Ergod. Th. & Dynam. Sys. (2004), **24**, 1725–1738
DOI: 10.1017/S0143385703000749
© *2004 Cambridge University Press*
Printed in the United Kingdom

New phenomena associated with homoclinic tangencies

SHELDON E. NEWHOUSE
Mathematics Department, Michigan State University, E. Lansing, MI 48824-1027, USA
(e-mail: sen@math.msu.edu)

(*Received* 21 *February* 2003 *and accepted in revised form* 15 *April* 2003)

In memory of Michael Herman

Abstract. We survey some recently obtained generic consequences of the existence of homoclinic tangencies in diffeomorphisms of surfaces. Among other things, it has been shown that they give rise to invariant topologically transitive sets with maximal Hausdorff dimension, that they prohibit the existence of various kinds of symbolic extensions and that they form an impediment to the existence of Sinai–Ruelle–Bowen (SRB) measures. The main new result described here, together with a positive answer to an as yet unproved conjecture of Palis, would prove that generically on surfaces, SRB measures only exist on uniformly hyperbolic attractors.

1. *Introduction*

The study of homoclinic motions forms an important part of the modern theory of dynamical systems. Such motions were first described by Poincaré in his geometric studies of the restricted three-body problem and, since that time, they have been studied by many authors. Transverse homoclinic motions produce complicated invariant sets which can be completely described in terms of certain symbolic systems called subshifts of finite type. Moreover, the dynamical properties of these invariant sets persist under small perturbations of the underlying dynamical system. However, homoclinic tangencies give rise to rich and varied dynamics under small perturbations. In this paper we will survey some of the phenomena associated to homoclinic tangencies. After reviewing some definitions and well-known facts, we describe some recent developments, and present a new result (Theorem 1.4) which shows that generically homoclinic tangencies form an impediment to the existence of Sinai–Ruelle–Bowen (SRB) measures.

For historical information and various interesting phenomena associated with homoclinic motions, we refer to the books by Palis and Takens [**29**] and Moser [**23**]. For general information on dynamical systems related to the concepts we study here, we refer to the books [**34**], [**13**] and [**26**].

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

While some of our discussion could be extended to systems of arbitrary finite dimension, the strongest results occur in two-dimensional systems. Hence, we will restrict ourselves to dimension two.

Let M be a compact C^∞ two-dimensional Riemannian manifold and, for $r \geq 1$, let $\mathcal{D}^r(M)$ be the space of C^r diffeomorphisms of M with the uniform C^r topology. It is known that $\mathcal{D}^r(M)$ is a Baire space: countable intersections of dense open sets are dense. For $f \in \mathcal{D}^r(M)$, a compact f-invariant set Λ is a *uniformly hyperbolic* set if there are constants $C > 0, \lambda > 1$ such that for each $x \in \Lambda$ there is a splitting $T_x M = E^s_x \oplus E^u_x$ such that, for $n \geq 0$,

$$\max(|Df^n_x \mid E^s_x|, |Df^{-n}_x \mid E^u_x|) \leq C\lambda^{-n}. \tag{1}$$

Here we let $|S \mid E|$ denote the norm of a linear map S restricted to a linear subspace E of the domain of S.

Let d denote the topological metric (distance function) on M induced by the Riemannian metric. It is well known that if the conditions in (1) hold, then the subspaces E^s_x, E^u_x are unique and depend continuously on $x \in \Lambda$. The *invariant manifold theorem* [**14**] guarantees that there are C^r injectively immersed manifolds $W^s(x)$, $W^u(x)$ tangent at x to E^s_x, E^u_x, respectively, which are defined by

$$W^s(x) = \{y \in M : d(f^n(y), f^n(x)) \to 0 \text{ as } n \to \infty\},$$

and

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

A uniformly hyperbolic set Λ is called a *hyperbolic basic set* if it has a dense orbit and there is a neighborhood U of Λ in M, such that $\bigcap_{n\in\mathbb{Z}} f^n(U) = \Lambda$. In this case, one calls the neighborhood U an *isolating* neighborhood of Λ.

If $\Lambda(f)$ is a hyperbolic basic set for f with an isolating neighborhood U, then there is a neighborhood $\mathcal{N}$ of f in $\mathcal{D}^r(M)$ such that if $g \in \mathcal{N}$, then $\Lambda(g) \stackrel{\text{def}}{=} \bigcap_{n\in\mathbb{Z}} g^n(U)$ is a hyperbolic basic set for g and there is a homeomorphism $h : \Lambda(g) \to \Lambda(f)$ such that $fh = hg$. It is common to call the set $\Lambda(g)$ the *continuation* of $\Lambda(f)$.

