SHADOWING AND STRUCTURAL STABILITY OF HYPERBOLIC SETS

THOMAS MATHEOS

Abstract. This paper gives an introductory glimpse into the study of discrete differentiable dynamical systems, which roughly means the study of the repeated action of a diffeomorphism on a manifold. As such, some familiarity with point-set topology and with calculus on manifolds will be helpful in understanding this paper (mainly, familiarity with the notions of homeomorphisms and of Riemannian metrics on the tangent bundle will be useful). Some basic concepts of dynamical systems are introduced, as well as the more advanced concepts of hyperbolicity and shadowing. Hyperbolicity describes a sort of invariant contraction and expansion of the system. Shadowing describes how a part of one dynamical system may be similar to a part of another. It turns out that hyperbolicity and shadowing are closely related; together they can describe how perturbations affect a dynamical system. In other words, they describe the 'stability' of a system. The relation between hyperbolicity and shadowing is formalized and proved in the second to last theorem of this paper, the Shadowing Theorem. This theorem is used to prove the final result of the paper: that hyperbolic dynamical systems are stable.

Contents

1. Introduction

A 'dynamical system' is, roughly speaking, a collection of things which change over time. For example, the interactions of air particles with the wing of an airplane could be modeled with a dynamical system. The properties of the system might indicate whether the shape of the wing allows the airplane to fly. If the shape does allow for flight, it could be asked if the airplane is 'barely' able to fly, that is, if a slight change in wing shape will prevent the airplane from flying. If the answer is no, then the shape might be called 'stable' in its flight ability, and 'unstable' if a slight change could prevent flight. Stability becomes especially important in this example when the inevitable imprecisions of human and machine craftsmanship is

Date: AUGUST 14, 2023.

considered – these are of little concern if a wing shape is stable in its flight ability, but are quite dangerous if the wing shape is unstable.

A similar notion of stability may be defined for other dynamical systems; it is called 'structural stability'. The definition of structural stability relies on a more precise understanding of the elements of a dynamical system. There are three main components: a space whose elements represent 'states' of the system, a notion of 'time', and a rule of 'time-evolution' which describes how the states of the system change with respect to time. Stability concerns 'small' changes in the system; to describe the stability of a system, there must be some notion of 'smallness' or 'distance' describing changes in the space or in the time-evolution rule. For this reason, the spaces in this paper will all be metric; as finer notions of distance are needed, the space and the time-evolution rule will be given differential structures.

The notions of time which this paper will focus on are discrete and reversible – that is, time will be in a natural correspondence to the integers. Time-evolving one step will be described by applying some function f to the space one time. Since time is reversible, f will be invertible, and applying f^{-1} can be thought of as backwards time-evolution. In order to ensure a consistency in the space of the systems considered, the domain and range of f will be the same and f will be a homeomorphism (remember from above that the space will always be metric and thus topological).

The airplane example may be made to fit these requirements. The metric space is the atmosphere, with the distance between air particles taken in the usual way (e.g. with a ruler). The measurements of the position of the air particles are made discretely, say with a video-camera. Forward time-evolution in this case is observed by advancing frame-by-frame in the video; backward time-evolution is observed by rewinding frame-by-frame. The time-evolution function is the function that describes the results of the laws of physics in action upon the atmosphere during each time period between successive frames of the video.

It turns out that compactness is fundamental in several of the results of this paper, and so the dynamical systems of this paper will all be on compact spaces. In §2, the systems considered and some basic concepts of dynamical systems will be precisely defined, including the notion of structural stability.

Once this is done, there are two remaining goals of the paper. The first is to describe a criterion for when a dynamical system is structurally stable, namely, that of 'hyperbolicity'. Roughly, a (part of a) system is hyperbolic if it has a certain notion of invariant contraction in one direction and expansion in another. Some intuition might be gained by thinking of the squishing and pulling of a sphere of Play-Doh into an ellipsoid: at each forward time-evolution, the top and bottom are squished together and the sides are pulled apart; at each backwards time-evolution, the sides are squished together and the top and bottom are pulled apart. Something analogous can happen in dynamical systems; in §3 this "something" is defined and some properties are discussed.

The final goal of the paper is to show the deep connection between the notion of 'shadowing' and the structural stability of hyperbolic systems. Shadowing roughly gives a notion of closeness between parts of systems. A common example is that of floating-point arithmetic in computers: the sequence of computations carried out by a computer might be different then the sequence of 'true' computations, because a computer might introduce some error at each step. Nevertheless, in certain situations, the computer can be programmed so that not only is the error term introduced at each step quite small, but also so that the difference between the final result of the computer and of the error-free computation are quite small. This floating-point arithmetic example is directly analogous to the connection between hyperbolicity and shadowing which is stated in the Shadowing Lemma. A powerful generalization of this connection is given by the Shadowing Theorem. Both are presented and proved in §4.

Finally, §5 uses the Shadowing Theorem to prove that hyperbolic parts of dynamical systems are structurally stable.

Throughout the paper, undefined or vague concepts are introduced in single quotes, and definitions are given in italics. If a part of the paper is restated directly or a phrase is being marked, it will be in double quotes. For instance, the first use of "dynamical system" was in single quotes and when it is defined below it will be italicized. The notions of the flight ability of a wing being "stable" or "unstable" are vague and so they were introduced with single quotes, but "strongly structurally stable" will be italicized when this notion is defined.

2. Basic Concepts

Let X be a compact metric space and let $f : X \to X$ be a homeomorphism. Then f^n denotes the composition of f with itself n times, f^{-n} denotes $(f^n)^{-1}$, and f^0 denotes the identity function Id. For any integer n, f^n is called an *iterate* of f. When n is positive f^n is called a *positive iterate* of f; when n is negative f^n is called a *negative iterate* of f. It is also said that $f^n(x)$ is a (*positive* or *negative*) *iterate* of x. Notice that f^n is a homeomorphism and that $f^n \circ f^m = f^{n+m}$ for any integers n and m. Sometimes, when there can be no confusion, fg will be written instead of $f \circ g$.

Each integer is called a 'point in time' and applying f to any element x in the space X is called 'time-evolving' x; the constituents of a dynamical system as discussed in the introduction are present. Indeed, the set of iterates $\{f^n\}_{n\in\mathbb{Z}}$ of f is called a *dynamical system*. (It may be more accurate to call the set of iterates of f a *discrete time dynamical system* but, since all systems in this paper are discrete, the 'discrete time' will always be omitted.) Sometimes the pair (X, f) is called a dynamical system and iteration as time-evolution is to be understood. Some authors refer to f itself as a dynamical system, as the space is inherent in the definition of f. It would be reasonable to let "dynamical system" refer to the iterates of a homeomorphism $f : X \to X$ where X is not compact; however, compactness is important for many results about hyperbolicity, which is an essentially uniform phenomenon, and therefore all dynamical systems in this paper act on compact spaces.

For any x in X the set $\{f^n(x)\}_{n\in\mathbb{Z}}$ of the iterates of f acting on x is called the *orbit* of x under f and is denoted $\mathcal{O}(x, f)$ or merely $\mathcal{O}(x)$. A set Λ is called *invariant* under f if $f(\Lambda) = \Lambda$, that is, if the orbit of any $x \in \Lambda$ is contained in Λ. The set of the positive (negative) iterates of f acting on x is called the *positive* (*negative*) *orbit* of x under f and is denoted $\mathcal{O}^+(x, f)$ ($\mathcal{O}^-(x, f)$) or merely $\mathcal{O}^+(x)$ ($\mathcal{O}^-(x)$). Sometimes the phrase 'under f' will be omitted. If there is an integer n such that $f^n(x) = x$ then the orbit of x is said to be *periodic* and $\mathcal{O}(x)$ is called a *periodic orbit*. In this case the point x is called a *periodic point* and x itself is also described as *periodic*. If m is the least natural number such that $f^m(x) = m$ then

x is said to have *period* m. If $m = 1$ then x is called a *fixed point*. Denote the set of periodic points of f by $P(f)$ and the set of fixed points of f by $F(x)$.

Example 2.1. Let $X = S^1$, the unit circle, and let $f : X \to X$ be the rotation by $2\pi\theta$ for some constant $\theta \in \mathbb{R}$. Then (X, f) is a dynamical system. If θ is rational and $\theta = \frac{m}{n}$ in lowest terms then $P(f) = S^1$ and every point has period n. If θ is irrational then no point of S^1 is periodic. In fact, the orbit of any point is dense in S^1 (see [4] for details).

Example 2.2. Let X be the one-point compactification of $\mathbb{R}$ and let $f : X \to X$ be the contraction $f(x) = \frac{1}{2}x$. Then (X, f) is a dynamical system. The orbit of any point $x \in \mathbb{R}$ is $\{(\frac{1}{2})^n x\}_{n\in\mathbb{Z}}$ and the only periodic points in X are 0 and ∞, which are fixed.

