Psi(x) - \Psi(y)||}{||x - y||} \mid x, y \in U\}$$

the *dispersion* of the graph.

Intuitively, on some splitting of a normed space E, we define a graph G as the set of the elements of the open set added to some continuous map of the open set onto the normed space. We also provide dispersion as a metric for the greatest slope of the continuous map of U onto E, or more simply, the greatest magnitude that must be added to points in our open set U to form G. We use these concepts to prove a more technical lemma which establishes a constant for which the image of a graph under a map F is also a graph with dispersions bounded by the same constant c.

Lemma 7.9. *Let $\lambda > \beta > 1, \alpha > 0, c > 0$. There exists a $\delta > 0$ that has the following property. If the following properties are satisfied:*

(1) if there exists a finite dimensional normed vector space $E = E_1 \oplus E_2$ s.t. $\gamma(E_1, E_2) \leq \alpha$
(2) and there exists F which is C^1 and acts on some open ball $B_r(0) \subset E$ s.t. $F : B_r(0) \to E'$, where E' is another Banach space,
(3) and if F has the following properties:
   (a) D_0F' is an isomorphism and $\gamma((D_0F)E_1, (D_0F)E_2)) < \alpha$
   (b) $||(D_0F) - (D_xF)|| \leq \delta$ for all $x \in B_r(0)$
   (c) $||(D_0F)v|| \geq \lambda||v||$ for all $v \in E_2$ and
   (d) $||(D_0F)/E_1|| \leq \beta$

Then for all (E_1, E_2) graphs G with dispersion $\leq c$ in $B_r(0)$, the image $F(G)$ is a $((D_0F)E_1, (D_0F)E_2)$ graph with dispersion $\leq c$.

Proof. We begin by notating F as follows:

$$F(x, y) = (\frac{D_0F}{E_1}x + p(x, y), \frac{D_0F}{E_2}y + q(x, y))$$

For clarity, let $L = \frac{D_0F}{E_1}, T = \frac{D_0F}{E_2}$, so the formula simplifies to

$$F(x, y) = (Lx + p(x, y), Ty + q(x, y))$$

where p and q are some functions such that the partial derivatives with respect to x, y have norm less than or equal to $\delta\alpha$. Let $U \subset E_2$ be open, and let $\Psi : U \to E$ be a map whose graph $\{(\Psi(v), v) \mid v \in U\}$ is G (such a function exists by the definition of the graph). Then, we have

$$F(G) = \{(L\Psi(v) + p(\Psi(v), v), Tv + q(\Psi(v), v)) \mid v \in U\}$$

Now define $\Phi : U \to (D_0F)E_2$ as $\Phi(v) = Tv + q(\psi(v, v))$. If $v, w \in U$, then

$$||\Phi(v) - \Phi(w)|| \geq ||T(v - w)|| - ||q(\Psi(v), v) - q(\psi(w), w)||$$

Combining this inequality with property (3) of F, we have

$$||\Phi(v)-\Phi(w)|| \geq \lambda||v-w||-\delta\alpha(||\Psi(v)-\Psi(w)||+||v-w||) \geq (\lambda-\delta\alpha(1+c))||v-w||$$

If $\lambda - \delta\alpha(1 + c) > 1$, then we have that Φ is a homeomorphism of U onto $\Phi(U)$ such that Φ^{-1} has a Lipschitz constant less than or equal to $\frac{1}{(\lambda-\delta\alpha(1+c))}$, and we have that $\Phi(U)$ is also open.