A point q is a *homoclinic point* of f if there is a hyperbolic basic set Λ and points $x \in \Lambda$, $y \in \Lambda$ such that

$$q \in (W^u(x) \setminus \Lambda) \bigcap (W^s(y) \setminus \Lambda).$$

Since we are dealing with two-dimensional manifolds, such a q can only exist if the sets $W^u(x)$, $W^s(y)$ are one-dimensional manifolds; i.e. C^r curves. The homoclinic point q is called a *transverse homoclinic point* if the curves $W^u(x)$, $W^s(y)$ are not tangent at q. Otherwise, we call q a *homoclinic tangency*.

Let $\mathcal{B}$ denote the collection of hyperbolic basic sets of a C^r diffeomorphism. There is an equivalence relation $\sim$ on $\mathcal{B}$ defined as follows. We say $\Lambda_1 \sim \Lambda_2$ if either $\Lambda_1 = \Lambda_2$ or there are points $x, y \in \Lambda_1$ and $x_1, y_1 \in \Lambda_2$ such that $W^u(x) \setminus \Lambda_1$ has a non-empty transverse intersection with $W^s(x_1) \setminus \Lambda_2$ and $W^s(y) \setminus \Lambda_1$ has a non-empty transverse intersection with $W^u(y_1) \setminus \Lambda_2$. This relation was considered in [**26**] for periodic points. If $\Lambda_1 \sim \Lambda_2$, we will say that Λ_1 is *homoclinically related* (or h-related) to Λ_2. We call an equivalence class a *homoclinic class* (or h-class). The closure of an h-class will be

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

called an h-closure. An h-closure containing more than a single periodic orbit will be called a *homoclinic set*. Let Λ be a homoclinic set which is the closure of the union of the elements of a particular h-class $\mathcal{C}$. We call Λ the *homoclinic set of* $\mathcal{C}$. If p is a periodic point whose orbit is in $\mathcal{C}$, we will say that p is *associated* to Λ and that Λ is the homoclinic set of p. We also say that Λ is the homoclinic closure of p. If a homoclinic set Λ contains a hyperbolic basic set Λ_1 with a homoclinic tangency, we will say that Λ contains a homoclinic tangency.

The following result is proved in [**26**].

- Every homoclinic set is an uncountable topologically transitive set containing a dense set of associated periodic points. Moreover, it coincides with the closure of the transverse homoclinic points of the orbits of any of its associated periodic points.

We remark that homoclinic sets have a lower semicontinuity property in the following sense. Let $\Lambda(f)$ be a homoclinic set for f and let $p = p(f)$ be an associated periodic point of minimal period. If g is C^r close to f, then the continuation $p(g)$ of p has an h-closure $\Lambda(g)$ which we call the *continuation* of $\Lambda(f)$. This is independent of the choice of p with minimal period provided that g is close enough to f.

- The map $g \to \Lambda(g)$ is lower semicontinuous as a map from a neighborhood of f in $\mathcal{D}^r(M)$ into the collection of compact subsets of M with the Hausdorff metric.

One can think of homoclinic sets as the basic 'building blocks' of systems with chaotic motion. We are interested in detailed information about homoclinic sets and the phenomena which are induced from them when the ambient diffeomorphism f varies with external parameters. We will see that there are several interesting phenomena which appear after small perturbations when a given map f has a homoclinic tangency. First, we recall a general situation in which there is a whole open set in $\mathcal{D}^r(M)$ whose elements have no homoclinic tangencies. An ϵ-chain is a sequence $x_0, x_1, \ldots, x_n$ in M such that $d(fx_i, x_{i+1}) < \epsilon$ for $0 \leq i < n$. A periodic ϵ-chain is an ϵ-chain $x_0, x_1, \ldots, x_n$ such that $x_0 = x_n$. A point x is *chain recurrent* if, for every $\epsilon > 0$, there is a periodic ϵ-chain passing through x. The collection of chain recurrent points is a non-empty, compact, f-invariant set called the *chain recurrent set* and is denoted $\mathcal{R}(f)$. We say that the diffeomorphism f is *hyperbolic* if $\mathcal{R}(f)$ is a hyperbolic set. It is well known that a diffeomorphism f is hyperbolic if and only if it satisfies Smale's axiom A and no cycle properties in which case the chain recurrent set equals the non-wandering set. Hyperbolic diffeomorphisms are *chain stable*. This means that there is a neighborhood $\mathcal{N}$ of f in $\mathcal{D}^r(M)$ such that if $g \in \mathcal{N}$, then there is a homeomorphism $h : \mathcal{R}(g) \to \mathcal{R}(f)$ such that $fh = hg$. A well-known result due to Palis [**27**] following work by Mane [**20**] states that, in $\mathcal{D}^1(M)$, hyperbolicity is equivalent to chain stability. A diffeomorphism is called *Anosov* if the whole manifold M is a hyperbolic set. In dimension two, M must be a torus and $\mathcal{R}(f) = M$.