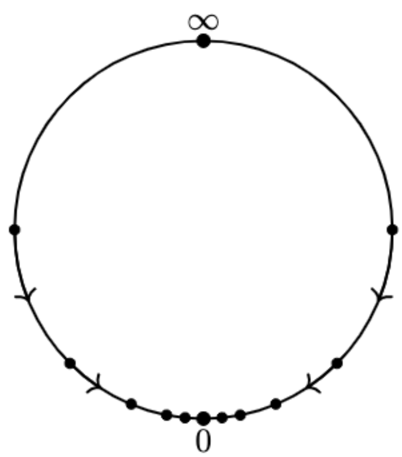

Figure 1. Contraction of the Unit Circle.

Example 2.3. The one-point compactification of $\mathbb{R}$ is homeomorphic to the unit circle. Identifying 0 with the bottom of the circle and ∞ with the top, the map in the previous example is depicted pictorially above.

One way to think of a periodic point is that it 'eventually returns to itself'. This idea is called recurrence, and there are other (weaker) forms of recurrence. For example, some subsequence of $\mathcal{O}^+(x)$ or of $\mathcal{O}^-(x)$ may converge to a point y. In the first case, y is called an ω-*limit point* of x. In the second case, y is called an α-*limit point* of x. The set of all ω-limit points of x is denoted $\omega(x)$ and the set of all α-limit points of x is denoted $\alpha(x)$. Roughly, the orbit of x keeps approaching nearer and nearer the points in these sets.

These limit point sets give rise to the notion of 'topological transitivity': a compact invariant set $\Lambda \subset X$ of f is *topologically transitive* if there exists $x \in \Lambda$ such that $\omega(x) = \Lambda$. This roughly means that for any $y \in \Lambda$ the positive orbit of x eventually gets arbitrarily close to y. Topological transitivity is often referred to merely as "transitivity", as is done in the rest of this paper. Transitivity is closely related to orbit density, as Birkhoff showed.

Theorem 2.4 (Birkhoff)**.** *If Λ is a compact invariant set then the following conditions are equivalent:*

(1) Λ is transitive.
(2) For any two open sets U and V in Λ, there is a natural number n such that $f^n(U) \cap V \neq \varnothing$.
(3) There is $x \in \Lambda$ such that the positive orbit of x is dense in Λ.

In the third condition, "positive orbit" can be replace by "negative orbit", but not by "orbit". A proof of the theorem is given in [4].

Example 2.5. If f is an irrational rotation of S^1, the unit circle, then the positive (and negative) orbit of any $x \in S^1$ is dense in S^1, and so (by Theorem 2.4) S^1 is transitive for f.

Another type of recurrence is found in 'periodic pseudo-orbits'. Let $a < b$ be elements of $\{-\infty\} \cup \mathbb{Z} \cup \{+\infty\}$. A set $P = \{x_n\}_{a<n<b}$ is called an *ε-pseudo-orbit* of f if $d(f(x_n), x_{n+1}) < \varepsilon$ for each $a < n < b-1$. If there is a whole number m such that $x_{m+n} = x_n$ for all $a < n < b-m$ then the pseudo-orbit is called *periodic*. In the case that a and b are integers, and so the ε-pseudo-orbit P is finite, P is called an *ε-chain from x_{a+1} to x_{b-1}*. Any chain may be re-indexed to start at zero and go through the natural numbers, so a chain is often presented as going from x_0 to x_n for some $n \in \mathbb{N}$. Finite periodic ε-pseudo-orbits are called *periodic ε-chains*.

A point $x \in X$ is called *chain recurrent* if, for any $\varepsilon > 0$, there is a periodic ε-chain which contains x. Chain recurrence forms an equivalence class. The set of these equivalence classes is called the *chain recurrent set* of f and is denoted $\mathrm{CR}(f)$.

Pseudo-orbits of f may be thought of as actual orbits of dynamical systems which are 'close' to f. Given an ε-pseudo-orbit P of f, a function f' can be constructed (by adding some term near each element of P) such that P is an actual orbit of f'. It turns out that when ε is small enough – and when P is close to a 'hyperbolic set', which will be defined in §3 – there is an orbit of f itself which has a contiguous subset close to P. This is the Shadowing Lemma presented at the beginning of §4. A generalization of these ideas is used to prove the final theorem of this paper.

Before motivating and defining structural stability, it should be noted that there are other notions of recurrence not mentioned in this paper. For more details, see [1] or [4].

Some dynamical systems which are in some sense 'close' to each other exhibit dramatically different orbital structures, while others exhibit 'the same' orbital structures. Consider Examples 2.1 and 2.2 above. Take f and X as in Example 2.2. Let $g : X \to X$ be a differentiable function with Dg close to $\frac{1}{2}$ everywhere and the distance between f and g small everywhere. Then g is a contraction of $\mathbb{R}$ and the only periodic points of (X, g) are ∞ and some x_0 close to 0; both are fixed points (see [1] for details).

The situation is quite different for rigid rotations of S^1 (Example 2.1). Since both the rationals and irrationals are dense in $\mathbb{R}$, very small changes in the angle of rotation take the system from consisting entirely of periodic points with uniform period to consisting entirely of points whose orbits are dense in S^1.

It is said that the contraction in Example 2.2 is structurally stable while rotations of the circle are not. The formal definition of structural stability relies on the notion of topological conjugacy and on imposing a topology on the space of C^r diffeomorphisms on a given space.

Definition 2.6. Two homeomorphisms $f : X \to X$ and $g : Y \to Y$ are *topologically conjugate* if there exists a homeomorphism $h : X \to Y$ such that $hf = gh$. If f and g are topologically congutate then it is said that f is *topologically conjugate to* g, and g to f. The homeomorphism h is called a *topological conjugacy* or merely a *conjugacy.*

Some intuition might be gained from the special case where f and g are linear transformations. In this case the conjugacy h is a linear change of coordinates. It is sometimes said that f and g are 'the same linear transformation up to a change in coordinates'. When f and g are not linear, h is a sort of nonlinear continuous change of coordinates, and it might be said that f and g are 'the same dynamical system up to a homeomorphism'.

Notice that $hf = gh$ implies $hf^n = g^n h$ for all integers n. This means a conjugacy h maps orbits of f to orbits of g, that is,

$$h(\mathcal{O}(x, f)) = \mathcal{O}(h(x), g)$$

for any $x \in X$. Further,

$$h(P(f)) = P(g), \quad h(\omega(x, f)) = \omega(h(x), g), \text{ and } \quad h(CR(f)) = CR(g).$$

Thus the orbit structures of two dynamical systems (X, f) and (Y, g) are topologically equivalent if f and g are topologically conjugate. Topological conjugacy is fundamental in C^r (r-times continuously differentiable) structural stability.

Since this paper is concerned with C^r structural stability, the spaces considered must have some differentiable structure. Accordingly, from this point onwards the spaces considered (unless explicitly indicated otherwise) will be compact C^∞ Riemannian manifolds without boundary, denoted M, and the functions of the dynamical systems will be C^1 diffeomorphisms (that is, the functions and their inverses are continuously differentiable). For a more thorough discussion of manifolds, see [2].

Denote by $\text{Diff}^r(M)$ the set of C^r diffeomorphisms of M endowed with the C^r topology. The C^r topology can be described with a metric as follows. Fix a finite cover of admissible coordinate neighborhoods (U_i, φ_i), $i = 1, 2, \ldots, N$ of M. Then the metric d_{C^r} is, for any f, g in $\text{Diff}^r(M)$,

$$\sup\Big\{|\varphi_j f \varphi_i^{-1}(x) - \varphi_j g \varphi_i^{-1}(x)|,\ |D(\varphi_j f \varphi_i^{-1})(x) - D(\varphi_j g \varphi_i^{-1})(x)|,$$
$$|D^2(\varphi_j f \varphi_i^{-1})(x) - D^2(\varphi_j g \varphi_i^{-1})(x)|, \ldots, |D^r(\varphi_j f \varphi_i^{-1})(x) - D^r(\varphi_j g \varphi_i^{-1})(x)|\Big\},$$

where the supremum is taken over all i, j, and x for which the expressions are well-defined. Notice that the choice of cover of charts does not affect the topology. If f and g are defined only on proper subsets of M then $d_{C^r}(f, g)$ is taken to be the same as above but with the supremum restricted to those x in the intersection of the domains of f and g.

For the rest of the section take r, m natural numbers with $m \geq r$. If a function is a C^m diffeomorphism then it is also a C^r diffeomorphism and therefore belongs to Diff^m and Diff^r.

Definition 2.7. A diffeomorphism $f \in \text{Diff}^m(M)$ is C^r *structurally stable* if there exists a C^r neighborhood $\mathcal{U}$ of f in $\text{Diff}^r(M)$ such that every $g \in \mathcal{U}$ is topologically conjugate to f.

There is a stronger version of structural stability, called *strong structural stability.*

Definition 2.8. A diffeomorphism $f \in \text{Diff}^m(M)$ is C^r *strongly structurally stable* if there exists a C^r neighborhood $\mathcal{U}$ of f in $\text{Diff}^r(M)$ such that every $g \in \mathcal{U}$ is topologically conjugate to f and the conjugacy h_g may be chosen such that both h_g and h_g^{-1} converge uniformly to the identity as g converges to f in the C^r topology.