Now, let $\hat{\Psi} := \Phi(U) \to (D_0F)(E_1)$ be defined as

$$\hat{\Psi}(v) = (L\Psi\Phi^{-1})v + p(\Psi(\Phi^{-1}v)), \Phi^{-1}v)$$

Then we have that

$$F(G) = \{(\hat{\Psi}(x), x) \mid x \in \Phi(U)\}$$

Since $\hat{\Psi} = \hat{\Psi} \circ \Phi^{-1}$, we have $\hat{\Psi}(w) = L\Psi(w) + p(\Psi(w), w)$, so we have a final inequality:

$$||\hat{\Psi}(x)-\hat{\Psi}(y)|| \leq \beta||\psi(x)-\psi(y)||+\delta\alpha(||\psi(x)-\psi(y)||+||x-y||) \leq (c\beta+\delta\alpha(1+c))||x-y||$$

so we have that the dispersion of $F(G)$ is less than or equal to the following:

$$C\frac{\beta+\frac{\delta\alpha(1+c)}{c}}{\lambda-\delta\alpha(1+c)}=\frac{c\beta}{\lambda}\frac{1+\frac{\delta\alpha(1+c)}{c\beta}}{1-\frac{\delta\alpha(1+c)}{\lambda}}<1$$

when δ is sufficiently small, since $c<1, \frac{\beta}{\lambda}<1$. □

Now let f be an mpt on a manifold M and let μ be absolutely continuous with respect to Lebesgue measure on M. Define

$$E^u(x)=\oplus\{E_j(x)\mid\lambda_j(x)>0\}, E^s(x)=\oplus\{E_j(x)\mid\lambda_j(x)\leq 0\}$$

In addition, write $\sum_j=\{x\mid\dim E^u(x)=j\}$, and $S=\{j\geq 0\mid\mu(\sum_j)>0\}$. If $j\in S$, define μ_j as the measure given by $\mu_j(A)=\frac{\mu(A\cap\sum_j)}{\mu(E_j)}$, where A is a Borel subset on M. We have then that $h_\mu(f)=\sum_{j\in S}\mu(\sum_j)h_{\mu j}(f)$, and we want to show that $h_{\mu j}(f)\geq\int_M\chi d\mu_j$. For clarity, we now refer to μ_j as μ and $\sum_j$ as $\sum$. Clearly this property holds for $j=0$, so assume $j>0$. Let $\epsilon>0$. By Theorem 3.6 and Theorem 2.17 (the Birkhoff Ergodic Theorem and Egorov's Theorem, respectively), we have that there exists a compact set $K\subset M$ such that $\mu(K)\geq 1-\epsilon$, and $E^s(x)\oplus E^u(x)=T_xM$ is continuous when x varies in K. Then for some $N>0$, there exists some $\lambda>\beta>1$ such that if $g=f^N$, the following holds:

(1) $||(D_xg^n)v||\geq\lambda^n||v||$
(2) $||(D_xg^n)\mid E^s(x)||\leq\beta$, and
(3) $\log|\frac{\det(D_xg^n)}{E^u(x)}|\geq N(\chi(x)-\epsilon)n$

for $x\in K, n\geq 0, v\in E^u(x)$.

If $x\in K$, define

$$D_r(x)=\{x+y_1+y_2\mid y_1\in E^s(x), y_2\in E^u(x), ||y_1||\leq r, ||y_2||\leq r\} \tag{7.10}$$

Let $K_2>K_1>0$, $r_1>0$ such that $B_{k,r}(x)\subset D_r(x)\subset B_{K_2r}(x)$ for all $x\in K, 0<r\leq r_1$.

For the duration of this section we will now assume M is a Euclidean space to avoid cumbersome changes in coordinates, but the proof holds on a Riemannian manifold with local coordinates. We now state and prove the final two lemmas required. The first we will show establishes a constant ζ such that for a graph established by stable and unstable subspaces for a periodic point x, the iterated map at that point is also a graph, bounded by the same dispersion constant c.

