INTRODUCTION TO HYPERBOLIC SETS

by

Oscar E. LANFORD III

Text of a series of lectures delivered at the International School of Mathematical Physics on Regular and Chaotic Motions in Dynamical Systems, (a NATO Advanced Study Institute) held at the E. Majorana International Center for Scientific Culture, Erice, Sicily, in July 1983.

Institut des Hautes Etudes Scientifiques
35, Route de Chartres
91440 Bures-sur-Yvette (France)

October 1984

IHES/M/84/48

Introduction to Hyperbolic Sets

*Oscar E. Lanford III**

IHES
91440 Bures-sur-Yvette
France

1. Preliminaries

One of the most illuminating general observations about dynamical systems is that it often happens that orbits starting very close together diverge exponentially. Exponential separation of orbits gives rise, notably, to the sensitive dependence on initial condition which accounts for the apparently stochastic behavior of deterministic dynamical systems. In these lectures, I will discuss systems which have a technically very strong version of the property of exponential separation of orbits, ones in which there are no neutral separations between nearby orbits so that each pair separates exponentially either forward or backward in time (and most separate both forward and backward) with strong uniformity assumptions on the rate of separation. Separation of orbits will not, however, be required everywhere in the state space of the system, but only in the neighborhood of some compact invariant set. Thus, the objects we will analyze are, roughly, compact invariant sets with the property that pairs of orbits starting out very near to each other and remaining near the set in question diverge exponentially in a uniform way. Such sets are called *hyperbolic sets*. (The above is intended only as a very general indication of what a hyperbolic set is and will be misleading if taken too literally; the formal definition is given in Section 3.)

The study of hyperbolic sets has led to a rich and deep mathematical theory which grew out of work of Anosov and Smale in the 1960's, with important contributions from many others. From the point of view of applications this investigation has not been as successful as might have been hoped; non-trivial attracting sets arising in examples studied so far generally do not satisfy the strong uniformity assumptions essential for this theory. This discouraging estimate of the practical usefulness of the theory of hyperbolic sets may change as more (and higher-dimensional) examples are studied. Even if it does not, the thoroughly non-trivial mathematical ideas developed in this theory are a promising starting point for the analysis of sensitive dependence on initial conditions in more general situations.

My objective in these notes is to present some of the basic elements of the theory of hyperbolic sets with a minimum of technical complexity and mathematical prerequisites. We will consider only mappings f and have nothing to say about flows despite the fact that the theory of hyperbolic flows is not an entirely routine extension of the theory for mappings. Our mappings will be assumed to be invertible and (at least)

*Preparation of these notes was supported in part by the National Science Foundation through Grant MCS81-07086 at the University of California, Berkeley.

continuously differentiable. We call the space on which they act M, and we will write all formulas for the case where M is an open set in m-dimensional Euclidean space, in order to avoid the distractions involved in the coordinate independent theory of more general manifolds. This convention will permit us to treat things like derivatives in a concrete and elementary way. In particular, $Df(x)$ will denote the $m \times m$ Jacobian matrix of first partial derivatives of the components of f at the point x.

I have not been able to include in these notes either historical remarks or an adequate bibliography. For these matters, as well as for a more systematic development of the theory sketched here, I refer the reader to Shub's excellent monograph [2].

2. Hyperbolic fixed points. Stable and unstable manifolds.

In this section we review some facts about hyperbolic fixed points of diffentiable mappings. In view of future applications, we call the reader's attention to the fact that the theory described in this section is valid for mappings on infinite dimensional spaces as well as finite dimensional ones. Indeed, beyond some care in the use of spectral theory, there is very little need to modify the treatment to gain for this useful extra generality.

Let L be a (bounded) linear operator on a Banach space E. We say that L is *hyperbolic* if the spectrum of L is disjoint from the unit circle, or, equivalently, if there is a direct-sum decomposition

$$E = E^s \oplus E^u$$

where E^s and E^u are closed linear subspaces, invariant under L, such that the spectrum of the restriction of L to $E^s(E^u)$ is strictly *inside (outside)* the unit circle. Intuitively, E^s is the space spanned by the eigenvectors (and generalized eigenvectors, if L is not diagonalizable) whose corresponding eigenvalues have modulus smaller than one. Note, however, that in order to get these eigenvectors it is generally necessary to extend the space E to allow for multiplication by complex numbers, whereas the above direct-sum decomposition requires no such extension.

We will refer to the vectors in E^s as *stable* or *contracting* vectors for L. E^s can be characterized as the set of vectors ξ such that

$$L^n \xi \to 0 \text{ as } n \to \infty,$$

or as the set of vectors ξ such that

$$\|L^n \xi\| \text{ remains bounded as } n \to \infty.$$

Furthermore, the convergence to zero of the $L^n \xi, \xi \in E^s$ is *exponential* and *uniform:* There are constants c, λ, with $\lambda < 1$, such that

$$\|L^n \xi\| \leqslant c \lambda^n \|\xi\| \text{ for all } \xi \in E^s.$$

It need not be true in general that

$$\|L\xi\| < \|\xi\| \text{ for all non-zero } \xi \in E^s,$$

and in this sense it is not quite precise to refer to E^s as the space of contracting vectors. It is however possible to replace the original norm with an equivalent one in such a way that, for some $\lambda < 1$ (not necessarily the same as the one above)

$$\|L\xi\| \leqslant \lambda \|\xi\| \text{ for all } \xi \in E^s. \tag{2.1S}$$

Similarly, assuming that L is invertible, E^u can be characterized as the set of vectors ξ

for which $\|L^{-n}\xi\|$ converges to zero, or, alternatively, remains bounded, and we can renorm so that

$$\|L^{-1}\xi\| \leqslant \lambda \|\xi\| \text{ for all } \xi \in E^u \tag{2.1U}$$

holds simultaneously with (2.1S). Note, however, that the set of vectors ξ such that $\|L^n\xi\| \to \infty$ as $n \to \infty$ is *not* E^u; it is rather the set of all vectors with non-zero E^u component.

A fixed point x_0 for the continuously differentiable mapping f is said to be a *hyperbolic* fixed point if $Df(x_0)$ is a hyperbolic linear operator. For the remainder of this section x_0 will denote a hyperbolic fixed point for f, and $E^s(E^u)$ the contracting (expanding) subspace for $Df(x_0)$. We will assume that the norm has been arranged so that there is a constant $\lambda < 1$ such that (2.1S),(2.1U) hold (with L replaced by $Df(x_0)$).

We now formulate a theorem asserting the existence of an invariant manifold for f, passing through x_0, which can be thought of as a non-linear generalization of the linear subspace

$$\{x_0+\xi^s : \xi^s \in E^s\}$$

invariant under the linearization $Df(x_0)$ of f at x_0, and on which this linearization is contractive. We will use systematically, as coordinates for a neighborhood of x_0, pairs (ξ^s, ξ^u) with $\xi^s \in E^s$, $\xi^u \in E^u$; the corresponding point x is $x_0+\xi^s+\xi^u$.

Theorem 2.1. Stable Manifold Theorem for Hyperbolic Fixed Points. *For $\epsilon > 0$ sufficiently small:*
1. To each $\xi^s \in E^s$ with $\|\xi^s\| < \epsilon$, there corresponds exactly one $\xi^u \equiv w_s(\xi^s) \in E^u$ such that, writing x for $x_0+\xi^s+\xi^u$,

$$d(f^n(x), x_0) < \epsilon \text{ for all } n > 0.$$

2. For x as in 1., the sequence $(f^n(x))_{n \geqslant 0}$ not only remains near x_0 but converges exponentially to it. There exists $\lambda_1 < 1$ such that

$$d(f^n(x), x_0) \leqslant \lambda_1{}^n d(x, x_0)$$

3. The mapping w_s (from a neighborhood of 0 in E^s into E^u) is continuously differentiable; both w_s and Dw_s vanish at 0. If f is r times continuously differentiable, so is w_s.

We will not prove the Stable Manifold Theorem here. A proof can be found in Chapter 5 of Shub [2], and a sketch in the spirit of these notes in Lanford [1].

Some of the content of the Stable Manifold Theorem can be put into more geometrical language as follows: For $\epsilon > 0$, define

$$W^s_\epsilon = \{x : d(f^n(x), x_0) < \epsilon \text{ for } n \geqslant 0\}.$$

It is immediate from the definition that f maps W^s_ϵ into itself. The theorem then says that, for sufficiently small ϵ, W^s_ϵ is a manifold, as smooth as f, passing through x_0, whose tangent space at x_0 is exactly E^s.

The sets W^s_ϵ (ϵ small) are called *local stable manifolds* for f at x_0. We can also define the *(global) stable manifold* as:

$$W^s = \{x \in M : f^n(x) \to x_0 \text{ as } n \to \infty\}.$$

If $x \in W^s$ and $\epsilon > 0$ then $f^n(x) \in W^s_\epsilon$ for some n and conversely, i.e.,

$$W^s = \bigcup_{n=0}^{\infty} f^{-n} W^s_\epsilon.$$

Thus, W^s is made up of countably many pieces each of which is an imbedded disk with the same dimension as E^s. Nevertheless, taken as a whole, W^s can (and often does) accumulate on itself in a complicated way. The following example gives some idea of what the possibilities are:

Example: The quotient space $\mathbf{R}^2/\mathbf{Z}^2$ is one of the standard ways of representing a two-dimensional torus. We define a (particular) *linear automorphism* of the two dimensional torus by passing to the quotient modulo $\mathbf{Z}^2$ of the mapping

$$(x_1,x_2)\rightarrow(2x_1+x_2,x_1+x_2)$$

on $\mathbf{R}^2$. The image of the origin under passage to quotients is a fixed point; the derivative of the mapping at this fixed point (and everywhere else) has matrix

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

Since neither of the eigenvalues $(3\pm\sqrt{5})/2$ of this matrix is on the unit circle, the fixed point is hyperbolic. The contracting space is one dimensional. In view of the simple linear form of the mapping, it is easy to see that the global stable manifold is the image under passage to quotients modulo $\mathbf{Z}^2$ of the linear subspace of $\mathbf{R}^2$ generated by the contracting eigenvector of this matrix, i.e., the line in the plane passing through the origin with slope $-(1+\sqrt{5})/2$. Since the slope is irrational, this image wraps densely around the torus (but without ever crossing itself.)

Local and global *unstable* manifolds are defined respectively as:

$$W^u_\epsilon=\{x:d(f^{-n}(x),x_0)<\epsilon \text{ for all } n\geqslant 0\}$$

$$W^u=\{x:f^{-n}(x)\rightarrow x_0 \text{ as } n\rightarrow\infty\}.$$

i.e., as the local and global stable manifolds for f^{-1}. By this last observation, the properties of unstable manifolds can easily be read off from the properties already formulated for stable manifolds.

3. Hyperbolic sets.

We consider a differentiable mapping f with differentiable inverse, acting on a state space M which we take to be an open subset of $\mathbf{R}^m$. For any n and any $x\in M$, $Df^n(x)$ is an $m\times m$ matrix. We will say that a vector $\xi\in\mathbf{R}^m$ is a *contracting* or *stable* vector for f at x if $\|Df^n(x)\xi\|$ goes to zero exponentially as n goes to infinity, i.e., if there are constants c,λ with $\lambda<1$ such that

$$\|Df^n(x)\xi\|\leqslant c\lambda^n\|\xi\| \text{ for all } n\geqslant 0.$$

(At this point, c and λ can depend in an arbitrary way on x and ξ.) For any fixed x, the set of contracting vectors forms a linear subspace of $\mathbf{R}^m$, which we will denote by $E^s(x)$. We now claim that c and λ may be taken to depend only on x, not on $\xi\in E^s(x)$. To see this: Let $\xi_1,\ldots,\xi_j$ be a basis for $E^s(x)$ with $\|\xi_i\|=1$ for each i from 1 to j. For each i, let c_i and $\lambda_i<1$ be such that

$$\|Df^n(x)\xi_i\|\leqslant c_i\lambda_i^n \text{ for all } n\geqslant 0.$$

Also, let D be a constant such that

$$\sum_{i=1}^{j}|a_i|\leqslant D\|\sum_{i=1}^{j}a_i\xi_i\|$$

for all $a_1, \ldots, a_j$. It is then easy to see that if we put

$$\lambda = \max\{\lambda_1, \ldots, \lambda_j\}$$
$$c = D \max\{c_1, \ldots, c_j\},$$

we have

$$\|Df^n(x)\xi\| \leqslant c\lambda^n \|\xi\| \text{ for all } n \geqslant 0 \text{ and all } \xi \in E^s(x)$$

Similarly, we define the space $E^u(x)$ of *expanding* or *unstable* vectors at x to be the set of vectors ξ such that there exist constants c, λ with $\lambda < 1$ such that

$$\|Df^{-n}(x)\xi\| \leqslant c\lambda^n \|\xi\| \text{ for all } n.$$

Again, $E^u(x)$ is a linear subspace of $\mathbf{R}^m$ and c, λ can be taken to depend only on x, not on $\xi \in E^u(x)$.