In other words, a C^m diffeomorphism f is C^r structurally stable if small C^r perturbations cannot change topologically the orbit structure of f. The term perturbation here is not precise; it sometimes refers to a C^r diffeomorphism g which is C^r close to f and it sometimes refers to the difference of f and g.

If $f \in \text{Diff}^{r+1}(M)$ is C^r structurally stable then it is also C^{r+1} structurally stable. Thus C^1 structural stability is the strongest; the phrase '(strong) structural stability' refers to C^1 (strong) structural stability. There is no notion of C^0 structural stability because C^0 perturbations are too damaging; see [4] or [1] for details.

In the next section the concept of a hyperbolic set is introduced. A set is only hyperbolic as part of a dynamical system, so any reference to a hyperbolic set Λ is implicitly a reference to a function f. When $f\big|_\Lambda$ (that is, f restricted to Λ) is (strongly) structurally stable, it is said that the set Λ itself is (strongly) structurally stable. The main result of this paper is that every hyperbolic set is strongly structurally stable.

3. Hyperbolicity

Hyperbolicity is the key to structural stability, and structural stability is an inherently differential notion. Accordingly, hyperbolicity imposes some structure on derivatives and tangent maps. Therefore, before defining hyperbolic sets and discussing some of their important properties, it will be helpful to discuss tangent bundles and the like, as well as some notation.

Unless explicitly indicated, M will always refer to a compact C^∞ Riemannian manifold without boundary and f will always refer to a C^1 diffeomorphism. For any $x \in M$, T_xM denotes the tangent space at x. If Λ is a subset of M then the disjoint union $\bigsqcup_{x\in\Lambda} T_xM$ is denoted $T_\Lambda M$. The tangent bundle $\bigsqcup_{x\in M} T_xM$ is denoted TM.

It will be important for the linear subspaces of the tangent spaces to have a topology. Let m be a natural number. The m-Grassmann space

$$G^m(M) = \{V \mid V \text{ is an } m\text{-dimensional linear subspace of } T_xM,\ x \in M\}$$

is given a topology in the following way. Let $\{x_k\} \subset M$ and $\{E(x_k)\} \subset G^m(M)$ be sequences such that $E(x_k)$ is an m-dimensional linear subspace of $T_{x_k}M$ for every k. Take some $x \in M$ and an m-dimensional linear subspace $E(x)$ of T_xM. If there is a basis $\{e^1_{x_k}, \dots, e^m_{x_k}\}$ of $E(x_k)$ for every k and a basis $\{e^1_x, \dots, e^m_x\}$ of $E(x)$ such that $e^j_{x_k} \to e^j_x$ for every $j \in 1, 2, \dots m$ then $\{E(x_k)\}$ is said to converge to $E(x)$. This definition of convergence gives a topology on G^m.

The notion of the convergence of $e^j_{x_k} \to e^j_x$ above comes from the Riemannian metric on TM. The norm induced by the Riemannain metric will be denoted $|\cdot|$, that is, $|v|$ automatically means $|v|_x$ for any $v \in T_xM$. Since the base point x is implicit in v, there is no ambiguity here in base point and so there is no ambiguity in the notation (as long as it is clear what v denotes, which will always be the case).

Similarly, the tangent map will be denoted Tf. Comparing to the notation with base points marked and to derivative notation gives

$$Tf(v) = T_xf(v) = Df(x)(v)$$

where $v \in T_xM$. As before, since the base point x is implicit in v, there is no ambiguity in the notation. Also, Tf^n will be written in place of $T(f^n) = (Tf)^n$.

The final note before defining hyperbolic sets concerns distance on M. A metric is induced on M by the Riemannian metric on TM by defining $d(x, y)$ to be the infimum of the lengths of piecewise differentiable curves joining x and y. If d is written without a subscript it refers to this metric; any subscript of d will be of the form C^r and d_{C^r} refers to the metric defined in §2.

Definition 3.1. Let $f : M \to M$ be a C^1 diffeomorphism. An invariant set $\Lambda \subset M$ of f is *hyperbolic* if

(1) for each $x \in \Lambda$ the tangent space T_xM splits into a direct sum

$$T_xM = E^s(x) \oplus E^u(x)$$

invariant in the sense that

$$Tf(E^s(x)) = E^s(f(x)), \quad Tf(E^u(x)) = E^u(f(x))$$

and

(2) there exist constants $C \geq 1$ and $0 < \lambda < 1$ such that

$$|Tf^n(v)| \leq C\lambda^n|v|, \qquad \forall x \in \Lambda, v \in E^s(x), n \geq 0,$$
$$|Tf^{-n}(v)| \leq C\lambda^n|v|, \quad \forall x \in \Lambda, v \in E^u(x), n \geq 0.$$

Sometimes, it is said that Λ is a hyperbolic set *of* or *for* f. It is also said that (Λ, f) is a *hyperbolic dynamical system.* If Λ is a single orbit then it is called a *hyperbolic orbit*.

Remark 3.2.

(1) Since M is compact, the hyperbolicity of a set is independent of the choice of Riemannian metric.
(2) If Λ is hyperbolic for f then it is also hyperbolic for f^{-1}.
(3) Any invariant subset of a hyperbolic set is hyperbolic.
(4) By invariance, the two inequalities also hold for negative iterates of v. That is, for any $x \in \Lambda$ and any (possibly negative) integer m,

$$|Tf^n(f^m(v))| \leq C\lambda^n|Tf^m(v)|, \qquad \forall x \in \Lambda, v \in E^s(x), n \geq 0,$$
$$|Tf^{-n}(f^m(v))| \leq C\lambda^n|Tf^m(v)|, \quad \forall x \in \Lambda, v \in E^u(x), n \geq 0.$$

In particular, taking $m = -n$ gives

$$|Tf^{-n}(v)| \geq C^{-1}(\lambda^{-1})^n|v|, \quad \forall x \in \Lambda, v \in E^s(x), n \geq 0$$

and taking $m = n$ gives

$$|Tf^n(v)| \geq C^{-1}(\lambda^{-1})^n|v|, \qquad \forall x \in \Lambda, v \in E^u(x), n \geq 0.$$

In other words, if $v_s \in E^s(x)$, $v_u \in E^u(y)$ for some $x, y \in \Lambda$, positive iterates of f contract v_s and expand v_u while negative iterates expand

v_s and contract v_u. The invariance of $E^s(x)$ and $E^u(x)$ means that v_u 'keeps on expanding (contracting)' under positive (negative) iterates of f. Similarly for v_s.

(5) Sometimes a hyperbolic set Λ will be called 'hyperbolic with splitting $T_\Lambda M = E^s \oplus E^u$'. In this situation,
$$E^s := \bigsqcup_{x\in\Lambda} E^s(x) \quad \text{and} \quad E^u := \bigsqcup_{x\in\Lambda} E^u(x),$$
and $T_\Lambda M = E^s \oplus E^u$ is called a *hyperbolic splitting.*

(6) Sometimes, to make explicit the system for which Λ is hyperbolic, $E^s(x)$ and $E^u(x)$ will be written as $E^s(x, f)$ and $E^u(x, f)$, respectively.

Before exploring properties of hyperbolic sets, some examples should be provided.

Example 3.3 (Hyperbolic Linear Isomorphism)**.** Let $A : E \to E$ be a linear isomorphism of a finite-dimensional vector space E such that no eigenvalue of A has magnitude 1. Such a map is called a *hyperbolic linear isomorphism.* It turns out (see [4]) that A is a hyperbolic linear isomorphism if and only if E admits a splitting $E = E^s \oplus E^u$ such that there exist constants $C \geq 1$, $0 < \lambda < 1$ such that
$$|A^n(v)| \leq C\lambda^n|v|, \qquad \forall x \in \Lambda, v \in E^s, n \geq 0,$$
$$|A^{-n}(v)| \leq C\lambda^n|v|, \quad \forall x \in \Lambda, v \in E^u, n \geq 0.$$
Since $DA = A$, E is hyperbolic for the dynamical system (E, A).

For instance, A could be diagonal with entries 1/2 and 2; then E^s and E^u would be the standard axes. In Figure 2, E^s and E^u are askew from the standard axes, as would happen if A were multiplied by a rotation matrix.

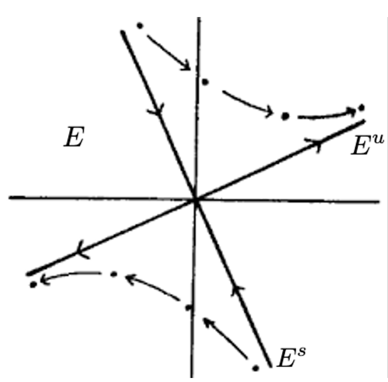

FIGURE 2. Hyperbolic splitting of E, with example orbits.