It is easy to see that the existence of a homoclinic tangency is an obstruction to hyperbolicity. Indeed, it can be shown that such a tangency is in the chain recurrent set and there can be no splitting as required for hyperbolicity.

We say that f has *persistent homoclinic tangencies* if there is a hyperbolic set $\Lambda(f)$ with a homoclinic tangency and there is a neighborhood $\mathcal{N}(f)$ such that if $g \in \mathcal{N}(f)$, then the continuation $\Lambda(g)$ of $\Lambda(f)$ is defined and also has some homoclinic tangency.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

Thus, diffeomorphisms with persistent homoclinic tangencies have neighborhoods none of whose elements are hyperbolic.

The following result [**25**] shows that on surfaces any diffeomorphism with a homoclinic tangency can be perturbed to create persistent homoclinic tangencies. Palis and Takens [**29**] have given a nice proof of this result.

THEOREM 1.1. *Let M be a compact C^∞ manifold, let $2 \leq r \leq \infty$ and let $f \in \mathcal{D}^r(M)$ have a homoclinic tangency. Then, given any neighborhood $\mathcal{N}$ of f in $\mathcal{D}^r(M)$, there is a $g \in \mathcal{N}$ which has persistent homoclinic tangencies.*

Homoclinic tangencies frequently give rise to infinitely many sinks. In fact, C^r generically with $r \geq 2$, a homoclinic set is in the closure of the periodic sinks provided that it contains a tangency and a dissipative periodic orbit. A periodic point p with $f^\tau(p) = p$ is called dissipative if $|\det Df^\tau(p)| < 1$.

It is known that if f is a hyperbolic diffeomorphism which is not Anosov, then the Hausdorff dimension of the chain recurrent set is strictly less than two. The next theorem, which was recently obtained in [**10**], shows that C^r generically with $r \geq 2$, each homoclinic set which contains a homoclinic tangency has maximal Hausdorff dimension. Recall that a subset $\mathcal{B}$ of a topological space is called *residual* if it contains a countable intersection of open dense sets. In $\mathcal{D}^r(M)$ such sets are, of course, dense.

THEOREM 1.2. *Let $2 \leq r \leq \infty$. There is a residual subset $\mathcal{B}$ of $\mathcal{D}^r(M)$ such that if $f \in \mathcal{B}$, and Λ is a hyperbolic basic set for f with a homoclinic tangency, then the homoclinic set of Λ has Hausdorff dimension two.*

We sketch the idea of the proof of Theorem 1.2 referring the reader to [**10**] for the details. The idea of the proof comes from that of a similar result for C^1 area-preserving diffeomorphisms which goes back to [**24**]. A given f with a homoclinic tangency is C^r perturbed to get persistent homoclinic tangencies. Then, one makes use of a result by Gonshenko *et al* to make a further C^r perturbation to obtain a periodic point p whose stable and unstable manifolds have a whole interval of homoclinic tangencies. Having this, a further perturbation gives special zero-dimensional hyperbolic basic sets Λ_1, Λ_2 homoclinically related to p such that the unstable Hausdorff dimension of Λ_1 is close to one and the stable Hausdorff dimension of Λ_2 is close to one. Since Λ_1 is homoclinically related to Λ_2, Lemma 8 in [**25**] gives another hyperbolic basic set Λ in the same h-class which contains them both. The set Λ will have Hausdorff dimension close to two. This proves that, given $n > 0$, there is a dense set $\mathcal{B}_n$ of diffeomorphisms with homoclinic tangencies having homoclinic sets of Hausdorff dimension greater than $2 - (1/n)$. This set is also open since the Hausdorff dimension of hyperbolic basic sets varies continuously [**30**]. Theorem 1.2 follows taking the intersection $\mathcal{B} = \bigcap_n \mathcal{B}_n$.

During the past 15 years or so, there has been much interest in the study of so-called SRB measures in area-decreasing planar diffeomorphisms. An f-invariant probability measure ν of a diffeomorphism will be called an SRB measure if it is ergodic, has compact support and has absolutely conditional measures on unstable manifolds.