Lemma 7.11. *For all $c>0$, there exists some $\zeta>0$ such that for some periodic $x\in K$, if a set $G\subset M$ is contained in the ball $B_\zeta m(x)$ and is an $(E^s(x), E^u(x))$-graph with dispersion $\leq c$, then $g^m(G)$ is a $((D_xg^m)E^s(x),(D_xg^m)E^u(x))$-graph with dispersion $\leq c$.*

Proof. Let $\alpha=\sup\{\gamma(E^s(x),E^u(x))\mid x\in K\}$, and let $\delta>0$ be given based on the parameters α,β,c,λ, as given by Lemma 7.9. We want to show that there exists some $C>0, 0<t\leq 1$ s.t.

$$||(D_xg^n)-(D_yg^n)||\leq C^n||x-y||^t, x,y\in M. \tag{7.12}$$

Doing so would allow us to manipulate the inequality to be less than δ, which would then allow application of Lemma 7.9.

Let $0 < t \leq 1, C_0 > 0$ such that $||(D_x g) - (D_y g)|| \leq C_0 ||x - y||^t$. g is Lipschitz continuous, i.e. for all $x, y \in M$,

$$\frac{|g(y) - g(x)|}{|y - x|} \leq A$$

where A is called the Lipschitz constant. Then $A \geq ||(D_x g)||$ for all $x \in M$. Now choose some C such that for all $n \geq 0$, we have that

$$C \geq A + C_0 \frac{A^{t+1}{}^n}{C} \tag{7.13}$$

For $n = 1$, we have that $||(D_x g) - (D_y g)|| \leq C_0 ||x - y||^t \leq C ||x - y||^t$, clearly. If it holds for $n = \{1, \ldots, m\}$, then we have that

$$\begin{aligned}
&||(D_x g^{m+1}) - (D_y g^{m+1})|| \\
&\leq ||(D_{g^m(x)} g) - (D_{g^m(y)} g)|| \cdot ||(D_y g^m)||) + (||(D_{g^m(x)} g)|| \cdot ||(D_x g^m) - (D_y g^m)||) \\
&\leq A^m C_0 ||g^m(x) - g^m(y)||^t + AC^m ||x - y||^t \\
&\leq A^m C_0 A^{mt} ||x - y||^t + AC^m ||x - y||^t \\
&\leq C^{m+1} ||x - y||^t
\end{aligned}$$

So using this inductive argument we have proven (7.12) holds. Now, take $\zeta > 0$ such that $(C\zeta^t)^n < \delta, n \geq 1$. Then

$$||(D_y g^m) - (D_x g^m)|| \leq C^m ||x - y||^t \leq C^m \zeta^{mt} < \delta, y \in B_\zeta m(x)$$

The result follows from Lemma 7.9. □

Lemma 7.14. *Let $\Lambda_n(y)$ be the set of all $w \in y + E^u(x)$ s.t. $g^j(w) \in D_{\rho \frac{g^j(x)}{K_1}}(g^j(x))$ for all $0 \leq j \leq n$. If $g^n(x) \in K$ and $y \in E^s(x)$, if $\Lambda_n(y) \neq \emptyset$, then $g^n(\Lambda_n(y))$ is an $(E^s(g^n(x)), E^u(g^n(x)))$-graph with dispersion $\leq c$.*

Proof. The lemma holds for $n = 0$. Assume it holds for some $m > 0$. If $g^{m'}(x) \in K, g^j(x) \notin K$, $m < j < m'$ and $\Lambda_{m'}(y) \neq \emptyset$, then we have that

$$g^{m'}(\Lambda_{m'}(y)) \subset g^{m'-m}(g^m(\Lambda_m(y))) \cap D_{\rho \frac{g^{m'}(x)}{K_1}}(g^{m'}(x))$$

We know that $g^m(\Lambda_m(y))$ is an $(E^s(g^m(x)), E^u(g^m(x)))$-graph with dispersion $\leq c$ and

$$\begin{aligned}
g^m(\Lambda_m(y)) &\subset D_{\rho \frac{g^m(x)}{K_1}}(g^m(x)) \\
&\subset B_{k_2 \rho \frac{g^m(x)}{K_1}}(g^m(x)) \subset B_{\zeta^{m'-m}}(g^m(x))
\end{aligned}$$

Since $N(g^m(x)) = m' - m$, by Lemma 7.9 it follows that $g^{m'}(\Lambda_m(y)) = g^{m'-m}(g^m(\Lambda_m(y)))$ is an $(E^s(g^{m'}(x)), E^u(g^{m'}(x)))$-graph with dispersion c, and thus so is its subset $g^{m'}(\Lambda_{m'}(y))$.