Example 3.4 (Hyperbolic Fixed Point)**.** Let (M, f) be a dynamical system where f is a C^1 diffeomorphism such that $p \in M$ is a fixed point of f and $Df(p)$ is a hyperbolic linear isomorphism. Then $\{p\}$ is a hyperbolic set and p is called a *hyperbolic fixed point* of f.

For instance, if $A : \mathbb{R}^2 \to \mathbb{R}^2$ is a diagonal matrix with entries 1/2 and 2 then A is a hyperbolic linear isomorphism, A is everywhere its own derivative, and $(0, 0)$ is a hyperbolic fixed point.

Example 3.5 (Hyperbolic Periodic Point)**.** Let (M, f) be a dynamical system where f is a C^1 diffeomorphism such that p is a fixed point of f^m and $Df^m(p)$ is a hyperbolic linear isomorphism. Notice that p is a periodic point of f. With this setup, $\mathcal{O}(p, f)$ is hyperbolic and $\mathcal{O}(p, f)$ is called a *hyperbolic periodic orbit*. Notice that p is a hyperbolic fixed point of f^m.

The notions of hyperbolic linear isomorphims, fixed points, and periodic points were around before the notion of hyperbolic sets given above. Smale, inspired by the work of Anosov and Piexoto, came up with this notion of hyperbolic set as a way to extend previous work done in structural stability to more general situations. One such situation is that of the Anosov toral automorphisms below.

Example 3.6 (Anosov Toral Automorphism)**.** A linear isomorphism $A : \mathbb{R}^2 \to \mathbb{R}^2$ is called an *Anosov automorphism* if A is hyperbolic, has integer entries, and $\det A = \pm 1$. The usual example is

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Since $|\det A| = 1$ and A^{-1}, the entries of A^{-1} are integers and $|\det A^{-1}| = 1$. Thus A^{-1} is an Anosov automorphism; the same argument shows that the inverse of any Anosov automorphism is an Anosov Automorphism.

It can be shown (for instance, in [4]) that the eigenvalues of any Anosov automorphism are two irrational numbers λ_1, λ_2 with $|\lambda_1| < 1 < |\lambda_2|$.

Anosov automorphisms induce an automorphism f on the torus $\mathbb{T}^2 = S^1 \times S^1$ as follows. Let A be an Anosov automorphism. Since A has integer entries, it maps $\mathbb{Z}^2$ to itself and $A(x + n) - A(x) \in \mathbb{Z}^2$ for any $x \in \mathbb{R}^2$ and any $n \in Z^2$. Then, if $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ is the projection that takes each component modulo 1, $f : T^2 \to T^2$ given by

$$\pi A = f\pi$$

is an automorphism; it is called an *Anosov toral automorphism*. Moreover, f is C^∞ and A^{-1} induces a C^∞ map of $\mathbb{T}^2$, which is f^{-1}. Thus Anosov toral automorphisms are diffeomorphisms. The entire torus is hyperbolic for an Anosov toral automorphism f since the Anosov automorphism which induced f is a hyperbolic linear isomorphism. Smale coined the term *Anosov diffeomorphsim* to refer to a diffeomorphism f on a manifold M such that M is hyperbolic for f.

The periodic points of Anosov toral automorphisms are dense in $\mathbb{T}^2$ and the whole torus $\mathbb{T}^2$ is transitive for an Anosov toral automorphism. A proof of this can be found in [4]. The toral automorphism induced by

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

is sometimes called *Arnold's cat map* because Arnold demonstrated its periodic and transitive natures by applying it to a picture of his cat.

Now, some properties of hyperbolic sets. Recall that a norm $|\cdot|$ is said to be C^r if, acting on every C^∞ local vector field, $|\cdot|^2$ is a C^r function. Recall also that on a compact manifold all C^0 norms are equivalent.

Theorem 3.7. *Let $\Lambda \subset M$ be a hyperbolic set of f with splitting $T_\Lambda = E^s \oplus E^u$. There is a C^∞ Riemannian metric of M, with induced norm $|\cdot|$, and there is a*

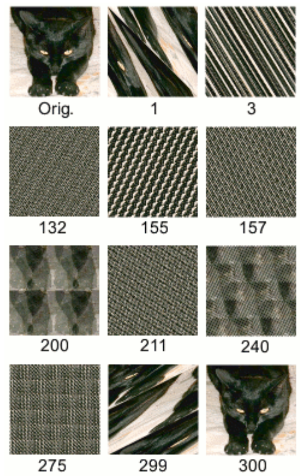

FIGURE 3. Arnold's cat map, applied to a cat. The (approximate) restoration of the cat is due to the density of the periodic orbits. The image is from [5].

constant $0 < \tau < 1$ *such that*

$$\begin{aligned} |Tf(v)| &\leq \tau|v|, && \forall v \in E^s, \\ |Tf^{-1}(v)| &\leq \tau|v|, && \forall v \in E^u. \end{aligned}$$

In short, there is a Riemannian norm for which the hyperbolic behavior is an immediate contraction and expansion. A proof is given in [4].

There are some useful characterizations of the splitting $T_xM = E^s(x) \oplus E^u(x)$ as sets. In particular, since they are characterized as sets, the splitting is unique. One characterization involves the γ-*cones about* $E^s(x)$ and $E^u(x)$, denoted $C_\gamma(E^s(x))$ and $C_\gamma(E^u(x))$, respectively. If for any $x \in \Lambda$ and any $v \in T_xM$, v_s denotes the projection of v onto $E^s(x)$ and v_u denotes the projection of v onto $E^u(x)$ then the cones are defined by

$$\begin{aligned} C_\gamma(E^s(x)) &:= \{v \in T_xM \big| \ |v_u| \leq \gamma|v_s|\}, \\ C_\gamma(E^u(x)) &:= \{v \in T_xM \big| \ |v_s| \leq \gamma|v_u|\}. \end{aligned}$$

Theorem 3.8. *Let $\Lambda \subset M$ be a hyperbolic set of f with splitting $T_\Lambda M = E^s \oplus E^u$. For any $x \in \Lambda$ and any $v \in T_xM$ let v_s denote the projection of v onto $E^s(x)$ and let v_u denote the projection of v onto $E^u(x)$. Then for any $x \in \Lambda$, $E^s(x)$ and $E^u(x)$ are characterized by*

$$\begin{aligned} E^s(x) &= \{v \in T_xM \ \big| \ |Tf^nv| \to 0 \text{ as } n \to \infty\} \\ &= \{v \in T_xM \ \big| \ \exists r > 0 \text{ such that } |Tf^nv| \leq r \ \forall n \geq 0\} \\ &= \{v \in T_xM \ \big| \ \exists \gamma > 0 \text{ such that } Tf^nv \in C_\gamma(E^s(f^n(x))) \ \forall n \geq 0\} \text{ and} \\ E^u(x) &= \{v \in T_xM \ \big| \ |Tf^{-n}v| \to 0 \text{ as } n \to \infty\} \\ &= \{v \in T_xM \ \big| \ \exists r > 0 \text{ such that } |Tf^{-n}v| \leq r \ \forall n \geq 0\} \\ &= \{v \in T_xM \ \big| \ \exists \gamma > 0 \text{ such that } Tf^{-n}v \in C_\gamma(E^s(f^{-n}(x))) \ \forall n \geq 0\}. \end{aligned}$$

In particular, the splitting is unique: if there is another splitting $T_xM = G^s(x) \oplus G^u(x)$, $x \in \Lambda$, then $G^s(x) = E^s(x)$ and $G^u(x) = E^u(x)$.

A proof can be found in [4]. That the hyperbolic splitting of $T_\Lambda M$ is unique allows $E^s(x)$ and $E^u(x)$ to be referred to without qualification, as is done in the next theorem.

Theorem 3.9. *Let $\Lambda \subset M$ be a hyperbolic set of f. Then, in terms of the Grassmann topology described earlier, $E^s(x)$ and $E^u(x)$ vary continuously with $x \in \Lambda$. In particular, the dimensions of $E^s(x)$ and $E^u(x)$ are locally constant and the closure $\overline{\Lambda}$ is a hyperbolic set of f.*

A proof of the theorem can be found in [4]. Since M is compact and the closures of hyperbolic sets are hyperbolic, for the rest of this paper hyperbolic sets will be assumed compact. (The theorems of this paper do not rely upon the boundary in any way; restricting to the interior will not invalidate them.)

A related notion is that of C^r subbundles. If there is a linear subspace $E(x) \subset T_xM$ for each x in some $\Lambda \subset M$, the union

$$E = \bigsqcup_{x \in \Lambda} E(x)$$

is an *m-dimensional C^r subbundle of $T_\Lambda M$* if for every $x \in \Lambda$ there is a neighborhood U of x in Λ and m linearly independent C^r vector fields $e_1, \dots, e_m : M \to TM$ such that the vectors $e_1(y), \dots, e_m(y)$ span $E(y)$ for every $y \in U$. In this case, $E(x)$ is called the *fiber* of E at x. It can be shown that E is a C^0 subbundle of $T_\Lambda M$ if and only if the $E(x)$ vary continuously with x in the Grassmanian topology. Thus, if Λ is hyperbolic, Theorem 3.9 guarantees E^s and E^u are C^0 subbundles of $T_\Lambda M$.