It follows almost immediately from the definition of $E^s(x)$ and $E^u(x)$ (and the chain rule) that

$$Df(x)E^s(x) = E^s(f(x))$$
$$Df(x)E^u(x) = E^u(f(x)).$$

Furthermore, a λ which works for x will also work for $f(x)$ (but it may be necessary to take c larger at $f(x)$ than at x.)

We say that a point x is *hyperbolic* if

$$\mathbf{R}^m = E^s(x) \oplus E^u(x),$$

i.e., if every vector can be decomposed uniquely as the sum of a vector expanding at x and one contracting at x. From the preceding remark, if x is hyperbolic, so are $f(x)$ and $f^{-1}(x)$.

The definitions given up to now in this section are not standard and should not be taken completely seriously. The following definition, on the other hand, is standard (although expressed in a slightly unconventional way): A set $\Lambda \subset M$ is said to be a *hyperbolic set* for f if it is compact, invariant for f (i.e., $f\Lambda = \Lambda$), if each point $x \in \Lambda$ is hyperbolic, and if the constants c, λ can be taken to be independent of x, i.e., if there exist c, λ, with $\lambda < 1$, such that

$$\|Df^n(x)\xi\| \leqslant c\lambda^n \|\xi\| \text{ for all } x \in \Lambda,\ \xi \in E^s(x),\ n \geqslant 0 \qquad (3.1S)$$

$$\|Df^{-n}(x)\xi\| \leqslant c\lambda^n \|\xi\| \text{ for all } x \in \Lambda,\ \xi \in E^u(x),\ n \geqslant 0 \qquad (3.1U)$$

The essential thing in this definition is the uniformity of the rate of decay of the contracting (respectively expanding) vectors under application of $Df^n(x)$ (respectively $Df^{-n}(x)$). It would therefore be more accurate to call such sets Λ *uniformly hyperbolic* rather than simply hyperbolic. In this connection, it may be illuminating to contrast uniform hyperbolicity with a weaker probabilistic kind of hyperbolicity. One natural generalization of the notion of hyperbolic set is that of a finite f-invariant measure μ such that μ-almost all points x are hyperbolic. This condition can be reformulated in the language of *Lyapunov characteristic exponents* as saying that all Lyapunov exponents are non-zero μ-almost everywhere. In contrast to the theory of uniform hyperbolicity, which is highly developed and rich and seems to be fairly complete, the investigation of probabilistic notions of hyperbolicity is just beginning and shows great promise for further development.

Historically, the investigation of uniformly hyperbolic sets began with the case in which the whole space M is hyperbolic (if we leave aside the relatively trivial case of hyperbolic fixed and periodic points.) If M is a hyperbolic set for f, we say that f is an *Anosov diffeomorphism.*

As we have already noted, the fields of subspaces $E^s(x)$ and $E^u(x)$ have a natural covariance property:

$$Df^{\pm 1}(x)E^s(x)=E^s(f^{\pm 1}x) \tag{3.2}$$

and similarly for $E^u(x)$. Because of this covariance, the uniform expansivity condition (3.1U) is equivalent to

$$\|Df^n(x)\xi\| \geqslant c^{-1}\lambda^{-n}\|\xi\| \text{ for all } x\in\Lambda,\ \xi\in E^u(x),\ n\geqslant 0 \tag{3.3}$$

which is a more direct-looking formulation of expansivity.

It turns out to be an automatic consequence of the definition of hyperbolic set that the subspaces $E^s(x)$ and $E^u(x)$ vary in a continuous way with x on any hyperbolic set. The notion of continuity of a field of subspaces needs to be defined, but this poses no problems; any reasonable definition will work. We adopt the following one: If x is a hyperbolic point for f, we let $P^s(x)$ denote the projection onto $E^s(x)$ along $E^u(x)$, i.e., the linear operator on $\mathbf{R}^m$ which is zero on $E^u(x)$ and the identity on $E^s(x)$, and we write $P^u(x)$ for $\mathbf{1}-P^s(x)$. We say that the splitting

$$\mathbf{R}^m=E^s(x)\oplus E^u(x)$$

varies continuously with x if $P^s(x)$ is a continuous matrix-valued function of x.

Proposition 3.1. *Let Λ be a hyperbolic set for f. Then $P^s(x)$ is a continuous function of x on Λ.*

Proof. We first argue that the condition

$$\|Df^n(x)\xi\| \leqslant c\lambda^n\|\xi\| \text{ for all } n \tag{3.4S}$$

characterizes vectors ξ in $E^s(x)$. If ξ is not in $E^s(x)$, then it is the sum of a component in $E^s(x)$ and a *non-zero* component in $E^u(x)$. By (3.3), applying $Df^n(x)$ to the $E^u(x)$ component gives an exponentially growing sequence of vectors; the $E^s(x)$ component gives an exponentially decreasing sequence; so $\|Df^n(x)\xi\|$ is (eventually) exponentially growing and so does not satisfy (3.4).

Thus

$$\{(x,\xi)\colon x\in\Lambda,\ \xi\in E^s(x)\}=\bigcap_{n\geqslant 0}\{(x,\xi)\colon x\in\Lambda,\ \|Df^n(x)\xi\|\leqslant c\lambda^n\|\xi\|\}.$$

The right-hand side is an intersection of closed sets in $\Lambda\times\mathbf{R}^m$, so the left-hand side is closed. Hence:

$$\text{If } x_n\to x_0 \text{ and } \xi_n\in E^s(x_n)\to\xi_0 \text{ then } \xi_0\in E^s(x_0). \tag{3.5}$$

A similar statement holds for the $E^u(x)$'s.

The idea is now as follows: Intuitively, what we have just shown says that the fields of subspaces $E^s(x)$ and $E^u(x)$ are upper semi-continuous; their only possible discontinuities are jumps where they become discontinuously bigger. But since their direct sum is always (on Λ) equal to $\mathbf{R}^m$, if one of them were to become bigger discontinuously, the other would have to become smaller discontinuously. As this is impossible, both must be continuous.

To make this argument precise, it suffices to show
1. If (x_n) is a sequence of points of Λ converging to a limit x_0, and if the $P^s(x_n)$ converge to a limit P_0, then $P_0=P^s(x_0)$.
2. $P^s(x)$ is bounded on Λ.
Continuity then follows from a standard compactness argument.

To prove 1., we first note that P_0 is a projection since

$$(P_0)^2=\lim_{n\to\infty}(P^s(x_n))^2=\lim_{n\to\infty}P^s(x_n)=P_0,$$

From (3.5),

$$P_0\xi=\lim_{n\to\infty}P^s(x_n)\xi\in E^s(x_0) \text{ for all } \xi,$$

so the space onto which P_0 projects is contained in $E^s(x_0)$. Similarly,

$$(1-P_0)\xi\in E^u(x_0) \text{ for all } \xi,$$

so the null space of P_0 is contained in $E^u(x_0)$. Since

$$\dim(E^s(x_0))+\dim(E^u(x_0))=m,$$

both inclusions must be equalities, i.e. $P_0=P^s(x_0)$, as desired.

To prove 2: If $P^s(x)$ is not bounded, there must exist a sequence of points (x_n) in Λ and a sequence ξ_n of non-zero vectors such that

$$\frac{\|P^s(x_n)\xi_n\|}{\|\xi_n\|}\to\infty.$$

We normalize the ξ_n's by requiring that

$$\|P^s(x_n)\xi_n\|=1;$$

then

$$\xi_n\to 0.$$

By passing to a subsequence, we can assume that the x_n converge to a limit x_0 and that the $P^s(x_n)\xi_n$ converge to a limit ξ_0. By the normalization condition, $\|\xi_0\|=1$. Since

$$P^s(x_n)\xi_n\in E^s(x_n)\to\xi_0,$$

(3.5) implies that $\xi_0\in E^s(x_0)$. On the other hand,

$$P^u(x_n)\xi_n=\xi_n-P^s(x_n)\xi_n\to-\xi_0$$

(since $\xi_n\to 0$), so $\xi_0\in E^u(x_0)$. Thus, ξ_0 is a non-zero vector belonging to both $E^s(x_0)$ and $E^u(x_0)$; this contradicts

$$\mathbf{R}^m=E^s(x_0)\oplus E^u(x_0),$$

and hence proves that $P^s(x)$ must be bounded.

In its dependence on x, $P^s(x)$ is slightly more regular than just continuous; it can be shown to be Hölder continuous (provided that $Df(x)$ is). No matter how smooth f is, however, $P^s(x)$ need not be more than Hölder continuous. There exist for example analytic Anosov diffeomorphisms for which $P^s(x)$ is nowhere differentiable in x.

The presence of the constant c in the condition

$$\|Df^n(x)\xi\|\leqslant c\lambda^n\|\xi\| \text{ for all } x\in\Lambda,\ \xi\in E^s(x),\ n\geqslant 0$$

is a technical nuisance; it is easier to work with the stronger condition

$$\|Df(x)\xi\| \leqslant \lambda \|\xi\| \text{ for all } x \in \Lambda,\ \xi \in E^s(x)$$

(which, by the chain rule, is equivalent to the preceding one with $c=1$). It is necessary to use the former condition to make the definition of hyperbolic set independent of coordinate system. It is possible, however, to make the latter condition hold by adjusting the norm used, at the expense, in general, of making the norm x-dependent.

Proposition 3.2. *Let Λ be a hyperbolic set for f. Then there exist a constant $\lambda<1$ and a Riemannian metric for M (i.e., an x-dependent positive-definite scalar product $(\ ,\ )_x$ with associated norm $\|\xi\|_x=[(\xi,\xi)_x]^{1/2}$) such that, for all $x\in\Lambda$*

$$\|Df(x)\xi\|_{f(x)} \leqslant \lambda \|\xi\|_x \text{ for } \xi \in E^s(x) \tag{3.6S}$$

$$\|Df^{-1}(x)\xi\|_{f^{-1}(x)} \leqslant \lambda \|\xi\|_x \text{ for } \xi \in E^u(x). \tag{3.6U}$$

Such a metric is called an *adapted* or *Lyapunov* metric. We omit the proof of the above proposition; it is not difficult.

To keep things elementary and to avoid complicating the notation unnecessarily, we will consider for the remainder of these notes only the case in which M is an open subset of $\mathbf{R}^m$ and in which (3.6) holds for the standard (x-independent) scalar product.

4. Examples.

A set consisting of a single hyperbolic fixed point is a simple example of a hyperbolic set. In this section we describe briefly some less trivial examples.

1. Linear automorphism of the two-dimensional torus. (This example was already introduced in Section 2.) Represent the two-dimensional torus as $\mathbf{R}^2/\mathbf{Z}^2$, and let f denote the mapping on the quotient space induced by

$$(x_1,x_2) \rightarrow (2x_1+x_2, x_1+x_2)$$

on $\mathbf{R}^2$. $Df(x)$ is everywhere equal to the 2×2 matrix

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

which has one eigenvalue greater than 1 and one between 0 and 1. The contracting and expanding subspaces $E^s(x)$, $E^u(x)$ do not vary with x; they are respectively the one-dimensional spaces spanned by the contracting and expanding eigenvectors of

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

The whole torus is a hyperbolic set, i.e., f is an Anosov diffeomorphism.