□

Theorem 7.15. *Let $g, f, M, K, E^u(x), E^s(x), \mu, \chi$ be defined as above. Then*

$$h_\mu(f) = \int_M \chi d\mu \tag{7.16}$$

Proof. Fix the constant c from Lemma 7.11 small enough such that there exists some $a > 0$ such that if $x \in K, y \in M$ with $|x-y| < a$, then for all $E \subset T_yM$ which is an $(E^s(x), E^u(x))$-graph with dispersion $\leq c$,

$$|\log|\frac{\det(D_y g)}{E}| - \log|\frac{\det(D_x g)}{E^u(x}|| \leq \epsilon$$

Recall that we defined D_r (7.10) as follows: $D_r(x) = \{x + y_1 + y_2 \mid y_1 \in E^s(x), y_2 \in E^u(x), ||y_1|| \leq r, ||y_2|| \leq r\}$, with $K_2 > K_1 > 0$, $r_1 > 0$ such that $B_{k,r}(x) \subset D_r(x) \subset B_{K_2 r}(x)$ for all $x \in K, 0 < r \leq r_1$.

Now we define a new function $N(x)$: If $x \in K$, define $N(x)$ the least value $N(x)$ such that $g^{N(x)}(x) \in K$.

We prove that $N(x)$ is both well-defined and integrable at a.e. x. If $W_j = \{x \in K \mid N(x) = j\}$, then $\bigcup_{n\geq 0} g^n(K) = \bigcup_{j=1}^{\infty} \bigcup_{i=0}^{j-1} g^i(W_j)$ mod 0. Sets on the righthand side of the union are pairwise disjoint, so

$$\begin{aligned} 1 > \mu(\bigcup_{n\geq 0} g^n(K)) &= \sum_{j=1}^{\infty} \sum_{i=1}^{j-1} \mu(g^i(W_j)) \\ &= \sum_{j=1}^{\infty} j\mu(W_j) = \int_M N d\mu \end{aligned}$$

Now we extend the domain of N to all of M by setting $N(x) = 0$ for $x \notin K$. Define $\rho : M \to (0,1)$ as $\rho(x) = \min(a, \frac{K_1}{K_2}\zeta^{N(x)})$. Since N is integrable, $\log \rho$ is also integrable. Let ν denote the Lebesgue measure on M.

We now show that there exists some $K' \subset K$ with $\mu(M \setminus K') \leq 2\sqrt{\epsilon}$ s.t.

$$h_\nu(g, \rho, x) \geq N(\chi(x) - \epsilon - 4C\epsilon) - \epsilon \tag{7.17}$$

for μ-a.e. x, where $C = \sup\{\log \frac{|\det(D_p(f)|}{E} \mid p \in M, E \subset T_pM\}$.

We have from the Birkhoff Ergodic Theorem (Theorem 3.6) that

$$\mu(\{x \mid \lim_{n\to\infty} \frac{1}{n}\#\{0 \leq j \leq n \mid g^j(x) \notin K\} \leq \sqrt{\epsilon}\}) \geq 1 - \sqrt{\epsilon}$$

Then there exists some compact set $K_1 \subset K$ such that $\mu(K_1) \geq 1 - 2\sqrt{\epsilon}, N_0 > 0$ s.t. for all $n \geq N_0$, $\#\{0 \leq j \leq n \mid g^j(x) \notin K\} \leq 2n\sqrt{\epsilon}$, for all $x \in K_1$. Let $K' = K \cap K_1$. Then $\mu(K \setminus K') \leq \mu(M \setminus K_1) \leq 2\sqrt{\epsilon}$. Fix any $x \in K'$. There exists some $B > 0$ s.t.

$$\nu(S_n(g, \rho, x)) = B \int_{E^s(x)} \nu[(y + E^u(x)) \cap S_n(g, \rho, x)] d\nu$$

for all $n > 0$.