Two C^0 subbundles E_1 and E_2 of $T_\Lambda M$ are said to form a direct sum, denoted $E_1 \oplus E_2$, if $E_1(x)$ and $E_2(x)$ form a direct sum at every $x \in \Lambda$. Thus, if Λ is hyperbolic, $E^s \oplus E^u$ is a direct sum and the notation in the Remark after 3.1 is justified.

Another piece of notation is needed for the next theorem, which describes the persistence of hyperbolicity for an invariant set. Given a set $U \subset M$, let $d(x, U)$ denote the infimum of $d(x, y)$ taken over all $y \in U$ and let $B(U, a)$ denote the set $\{x \in M \mid d(x, U) < a\}$.

Theorem 3.10 (Persistence of hyperbolicity for an invariant set)**.** *Let $\Lambda \subset M$ be a hyperbolic set of f. There is a neighborhood $\mathcal{U}$ of f in $\mathrm{Diff}^1(M)$ and a number $a > 0$ such that for any $g \in \mathcal{U}$, every compact g-invariant set Δ which is contained in $B(\Lambda, a)$ is hyperbolic. Further, as g approaches f in $\mathrm{Diff}^1(M)$ and $x \in \Delta$ approaches $y \in \Delta$, $E^s(x, g)$ approaches $E^s(x, f)$ and $E^u(x, g)$ approaches $E^u(x, f)$.*

This is similar to structural stability, though the existence of a nonempty g-invariant set close to Λ is not guaranteed here.

The proof of Theorem 3.10 relies on a lemma whose statement is rather verbose. It deals with a 'fiber-preserving' map; recall that, if $\Lambda \subset M$ is f-invariant then a map F between two C^0 subbundles of $T_\Lambda M$ is called *fiber-preserving over f* if

$$\pi F = f \pi$$

where $\pi : TM \to M$ is the bundle projection. If also F is continuous and $F\big|_{E(x)}$ is linear for every $x \in \Lambda$ then F is a C^0 *bundle homomorphism.* If in addition $F\big|_{E(x)}$ is an isomorphism for every $x \in \Lambda$ then F is a C^0 *bundle isomorphism.*

Lemma 3.11. *Let $g : M \to M$ be a diffeomorphism, let $\Delta \subset M$ be a g-invariant set, and let $B : T_\Delta M \to T_\Delta M$ be a C^0 bundle isomorphism which is fiber-preserving over g. Let $T_\Delta M = E_1 \oplus E_2$ be a direct sum and let B_{ij} denote the projection onto E_i of $B\big|_{E_j}$, so that B can be written*

$$\begin{pmatrix} B_{11} & B_{22} \\ B_{21} & B_{22}. \end{pmatrix}$$

If there are constants $\lambda > 0$ and $\varepsilon > 0$ such that

$$\max\{|B_{11}^{-1}|, |B_{22}|\} < \lambda,$$
$$\max\{|B_{12}|, |B_{21}|\} < \varepsilon, \text{ and}$$
$$\lambda + \varepsilon < 1,$$

then there is a unique C^0 bundle homomorphism $P_B : E_1 \to E_2$ over Id *such that $|P_B| \leq 1$, the graph* $\mathrm{gr}(P_B)$ *is a B-invariant C^0 subbundle, and $B_x\big|_{\mathrm{gr}(P_x)}$ is expanding. Further, P_B, and hence* $\mathrm{gr}(P_B)$*, depends continuously on B.*

Remark 3.12. Here, that $\mathrm{gr}(P_B)$ is B-invariant means $B_x(\mathrm{gr}(P_x)) = \mathrm{gr}(P_{g(x)})$, where B_x denotes $B\big|_{T_xM}$ and P_x denotes $P_B\big|_{T_xM}$.

Details for the proof of this lemma and the preceding theorem can be found in [4]. Since the details of the proof are not enlightening but the strategy is important and general in the study of hyperbolic dynamics, only an outline is given in this paper. The main technique in the lemma, which also is used in the proof of the Shadowing Theorem, is to discover a way to apply the Contraction Mapping Principle to produce the desired function.

Theorem 3.13 (Contraction Mapping Principle)**.** *Let X be a complete metric space with metric d and let $\varphi : X \to X$ be a function such that there exists $c < 1$ for which $d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y)$ for all $x, y \in X$. Then there is a unique $x \in X$ such that $\varphi(x) = x$.*

A proof of this theorem can be found, for instance, in [7]. It is possible to apply the Contracion Mapping Principle in this lemma once the B-invariance of P_B is expressed in an equation of the form

$$P_{g(x)} = [\text{expression dependent on } P_B]$$

which holds for all $x \in \Delta$. This allows the construction of a map T acting on the space of C^0 bundle homomorphisms from E_1 to E_2 such that a fixed point of T is a B-invariant bundle homomorphim. The rest of the proof of the lemma consists in verifying that the Contraction Mapping Principle applies to T, and then doing some size checking to establish that $B_x\big|_{\mathrm{gr}(P_x)}$ is expanding.

The lemma establishes the theorem in the following way. The continuity of $E^s(x)$ and $E^u(x)$ (and some properties of hyperbolic sets) allows the lemma to be applied with (1) g and Δ in the lemma taken to be the same as g and Δ in the theorem, (2) E_1 and E_2 in the lemma taken to be local continuations G^s and G^u of E^s and E^u, respectively, and (3) B in the lemma taken to be either Dg or Dg^{-1}. Further, restricting a to be sufficiently small, the lemma can be applied with the same constants λ and ε whether E_1 is taken to be G^s, E_2 to be G^u, and B to be Dg or E_1 is taken to be G^u, E_2 to be G^s, and B to be Dg^{-1}. If a space is Dg^{-1}-invariant than it is also Dg-invariant and if Dg^{-1} is expanding on that space then Dg is contracting on it. Thus the lemma gives two Dg-invariant spaces Q and W where Dg is expanding on Q and contracting on W. Checking the dimensions of Q and W shows that $Q \oplus W = T_\Delta M$; therefore Δ is hyperbolic.

The technique of using the linear structure of the derivative to apply the Contraction Mapping Principle is crucial to the Shadowing Theorem and to the structural stability of hyperbolic sets. In Euclidean space, the linear structure is often utilized by taking the difference $f - Df(0)$. On a manifold this subtraction does not make sense in general, but using the exponential map to 'lift' functions locally to the tangent bundle allows an analogous subtraction.

Recall that, for $x \in M$, the *exponential map at x*

$$\exp_x : T_x M \to M$$

is defined to be

$$\exp_x(v) = \sigma_v(t)$$

where $\sigma_v(t)$ is the geodesic, determined by the Riemannian metric of M, which goes through x at $t = 0$ with velocity v.

To make discussing the properties of the exponential map easier, let the balls of radius ρ on $T_x M$ and on TM be denoted $T_x M(\rho)$ and $TM(\rho)$, respectively. For any $x \in M$, let 0_x denote the origin of $T_x M$. The next theorem reviews some properties of the exponential map; it can be found in [6].

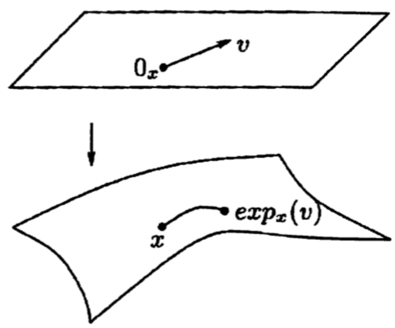

FIGURE 4. The exponential map. The image is from [4].

Theorem 3.14 (Properties of exp)**.**

(1) $\exp_x(0_x) = x$.
(2) $D(\exp_x)(0_x) : T_x M \to T_x M = \mathrm{Id}$.
(3) $\exists \rho > 0$ *such that, for any* $x \in M$, $\exp_x : T_x M(\rho) \to M$ *is a* C^∞ *embedding.*
(4) $d(x, exp_x(v)) = |v|$ *for all* $v \in T_x M(\rho)$*, where* ρ *is as in (3) and* d *and* $|\cdot|$ *are both induced by the given Riemannian metric of* M*.*
(5) $\exp : TM \to M$*, given by* $\exp(v) = \exp_x(v)$ *when* $v \in T_x M$*, is* C^∞*.*

A noteworthy consequence of (3) and (4) is that, taking x as the base, any point $y \in B(x, \rho)$ determines a vector $\exp_x^{-1} y \in T_x M$ with $|\exp_x^{-1} y| = d(x, y)$. In a Euclidean space, $\exp_x^{-1} y$ is just the vector $y - x$.

The exponential map can be used to define a 'self-lifting' of a function f to a map on the tangent bundle which is fiber-preserving over f, though it is not linear on fibers. Precisely, given a C^1 diffeomorphism $f : M \to M$, take some $0 < r < \rho$ such that, for any $x, y \in M$, $d(x, y) < r$ implies $d(f(x), f(y)) < \rho$. Define the *self-lifting of* f to be the function $F_f : TM((r) \to TM$ given by

$$F_f(v) = \exp_{f(x)}^{-1} f \exp_x(v),$$

where $v \in T_xM$. Then F_f is fiber-preserving over f and since f is C^1, so is F_f. A similar but more intricate lifting will be used in the proof of the Shadowing Theorem.