2. Smale horseshoe. The idea here is to look, not at the full structure of a mapping but at its action on a square Δ somewhere in the space (taken to be two-dimensional) on which the mapping acts. The mapping does not send the square into itself, but rather compresses it horizontally, stretches it vertically, bends it into a horseshoe shape, and lays it back over itself (extending out of top and bottom) as indicated in Figure 4.1. The set of points in Δ mapped back into Δ by one application of f consists of two roughly horizontal strips which we call Δ_0 and Δ_1.

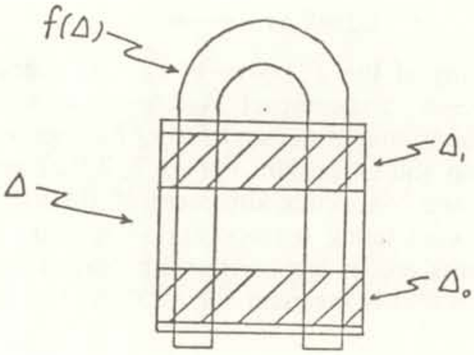

Figure 4.1. The Smale horseshoe

We now ask: What does the set Λ of points x such that $f^n(x) \in \Delta$ for all n (negative as well as positive) look like? It is not possible to answer this question on the basis of the qualitative description of f we have given, but it does become possible to answer it if we make the special assumption that f acts in a strictly affine way, with vertical expansion and horizontal contraction, on each of Δ_0 and Δ_1 (with, in addition, rotation through 180° on Δ_1). Under this assumption, there is a one-one correspondence between points of Λ and sequences

$$\mathbf{i} = (\cdots, i_{-1}, i_0, i_1, \cdots)$$

of 0's and 1's; the point x corresponding to the sequence $\mathbf{i}$ is determined by the condition

$$f^n(x) \in \Delta_{i_n} \text{ for all } n.$$

The set $\Lambda \subset \Delta$ is nondenumerable, closed, and totally disconnected. The orbit corresponding to a sequence $\mathbf{i}$ is periodic if and only if $\mathbf{i}$ is; since there are infinitely many periodic sequences, f has infinitely many periodic orbits in Λ. For each $x \in \Lambda$, $E^s(x)$ is the space of horizontal vectors and $E^u(x)$ the space of vertical vectors; it is not difficult to see that Λ is a hyperbolic set for f.

All these statements are proved in a straightforward and elementary way from the assumption that f is exactly affine on each of Δ_0 and Δ_1; the details can be found, for example, in Section 2 of Lanford [1]. It follows, however, from general theory we are going to develop that the qualitative properties of Λ (hyperbolicity, correspondence with sequences $\mathbf{i}$, existence of infinitely many periodic orbits) persist under small perturbations of f and hence are not specific to the assumed piecewise affine form.

3. Transverse homoclinic orbit. Let $\bar{x}$ be a hyperbolic fixed point for a differentiable mapping f on a two-dimensional space. Assume that $Df(\bar{x})$ has one-dimensional expanding and contracting subspaces, so W^s and W^u are both one-dimensional, i.e., are curves. Assume further that these curves intersect transversally (i.e., at non-zero angle) at some point $x_0 \neq \bar{x}$. (In general, a point x_0 other than $\bar{x}$ belonging to both W^s and W^u is called a *homoclinic point.* If, as we are assuming here, W^s and W^u intersect transversally at x_0, we speak of a *transverse* homoclinic point.) Since $x_0 \in W^s$,

$$f^n(x_0)\rightarrow\bar{x} \text{ as } n\rightarrow\infty,$$

and since $x_0\in W^u$,

$$f^n(x_0)\rightarrow\bar{x} \text{ as } n\rightarrow-\infty.$$

Let Λ denote the set consisting of the $f^n(x_0)$, $-\infty<n<\infty$, together with $\bar{x}$. Because $f^n(x_0)$ converges to $\bar{x}$ as $n\rightarrow\pm\infty$, Λ is compact. We claim that Λ is a hyperbolic set. To check this, it is necessary to determine $E^s(x)$ and $E^u(x)$ for a general point $x\in\Lambda$ and to prove uniformity of contraction and expansion. For $x=\bar{x}$, $E^s(x)$ and $E^u(x)$ are just the (one-dimensional) expanding and contracting subspaces of $Df(\bar{x})$. For $x=x_0$, it is easy to see that the only possible contracting (expanding) vectors are those tangent to W^s (W^u) at x_0. That these vectors are in fact contracting (expanding) is not difficult to show, using the fact that W^s (W^u) is invariant for f. Thus, the point x_0 is hyperbolic. From the general relations:

$$Df(x)E^s(x)=E^s(f(x))$$

$$Df(x)E^u(x)=E^u(f(x))$$

it follows that each point $x_n=f^n(x_0)$ is hyperbolic. To complete the proof that Λ is a hyperbolic set, it is only necessary to show that the contraction and expansion are sufficiently uniform. This follows from a local analysis of the action of f in a neighborhood of $\bar{x}$ which we will leave as an exercise.

This example has a quite different flavor from the Smale horseshoe example discussed earlier. In this case, Λ contains no periodic orbit except for the fixed point $\bar{x}$, while the hyperbolic set obtained in the Smale horseshoe construction contains infinitely many periodic orbits. On the other hand, the set obtained from the Smale horseshoe construction is isolated in the sense that any orbit not in it must at some time leave Δ while, as we will see in the next section, there are periodic orbits which stay arbitrarily near to any transverse homoclinic orbit. Intuitively, the Smale horseshoe example seems to be "complete", and the homoclinic orbit example seems to be only a small piece of a larger hyperbolic set. To some extent, this intuitive distinction is made precise in the notion of locally maximal hyperbolic set to be developed in Section 12.

5. Shadowing

Let $\delta>0$. A δ *pseudo-orbit* for f means a sequence of points (x_i) such that

$$d(x_{i+1},f(x_i))<\delta$$

for all (relevant) i. In other words, x_{i+1} is obtained from x_i by applying f, then making a jump which has length less than δ but which is otherwise arbitrary. (In the absence of a explicit statement to the contrary, our pseudo-orbits will always to be assumed to be defined for all integers i, negative as well as positive. The notion of pseudo-orbit does however make sense if the index set is any subinterval of **Z**.)

One of the keys to the analysis of hyperbolic sets is the fact that, if δ is small enough, every δ pseudo-orbit near a hyperbolic set is followed for all i by a true orbit. The precise statement is as follows:

Theorem 5.1. Shadowing Theorem. *Let Λ be a hyperbolic set for f, and let $\epsilon>0$. Then there exists $\delta>0$ such that, for every δ pseudo-orbit (x_i) with $d(x_i,\Lambda)<\delta$ for all i, there is a y_0 such that*

$$d(x_i,f^i(y_0))<\epsilon \text{ for all } i\in\mathbf{Z}. \qquad (5.1)$$

Furthermore, if ϵ is small enough, y_0 is uniquely determined.

If (5.1) holds, one says that the orbit $f^i(y_0)$ ϵ*-shadows* the pseudo-orbit (x_i). The theorem as formulated can easily be deduced from the apparently weaker result in which the pseudo-orbit is required to be contained in Λ (rather than simply near it). On the other hand, it is a delicate matter whether the shadowing orbit $(f^i(y_0))$ is necessarily contained in Λ or not; this turns out to be the case if and only if Λ has a property called *local product structure* to be discussed in Section 12.

Before giving the proof of the Shadowing Theorem, we will indicate some of its uses. One immediate consequence of the uniqueness assertion of the theorem is

Corollary 5.2 *Let Λ be a hyperbolic set for f. There exists $\epsilon>0$ such that, if*

$$d(f^i(x_0),\Lambda)<\epsilon \text{ for all } i$$

and if $\bar{x}_0\neq x_0$, then

$$d(f^i(x_0),f^i(\bar{x}^0))>\epsilon \text{ for some } i.$$

In other words, two orbits near Λ cannot stay arbitrarily close together for all time. This property is called *expansivity*; a number ϵ with the indicated property is called an *expansivity constant*.

A typical use of the Shadowing Theorem is to prove the existence of periodic orbits. The idea is simply that it is often easy to construct directly periodic pseudo-orbits with arbitrarily small δ and that, if (x_i) is a periodic pseudo-orbit with a unique shadowing orbit, then the shadowing orbit must also be periodic. For a concrete example, consider the situation described in Example 3 of the preceding section. That is: Λ consists of a hyperbolic fixed point $\bar{x}$ of saddle type (in two dimensions) together with a single orbit $(f^i(x_0))$ where x_0 is a transverse homoclinic point for $\bar{x}$:

$$f^n(x_0)\to\bar{x} \text{ as } n\to\infty, \text{ so } x_0\in W^s(\bar{x})$$

$$f^n(x_0)\to\bar{x} \text{ as } n\to-\infty, \text{ so } x_0\in W^u(\bar{x})$$

$$W^s(\bar{x}) \text{ intersects } W^u(\bar{x}) \text{ transversally at } x_0.$$

We will use the Shadowing Theorem to show that there are periodic orbits arbitrarily near to x_0. For any large N we construct a pseudo-orbit $(z_n), 0\leqslant n\leqslant N$ as follows: Put $z_0=x_0$, and make $z_n=f(z_{n-1})$ until $n=n_0\approx N/2$. At this point, z_n is very near to $\bar{x}$. Thus, by making only a small jump, we can take $z_{n_0+1}=f^{-(N-n_0-1)}(x_0)$, which is also very near to $\bar{x}$. Now go back to making $z_n=f(z_{n-1})$; this leads us back to $z_N=x_0$. We thus get a finite pseudo-orbit starting and ending at x_0; we repeat it periodically to get a periodic pseudo-orbit defined for all n. Given any $\epsilon>0$, we can by taking N big enough arrange that the pseudo-orbit constructed in this way is a δ pseudo-orbit for the δ related to ϵ as in the Shadowing Theorem. Let y_0 be a point whose orbit ϵ-shadows (z_n). Since (z_n) is periodic (with period N), the orbit of $f^N(y_0)$ shadows it exactly as well as does that of y_0, so, if ϵ is small enough to guarantee uniqueness of the shadowing orbit, it must be the case that

$$f^N(y_0)=y_0,$$

i.e., the shadowing orbit is also periodic. Since $z_0=x_0$, ϵ-shadowing implies that

$$d(x_0,y_0)<\epsilon.$$

Thus, as asserted, there are periodic points arbitrarily close to x_0

The above argument can easily be extended to prove the following: *For any* $\epsilon>0$, *there exists* N *such that, for any sequence* $\mathbf{i}=(\cdots,i_{-1},i_0,i_1,\cdots)$ *of 0's and 1's, there is a* y_0 *such that*

$$d(f^{nN}(y_0),x_0)<\epsilon \text{ if } i_n=1$$

and

$$d(f^{nN}(y_0),\bar{x})<\epsilon \text{ otherwise.}$$

In other words: There is an initial point y_0 whose orbit under f^N jumps back and forth from the vicinity of x_0 to the vicinity of $\bar{x}$ in any specified way. The preceding condition need not determine y_0 uniquely, but it is not difficult to specify a z_0 corresponding to each sequence $\mathbf{i}$ in such a way that $f^N(z_0)$ corresponds to the sequence obtained by shifting $\mathbf{i}$ one place to the left (and also in such a way that z_0 varies continuously with $\mathbf{i}$.) Thus we get

Proposition 5.3. *Let* $\epsilon>0$. *Then for any sufficiently large integer* N *there is a compact* $\Xi\subset\{x:d(x,\bar{x})<\epsilon \text{ or } d(x,x_0)<\epsilon\}$, *invariant for* f^N, *such that* f^N *on* Ξ *is conjugate to the left shift on the space* $\{0,1\}^{\mathbf{Z}}$ *of sequences of 0's and 1's.*

Such a Ξ is often called an *imbedded 2-shift.*

We are going to prove the Shadowing Theorem by reducing it to a fixed point problem in a Banach space. The idea is roughly as follows: Define an operator A on the space of sequences $\mathbf{x}$ in M by

$$A(\mathbf{x})_n=f(x_{n-1}).$$

A sequence $\mathbf{x}$ is an orbit if and only if

$$A(\mathbf{x})=\mathbf{x},$$

i.e., if and only if $\mathbf{x}$ is a fixed point for A. Similarly, if $\mathbf{x}$ is a δ pseudo-orbit,

$$\|A(\mathbf{x})-\mathbf{x}\|\equiv\sup_n\|A(\mathbf{x})_n-x_n\|\leqslant\delta.$$

What we need in order to prove the Shadowing Theorem is a result saying that near every point left approximately fixed by A there is a point left *exactly* fixed. We formulate in the next section a simple but powerful theorem along these lines, and in the following section we show how it can be applied to prove the Shadowing Theorem.