Thus we wish to show that

$$\limsup_{n\to\infty} \inf_{y\in E^s(x)} \frac{1}{n}[-\log \nu((y+E^u(x))\cap S^n(g,\rho,x))] \geq N(\chi(x)-\epsilon-4C\sqrt{\epsilon})-\epsilon$$

Let $\Lambda_n y$ be as defined in Lemma 7.14. Since $B_{K,r}(x)\subset D_r(x)\subset B_{K_2 r}(x)$, we have that

$$\Lambda_n(y)\supset (y+E^u(x))\cap S_n(g,\rho,x)$$

and thus our claim is reduced to showing

$$\limsup_{n\to\infty} \inf_{y\in E^s(x)} \frac{1}{n}[-\log \nu(\Lambda_n(y))] \geq N(\chi(x)-\epsilon-4C\sqrt{\epsilon})-\epsilon \tag{7.18}$$

To prove this claim take $D>0$ s.t. $D>\text{volume}(G)$, for all $(E^s(w),E^u(w))$ graphs G with dispersions $\leq c$ in $D_{\rho(w)}(w), w\in K$. If $g^n(x)\in K$, $y\in E^s(x)$,

$$D>\text{vol}(g^n(\Lambda_n(y)))=\int_{\Lambda_n(y)}|\frac{\det(D_z g^n)}{T_z\Lambda_n(y)}|d\nu$$

Let $F_n=\{0\leq j<n \mid g^j(x)\in K\}$. If $n\geq N_0$, then

$$\begin{aligned}
&\log|\frac{\det(D_z g^n)}{T_z\Lambda_n(y)}| = \sum_{j=0}^{n-1}\log|\frac{\det(D_{g^j(z)}(g)}{T_{g^j(z)}g^j(\Lambda_n(y))}|\\
&\geq \sum_{j\in F_n}\log|\frac{\det(D_{g^j(z)}(g))}{T_{g^j(z)}g^j(\Lambda_n(y))}| - NC(n-\#F_n)\\
&\geq \sum_{j\in F_n}\log|\frac{\det(D_{g^j(z)}(g))}{E^u(g^j(x))}| - \epsilon n - NC(n-\#F_n)\\
&\geq \sum_{j\in F_n}\log|\frac{\det(D_{g^j(z)}(g))}{E^u(g^j(x))}| - \epsilon n - 2NC(n-\#F_n)\\
&\geq nN(\chi(x)-\epsilon)-\epsilon n-2NC(n-\#F_n)\\
&\geq nN(\chi(x)-\epsilon)-\epsilon n-4NCn(n-\#F_n)
\end{aligned}$$

Thus we have that $D>\nu(\Lambda_n(y))\cdot e^{n(N(\chi(x)-\epsilon-4C\epsilon))}$ for all $y\in E^s(x), n\geq N_0$, proving (7.18).

Now we have shown there exists some $K'\subset K$ which follows (7.17), using Proposition 7.7, we have

$$\begin{aligned}
&h_\mu(g)\geq\int_M h_\nu(g,\rho,x)d\mu\\
&\geq\int_{K'}h_\nu(x)\\
&\geq N\int_{K'}\chi\mu - N(\epsilon+4C\sqrt{\epsilon})-\epsilon\\
&\geq N\int_M\chi d\mu - N(\epsilon+6C\sqrt{\epsilon})-\epsilon\\
&\implies h_\mu(f)\geq\frac{1}{N}h_\mu(g)\geq\int_M\chi d\mu-(\epsilon+6C\sqrt{\epsilon})-\frac{\epsilon}{N}
\end{aligned}$$

ϵ is arbitrarily small, so assuming Ruelle's inequality holds, we achieve the desired equality

$$h_\mu(f) = \int_M \chi d\mu$$

□

8. SRB Measures

Here we describe a specific conservative invariant probability measure. These were introduced by Sinai, Ruelle, and Brown, and are thus titled SRB measures. We cite surveys by Young and Wilkinson [5], [14], and [16].