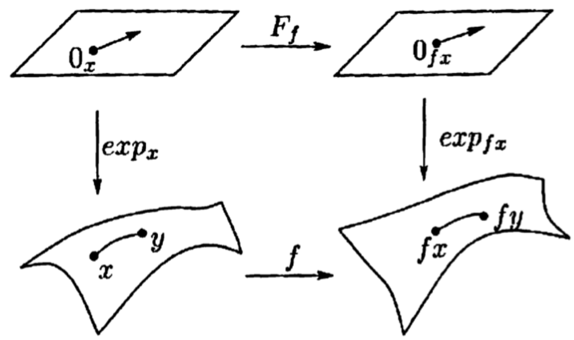

FIGURE 5. The self-lifting map. The image is from [4].

4. SHADOWING

Now, shadowing will be defined. The most immediate relation between shadowing and hyperbolicity happens through pseudo-orbits, and for this reason the definition of pseudo-orbits is restated here.

Definition 4.1 (Shadowing)**.** Let (X, f) be a dynamical system. Let $a \in \mathbb{Z} \cup \{-\infty\}$ and let $b \in \mathbb{Z} \cup \{\infty\}$ with $a < b$. A sequence $\{x_n\}_{a<n<b} \subset X$ is said to be *δ-shadowed* by the orbit $\mathcal{O}(x)$ of $x \in X$ if $d(x_n, f^n(x)) < \delta$ for all $a < n < b$. In this case we say also that $\mathcal{O}(x)$ *δ-shadows* the sequence.

Definition 4.2 (Pseudo-orbit)**.** Let (X, f) be a dynamical system. Let $a \in \mathbb{Z} \cup \{-\infty\}$ and let $b \in \mathbb{Z} \cup \{\infty\}$ with $a < b$. A sequence $\{x_n\}_{a<n\leq b} \subset X$ is called an *ε-pseudo-orbit* if $d(x_{n+1}, f(x_n)) < \varepsilon$ for all $a < n < b$. If a pseudo-orbit P is a subset of A, it is said that P is a *pseudo-orbit in* A.

The "immediate relation" mentioned above which links shadowing and hyperbolic sets is the following: pseudo-orbits near hyperbolic sets are always shadowed by an actual orbit.

Theorem 4.3. (Shadowing Lemma) *Let $\Lambda \subset M$ be a hyperbolic set for the dynamical system (M, f). There exists a neighborhood $U(\Lambda)$ of Λ such that for any $\delta > 0$ there exists an $\varepsilon > 0$ such that every ε-pseudo-orbit in $U(\Lambda)$ is δ-shadowed by an orbit of f. Moreover, the orbit is unique if δ is sufficiently small.*

This theorem follows from the more general Shadowing Theorem proved below. For a direct proof, see [4].

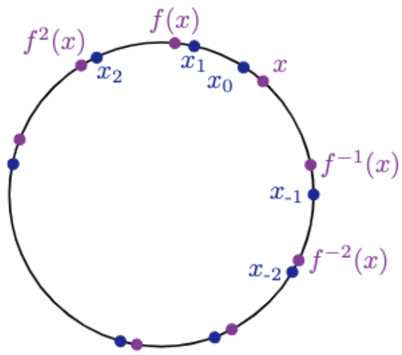

FIGURE 6. Pseudo-orbit (blue) shadowed by an orbit (purple).

The Shadowing Lemma gives a criterion under which the difference between pseudo-orbits and orbits is small. This situation can be viewed from a different angle. Given some hyperbolic set Λ and some $\delta > 0$, take $\varepsilon > 0$ as given by the Shadowing Lemma. Take any ε-pseudo-orbit P in Λ. Label each element of P as x_n in the same way as was done in Definition 4.2. The Shadowing Lemma ensures the existence of a point $y \in M$ such that $\mathcal{O}(y)$ δ-shadows P. Label $f^n(y)$ as y_n for all integers n. These labelings induce a function α from the integers to P and a function β from the integers to $\mathcal{O}(y)$. That $\mathcal{O}(y)$ δ-shadows P means that $d(\alpha(n), \beta(n)) < \delta$ for all integers n.

In effect, adopting this view makes two changes to the Shadowing Lemma: the role of the integers is emphasized and the result of the theorem has become one of functions rather than of sequences. It is stated with these changes.

Theorem 4.4. (Shadowing Lemma rephrased) *Let $\Lambda \subset M$ be a hyperbolic set for the dynamical system (M, f). There exists a neighborhood $U(\Lambda)$ of Λ such that for any $\delta > 0$ there exists an $\varepsilon > 0$ with the following property:*

If

(1) *$P = \{x_n\}_{n \in Z}$ is an ε-pseudo-orbit,*
(2) *$f' : U(\Lambda) \to M$ has $f'(x_n) = x_{n+1}$,*
(3) *$g : \mathbb{Z} \to \mathbb{Z}$ is given by $g(n) = n + 1$, and*
(4) *$\alpha : \mathbb{Z} \to P$ is given by $\alpha(n) = x_n$*

then there exists a point $x \in M$ and a function $\beta : \mathbb{Z} \to U(\Lambda)$ given by $\beta(n) = f^n(x)$ such that

(1) *$\beta g = f\beta$,*
(2) *$d(\alpha(n), \beta(n)) < \delta$ for all n.*

Moreover, the point x and the function β are unique if δ is sufficiently small.

Remark 4.5. A few statements of clarification should be given.

(1) For word economy, the theorem has been stated for a pseudo-orbit indexed by the integers. To account for pseudo-orbits indexed by a proper subset of the integers, some additional specifications for the domains of α and β must be made. Once they are, the theorem holds in as general a situation as in the original formulation.

(2) The codomain $U(\Lambda)$ of β was not mentioned in the exposition above, but it plays a prominent role in both phrasings of the theorem. The idea is that in order for the hyperbolicity of Λ to have any effect on a pseudo-orbit, the pseudo-orbit must be close to Λ.
(3) Two functions, g and f', have been introduced in the new statement. Their roles are to make explicit the transition from one element in the pesudo-orbit to the next element, which in the exposition above was implicit in the notions of 'iterating' over f and over the integers. If f' is a diffeomorphism, then iterating over the pseudo-orbit (of f) is the same as iterating over an actual orbit of f'. This perspective allows the lemma to be generalized to apply to perturbations of f instead of pesudo-orbits, as is done in the Shadowing Theorem below.
(4) If the integers are given the discrete topology then g becomes a homeomorphism, α and β become continuous functions, and $d_{C^0}(\alpha, \beta) < \delta$. When the result is generalized below in the Shadowing Theorem, the domain of g becomes an arbitrary topological space.

While this phrasing of the Shadowing Lemma is more cumbersome than the first, it is more easily generalized. Instead of referencing elements of sets by integer indexes (as is done with pseudo-orbits), it is often useful to reference elements of sets by points in a topological space (for instance, elements of subbundles and sections of a manifold are referenced by points of the manifold). A pseudo-orbit of f is just an orbit of some perturbation of f (recall the discussion of pseudo-orbits in §2). Instead of defining a specific perturbation to iterate over a particular pseudo-orbit, perturbations themselves may be studied. These considerations lead to a generalized version of the Shadowing Lemma, called the Shadowing Theorem.

Theorem 4.6. (Shadowing Theorem) *Let $f : M \to M$ be a diffeomorphism and let $\Lambda \subset M$ be a hyperbolic set for (M, f). There exists a neighborhood $U(\Lambda)$ of Λ and $\varepsilon_0, \delta_0 > 0$ such that for any $\delta > 0$ there is an ε with the following property:*
If

(1) $f' : U(\Lambda) \to M$ is a C^2 diffeomorphism with $d_{C^1}(f', f) < \varepsilon_0$,
(2) Y is a topological space and $g : Y \to Y$ is a homeomorphism, and
(3) $\alpha : Y \to U(\Lambda)$ is a continuous function such that $d_{C^0}(\alpha g, f'\alpha) < \varepsilon$,

then there exists a continuous function $\beta : Y \to U(\Lambda)$ such that

(1) $\beta g = f'\beta$, and
(2) $d_{C^0}(\alpha, \beta) < \delta$.

Moreover, the function β is unique when δ is sufficiently small: if $\bar{\beta} g = f'\bar{\beta}$ and $d_{C^0}(\alpha, \bar{\beta}), d_{C^0}(\alpha, \beta) < \delta_0$ then $\bar{\beta} = \beta$.

Remark 4.7.