6. A fixed point theorem

Proposition 6.1. *Let A be a differentiable mapping defined on a neighborhood of* 0 *in a Banach space X, with values in X, and let Γ be a linear operator on X such that $(\Gamma-1)^{-1}$ is bounded. Assume that there are positive constants ρ, κ with $\kappa<1$ such that*
1. For $\|\xi\|\leqslant\rho$, $\|(\Gamma-1)^{-1}\|\cdot\|DA(\xi)-\Gamma\|\leqslant\kappa$.
2. $\|(\Gamma-1)^{-1}\|\cdot\|A(0)\|\leqslant(1-\kappa)\rho$.
Then A has exactly one fixed point in $\{\|\xi\|\leqslant\rho\}$.

Proof. We will apply the contraction mapping principle to

$$\Phi(\xi)\equiv\xi-(\Gamma-1)^{-1}[A(\xi)-\xi]$$

which has the same fixed points as A does. Condition 1. will imply that Φ is contractive on $\{\|\xi\|\leqslant\rho\}$; condition 2. that Φ maps this ball into itself. The verifications are almost immediate:

$$D\Phi(\xi)=1-(\Gamma-1)^{-1}(DA(\xi)-1)$$
$$=1-(\Gamma-1)^{-1}[(\Gamma-1)+(DA(\xi)-\Gamma)]=-(\Gamma-1)^{-1}(DA(\xi)-\Gamma)$$

From this and 1.,

$$\|D\Phi(\xi)\|\leqslant\kappa \text{ for } \|\xi\|\leqslant\rho,$$

which proves contractivity.
Now let $\|\xi\|\leqslant\rho$. Then

$$\|\Phi(\xi)\|\leqslant\|\Phi(\xi)-\Phi(0)\|+\|\Phi(0)\|$$
$$\leqslant\kappa\|\xi\|+\|(\Gamma-1)^{-1}\|\|A(0)\|\leqslant\kappa\rho+(1-\kappa)\rho=\rho.$$

7. Proof of the Shadowing Theorem

The proof of the Shadowing Theorem itself involves a certain amount of messy detail which distracts attention from the main ideas. We will therefore begin by proving a preliminary result which, although weaker than what we will eventually prove, displays the underlying argument in a clean way.

Proposition 7.1 *Let Λ be a hyperbolic set for f. Then there exists an $\epsilon>0$ such that, if $x_0\in\Lambda$ and if $x_0'\neq x_0$, then $d(f^n(x_0),f^n(x_0'))>\epsilon$ for some n.*

Note that this proposition proves less than expansivity since x_0 has to be in Λ (whereas, in Corollary 5.2, it is only required that $f^n(x_0)$ remain near Λ for all n).
Proof. Let $x_0\in\Lambda$; $x_n\equiv f^n(x_0)$, and let X denote the space of all bounded sequences $\xi=(\xi_n)_{n\in Z}$ in $\mathbf{R}^m$, equipped with the norm

$$\|\xi\|=\sup_n\|\xi^n\|$$

Define an operator A on X by

$$A(\xi)_n=f(x_{n-1}+\xi_{n-1})-x_n.$$

We don't need to worry about the domain of definition of A beyond noting that it contains a neighborhood of 0. A sequence ξ is a fixed point for A if and only if $(x_n+\xi_n)$ is an orbit for f. In particular, 0 is a fixed point for A. We are going to use the uniqueness statement of the fixed point theorem of the preceding section to show that there is an $\epsilon>0$ such that A has no other fixed point of norm no larger than ϵ. *It is an essential part* *of the argument that, although A depends on the orbit* (x_n), *we show that there is an* ϵ *which works for all orbits.*

To apply the fixed point theorem, we have first to check that A is differentiable on a neighborhood of $\mathbf{0}$ in X. This is an elementary verification, which we omit. The derivative of A at ξ is the linear operator on X given by

$$(DA(\xi)\eta)_n = Df(x_{n-1}+\xi_{n-1})\eta_{n-1} \tag{7.1}$$

The next step is to choose a Γ, and in this case we can take it to be simply $DA(\mathbf{0})$. (In later applications of these ideas, finding the right Γ will be considerably more complicated.) The main thing we have to do is to establish a bound on $\|(\Gamma-\mathbf{1})^{-1}\|$ which is independent of (x_n); the rest of the argument will be relatively routine.

Let X^s be the subspace of X consisting of sequences η with $\eta_n \in E^s(x_n)$ for all n; similarly, let X^u be the space of sequences with $\eta_n \in E^u(x_n)$. We will refer to X^s (respectively X^u) as the space of *contracting* (respectively *expanding*) sequences. From (7.1) (with $\xi=0$),

$$(\Gamma\eta)_{n+1} = Df(x_n)\eta_n,$$

and, from the covariance of the splitting, Γ maps each of X^s and X^u to itself. Furthermore, if $\eta \in X^s$,

$$\|\Gamma\eta\| = \sup_n \|(\Gamma\eta)_{n+1}\| = \sup_n \|Df(x_n)\eta_n\|$$

$$\leqslant \lambda\cdot\sup_n \|\eta_n\| = \lambda\|\eta\|,$$

i.e.,

$$\|\Gamma|_{X^s}\| \leqslant \lambda < 1.$$

Thus, on X^s, $(\Gamma-\mathbf{1})^{-1} = -\sum_{n=0}^{\infty}\Gamma^n$ and so $\|(\Gamma-\mathbf{1})^{-1}\| \leqslant (1-\lambda)^{-1}$. Similarly, for $\eta \in X^u$,

$$(\Gamma^{-1}\eta)_n = Df^{-1}(x_{n+1})\eta_{n+1},$$

so

$$\|\Gamma^{-1}\eta\| \leqslant \lambda\|\eta\|,$$

so, on X^u,

$$(\Gamma-\mathbf{1})^{-1} = \Gamma^{-1}(\mathbf{1}-\Gamma^{-1})^{-1} = \sum_{n=1}^{\infty}\Gamma^{-n}$$

and

$$\|(\Gamma-\mathbf{1})^{-1}\| \leqslant \lambda/(1-\lambda) < (1-\lambda)^{-1}.$$

Since $\mathbf{R}^m = E^s(x_n) \oplus E^u(x_n)$ for all n, $X = X^s \oplus X^u$, and since $\Gamma-\mathbf{1}$ is invertible on each of X^s and X^u, it is invertible on X. Now comes an irritating complication. In general, $E^s(x_n)$ and $E^u(x_n)$ need not be orthogonal, so it need not be true that the norm of $(\Gamma-\mathbf{1})^{-1}$ on X is the larger of the norms of its restrictions to X^s and X^u. To cope with this, we let

$$D = \sup_{x\in\Lambda} \max\{\|P^s(x)\|, \|P^u(x)\|\}.$$

Then

$$\max\{\|\xi^s\|,\|\xi^u\|\}\leqslant D\|\xi^s+\xi^u\|$$

for all $\xi^s\in E^s(x)$, $\xi^u\in E^u(x)$, and for all $x\in\Lambda$. From this it follows at once that, if $\xi^s\in X^s,\xi^u\in X^u$,

$$D^{-1}\max\{\|\xi^s\|,\|\xi^u\|\}\leqslant\|\xi^s+\xi^u\|\leqslant 2\max\{\|\xi^s\|,\|\xi^u\|\}$$

and hence that

$$\|(\Gamma-1)^{-1}\|\leqslant 2D\max\{\|(\Gamma-1)^{-1}|_{X^s}\|,\|(\Gamma-1)^{-1}|_{X^u}\|\}\leqslant 2D/(1-\lambda)\equiv B. \tag{7.2}$$

This is what we needed: a bound on $\|(\Gamma-1)^{-1}\|$ independent of the orbit (x_n).

To complete the proof, we have to show how to choose ρ,κ with $\kappa<1$ so that

$$\|(\Gamma-1)^{-1}\|\cdot\|DA(\xi)-\Gamma\|\leqslant\kappa$$

if $\|\xi\|\leqslant\rho$. (This is condition 1 of Proposition 6.1. In the case at hand, condition 2 is immediate since $A(0)=0$.) Any $\kappa<1$ will do; suppose one has been chosen. Then take ρ small enough so that, if $d(x,x')\leqslant\rho$

$$\|Df(x)-Df(x')\|\leqslant\kappa/B$$

(using the uniform continuity of Df on a neighborhood of Λ; B is as defined in (7.2).) From

$$(DA(\xi)\eta)_{n+1}=Df(x_n+\xi_n)\eta_n$$

(and using $\Gamma=DA(0)$) it follows that

$$\|DA(\xi)-\Gamma\|\leqslant\sup_n\|DA(x_n+\xi_n)-DA(x_n)\|\leqslant\kappa/B\leqslant\kappa/\|(\Gamma-1)^{-1}\|$$

provided $\|\xi\|\leqslant\rho$.

We have thus verified the hypotheses of the fixed point theorem of the preceding section; that theorem implies that A has no fixed point other that 0 in $\{\|\xi\|\leqslant\rho\}$ so the proposition is proved with $\epsilon=\rho$.

The proof of the Shadowing Theorem begins in the same way as the above proof: Given a pseudo-orbit (x_n), we introduce the space X and the operator A exactly as before, i.e.,

$$A(\xi)_{n+1}=f(x_n+\xi_n)-x_{n+1}.$$

Again:

$$(DA(\xi)\eta)_{n+1}=Df(x_n+\xi_n)\eta_n. \tag{7.3}$$

The main problem is to construct an appropriate Γ, i.e., an operator approximating $DA(0)$ and for which we can estimate $\|(\Gamma-1)^{-1}\|$. Informally, the argument goes as follows: The first step is to construct a continuous extension of the splitting

$$\mathbf{R}^m=E^s(x)\oplus E^u(x)$$

to a neighborhood U of Λ. If, then, (x_n) is a pseudo-orbit with a very small δ staying very close the Λ, it will follow from continuity that, for each n, $Df(x_n)$ is contractive on $E^s(x_n)$ and expansive on $E^u(x_n)$. It need not, however, map $E^s(x_n)$ (respectively $E^u(x_n)$) to $E^s(x_{n+1})$ ($E^u(x_{n+1})$). We fix this by modifying $Df(x_n)$ slightly to get a new linear mapping $\Gamma^{x_n}_{x_{n+1}}$ which does preserve the splitting and show that the necessary modification can be made small enough so that $\Gamma^{x_n}_{x_{n+1}}$ is still contractive on $E^s(x_n)$ and expansive on $E^u(x_n)$. We take Γ to be defined by

$$(\Gamma\eta)_{n+1}=\Gamma^{x_n}_{x_{n+1}}\eta_n;$$

then Γ (respectively Γ^{-1}) maps X^s (respectively X^u) contractively to itself, and so $\|(\Gamma-\mathbf{1})^{-1}\|$ can be estimated as in the proof of Proposition 7.1. Since $\Gamma^{x_n}_{x_{n+1}}\approx Df(x_n)$, $\Gamma\approx DA(0)$. Further, if $\|\xi\|$ is small, $\|DA(\xi)-DA(0)\|$ will also be small, so we can satisfy condition 1 of the fixed point theorem on a small ball about 0. Condition 2 of that theorem will also be satisfied if δ is small enough, and the Shadowing Theorem will follow.

We turn now to the formal details.