A family $(a, b) \to f_{a,b}$ of diffeomorphisms which is near the Henon family $H_{a,b}(x, y) = (1 + y - ax^2, bx)$ is called a *Henon-like* family. Diffeomorphisms $f_{a,b}$

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

in Henon-like families are called *Henon-like* diffeomorphisms. It is a consequence of the work of many authors (e.g. [**2, 3, 22, 35**]) that for Henon-like families with $|b|$ small there is a positive Lebesgue measure set of parameters $A(b)$ such that $f_{a,b}$ has an SRB measure for $a \in A(b)$. Thus, such measures exist with positive probability in parameters in Henon-like families. At the present time, all known SRB measures in Henon-like diffeomorphisms are supported on homoclinic sets with tangencies.

To be more precise, in a Henon-like family $f_{a,b}$, let us call a pair of parameters (a, b) *good* if $f_{a,b}$ has a hyperbolic periodic saddle point $p_{a,b}$ and an SRB measure $\nu_{a,b}$ supported on the homoclinic set of $p_{a,b}$. All presently known good pairs (a_0, b_0) have the following property. There is a sequence $a_1, a_2, \ldots$ of real parameters converging to a_0 such that each f_{a_i,b_0} contains persistent homoclinic tangencies. It may be that the techniques in the proof of Theorem 1.4 given later can be extended to show that, for any good pair (a_0, b_0), one can find a sequence $a_1', a_2', \ldots$ converging to a_0 such that there is no SRB measure supported on the homoclinic set of p_{a_i',b_0}. This prompts us to make the following conjecture.

CONJECTURE 1.3. *Let $H_{a,b}$ be the Henon family of planar maps. Then, for each parameter b, there is a residual set of parameters a such that $H_{a,b}$ has no SRB measure.*

For $b = 0$, this conjecture is true since the maps essentially reduce to the familiar logistic family $a \to f_a(x) = ax(1 - x)$ and one may apply the well-known results of Graczyk and Swiatek [**12**] and Ljubich [**19**]. These imply that there is a dense open set of parameters a such that f_a has an attracting periodic orbit whose basin contains a set of full measure. For $|b| = 1$, the map $H_{a,b}$ is area preserving and the conjecture implies that, for a residual set of parameters a, one has that $H_{a,b}$ has zero measure-theoretic entropy on any invariant set of finite area. Indeed, since the non-wandering set of $H_{a,b}$ is compact, any invariant probability measure has compact support. Further, using the Pesin entropy formula and results from Ledrappier [**17**] (or, more generally, those in [**18**]), any ergodic component of normalized Lebesgue measure is an SRB measure.

We us proceed to state the main new result in the present paper.

If ν is an invariant probability measure for f, we say that a subset Λ *carries* the measure ν if $\nu(\Lambda) = 1$.

THEOREM 1.4. *There is a residual subset $\mathcal{U}$ in $\mathcal{D}^r(M)$ for $2 \leq r \leq \infty$ such that if $f \in \mathcal{U}$ and Λ is a homoclinic set for f which contains a tangency and has an associated dissipative periodic point, then Λ does not carry an SRB measure.*

Theorem 1.4 has interesting relations to a conjecture by Palis [**28**]. This conjecture states that there is a dense subset $\mathcal{A}$ in the space $\mathcal{D}^r(M)$ of C^r diffeomorphisms of a compact surface M such that if $f \in \mathcal{A}$, then either f is hyperbolic or f has a homoclinic tangency. For $r = 1$, this conjecture has recently been proved by Pujals and Samborino [**32**]. It is still unproved for $r \geq 2$. It follows from Theorem 1.1 that if the conjecture is true for $r \geq 2$, then the set $\mathcal{A}$ can actually be chosen to be dense and open and the homoclinic tangency occurs in some homoclinic set.

If one wants information about all homoclinic sets, one has to pass from dense and open to residual. Thus, we state the following conjecture.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

CONJECTURE 1.5. (Weak Palis homoclinic conjecture) *Let $r \geq 1$. There is a residual subset $\mathcal{A}$ in $\mathcal{D}^r(M)$ such that, for $f \in \mathcal{A}$, each homoclinic set for f is either uniformly hyperbolic or has a homoclinic tangency.*

In view of Theorem 1.4, we have the following statement.