(1) In this theorem, f' is a perturbation of f, Y is the topological space used to reference subsets of M, and α and β are the means of reference.
(2) In the Shadowing Lemma, the pseudo-orbit iterating function f' may without loss of generality be chosen to be a C^2 diffeomorphism. Taking ε in the Shadowing Lemma small enough ensures $d_{C^1}(f', f) < \varepsilon_0$, where ε_0 is as in the Shadowing Theorem. Then, taking Y to be the integers endowed with the discrete topology, taking g to be iteration over the integers, and taking α to be iteration over some ε-pseudo-orbit turns the Shadowing Theorem

into the Shadowing Lemma. In other words, the Shadowing Lemma follows directly from the Shadowing Theorem.

The strategy for the proof of the Shadowing Theorem, like that of Theorem 3.10, is to transform the theorem into a statement about fixed points and then to apply the Contraction Mapping Principle. The fixed-point-transformation follows from the fact that g is a homeomorphism and therefore invertible: the desired β is a fixed point of $F \in C^0(Y, U(\Lambda))$ given by

$$F(\zeta) = f \circ \beta \circ g^{-1},$$

where $C^0(Y, U(\Lambda))$ is the space of continuous functions from Y to $U(\Lambda)$. In order to construct a contraction mapping, it is helpful to access the structure that hyperbolicity imposes on the linear part of f. Composing F with the exponential map in the following way helps utilize this linear structure.

Let B be a ball around α in $C^0(Y, U(\Lambda))$ with a radius θ small enough that on B, the function $\mathcal{A}$ given by $\mathcal{A}\beta(y) = \exp^{-1}_{\alpha(y)} \beta(y)$ is well-defined. By Theorem 3.14, θ is independent of Y, g, and α.

Then $\mathcal{A}$ can be used to 'lift' F to act on maps from Y to TM instead of on maps from Y to M. Let $C^0_\alpha(Y, TM)$ denote

$$\{v \in C^0(Y, TM) \mid v(y) \in T_{\alpha(y)}M \quad (y \in Y)\},$$

the vectorfields along α, and let $B^\alpha_r(0)$ denote the ball of radius r centered at 0 in $C^0_\alpha(Y, U(\Lambda))$. Let

$$F^\alpha : B^\alpha_\theta \to C^0_\alpha(Y, TM)$$

be given by

$$F^\alpha = \mathcal{A}F\mathcal{A}^{-1}.$$

Then, if v a fixed point of F^α, $\mathcal{A}^{-1}v$ is a fixed point of f. Thus the proof of the Shadowing Theorem consists in showing that F^α is a contraction mapping and that the fixed point β of F has the properties desired. For this, a sort of bound on DF^α is needed.

Lemma 4.8. *Let $f : M \to M$ be a diffeomorphism and let $\Lambda \subset M$ be a hyperbolic set for (M, f). Let Y be a topological space and let $g : Y \to Y$ be a homeomorphism*

There exists a neighborhood $U(\Lambda)$ of Λ and constants $\varepsilon_0, \varepsilon > 0$ such that for any continuous function $\alpha : Y \to U(\Lambda)$ there exists a constant $R > 0$ independent of Y, g, and α such that $\|((DF^\alpha)_0 - \mathrm{Id})^{-1}\| < R$ whenever f' is a C^2 diffeomorphism, $d_{C^1}(f, f') < \varepsilon_0$, $d_{C^0}(\alpha g, f'\alpha) < \varepsilon$, $||\cdot||$ is the norm of convergence, and F^α is defined as it is above.

The idea is roughly this: Since f is hyperbolic, Df can be bounded with respect to E^s and E^u. In particular, the bounding is not dependent on specific points in Λ. By the continuity of E^s and E^u and the persistence of hyperbolicity, a similar bounding can be found for Df' near E^s and E^u. Finally, since $(D\exp^{-1}_x)\big|_x = \mathrm{Id}$, restricting $d_{C^0}(\alpha, f'\alpha g^{-1})$ to be small makes (DF^α) close to Df' and thus $(DF^\alpha)_0$ may be bounded as described in the lemma. Many of these steps are more easily seen when referencing the explicit formula for DF^α; the formula is written below. For a full proof, see [1].

Much of what is discussed above will be restated in the proof of the Shadowing Theorem, which is given now.

Proof of the Shadowing Theorem. First, the groundwork needed to utilize Lemma 4.8 is set.

Let $C^0_\alpha(Y, TM)$ denote $\{v \in C^0(Y, TM) \mid v(y) \in T_{\alpha(y)}M \quad (y \in Y)\}$, the vector-fields along α.

Let $\|\cdot\|$ denote the norm of uniform convergence on $C^0_\alpha(Y, U(\Lambda))$.

Let $B_r(\alpha)$ denote the ball of radius r centered at α in $C^0(Y, U(\Lambda))$. Let $B^\alpha_r(0)$ denote the ball of radius r centered at 0 in $C^0_\alpha(Y, U(\Lambda))$.

Let $\theta > 0$ be a constant independent of Y, g, and α and small enough that the map $\mathcal{A} : B_\theta(\alpha) \to C^0_\alpha(Y, TM)$ given by $\mathcal{A}\beta(y) = \exp^{-1}_{\alpha(y)} \beta(y)$ is well-defined. Notice that $\mathcal{A}$ is a homeomorphism onto $B^\alpha_\theta(0)$.

Let the map $F \in C^0(Y, U(\Lambda))$ be given by $F(\zeta) = f' \circ \zeta \circ g^{-1}$.

Let the map $F^\alpha : B^\alpha_\theta \to C^0_\alpha(Y, TM)$ be given by $\mathcal{A}F\mathcal{A}^{-1}$. More explicitly,

$$F^\alpha(v)(y) = \exp^{-1}_{\alpha(y)}(f'(\exp_{\alpha(g^{-1})} v(g^{-1}(y)))).$$

The groundwork has been set and the theorem can now be proved. The desired map β is a fixed point of F. If v is a fixed point of F^α then $\mathcal{A}^{-1}v$ is a fixed point of F, so it suffices to find a fixed point of F^α. For this it is important that F^α is smooth in v and its derivative

$$\begin{aligned}((DF^\alpha)_v\xi)(y) = {} & (D\exp^{-1}_{\alpha(y)})\big|_{f'\exp_{\alpha(g^{-1}(y))} v(g^{-1}(y))} \\ & \cdot (Df')\big|_{\exp_{\alpha(g^{-1}(y))} v(g^{-1}(y))} \cdot (D\exp_{\alpha(g^{-1}(y))})\big|_{v(g^{-1}(y))} \cdot \xi(g^{-1}(y))\end{aligned}$$

is Lipschitz in v. Note that f' being C^2 is essential for this derivative being Lipschitz.

Writing $F^\alpha(v) = (DF^\alpha)_0 v + H(v)$ gives

$$\begin{aligned} v = F^\alpha(v) &= (DF^\alpha)_0 v + H(v) \\ \iff -((DF^\alpha)_0 - \mathrm{Id})v &= H(v) \\ \iff v = -((DF^\alpha)_0 - \mathrm{Id})^{-1}H(v) &:= T(v)\end{aligned}$$

That is, v is a fixed point of F^α if and only if v is a fixed point of T. The Contraction Mapping Principle will be applied to T.

By the linearity of the derivative, that DF^α is Lipschitz in v means that DH is Lipschitz in v as well, say with constant K. Since DH is Lipschitz, it is bounded on B^α_θ. Then, by the generalized mean value theorem, on any ball around 0 in B^α_θ there exists w such that

$$\begin{aligned}||H(v_1) - H(v_2)|| \le ||DH(w)||\; ||v_1 - v_2|| &\le K||w - 0||||v_1 - v_2|| \\ &\le K||w||\; ||v_1 - v_2||\end{aligned}$$

for any v_1, v_2 in the ball. Combining all this, taking R as in Lemma 4.8 shows that for any $v_1, v_2 \in B^\alpha_\theta$

$$\|T(v_1) - T(v_2)\| < RK \max(\|v_1\|, \|v_2\|)\|v_1 - v_2\|.$$

Thus T is a contraction near 0.

The constants of the theorem are now specified. Take $\delta_0 = \frac{1}{2RK}$. Take θ such that, in addition to making $\mathcal{A}$ well-defined, $\theta \le \min(\delta, \delta_0)$. (Remember that the theorem provides a result which holds "for any $\delta > 0\ldots$".) Take ε and ε_0 so that the Lemma 4.8 holds and so that $\varepsilon < \frac{\theta}{2R}$.

With these constants, restricting T to $B_\theta^\alpha(0)$ gives

$$\|T(v_1) - T(v_2)\| < RK\max(\|v_1\|, \|v_2\|)\|v_1 - v_2\| \leq \frac{1}{2}\|v_1 - v_2\|$$

for any $v_1, v_2 \in B_\theta^\alpha(0)$. Additionally, that

$$H(0)(y) = F^\alpha(0)(y) = \exp_{\alpha(y)}^{-1} f'(\alpha(g^{-1}(y)) = d_{C^0}(f'\alpha g^{-1}(y))$$

means

$$\|T(0)\| < R||H(0)|| < \frac{\theta}{2}$$

whenever α is such that $d_{C^0}(\alpha g, f'\alpha) < \varepsilon$. Thus, for all $v \in B_\theta^\alpha(0)$,

$$\|T(v) - T(0)\| \leq \|T(v)\| + \|T(0)\| < \frac{1}{2}\|v\| + \frac{1}{2}\theta \leq \theta$$

and so $T(B_\theta^\alpha(0)) \subset B_\theta^\alpha(0)$. Thus T is a contraction map.