1. *There is a neighborhood U of Λ and a bounded continuous projection-valued extension of $P^s(x)$ to U.* We will also denote the extension by $P^s(x)$.

Proof. $P^s(x)$ is a continuous matrix-valued function defined on Λ satisfying

$$(P^s(x))^2=P^s(x) \tag{7.4}$$

(equivalent to the fact that $P^s(x)$ is a projection.) First use the Tietze Extension Theorem to extend each of the matrix elements of $P^s(x)$ to a bounded continuous function defined on all of M. This gives a continuous matrix-valued extension $\tilde{P}^s(x)$ to all of M, but the projection condition (7.4) need not hold except on Λ. For $x\in\Lambda$, $\tilde{P}^s(x)$ *is* a projection, so its spectrum consists of the two points 0 and 1. By elementary perturbation theory, the spectrum of $\tilde{P}^x(x)$ varies continuously with x. Let U be a neighborhood of Λ on which the spectrum is contained in the union of two open disks, each of radius 1/3, centered respectively at 0 and 1. We can take, as our extension of $P^s(x)$ to U, the spectral projection for $\tilde{P}^s(x)$ corresponding to the part of the spectrum inside the disk of radius 1/3 and center 1.

We will also call the extension $P^s(x)$, and for general $x\in U$, we will write

$P^u(x)$ for $\mathbf{1}-P^s(x)$

$E^s(x)$ for the range of $P^s(x)$

$E^u(x)$ for the range of $P^u(x)$.

Warning: This notation is not consistent with the definition given in Section 3 of $E^s(x)$ and $E^u(x)$ for a general $x\in M$. The latter notation will not be used in this section.

2. For $x,x'\in U$, put

$$\Gamma^{x'}_x=P^s(x)Df(x')P^s(x')+P^u(x)Df(x')P^u(x'). \tag{7.5}$$

Note that

a. $\Gamma^{x'}_x$ *maps* $E^s(x')$ *to* $E^s(x)$ *and* $E^u(x')$ *to* $E^u(x)$.

b. $\Gamma^{x'}_x$ *is a continuous function of* x,x'.

c. If $x'\in\Lambda$ *and* $x=f(x')$ *then* $\Gamma^{x'}_x$ *reduces to* $Df(x')$.

3. Pick a λ_1 with $\lambda<\lambda_1<1$. *There exists $\delta_1>0$ such that, if $d(x',\Lambda)\leqslant\delta_1$ and $d(f(x'),x)\leqslant\delta_1$ then*

$$\|\Gamma^{x'}_x\xi\|\leqslant\lambda_1\|\xi\| \text{ for } \xi\in E^s(x')$$

$$\|(\Gamma^{x'}_x)^{-1}\xi\|\leqslant\lambda_1\|\xi\| \text{ for } \xi\in E^u(x)$$

Proof. This follows easily from 2a.,2b., the continuity of $E^s(x')$ and $E^u(x)$, and

$$\|Df(z)\xi\|\leqslant\lambda\|\xi\| \text{ for } \xi\in E^s(z),\ z\in\Lambda$$

$$\|(Df(z))^{-1}\xi\| \leqslant \lambda \|\xi\| \text{ for } \xi \in E^u(f(z)),\ z \in \Lambda.$$

4. Let (x_n) be a pseudo-orbit in U. Define an operator Γ by

$$(\Gamma\eta)_{n+1} = \Gamma_{x_{n+1}}^{x_n} \eta_n, \tag{7.6}$$

and let

$$X^s = \{\eta \in X : \eta_n \in E^s(x_n) \text{ for all } n\}$$

$$X^u = \{\eta \in X : \eta_n \in E^u(x_n) \text{ for all } n\}$$

By 2a., Γ maps each of X^s and X^u to itself. Let λ_1, δ_1 be as in 3. If (x_n) is a δ_1 pseudo-orbit, and if $d(x_n, \Lambda) \leqslant \delta_1$ for all n, then 3. and the definition of Γ imply

$$\|\Gamma\| \leqslant \lambda_1 \text{ on } X^s$$

$$\|\Gamma^{-1}\| \leqslant \lambda_1 \text{ on } X^u.$$

By the argument used in the proof of Proposition 7.1, there is a constant D such that, for any pseudo-orbit as above,

$$\|(\Gamma - \mathbf{1})^{-1}\| \leqslant 2D/(1-\lambda_1) \equiv B \tag{7.7}$$

It is now easy to complete the proof of the Shadowing Theorem. We suppose, thus, that we are given $\epsilon > 0$; we want to find δ so that any δ pseudo-orbit staying within a distance δ of Λ is ϵ-shadowed by a true orbit. We will take δ smaller than the δ_1 of 3. and also small enough so that the pseudo-orbits we consider are necessarily in U; then (7.7) holds. Take any κ with $0 < \kappa < 1$. From the definition (7.6) of Γ, the formula (7.3) for $DA(\xi)$, and the continuity of $Df(x)$, it follows that if we take δ small enough we can guarantee that

$$\|\Gamma - DA(0)\| \leqslant \kappa/3B \tag{7.8}$$

Similarly, we can choose $\epsilon_1 < \epsilon$ so that $\|Df(z_1) - Df(z_2)\| \leqslant \kappa/3B$ whenever $d(z_1, z_2) \leqslant \epsilon_1$; then it follows from (7.3) that

$$\|DA(\xi) - DA(0)\| \leqslant \kappa/3B \tag{7.9}$$

for $\|\xi\| \leqslant \epsilon_1$.

Combining (7.7), (7.8), and (7.9), we get

$$\|(\Gamma - \mathbf{1})^{-1}\| \cdot \|\Gamma - DA(\xi)\| \leqslant B(\kappa/3B + \kappa/3B) < \kappa$$

for $\|\xi\| \leqslant \epsilon_1$, so condition 1. of the fixed point theorem of Section 6 is verified. If, finally, we also take δ small enough so that

$$\delta < \epsilon_1(1-\kappa)/B \tag{7.10}$$

we get

$$\|(\Gamma - \mathbf{1})^{-1}\| \cdot \|A(0)\| \leqslant B\delta < \epsilon_1(1-\kappa),$$

which is condition 2 of the fixed point theorem (with $\rho = \epsilon_1$). Thus, A has exactly one fixed point with norm no larger than ϵ_1, i.e., there is one and only one y_0 such that

$$d(f^n(y_0), x_n) \leqslant \epsilon_1$$

for all n.

8. Extensions. Structural stability of hyperbolic sets.

There is much more to be extracted from the methods developed in the preceding section. One major extension starts from the observation that these methods imply the existence of shadowing orbits *not just for f itself but for any $\tilde{f}$ near enough to f.*

Theorem 8.1. *Let Λ be a hyperbolic set for f, and let $\epsilon>0$. Then there is a neighborhood W of f in the space of C^1 diffeomorphisms of M and a $\delta>0$ such that, if (x_n) is a δ pseudo-orbit for f with $d(x_n,\Lambda)\leqslant\delta$ for all n, and if $\tilde{f}\in W$, then there is a y_0 such that*

$$d(\tilde{f}^n(y_0),x_n)\leqslant\epsilon$$

for all $n\in\mathbf{Z}$. If ϵ is small enough, y_0 is uniquely determined.

Proof. The argument is a simple extension of the proof of the Shadowing Theorem; we indicate only the changes. We use the same Γ, and again take δ small enough so that (7.8) holds and ϵ_1 small enough so that (7.9) holds. We take W small enough so that

$$\|Df(x)-D\tilde{f}(x)\|\leqslant\kappa/3B$$

for all x if $\tilde{f}\in W$ Thus, if we define

$$\tilde{A}(\xi)_{n+1}=\tilde{f}(x_n+\xi_n)-x_{n+1}$$

we get

$$\|D\tilde{A}(\xi)-DA(\xi)\|\leqslant\kappa/3B$$

Combining estimates, we thus have

$$\|(\Gamma-1)^{-1}\|\cdot\|D\tilde{A}(\xi)-\Gamma\|\leqslant\kappa \text{ for } \|\xi\|\leqslant\epsilon_1.$$

Finally, instead of (7.10) we require

$$\delta\leqslant\epsilon_1(1-\kappa)/2B$$

and we require that W be small enough to guarantee that, if $\tilde{f}\in W$, then $d(\tilde{f}(x),f(x))\leqslant\delta$ for all x. From the latter condition,

$$\|\tilde{A}(0)-A(0)\|\leqslant\delta$$

and, as usual,

$$\|A(0)\|\leqslant\delta.$$

Hence

$$\|(\Gamma-1)^{-1}\|\cdot\|\tilde{A}(0)\|\leqslant B\cdot 2\cdot\epsilon_1(1-\kappa)/2B=\epsilon_1(1-\kappa),$$

and Theorem 8.1 follows from the fixed point theorem of Section 6.

We get a striking result if we apply Theorem 8.1 to *exact* f-orbits on Λ: If ϵ is small enough, if W corresponds to ϵ as in the theorem, if $\tilde{f}\in W$, and if $x_0\in\Lambda$, then there is a unique y_0 such that

$$d(\tilde{f}^n(y_0),f^n(x_0))\leqslant\epsilon.$$

In other words, every f-orbit on Λ is ϵ-shadowed by an $\tilde{f}$-orbit. We write

$$y_0=\tilde{h}(x_0);$$

$\tilde{h}$ is a mapping of Λ into M satisfying $d(\tilde{h}(x),x)\leqslant\epsilon$ for all x. From the uniqueness of y_0, it follows immediately that

$$\tilde{h}(f(x))=\tilde{f}(\tilde{h}(x)) \tag{8.1}$$

We will argue shortly that $\tilde{h}$ is one-one and continuous, and hence maps Λ homeomorphically onto a compact set $\tilde{\Lambda}$. The intertwining relation (8.1) implies that $\tilde{\Lambda}$ is invariant for $\tilde{f}$ and that $\tilde{f}$ on $\tilde{\Lambda}$ is topologically conjugate to f on Λ Thus:

Theorem 8.2. *Let Λ be a hyperbolic set for f. Then, for any $\tilde{f}$ sufficiently close to f in the C^1 topology, there is a homeomorphism $\tilde{h}$ of Λ onto a set $\tilde{\Lambda}$ invariant for $\tilde{f}$ such that*

$$\tilde{f}=\tilde{h}f\tilde{h}^{-1} \text{ on } \tilde{\Lambda}.$$

$\tilde{h}$ can be made as close as desired to the identity by making $\tilde{f}$ close to f.

Proof. It remains only to show that $\tilde{h}$, constructed above, is one-one and continuous. Take ϵ small enough so that any two distinct f orbits on Λ are somewhere separated by more than 2ϵ, and let $\tilde{f}$ belong to the corresponding W. Suppose

$$\tilde{h}(x_0)=\tilde{h}(x_0')\equiv y_0.$$

By the definition of $\tilde{h}$,

$$d(f^n(x_0),\tilde{f}^n(y_0))\leqslant\epsilon$$
$$d(f^n(x_0'),\tilde{f}^n(y_0))\leqslant\epsilon;$$

hence,

$$d(f^n(x_0),f^n(x_0'))\leqslant 2\epsilon$$

for all n. By the choice of ϵ, this implies $x_0=x_0'$.

To prove that $\tilde{h}$ is continuous, it suffices, using compactness, to prove that if $x^{(j)}\to x$ and $y^{(j)}\equiv\tilde{h}(x^{(j)})\to y$ then $\tilde{h}(x)=y$. For any j and any n,

$$d(f^n(x^{(j)}),\tilde{f}^n(y^{(j)}))\leqslant\epsilon.$$

Letting j go to infinity with n fixed, we get

$$d(f^n(x),\tilde{f}^n(y))\leqslant\epsilon \text{ for all } n,$$

which means exactly that

$$y=\tilde{h}(x),$$

as desired.