Theorem 8.1. *Let $(f, \mathcal{A}, \mu)$ be an mpt on a manifold M with positive Lyapunov exponents almost everywhere. Then Pesin's formula (Theorem 7.15) holds if and only if μ has absolutely continuous conditional measures on unstable manifolds.*

The proof can be found in [5]. We use measures which fulfill this criteria as one definition for SRB measures:

Definition 8.2. A measure which fulfills Theorem 8.1 is titled an *SRB measure*.

This definition is relatively unambiguous; an SRB measure is one which fulfills Pesin's formula (since one of the assumptions of Pesin's formula is that μ is absolutely continuous with respect to f and thus with respect to the unstable submanifolds of f.) Interestingly, SRB measures were not discovered in the context of Pesin's formula; rather, they were derived for *Anosov diffeomorphisms* and *Axiom A attractors*, defined as follows (respectively). Let $f : M \to M$ be a diffeomorphism of a compact Riemannian manifold M.

Definition 8.3. We call f an *Anosov diffeomorphism* if for each $x \in M$, $T_xM = E^u(x) \oplus E^s(x)$, where E^u and E^s are Df-invariant, $Df|_{E^u}$ is uniformly expanding and $Df|_{E^s}$ is uniformly contracting.

Definition 8.4. Let Λ be an attractor of f (recall the definition of *attractor* given in Definition 4.3). We call f an *Axiom A attractor* if $T_x\Lambda = E^s \oplus E^u$, where E^s and E^u have the same properties as above.

In this context we introduce another equivalent method of characterizing an SRB measure for an Axiom A attractor:

Theorem 8.5. *Let f be a C^2 diffeomorphism with Axiom A attractor Λ. Then there is a unique f-invariant Borel probability measure μ on Λ such that for every continuous ϕ on U, almost everywhere,*

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

Thus the convergence of the average of each continuous function composed with the iterated map yields the integral of that continuous function. Finding SRB measures for systems with attractors, absolutely continuous measures, and uniform hyperbolicity is possible (we have studied them without naming them in our proof of the Pesin entropy formula.) However, there is a case in which the existence of SRB measures is highly non-trivial, and this is for systems with *nonuniform hyperbolicity.*

Definition 8.6. A conservative diffeomorphism $f : M \to M$ is *nonuniformly hyperbolic* if the invariant measure representing volume is hyperbolic. In this sense, uniformly hyperbolic systems (those with Lyapunov exponents not equal to zero) are also nonuniformly hyperbolic.

There are very few cases in which SRB measures for nonuniformly hyperbolic systems without uniform hyperbolicity have been found. One such example is for the *Henon mappings.* These are a two-dimensional family of mappings in the plane, defined as follows:

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

For $a = 2, b = 0$, Benedicks and Young constructed an SRB measure for this family of attractors, given as follows.

Theorem 8.7. *With $T_{a,b}$ defined as above, there exists a rectangle $\Delta = (a_0, a_1) \times (0, b_1)$ in parameter space such that for all $(a, b) \in \Delta$, $T = T_{a,b}$ has an attractor Λ. In addition, for all sufficiently small $b > 0$, there exists a positive measure set Δ_b such that for all $a \in \Delta_b$, $T = T_{a,b}$ admits a unique SRB measure μ which is supported on the entire attractor and (T, μ) is Bernoulli.*

We omit the proof, but state it to demonstrate that there is some argument for the existence of SRB measures for nonuniformly hyperbolic systems. Both proofs detailing these SRB measures (the preceding theorem and that the Lorenz attractor is mixing, Theorem 6.7), use invariant cone fields in their results. More general results constructing SRB measures are being studied, in contexts of both systems with and without uniform hyperbolicity. Their practicality when describing chaotic dynamical systems is apparent, however, and thus worth further investigation.