> The weak Palis homoclinic conjecture implies that there is a residual subset $\mathcal{A}$ of $\mathcal{D}^r(M)$ such that if $f \in \mathcal{A}$, then any SRB measure whose support contains a dissipative periodic point must be supported on a *uniformly hyperbolic attractor*.

This conjecture would imply that, while SRB measures might occur Lebesgue typically in non-hyperbolic systems, they would be exceptional from the topological point of view.

In passing, it is worth noting that there are several interesting recent papers dealing with the C^1 generic properties of diffeomorphisms. See [**1**], [**4**], and [**5**].

Next, we wish to survey some connections between homoclinic tangencies and the existence of symbolic extensions.

Let $\mathbb{Z}$ denote the set of integers, $N \geq 2$ be a positive integer, $J = \{1, \ldots, N\}$ and let $\Sigma_N = J^{\mathbb{Z}}$ be the set of doubly infinite sequences of symbols in the alphabet J. Elements $\mathbf{a}$ in Σ_N are denoted as follows:

$$\mathbf{a} = (a_i), \quad i \in \mathbb{Z}.$$

We give Σ_N the standard metric

$$d(\mathbf{a}, \mathbf{b}) = \sum_{i \in \mathbb{Z}} \frac{|a_i - b_i|}{2^{|i|}},$$

making (Σ_N, d) into a compact metric space. The shift automorphism is the map $\sigma = \sigma_N : \Sigma_N \to \Sigma_N$ defined by

$$\sigma_N(\mathbf{a})_i = \mathbf{a}_{i+1} \quad \text{for } i \in \mathbb{Z}.$$

A *subshift* or *symbolic system* is a pair (S, X) where X is a closed σ_N-invariant subset of σ_N and S is the restriction of σ_N to X for some N.

Subshifts (S, X) are known to have many special dynamical properties. Let us list some of these.

(1) They are expansive. That is, there is a positive real number $\epsilon > 0$ such that if $x \neq y$ in X, then there is an integer n such that $d(S^n x, S^n y) > \epsilon$. This implies, for instance, that if ν is an S-invariant probability measure, then its metric entropy can be computed as the mean entropy $h_\nu(\alpha)$ for any finite partition α whose elements have diameter less than ϵ.

(2) The metric entropy function $\nu \to h_\nu(S)$ is upper semicontinuous as a function of ν. Hence, any subshift has measures of maximal entropy.

(3) The topological entropy $h_{\text{top}}(S)$ can be computed as

$$h_{\text{top}}(S) = \lim_{n \to \infty} \frac{1}{n} \log \, card \, B_n$$

where $card \, B_n$ is the number of n-blocks appearing in X.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

If a given system (f, M) could be 'modelled' in some sense by a subshift, then one would have much information about the dynamics of f. The most useful notion of 'modelling' is that of topological conjugacy. However, since subshift spaces have topological dimension zero, most systems cannot be topologically conjugate to subshifts. Thus, one would like to weaken the notion of topologically conjugacy somewhat to use symbolic systems as models.

Let $f : X \to X$ and $g : Y \to Y$ be homeomorphisms of the compact metric spaces X and Y, respectively. We say that (g, Y) is an extension of (f, X) if there is a continuous surjection $\pi : Y \to X$ such that $\pi g = f\pi$. In this case, we call the triple (g, Y, π) an extension triple of (f, X).

It is natural to ask when a given system (f, X) has a symbolic extension, i.e. when can we find an extension (g, Y) which is a symbolic system. An obvious necessary condition is that $h_{\text{top}}(f) < \infty$. Around 1988, Auslander asked the converse question:

> *Does every homeomorphism of a compact metric space with finite topological entropy have a symbolic extension?*

This question turned out to be highly non-trivial. In 1990, Boyle found a counter-example using an inverse limit construction which led to a certain zero-dimensional system with no symbolic extension.

One can ask for more precise information regarding the kinds of extensions a system may or may not have. One says that an extension triple (g, Y, π) is a *principal extension* of (f, X) if the extension map π simultaneously preserves entropies of all invariant probability measures. That is, for every g-invariant probability measure ν, we have $h_{\pi_\star \nu}(f) = h_\nu(g)$ (here, $\pi_\star \nu = \nu \circ \pi^{-1}$). In the sense of information theory, a principal extension of a given system is indistinguishable from that system.