Intuitively, this theorem expresses a *structural stability* property of hyperbolic sets. It says that, if Λ is a hyperbolic set for f and if $\tilde{f}$ is obtained by perturbing f slightly, then there is a set $\tilde{\Lambda}$ near Λ, invariant for $\tilde{f}$, such that the action of $\tilde{f}$ on $\tilde{\Lambda}$ is indistinguishable, from a topological point of view, from the action of f on Λ. Note that we have *not* proved, however, that $\tilde{\Lambda}$ is a hyperbolic set for $\tilde{f}$. This is in fact true, and will follow from a result to be proved in the next two sections.

If f is an Anosov diffeomorphism, i.e., if the whole state space M is a hyperbolic set for f, then the preceding result can easily be extended to show that f is structurally stable in the strict sense.

Corollary 8.3. Structural stability of Anosov diffeomorphisms. *Let f be an Anosov diffeomorphism. For any $\tilde{f}$ which is close enough to f in the C^1 topology, there is a* *homeomorphism $\bar{h}$ of M onto itself such that*

$$\bar{f}=\bar{h}f\bar{h}^{-1}.$$

Proof. All that needs to be shown is that the image of $\bar{h}$ is all of M. This follows from general (if high-powered) topological considerations, but we can also prove it directly from shadowing. Given y_0, we have to show that there is an x_0 such that

$$d(f^n(x_0),\bar{f}^n(y_0))\leqslant\epsilon \text{ for all } n,$$

i.e., such that the f-orbit of x_0 ϵ-shadows the $\bar{f}$-orbit of y_0. But by making $\bar{f}$ close enough to f in the C^0 topology, we can guarantee that the $\bar{f}$ orbit of any y_0 is a δ pseudo-orbit (for f) with any pre-assigned δ, so the existence of x_0 follows directly from the Shadowing Theorem.

9. A Banach-space characterization of hyperbolicity.

One of the keys to the analysis of hyperbolic sets is the systematic exploitation of the notion of pseudo-orbit and shadowing. A second, which we have already seen in the proof of the Shadowing Theorem, is the translation of geometrical (or dynamical) questions in the finite-dimensional space M into questions about differentiable operators on Banach spaces. In this section, we develop this line of attack further by showing how to recognize a hyperbolic set in terms of properties of a related operator on a Banach space.

Let Λ be a compact invariant set for the C^1 diffeomorphism f. We let $X(\Lambda)$ denote the space of bounded but otherwise arbitrary mappings ξ from Λ to $\mathbf{R}^m$, equipped with the supremum norm, and we define an operator $f_*(\Lambda)$ on $X(\Lambda)$ by

$$(f_*(\Lambda)\xi)(x)=Df(f^{-1}(x))\xi(f^{-1}(x)).$$

We will refer to elements of $X(\Lambda)$ as *(bounded) vector fields* on Λ. It will usually be clear what set Λ we have in mind, and in this case we will frequently write X and f_* for $X(\Lambda)$ and $f_*(\Lambda)$ respectively. Recall that we defined in Section 2 a linear operator on a Banach space to be hyperbolic if its spectrum does not intersect the unit circle.

Theorem 9.1. *Λ is a hyperbolic set for f if and only if $f_*(\Lambda)$ is a hyperbolic linear operator on $X(\Lambda)$.*

Remark. The theorem remains true if $X(\Lambda)$ is replaced by the space of continuous vector fields of Λ. We will not give the proof of this stronger result. The main ideas involved are the same as in the proof of Theorem 9.1, but a bit more technical work is needed.

Proof. For the proof of this theorem, we need to work with the general, co-ordinate independent, condition

$$\|Df^n(x)\xi\|\leqslant c\lambda^n\|\xi\| \text{ for } \xi\in E^s(x)$$

and similarly for $E^u(x)$, rather than with the more specialized condition with $c=1$.

The proof that f_* is hyperbolic if Λ is a hyperbolic set is almost immediate. Assume Λ is hyperbolic, and write

$$X^s(\Lambda)=\{\xi\in X(\Lambda):\xi(x)\in E^s(x) \text{ for all } x\} \tag{9.1}$$

$$X^u(\Lambda)=\{\xi\in X(\Lambda):\xi(x)\in E^u(x) \text{ for all } x\}.$$

Then

$$X=X^s\oplus X^u,$$

and X^s and X^u are invariant subspaces for f_* (by the covariance of the splitting $\mathbf{R}^m = E^s(x) \oplus E^u(x)$.) By the chain rule,

$$(f_*^n \xi)(f^n(x)) = Df^n(x)\xi(x)$$

(where n may be negative as well as positive.) Hence, for $\xi \in X^s$,

$$\|f_*^n \xi\| = \sup_X \|f_*^n \xi(f^n(x))\| = \sup_X \|Df^n(x)\xi(x)\|$$

$$\leqslant c\lambda^n \sup_X \|\xi(x)\| = c\lambda^n \|\xi\|.$$

Thus,

$$\|(f_*|_{X^s})^n\| \leqslant c\lambda^n,$$

so the spectrum of $f_*|_{X^s}$ is contained in $\{|z| \leqslant \lambda\}$. Similarly, the spectrum of $f_*|_{X^u}$ is contained in $\{|z| \geqslant \lambda^{-1}\}$, so f_* is a hyperbolic operator on X.

Conversely, suppose f_* is a hyperbolic operator on X. Then there is a splitting

$$X = X^s \oplus X^u$$

into f_*-invariant subspaces and constants c, λ, with $0 < \lambda < 1$, such that

$$\|(f_*|_{X^s})^n\| \leqslant c\lambda^n$$

$$\|(f_*|_{X^u})^{-n}\| \leqslant c\lambda^n.$$

What we want to show, in essence, is that this splitting has the "local form" of (9.1).

Let $\xi \in X^s$. Then

$$c\lambda^n \|\xi\| \geqslant \|(f_*)^n \xi\| = \sup_X \|Df^n(x)\xi(x)\|.$$

In particular: For any fixed x, $\|Df^n(x)\xi(x)\|$ goes to zero exponentially with n, i.e., $\xi(x) \in E^s(x)$. Similarly, if $\xi \in X^u$, $\xi(x) \in E^u(x)$ for all x. Since X^s and X^u span X, it follows that $E^s(x)$ and $E^u(x)$ span $\mathbf{R}^m$ for every x.

Now let $x_0 \in \Lambda$ and let $\xi_0 \in E^s(x_0)$. Define a vector field ξ by

$$\xi(x) = 0, \; x \neq x_0$$

$$= \xi_0, \; x = x_0.$$

Then

$$\|f_*^n \xi\| = \|Df^n(x_0)\xi_0\|$$

(where the norm on the left is the norm on X while the norm on the right is on $\mathbf{R}^m$.) Since the right-hand side goes to zero as n goes to infinity, $\xi \in X^s$, so

$$\|Df^n(x_0)\xi_0\| = \|f_*^n \xi\| \leqslant c\lambda^n \|\xi\| = c\lambda^n \|\xi_0\|,$$

which establishes the uniformity in x of the rate of decay of vectors in $E^s(x)$ under application of $Df^n(x)$. A similar argument establishes the uniformity in the decay of vectors in $E^u(x)$ under the application of $Df^{-n}(x)$.

It remains to show that, for all x,

$$E^s(x) \cap E^u(x) = \{0\}.$$

This is proved by the same sort of argument: Let $\xi_0 \in E^s(x_0) \cap E^u(x_0)$; let $\xi(x)$ be ξ_0 for $x = x_0$ and 0 otherwise. Then $\|f_*^n \xi\|$ goes to zero as n goes to ∞, so $\xi \in X^s$, but also goes to zero as n goes to $-\infty$ so $\xi \in X^u$. Hence

$$\xi \in X^s \bigcap X^u = \{0\},$$

so $\xi_0 = 0$, as desired.

10. Stability of hyperbolicity

In Section 8, we saw that, if Λ is a hyperbolic set for f and if $\tilde{f}$ is near enough to f in the C^1 topology, then there is a set $\tilde{\Lambda}$ near Λ, invariant for $\tilde{f}$, such that $\tilde{f}$ on $\tilde{\Lambda}$ is topologically conjugate to f on Λ. We will show here, using the characterization of hyperbolic set established in the preceding section, that $\tilde{\Lambda}$ is also a hyperbolic set for $\tilde{f}$. In fact, we will prove something considerably more comprehensive: If $\tilde{f}$ is near enough to f, then *any* invariant set for $\tilde{f}$ which is near enough to Λ is hyperbolic. Besides the application mentioned above, this theorem can be used, for example, to prove hyperbolicity of the imbedded 2-shifts which, as we saw in Section 5, exist arbitrarily near to a transverse homoclinic orbit.

Theorem 10.1. *Let Λ be a hyperbolic set for f. Then there exists a C^1 neighborhood W of f and a neighborhood U of Λ such that, if $\tilde{f} \in W$ and if Ξ is a compact subset of U invariant for $\tilde{f}$, then Ξ is a hyperbolic set for $\tilde{f}$.*

Proof. We will combine the characterization of hyperbolic sets given in the preceding section, the constructions used in the proof of the Shadowing Theorem, and the following simple result from operator theory.

Proposition 10.2. *Let Γ be a hyperbolic linear operator and let*

$$B \equiv \sup_{\theta \text{real}} \| (\Gamma - e^{i\theta}\mathbf{1})^{-1} \|$$

Then any linear operator Γ_1 with

$$\|\Gamma_1 - \Gamma\| < B^{-1}$$

is also hyperbolic.

Proof of Proposition 10.2. Write

$$\Gamma_1 - e^{i\theta}\mathbf{1} = [\Gamma - e^{i\theta}\mathbf{1}]\cdot[\mathbf{1} + (\Gamma - e^{i\theta}\mathbf{1})^{-1}(\Gamma_1 - \Gamma)].$$

For any real θ, the first factor on the right is invertible by assumption and the second because

$$\|(\Gamma - e^{i\theta}\mathbf{1})^{-1}(\Gamma_1 - \Gamma)\| < B \cdot B^{-1} = 1$$

Returning to the proof of the theorem: As in the proof of the Shadowing Theorem, we find a continuous bounded projection-valued extension of $P^s(x)$ to a neighborhood U_1 of Λ. We define

$$P^u(x) = \mathbf{1} - P^s(x)$$

$$\Gamma_{x_2}^{x_1} = P^s(x_2) Df(x_1) P^s(x_1) + P^u(x_2) Df(x_1) P^u(x_1).$$

Pick λ_1 between λ and 1, and recall from step 3 of the proof of the Shadowing Theorem that there is a neighborhood U_2 of Λ, contained in U_1, and a constant $\delta_1 > 0$ such that, if x_1, x_2 are in U_2 and if $d(x_2, f(x_1)) \leqslant \delta_1$ then

$$\|\Gamma_{x_2}^{x_1}\xi\| \leqslant \lambda_1 \|\xi\| \text{ for } \xi \text{ in the range of } E^s(x_1) \tag{10.1}$$

$$\|(\Gamma_{x_2}^{x_1})^{-1}\xi\| \leqslant \lambda_1\|\xi\| \text{ for } \xi \text{ in the range of } E^u(x_1).$$

Now let $\bar{f}$ be a diffeomorphism such that

$$d(\bar{f}(x), f(x)) \leqslant \delta_1 \text{ for } x \in U_2.$$

For any compact $\bar{f}$-invariant set $\Xi \subset U_2$, we put

$$X = X(\Xi)$$

$$X^s = \{\xi \in X : P^s(x)\xi(x) = \xi(x) \text{ for all } x \in \Xi\}$$

$$X^u = \{\xi \in X : P^u(x)\xi(x) = \xi(x) \text{ for all } x \in \Xi\}$$

and we define an operator Γ on X by

$$(\Gamma\xi)(\bar{f}(x)) = \Gamma_{\bar{f}(x)}^{x}\xi(x). \tag{10.2}$$

Then X^s and X^u are invariant for Γ and, from (10.1),

$$\|\Gamma\xi\| \leqslant \lambda_1\|\xi\| \text{ for } \xi \in X^s$$

$$\|\Gamma^{-1}\xi\| \leqslant \lambda_1\|\xi\| \text{ for } \xi \in X^u$$

An argument used in the proof of Proposition 7.1 shows, from these bounds, that $\Gamma - \mathbf{1}$ is invertible on each of X^s and X^u and that both inverses have norm no larger than $(1-\lambda_1)^{-1}$. Applying that argument with Γ replaced by $e^{-i\theta}\Gamma$ shows that the same is true for $\Gamma - e^{i\theta}\mathbf{1}$ for any real θ.