9. Conclusion

What preceded was a brief introduction to differentiable dynamical systems, smooth ergodic theory, hyperbolicity, and a hint of chaos theory. Ultimately, this culminated in a proof of the Pesin entropy formula.

Acknowledgments

I want to give thanks to my mentor, DeVon Ingram, who's taught me more math than I anticipated knowing this summer, and always pushed me to do better than my expectations. In addition, thanks to Profs. Rudenko and Babai for excellent apprentice lectures and introducing me to a variety of foundational topics in mathematics. Thanks to Prof. Peter May for organizing the REU and providing the opportunity to learn more about math. And finally, thanks to Dr. Jason MacLean and Hal Rockwell for supporting my academic and research endeavors this summer.

References

[1] Holmes, P. (2007a). History of dynamical systems. *Scholarpedia*, 2(5), 1843. https://doi.org/10.4249/scholarpedia.1843

[2] Bass, Richard (n.d.) *Real analysis for graduate students*. Retrieved August 7, 2023, from https://www.math.wustl.edu/ victor/classes/ma5051/rags100514.pdf

[3] Algoet, P. H., & Cover, T. M. (1988). A Sandwich Proof of the Shannon-McMillan-Breiman Theorem. *The Annals of Probability*, 16(2), 899–909. https://doi.org/10.1214/aop/1176991794

[4] Hasselblatt, B., & Katok, A. (2002), Handbook of Dynamical Systems. (Handbook of Dynamical Systems; Vol. 1, No. PART A). https://doi.org/10.1016/S1874-575X(06)80025-1
[5] Young, L.-S (n.d.). *Ergodic Theory of Differentiable Dynamical Systems.* https://cims.nyu.edu/ lsy/papers/lecturenotes93.pdf
[6] Teschl, G. (n.d.). *Ordinary Differential Equations and Dynamical Systems.* https://www.mat.univie.ac.at/ gerald/ftp/book-ode/ode.pdf
[7] Obata, D (n.d.). *A Simple Proof of Birkhoff's Ergodic Theorem.*
[8] Lee, J.M. *Smooth Manifolds.* https://link.springer.com/chapter/10.1007/978-0-387-21752-9_1
[9] Filip, S (2017). Notes on the Multiplicative Ergodic Theorem. *Ergod. Th. Dynam. Sys* 39: 1153-1189. https://doi.org/10.1017/etds.2017.68.
[10] Schecter, S (n.d.). *Proof of the Stable Manifold Theorem.* https://schecter.math.ncsu.edu/ma_732_sp13/stable.pdf
[11] Luzzatto, S., Melbourne, I., & Paccaut, F. (2005). The Lorenz attractor is mixing. *Communications in Mathematical Physics*, 260(2), 393–401. https://doi.org/10.1007/s00220-005-1411-9
[12] https://users.math.msu.edu/users/newhous7/papers/cones/cones-katok-final.pdf
[13] *Strange Attractors—Chaos & Fractals.* (n.d.). Retrieved August 9, 2023, from https://www.stsci.edu/ lbradley/seminar/attractors.html
[14] Wilkinson, A. (n.d.). *SMOOTH ERGODIC THEORY.*
[15] Mañé, R. (1981). A proof of Pesin's formula. *Ergodic Theory and Dynamical Systems*, 1(1), 95–102. https://doi.org/10.1017/S0143385700001188
[16] Young, L.-S. (n.d.). *What are SRB measures, and which dynamical systems have them?*