Around the time of the discovery of Boyle's example mentioned earlier, several authors proceeded to study the existence of symbolic and principal symbolic extensions. Downarowicz [**9**] gave a characterization of finite entropy zero-dimensional systems (f, X) which have principal symbolic extensions in terms of certain upper-semicontinuous functions on the space of invariant probability measures. It turned out that a necessary and sufficient condition was that f be *asymptotically h-expansive* in the sense of Misiurewicz [**21**]. Later, Boyle *et al* [**7**] extended this to the general case, i.e. a finite entropy system has a principal symbolic extension if and only if it is asymptotically h-expansive. However, Buzzi [**8**] has proved that every C^∞ map is asymptotically h-expansive. So, these results together yield the following striking result.

THEOREM 1.6. *Every C^∞ diffeomorphism of a compact manifold has a principal symbolic extension.*

It is natural to ask about systems with a finite amount of smoothness. Since every C^1 self-map of a compact manifold has finite topological entropy, the question simply becomes when does a C^r map on a compact manifold with $1 \leq r < \infty$ have a symbolic or principal symbolic extension. In this connection, recently Downarowicz and the author [**10**] have proved the following results.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

THEOREM 1.7.

(1) *Given any compact C^∞ manifold M, there is a C^1 diffeomorphism f on M with no symbolic extension at all.*
(2) *There is a residual subset $\mathcal{A}$ in the space of C^1 symplectic diffeomorphisms on a compact orientable two-dimensional manifold such that if $f \in \mathcal{A}$, then either f is Anosov or f has no symbolic extension at all.*
(3) *For any $2 \leq r < \infty$ and any compact two-dimensional manifold M^2, there exists a residual subset $\mathcal{A}$ of an open subset $\mathcal{U} \in \mathcal{D}^r(M^2)$ such that if $f \in \mathcal{A}$, then f has no principal symbolic extension.*

Regarding C^r maps with $2 \leq r < \infty$, we have the following problem.

PROBLEM. *Let $2 \leq r < \infty$. Does every C^r self-map of a compact C^∞ manifold have at least one symbolic extension?*

It turns out that intervals of homoclinic tangencies again play a fundamental role in the proof of Theorem 1.7. After making certain perturbations, one gets diffeomorphisms such that the entropy functions of invariant probability measures exhibit pathological continuity properties. Then one applies the results from Boyle and Downarowicz [**6**] concerning the existence of various types of symbolic extensions.

During our lecture at the conference, Dennis Sullivan made the following interesting observation. Note that if a system (f, X) possesses a (principal) symbolic extension, then so does any system (f_1, X_1) which is topologically conjugate to (f, X). Thus, these results have some interesting new corollaries. For instance, a generic C^1 non-Anosov area-preserving diffeomorphism on a surface is not topologically conjugate to *any* C^∞ diffeomorphism.

2. *Proof of Theorem 1.4*

We begin with a proposition which states that if a homoclinic set carries an SRB measure ν, then the support ν contains the full unstable manifold of the orbit of any associated periodic point.

PROPOSITION 2.1. *Suppose that Λ is a homoclinic set which carries an SRB measure ν. Then, the support of ν equals the closure of the orbit of the unstable manifold of any periodic point associated to Λ.*

Proof. Since ν is ergodic, Λ is f-invariant, and $\nu(\Lambda) > 0$, we have, in fact, that $\nu(\Lambda) = 1$. Also, since ν has non-zero Lyapunov exponents, there is an invariant set $\Lambda_1 \subset \Lambda$ of full ν-measure such that each $x \in \Lambda_1$ has a C^r unstable manifold $W^u(x)$. This is the set of points y such that $d(f^{-n}x, f^{-n}y)$ approaches zero exponentially fast as $n \to \infty$. Since ν has absolutely continuous conditional measures along unstable manifolds, it is known that one can choose Λ_1 so that if $x \in \Lambda_1$, then $W^u(x)$ is completely contained in the support of ν. We refer to standard places (e.g. [**11, 18, 31**]) for proofs of these facts. Thus, μ is supported on a set which contains many full unstable manifolds. We wish to show that every unstable manifold of a periodic orbit associated to Λ is also in the support of ν.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

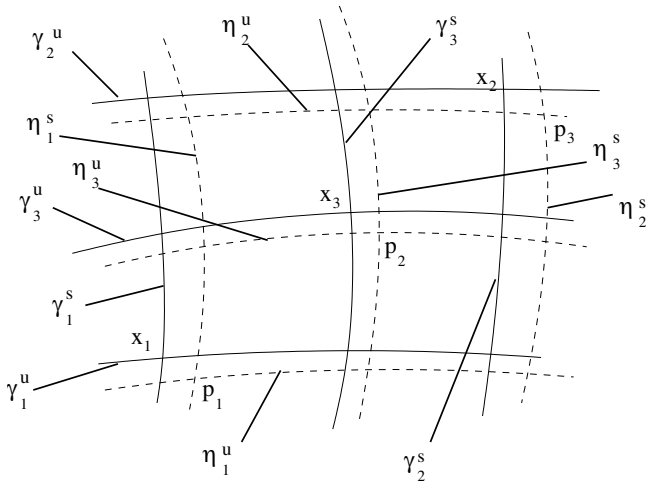


FIGURE 1. Full curves correspond to x_i, broken curves correspond to p_i.