Let

$$D = \sup_{x \in U_2}\left(\max\{\|P^s(x)\|, \|P^u(x)\|\}\right).$$

As in the proof of Proposition 7.1,

$$\|(\Gamma - e^{i\theta}\mathbf{1})^{-1}\| \leqslant 2D/(1-\lambda_1) \equiv B.$$

Now choose $U \subset U_2$ and $\delta_2 \leqslant \delta_1$ so that if $x_1 \in U$ and $d(x_2, f(x_1)) < \delta_2$, then

$$\|\Gamma_{x_2}^{x_1} - Df(x_1)\| \leqslant 1/3B.$$

(This is possible by step 2 of the proof of the Shadowing Theorem.) Finally, let W be a C^1 neighborhood of f small enough to ensure that any $\bar{f} \in W$ satisfies

$$d(f(x), \bar{f}(x)) \leqslant \delta_2 \text{ and } \|Df(x) - D\bar{f}(x)\| \leqslant 1/3B \text{ for all } x.$$

If $\bar{f} \in W$ and if Ξ is a compact $\bar{f}$-invariant subset of U, then, with Γ defined as in (10.2), we have

$$\|(\Gamma - e^{i\theta}\mathbf{1})^{-1}\| \leqslant B \text{ for all real } \theta.$$

$$\|\Gamma - \bar{f}_*(\Xi)\| = \sup_{x \in \Xi}\|\Gamma_{\bar{f}(x)}^{x} - D\bar{f}(x)\|$$

$$\leqslant \sup_{x \in \Xi}\|\Gamma_{\bar{f}(x)}^{x} - Df(x)\| + \sup_{x \in \Xi}\|Df(x) - D\bar{f}(x)\|$$

$$\leqslant 1/3B + 1/3B < 1/B$$

Proposition 10.2 now applies and shows that $\bar{f}_*(\Xi)$ is hyperbolic, as desired.

11. Stable and unstable manifolds for hyperbolic sets.

We recall the principal results of Section 2. If x_0 is a hyperbolic fixed point for f, the *stable manifold* W^s of x_0 means the set of points x such that $f^n(x)$ converges to x_0 as $n \to \infty$. This set is a submanifold of M with dimension equal to that of $E^s(x_0)$. A useful way of characterizing a local piece of W^s is as the set of points x such that $d(f^n(x), x_0) < \epsilon$ for all $n \geqslant 0$, with ϵ a sufficiently small positive number.

These considerations generalize with straightforward changes to points x_0 which are not fixed or periodic but which do lie in a hyperbolic set. The stable manifold for such a point will be the set of x's such that

$$d(f^n(x), f^n(x_0)) \to 0 \text{ as } n \to \infty,$$

i.e., points whose forward orbits are asymptotic to the not-necessarily-constant forward orbit of x_0. Note that this means that stable manifolds are not, in general, invariant, but rather *covariant:* f maps the stable manifold of x_0 to the stable manifold for $f(x_0)$. As with fixed points, the main task is to analyze *local* stable manifolds.

Theorem 11.1. Stable Manifold Theorem for Hyperbolic Sets. *Let* Λ *be a hyperbolic set for* f. *For sufficiently small* ϵ:

1. For any $x_0 \in \Lambda$ *and any* $\xi^s \in E^s(x_0)$ *with* $\|\xi^s\| < \epsilon$, *there is a unique element* $\xi^u \equiv w_s(x_0, \xi^s)$ *of* $E^u(x_0)$ *such that, writing* x *for* $x_0 + \xi^s + \xi^u$,

$$d(f^n(x), f^n(x_0)) < \epsilon \text{ for all } n > 0.$$

2.

$$d(f^n(x), f^n(x_0)) \to 0 \text{ exponentially as } n \to \infty.$$

3. For fixed x_0, $w_s(x_0, \xi^s)$ *is a continuously differentiable function of* ξ^s *which vanishes, together with its first derivative, at* 0. *If* f *is* r *times continuously differentiable,* $w_s(x_0, \xi^s)$ *is* r *times continuously differentiable in* ξ^s *and the partial derivatives with respect to* ξ^s *are jointly continuous in* x_0, ξ^s.

Proof. We will not give a complete proof but only sketch an argument showing how to reduce this theorem to the Stable Manifold Theorem for Hyperbolic Fixed Points in an infinite dimensional space. The construction is, by now, familiar: Let X denote the space of bounded vector fields ξ on Λ. Define an operator A on X by

$$A(\xi)(f(x)) = f(x + \xi(x)) - f(x).$$

(Intuitively: Write $h(x) = x + \xi(x)$; then the h corresponding to $A(\xi)$ is fhf^{-1}.) The zero vector field is evidently a fixed point for A, and it is easy to check that A is continuously differentiable on a neighborhood of $\mathbf{0}$ in X and that

$$DA(0) = f_*(\Lambda).$$

Since Λ is a hyperbolic set, we know from Theorem 9.1 that $f_*(\Lambda)$ is a hyperbolic linear operator, i.e., that $\mathbf{0}$ is a *hyperbolic* fixed point for A. Furthermore, from the proof of Theorem 9.1 we know what the contracting and expanding subspaces of $DA(\mathbf{0})$ are; they are respectively the spaces X^s (X^u) of vector fields with $\xi(x) \in E^s(x)$ ($E^u(x)$) for all x.

We apply the Stable Manifold Theorem for Hyperbolic Fixed Points to A. Thus, for ϵ sufficiently small, to every $\xi^u \in X^s$ with $\|\xi^s\| < \epsilon$, there corresponds a unique $\xi^u \equiv w_s(\xi^s) \in X^u$ such that

$$\|A^n(\xi^s + \xi^u)\| < \epsilon \text{ for all } n > 0. \tag{11.1}$$

It is easy to check that

$$A^u(\xi)(f^n(x))=f^n(x+\xi(x))-f^n(x),$$

so (11.1) can be rewritten as

$$d(f^n(x+\xi^s(x)+\xi^u(x)),f^n(x))<\epsilon \text{ for all } n>0 \text{ and all } x\in\Lambda. \qquad (11.2)$$

In principle, the value of ξ^u at a particular x depends on the values of ξ^s on all of Λ. It is clear from the form of the determining equation (11.2), however, that different x's are completely independent. It thus follows from the above that, for any $x_0 \in \Lambda$ and any $\xi^s \in E^s(x_0)$ with $\|\xi^s\|<\epsilon$, there is a unique $\xi^u \equiv w_s(x_0,\xi^s)\in E^u(x_0)$ such that, writing

$$x=x_0+\xi^s+\xi^u,$$

we have

$$d(f^n(x),f^n(x_0))<\epsilon \text{ for all } n>0,$$

and

$$w_s(\xi^s)(x)=w_s(x,\xi^s(x)). \qquad (11.3)$$

This proves the first statement of the theorem.

If f is r times continuously differentiable, then so is A and hence, by the Stable Manifold Theorem for Hyperbolic Fixed Points, so is w_s. It is not hard to show, using this fact and the form (11.3) of w_s that, for x_0 fixed, $w_s(x_0,\xi^s)$ is r times continuously differentiable in ξ^s.

It remains only to prove the joint continuity of the partial derivatives of $w_s(x_0,\xi^s)$ with respect to ξ^s. This can be done by repeating the above analysis for the operator obtained by letting A act on the space of *continuous* vector fields (which also has 0 as a hyperbolic fixed point).

As for the fixed point case, we can reformulate part of the above theorem in terms of the sets

$$W^s_\epsilon(x_0)=\{x\in M:d(f^n(x),f^n(x_0))<\epsilon \text{ for all } n\geqslant 0\}.$$

The theorem tells us that, if ϵ is sufficiently small, then for any $x_0\in\Lambda$, $W^s_\epsilon(x_0)$ is a submanifold of M, as smooth as f is, with dimension equal to that of $E^s(x_0)$, passing through x_0 and tangent there to $E^s(x_0)$; furthermore, in a natural sense, $W^s_\epsilon(x_0)$ varies in a continuous way with x_0.

We also define *global* stable manifolds:

$$W^s(x_0)=\{x\in M;d(f^n(x),f^n(x_0))\rightarrow 0 \text{ as } n\rightarrow\infty\}.$$

It is easy to show, by the same argument as we used in the fixed-point case, that $W^s(x_0)$ is a countable union of pieces each of which is a well-behaved submanifold of M. Also as in the fixed-point case, the global manifold can accumulate on itself in a complicated way.

We define local and global *unstable* manifolds $W^u_\epsilon(x_0)$ and $W^u(x_0)$ by

$$W^u_\epsilon(x_0)=\{x\in M:d(f^{-n}(x),f^{-n}(x_0)<\epsilon \text{ for all } n\geqslant 0\}$$

$$W^u(x_0)=\{x\in M:d(f^{-n}(x),f^{-n}(x_0))\rightarrow 0 \text{ as } n\rightarrow\infty\}$$

The theory of these objects is easily deduced by applying the theory of stable manifolds to f^{-1}.

12. Local product structure and local maximality

Let Λ be a hyperbolic set, ϵ a very small positive number, and let x,y be two points of Λ whose separation is much smaller than ϵ. We want to argue that, in this situation, $W^s_\epsilon(x)$ and $W^u_\epsilon(y)$ intersect in a single point. This is easy to see if x and y are the same; $W^s_\epsilon(x)$ and $W^u_\epsilon(y)$ are submanifolds of complementary dimension (i.e., whose dimensions add up to the dimension of the ambient space M) which intersect "at non-zero angle" at x because their tangent spaces there ($E^s(x)$ and $E^u(x)$, respectively) have only the zero vector in common. Thus, the intersection at x is isolated, and, if ϵ is small enough, the manifolds cannot bend enough to cross again. To deal with the case of y near but not equal to x, one uses an argument similar to one used in the proof of the Inverse Function Theorem (together with the continuity of the dependence of $W^u_\epsilon(y)$ on y).

The above argument is a sketch of a proof of the following result:

Proposition 12.1.
1. There exists $\eta>0$ such that, for any pair of points $x,y\in\Lambda$ $W^s_\eta(x)$ and $W^u_\eta(y)$ have at most one point in common.

2. Given η as in 1., there exists $\epsilon>0$ such that, for any pair of points x,y with $d(x,y)<\epsilon$, $W^s_\eta(x)$ and $W^u_\eta(y)$ do intersect; the point of intersection varies continuously with x,y.

We will write $[x,y]$ for the point of intersection of $W^s_\eta(x)$ and $W^u_\eta(y)$, with the understanding that it is defined only for pairs x,y with $d(x,y)<\epsilon$, where ϵ,η are as in the proposition. For a general pair of nearby points x,y in a general hyperbolic set Λ, $[x,y]$ may or may not be in Λ. We say that Λ *has local product structure* if $[x,y]\in\Lambda$ for every pair of points x,y with $d(x,y)$ sufficiently small.

The reason for this terminology is as follows: Suppose Λ has local product structure; pick $x_0\in\Lambda$; let $\epsilon>0$ be sufficiently small; and put

$$X=W^u_\epsilon(x_0)\cap\Lambda,$$

$$Y=W^s_\epsilon(x_0)\cap\Lambda.$$

Then the mapping

$$(x,y)\to[x,y]$$

maps $X\times Y$ homeomorphically onto a neighborhood of x_0 in Λ. Thus, in a neighborhood of each of its points, Λ admits local coordinates which map it into the product of a piece of stable manifold with a piece of unstable manifold.

We now turn to some apparently unrelated concepts. If f is any homeomorphism of a topological space to itself, and if Λ is a set invariant for f, we will say that Λ is a *locally maximal* invariant set if there is an open set U containing Λ such that there is no invariant set properly containing Λ and contained in U. The relation between U and Λ in this case is easily seen to be equivalent to the statement that any f-orbit which remains in U for all time must be in Λ, or that

$$\Lambda=\bigcap_{n=-\infty}^{\infty} f^{-n}U$$

i.e., that U is an *isolating neighborhood* for Λ.