By the Contraction Mapping Principle, T has a unique fixed point $v \in B_\theta^\alpha(0) \subset B_{\delta_0}^\alpha(0)$. As discussed above, v is also a fixed point of F^α. Hence there is a fixed point $\beta = \mathcal{A}^{-1}v$ of F, i.e. $\beta g = f'\beta$. Since $\mathcal{A}$ is a homeomorphism, $\beta \in B_\theta(\alpha) \subset B_\delta(\alpha)$. Thus β is as required and is unique in $B_{\delta_0}(\alpha)$. ☐

The Shadowing Theorem has been presented in its full generality, though only very special cases will be needed to prove that hyperbolic sets are structurally stable. In these cases, Y will be taken to be a subset of M, g will be either f or a small perturbation of f, and α will be the identity on Y. These simplifications allow the Shadowing Theorem to hold for perturbations f' which are C^1 diffeomorphisms close to f, not just for C^2 diffeomorphisms. The reason is roughly that the function F^α in the proof of the Shadowing Theorem is simplified, and so DF^α is Lipschitz in v even when f' is merely C^1. See [4] for details.

5. Structural Stability of Hyperbolic Sets

Theorem 5.1. (Structural Stability of Hyperbolic Sets) *Let $f : M \to M$ be a diffeomorphism, let $U \subset M$ be an open set, and let $\Lambda \subset U$ be a hyperbolic set for (M, f). Then for any open neighborhood $V \subset U$ of Λ and every $\rho > 0$ there exists an $\eta > 0$ such that if $f' : U \to M$ is a C^1 diffeomorphism with $d_{C^1}(f\big|_V, f') < \eta$ then*

(1) there exists a (nonempty) set $\Lambda' \subset M$ which is hyperbolic for f';
(2) there exists a homeomorphism $h : \Lambda' \to \Lambda$ such that $h \circ f\big|_{\Lambda'} = f\big|_\Lambda \circ h$ and $d_{C^0}(\mathrm{Id}, h) + d_{C^0}(\mathrm{Id}, h^{-1}) < \rho$.

Moreover, h is unique when ρ is sufficiently small.

Remark 5.2.

(1) The structural stability here is strong C^1 structural stability (Definition 2.8).
(2) That Λ' is nonempty is becuase it is constructed as the range of a homeomorphism on Λ, which is nonempty.
(3) If $\Lambda = M$, then f is an Anosov diffeomorphism. Thus, as promised in Example 3.6, this theorem shows that Anosov diffeomorphism are structurally stable.
(4) The constants ρ and η have been used instead of the traditional δ and ε to avoid confusion when the Shadowing Theorem is referenced.

Proof. Take any C^1 diffeomorphism f' with $d_{C^1}(f, f') < \eta$ with η to be determined later. The Shadowing Theorem is used three times. First it finds β such that $\beta \circ f = f' \circ \beta$, then it finds h such that $h \circ f' = f \circ h$, and finally it shows that h is a homeomorphism. When $Y, g, \alpha, \varepsilon_0, \delta_0$, and ε are referenced, they are as introduced in the Shadowing Theorem.

For the first application, take $Y = \Lambda$, $g = f$, and $\alpha = \mathrm{Id}\big|_\Lambda$. The Shadowing Theorem ensures the existence of a neighborhood $U(\Lambda)$ of Λ, constants $\delta_0, \varepsilon_0 > 0$, and a constant $\varepsilon > 0$ dependent on some $\delta > 0$ (to be determined later) such that there exists a continuous function $\beta : \Lambda \to U(\Lambda)$ with $\beta \circ f = f' \circ \beta$ and $d_{C^0}(\mathrm{Id}\big|_\Lambda, \beta) < \delta$ whenever $d_{C^1}(f, f') < \varepsilon_0$ and $d_{C^0}(f, f') < \varepsilon$.

Notice that already η must be at least as small as $\min(\varepsilon_0, \varepsilon)$, a condition dependent on the as yet undetermined δ mentioned above. A restriction on δ and further restrictions on η are introduced now.

Let Λ' denote $\beta(\Lambda)$. The goal is to apply the Shadowing Theorem to f' with $Y = \Lambda'$, $\alpha = \mathrm{Id}\big|_{\Lambda'}$, and $g = f'$ to obtain the function h promised in the current theorem. In order to do this, Λ' must be hyperbolic for f' and f' must be close enough to f that the Shadowing Theorem can be applied to f'.

Theorem 3.10 is used to gurantee Λ' is hyperbolic. Since β is continuous and Λ is compact, $\Lambda' = \beta(\Lambda)$ is compact. That Λ is invariant under f means Λ' is invariant under f', since

$$f'(\Lambda') = f' \circ \beta(\Lambda) = \beta \circ f(\Lambda) = \beta(\Lambda) = \Lambda'.$$

Restricting δ to be sufficiently small guarantees Λ' is contained in the set $B(\Lambda, a)$ from Theorem 3.10. It will be important later that δ is also chosen to be less than $\rho/2$. Taking η sufficiently small satisfies the final requirement of Theorem 3.10 and thus ensures Λ' is hyperbolic for f'.

Now the reason the roles of f and f' in the Shadowing Theorem can be switched will be explained. If $d_{C^1}(f, f')$ is sufficiently small then Lemma 4.8 holds after switching the roles of f and f'. Since Lemma 4.8 is what determines ε and ε_0 in the Shadowing Theorem, shrinking η again (to be less than the minimum of the required ε and ε_0) allows the Shadowing Theorem to be applied with the roles of f and f' switched.

Thus applying the Shadowing Theorem with $Y = \Lambda'$, $\alpha = \mathrm{Id}\big|_{\Lambda'}$, and $g = f'$ gives a continuous function h on Λ' such that $h \circ f' = f \circ h$ and $d_{C^0}(h, \mathrm{Id}\big|_{\Lambda'}) < \rho/2$.

The last application of the Shadowing Theorem will show $h \circ \beta$ is the identity on Λ. Then β is injective, since it has a left inverse, and β is surjective onto $\beta(\Lambda) = \Lambda'$. This means β is a homeomorphism and so its inverse h is also (see, for instance, [3]). As a side note, this means that the equation in the previous paragraph really should be written $h \circ f\big|_{\Lambda'} = f\big|_\Lambda \circ h$, as it is in the statement of the theorem.

The important part of the final application of the Shadowing Theorem is the uniqueness clause. Take the '$f' = f$' case, where $Y = \Lambda$, $g = f$, and $\alpha = \mathrm{Id}$. Then there exists some δ_0 and a unique continuous function γ such that $\gamma \circ f = f \circ \gamma$ and $d_{C^0}(\mathrm{Id}, \gamma) < \delta_0$. Clearly Id, the identity, works, and so (by uniqueness) $\gamma = \mathrm{Id}$. On the other hand, taking $\rho < \delta_0/2$ gives $d_{C^0}(\mathrm{Id}, h \circ \beta) < \delta_0$ and

$$h \circ \beta \circ f = h \circ f' \circ \beta = f \circ h \circ \beta,$$

so that $h \circ \beta = \gamma = \mathrm{Id}$, as desired.

Finally, $d_{C^0}(\mathrm{Id}, h) + d_{C^0}(\mathrm{Id}, h^{-1}) = d_{C^0}(\mathrm{Id}, h) + d_{C^0}(\mathrm{Id}, \beta) < \rho/2 + \rho/2 = \rho$. $\square$

Acknowledgments

I would like to thank my REU mentor Meg Doucette for introducing me to the field of dynamical systems and for guiding me throughout the program. I would also like to extend my gratitude to my brothers George and Niko Matheos for editing this paper and to Samuel J. Reynolds for discussing dynamical systems with me. Finally, I would like to thank Professor May for organizing and running the REU, and for providing the opportunity for me to participate in the program.

References

[1] Anatole Katok and Boris Hasselblatt. Introduction to The Modern Theory of Dynamical Systems. Cambridge University Press. 1995. Reprinted 1999.
[2] John Lee. Introduction to Smooth Manifolds. 2^{nd} ed. Springer Science+Business Media. 2013.
[3] James Munkres. Topology. 2^{nd} ed. Pearson Education. 2000. Reissued 2018.
[4] Lan Wen. Differentiable Dynamical Systems: An Introduction to Structural Stability and Hyperbolicity. The American Mathematical Society. 2016.
[5] Wikipedia. https://en.wikipedia.org/wiki/ArnoldWikimedia Foundation, Inc. March 16, 2023.
[6] William Boothby. An Introduction to Differentiable Manifolds and Riemannian Geometry. Academic Press. 1957.
[7] Walter Rudin. Principles of Mathematical Analysis. 3^{rd} ed. McGraw-Hill, Inc. 1976.