Note that, obviously, Λ_1 contains no periodic orbits. Let p be a hyperbolic periodic point associated to Λ.

For any point x, let $O(x) = \{f^n(x) : n \in \mathbb{Z}\}$ denote the orbit of x.

Let $x \in \Lambda_1$, and let $y \in \Lambda$ be an ω-limit point of x. Thus, there are three distinct iterates $x_1 = f^{n_1}x, x_2 = f^{n_2}x, x_3 = f^{n_3}x$ near y so that:

(1) there is a small curvilinear rectangle D such that $f(D) \bigcap D = \emptyset$;

(2) the boundary of D consists of pieces $\gamma_1^u \subset W^u(x_1), \gamma_1^s \subset W^s(x_1), \gamma_2^u \subset W^u(x_2), \gamma_2^s \subset W^s(x_2)$;

(3) x_3 is contained in the interior of D; and

(4) $O(p) \bigcap D = \emptyset$.

Let us recall a result due to Katok in [**16**]. He proves that there is a sequence $\mu_1, \mu_2, \ldots$ of measures supported on hyperbolic periodic orbits which converges weakly to ν. Indeed, his proof shows that we may find three hyperbolic periodic points p_i arbitrarily close to x_i, such that there are curves $\eta_i^u \subset W^u(p_i), \eta_i^s \subset W^s(p_i)$ such that

(5) η_i^u is C^r close to γ_i^u and η_i^s is C^r close to γ_i^s for $i = 1, 2, 3$;

(6) there is a rectangle $\tilde{D}$ close to D and bounded by the curves $\eta_1^u, \eta_1^s, \eta_2^u, \eta_2^s$ such that $p_3 \in \mathit{interior}(\tilde{D})$; and

(7) $O(p) \bigcap \tilde{D} = \emptyset$.

These properties are illustrated in Figure 1.

It now follows that each $W^s(p_i)$ has some non-empty transverse intersections with $W^u(O(x))$, so the inclination lemma [**29**, Theorem 2, p. 155] implies that

$$W^u(O(p_i)) \subseteq \mathit{Closure}(W^u(O(x))) \subseteq \Lambda \quad \text{for each } i. \tag{2}$$

Furthermore, since Λ is the closure of the set of transverse homoclinic points of p, we have that $W^s(O(p))$ accumulates on x_3. Since $O(p) \bigcap \tilde{D} = \emptyset$, we may find a $q \in O(p)$ such that $W^s(q)$ meets both the interior and exterior of $\tilde{D}$. Hence, it must also cross either η_1^u or η_2^u. Assume first that it crosses η_1^u. If the intersection is

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

transverse, then the inclination lemma shows that $W^u(O(p)) \subset \mathit{Closure}(W^u(O(p_1)))$. Even if the intersection is not transverse, since $W^s(q)$ goes from one side of η_1^u to the other, we can get the same conclusion. A similar argument in case $W^s(q)$ crosses η_2^u gives $W^u(O(p)) \subset \mathit{Closure}(W^u(O(p_2)))$. Now (2) implies the required statement that $W^u(O(p)) \subset \Lambda$. □

The next lemma, which gives new information about the typical way in which sinks approach a homoclinic set, is the key new ingredient of the proof of Theorem 1.4.

Let p be a hyperbolic saddle periodic point and q be a hyperbolic sink (attracting periodic point). We say that q is u-related to p (short for *unstably related*) if

$$W^u(O(p)) \bigcap W^s(O(q)) \neq \emptyset.$$

We also say that q is u-related to a homoclinic set Λ if there is a periodic saddle point p associated to Λ such that q is u-related to p.

LEMMA 2.2. *Let $r \geq 1$, let f be a C^r diffeomorphism of a compact surface and let $p(f)$ be a dissipative hyperbolic periodic point of f whose homoclinic closure contains a tangency. Then, arbitrarily C^r close to f one can find a g such that $p(g)$ has a u-related sink.*

Proof. Observe that the orbit of a homoclinic tangency consists of homoclinic tangencies and the stable and unstable manifolds of p accumulate near the orbit of the homoclinic tangency in $\Lambda(f)$. Thus, the stable and unstable manifolds of p accumulate on some homoclinic tangency in $\Lambda(f)$. After a small perturbation, we may assume that $W^u(p)$ has a point of tangency with $W^s(p)$.