At this point, the notion of local maximality may look very restrictive. It is, however, not vacuous; for example, a set consisting of a single hyperbolic fixed point is locally maximal. We note one immediate consequence of local maximality: If Λ is a locally maximal hyperbolic set, then the Shadowing Theorem can be sharpened to say that the shadowing orbit $(f^n(y_0))$ is actually *in* Λ if ϵ is small enough. This follows because, by definition, the shadowing orbit stays near Λ for all time, and an orbit which stays near a locally maximal set must be in that set.

The notion of local maximality connects in a very neat way with local product structure:

Theorem 12.2. *A hyperbolic set is locally maximal if and only if it has local product structure.*

It is almost immediate that a locally maximal hyperbolic set has local product structure. We will not prove the (more difficult) converse statement; for a proof, see Proposition 8.20 of Shub [2]. The theorem is surprising in the following respect: Local product structure (for a hyperbolic set) is a purely intrinsic property of the set and the action of the mapping on the set. To make this more evident, note that local product structure for a hyperbolic set Λ is equivalent to the following:

For each $\eta>0$, there is an $\epsilon>0$ such that, for any two points $x,y\in\Lambda$ with $d(x,y)<\epsilon$, there is a point $z\in\Lambda$ with

$$d(f^n(z),f^n(x))<\eta \text{ for all } n\geqslant 0$$

$$d(f^{-n}(z),f^{-n}(y))<\eta \text{ for all } n\geqslant 0.)$$

(Here, d can be any metric, not necessarily smooth or even Riemannian, inducing the topology of Λ.) Local maximality, on the other hand, is not an intrinsic property; it depends on how the set is imbedded in the ambient space. There is no real paradox, however; the property of hyperbolicity, necessary to deduce local maximality from local product structure, is a strong restriction on how the mapping acts on the ambient space in the vicinity of Λ.

13. Recurrence

In analyzing dynamical systems, it is useful to distinguish between orbits which are *transient*, i.e., which go away and never come back, and those which are in some sense recurrent. Unfortunately, there are a number of inequivalent ways of making this distinction precise, with different advantages and disadvantages. The elementary notion of recurrence—i.e., a point x is recurrent if it is an accumulation point of its forward orbit—does not seems to be very useful. We will discuss here three less obvious ways of making the distinction between recurrence and transience. For the purposes of this section, we will assume that that the space M on which our mapping f acts is compact (although this is inconsistent with our standing convention that M is an open subset of Euclidean space.)

1. Let x be a point of M. We define the ω *limit set* of x (written $\omega(x;f)$) to be the set of all accumulation points of the forward orbit of x, and we denote by $\tilde{\Omega}(f)$ the closure of the union over $x\in M$ of $\omega(x;f)$. It is easy to see that all periodic points of f are in $\tilde{\Omega}$ and that every forward orbit converges to $\tilde{\Omega}$. To say that x is in $\tilde{\Omega}$ is to say that the orbit of x is recurrent in a relatively strong sense, and hence to say that x is not in $\tilde{\Omega}$ is to say that the orbit of x is transient in a relatively weak sense.

2. A point $x\in M$ is *wandering* if there is an open set U containing x such that

$$f^{-n}U\cap U=\varnothing \text{ for all } n>0,$$

and a *non-wandering* point is one which is not wandering. We will denote by $\Omega(f)$ the set of all non-wandering points for f. By definition, the set of wandering points is open, so $\Omega(f)$ is closed. It is easy to see that

$$\tilde{\Omega}(f) \subset \Omega(f)$$

3. We say that a point x is *chain recurrent* if for every $\delta > 0$ there is a (finite) δ pseudo-orbit starting and ending at x, or, equivalently, if x lies on a periodic δ pseudo-orbit. We let $R(f)$ denote the set of all chain-recurrent points for f. It is straightforward to prove that $R(f)$ is closed and contains $\Omega(f)$. To say that a point x is *not* chain recurrent is to say that x is transient in the strong sense that the orbit of x and the orbits of all points sufficiently near to it go away and never return, not only for the motion induced by the mapping f but even for motions obtained by small (possibly non-deterministic) perturbations on f.

The notion of chain recurrence is more subtle than might at first appear. Notably, we have

Proposition 13.1 *If $x \in R(f)$, then for each $\delta > 0$ there is a δ pseudo-orbit in $R(f)$ (not just in the ambient space M) starting and ending at x.*

This proposition can be reformulated as the equality

$$R(f|_{R(f)}) = R(f).$$

The corresponding equality for the non-wandering set does not hold in general.

Proof. We need some notation. For $\delta > 0$ and $x \in M$, we write $R_\delta(x)$ for the set of points y such that there exist δ pseudo-orbits from x to y and from y to x. We also write $R(x;f)$ for the set of points y such that, for each $\delta > 0$, there exist δ pseudo-orbits from x to y and from y to x. It is immediate that x is chain recurrent if and only if $x \in R(x;f)$ and that every point of $R(x;f)$ is chain recurrent. It is not difficult to see that, for any $\delta > 0$ and any $x \in M$, $R_\delta(x)$ is open (Recall that a δ pseudo-orbit is defined by the condition that each $d(x_{n+1}, f(x_n))$ is *strictly* less than δ.) and that each $R(x;f)$ is closed.

Lemma 13.2. *As δ decreases to 0, $R_\delta(x)$ decreases to $R(x;f)$, i.e., if U is any open set containing $R(x;f)$, then $R_\delta(x) \subset U$ for all sufficiently small δ.*

Proof. Suppose not. Then there is an open set $U \supset R(x;f)$, a sequence (δ_n) decreasing to zero, and a sequence

$$y_n \in R_{\delta_n}(x) \setminus U.$$

By passing to a subsequence if necessary we can assume that the y_n converge to a limit y, and, since U is assumed to be open, y is not in U. We are going to argue that $y \in R(x;f)$; this will contradict $U \supset R(x;f)$ and thus establish the lemma.

What we have to do is to show that, for any $\delta > 0$, there are δ pseudo-orbits from x to y and from y to x. We will show how to construct the second of these; the construction of the first is similar but slightly easier. Take $\epsilon > 0$ small enough so that $d(f(z_1), f(z_2)) < \delta/2$ whenever $d(z_1, z_2) < \epsilon$; then take n large enough so that $d(y_n, y) < \epsilon$ and also so that $\delta_n < \delta/2$. Then if

$$z_0 = y_n, z_1, \ldots, z_j = x$$

is a δ_n pseudo-orbit from y_n to x,

$$y, z_1, \ldots, z_j$$

is a $\delta_n + \delta/2 < \delta$ pseudo-orbit from y to x.

We can now complete the proof of Proposition 13.1. The idea is as follows: Given $x \in R(f)$ and $\delta > 0$, we want to show that there is a δ pseudo-orbit in $R(f)$ starting and ending at x. Since $x \in R(f)$, there is such a pseudo-orbit in M, and, by the definition of $R_\delta(x)$, this pseudo-orbit lies in fact in $R_\delta(x)$. But Lemma 13.2 shows that, for δ small, $R_\delta(x)$ is not much larger than $R(x;f) \subset R(x)$ so the pseudo-orbit can be moved into $R(x)$ by a small perturbation.

To fill in the details, we first choose $\delta_1 < \delta/3$ so that

$$d(z_1,z_2) < \delta_1 \text{ implies } d(f(z_1),f(z_2)) < \delta/3.$$

Then, using Lemma 13.2, choose $\delta_2 < \delta/3$ so that

$$y \in R_{\delta_2}(x) \text{ implies } d(y,R(x;f)) < \delta_1.$$

Let

$$y'_0 = x, y'_1, \ldots, y'_n = x$$

be a δ_2 pseudo-orbit starting and ending at x. For each $j = 1, \ldots, n-1$, choose $y_j \in R(x)$ such that $d(y_j, y'_j) < \delta_1$, and put $y_0 = y_n = x$. With these choices,

$$\begin{aligned} d(y_{j+1}, f(y_j)) &\leqslant d(y_{j+1}, y'_{j+1}) + d(y'_{j+1}, f(y'_j)) + d(f(y'_j), f(y_j)) \\ &< \delta_1 + \delta_2 + \delta/3 < \delta/3 + \delta/3 + \delta/3 = \delta, \end{aligned}$$

so $y_0, \ldots, y_n$ is the desired δ pseudo-orbit in $R(x)$ starting and ending at x.

14. Global stability.

Up to now, we have studied hyperbolic sets semi-locally, i.e., without worrying about what is happening elsewhere in the state space M. We now take up the more global question: What can we say about a mapping f if we know that the set of all its recurrent points is hyperbolic? As we indicated in the last section, there are several alternative notions of recurrence and hence several variants of this question. In 1967 Smale formulated a condition on a diffeomorphism f, which he called *Axiom A*, the main content of which was the requirement that the non-wandering set of f is hyperbolic. He also found it necessary to require separately that the periodic points for f are dense in the non-wandering set. Even with this condition added, Axiom A was not strong enough to imply any reasonable kind of global stability for f; something further, called the *no-cycles condition*, was needed. It was observed after the fact that the two conditions in Axiom A and the no-cycles condition were, all taken together, equivalent to the single condition that the chain recurrent set $R(f)$ is hyperbolic. We will sketch here some of the consequences of this condition.

Proposition 14.1. *If $R(f)$ is hyperbolic, then periodic points are dense in $R(f)$*

Proof. By Proposition 13.1, each point $x \in R(f)$ lies on a periodic δ pseudo-orbit in $R(f)$ with δ as small as we like. Hence, by the Shadowing Theorem, there is a periodic orbit as close as we like to x. This periodic orbit is in $R(f)$ because *all* periodic points are chain recurrent.

Proposition 14.2. *If the closure of the set of periodic points of f is a hyperbolic set, it has local product structure.*

We make only a few remarks about the proof. The first idea is that, to prove local product structure, it is only necessary (by continuity) to show that $[x,y]$ is a limit of periodic points *for x,y themselves periodic points with d(x,y) small.* This latter statement is proved by a variant of the argument given in Section 5 to show that a transverse homoclinic point is a limit of periodic points. The construction of periodic pseudo-orbits passing through $[x,y]$ is done using that fact that, not only do $W^s_\eta(x)$ with $W^u_\eta(y)$ intersect (at $[x,y]$), but $W^u_\eta(x)$ with $W^s_\eta(y)$ also intersect (at $[y,x]$). The pseudo-orbits go from $[x,y]$ to x along $W^s_\eta(x)$, then from x to $[y,x]$ along $W^u_\eta(x)$, then from $[y,x]$ to y along $W^s_\eta(y)$, then back to $[x,y]$ along $W^u_\eta(y)$.

Putting together the above propositions, we see that, if $R(f)$ is hyperbolic, it has local product structure.

To close these notes we cite one of the high points of this circle of ideas:

Theorem 14.3. Ω-Stability Theorem. *Let f be a diffeomorphism of the compact manifold M such that $R(f)$ is hyperbolic. Then for any $\tilde{f}$ sufficiently close to f in the C^1 topology:*
1. $R(\tilde{f})$ is a hyperbolic set for $\tilde{f}$.
2. $R(\tilde{f})$ is close to $R(f)$.
3. The restriction of $\tilde{f}$ to $R(\tilde{f})$ is topologically conjugate to the restriction of f to $R(f)$.

Most of the elements of the proof are contained in the results we established in Sections 8 and 10 about structural stabiliy of hyperbolic sets and stability of hyperbolicity. To complete the argument, it is necessary to show that $R(\tilde{f})$ cannot be much larger than $R(f)$. This requires methods which we have not developed and for which we refer again to Shub's monograph.

References.

1. O.E. Lanford. Introduction to the mathematical theory of dynamical systems. To appear in the proceedings of the 1981 Les Houches school on Chaotic Behavior of Dynamical Systems, edited by R. Helleman and G. Iooss, to be published by North Holland.
2. M. Shub. Stabilité globale des systèmes dynamiques. Astérisque 56 (1978).