Let $\tau > 0$ be the period of p. Next, we can find a g_1 C^r near f such that:

(1) g_1 is C^∞;
(2) the continuation $p(g_1)$ is defined and hyperbolic;
(3) there is a C^2 linearization of g_1^τ near $p(g_1)$; and
(4) $W^u(p(g_1))$ and $W^s(p(g_1))$ have a quadratic tangency at a point q_1.

Replacing g_1 by g_1^τ, we may assume that $p(g_1)$ is a fixed saddle point of g_1.

Let λ, σ denote the eigenvalues of $Dg_1^\tau(p(g_1))$ with $0 \leq |\lambda| < 1 < |\sigma|$. Replacing g_1 by g_1^2, we may assume that λ and σ are positive.

Next, we embed g_1 in a one-parameter family $\mu \to f_\mu$ of diffeomorphisms defined for μ near 0 so that the family creates a non-degenerate tangency of $W^u(p_\mu)$ and $W^s(p_\mu)$ at q_1. Here, $f_0 = g_1$ and p_μ is the continuation of $p(g_1)$.

Now, we use the techniques of the proof of Theorem 1 on page 47 of the book [**29**] by Palis and Takens, to show that, for some μ's near zero, there are curvilinear rectangles D_μ near q_1 whose first return maps F_μ of f_μ are smoothly conjugate to small perturbations of certain Henon maps $H_{\bar\mu,b}$ defined on the square $Q = [-1, 1] \times [-1, 1]$ where $\bar\mu = \bar\mu(\mu)$ depends smoothly on μ. We will see that the maps $H_{\bar\mu,b}$ can be chosen to have fixed sinks $z_0(\bar\mu)$ whose stable sets contain a full-width strip of fixed height in Q. Thus, the corresponding maps f_μ will have periodic sinks $w_0(\mu)$ whose stable manifolds contain full-width subrectangles of D_μ. The return maps F_μ will contract the area very sharply for small μ on a larger rectangle $D'_\mu \supset D_\mu$ which has a part of $W^u(p_1)$ as a full-height curve. This will imply that p_1 is u-related to the sinks $w_0(\mu)$.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

Let us proceed to the details.

We find a one-parameter family $\mu \to f_\mu$ of C^r diffeomorphisms on M such that the following conditions hold.

(1) $f_0 = g_1$.

(2) The family $\{f_\mu\}$ creates a non-degenerate homoclinic tangency between $W^u(g_1)$ and $W^s(g_1)$ at q_1.

(3) There is a μ-dependent C^2 linearization ψ_μ of f_μ in a neighborhood of a compact curve $\gamma^s(\mu) \subset W^s(p_\mu)$ such that $\gamma^s(0)$ contains q_1 and $p(g_1)$. (*Note*. The C^2 smoothness of the linearization follows as in [**29**]: using a preliminary approximation, we may assume that f is C^∞ and the eigenvalues satisfy the Sternberg linearization non-resonance conditions.)

(4) There is a sequence J_n of parameter intervals converging to zero as $n \to \infty$ so that if $\mu \in J_n$, then one can find a rectangle D_μ near q_1 and a C^r diffeomorphism $\Phi_\mu : D_\mu \to Q$ such that, letting F_μ be the first return map of f_μ to D_μ and $\lambda = \lambda_\mu, \sigma = \sigma_\mu$ be the eigenvalues of Df_μ at p_μ, we have:

   (a) $\Phi_\mu F_\mu \Phi_\mu^{-1} \stackrel{\text{def}}{=} G_{n,\mu}$ is defined on Q and has the form

$$G_{n,\mu}(x, y) = (y, \bar{\mu} + y^2 - \lambda_\mu^n \sigma_\mu^n x) + S(\mu, x, y)$$

   where $S(\mu, x, y) \to 0$ as $n \to \infty$ and $\bar{\mu} \sim \mu \cdot \sigma_\mu^{2n}$;

   (b) $\bar{\mu}$ crosses all of $[0, \frac{1}{2}]$ as μ crosses J_n; and

   (c) denoting the height of D_μ by $height(D_\mu)$ and the width of D_μ by $width(D_\mu)$, we have $height(D_\mu) \sim \sigma_\mu^{-2n}$, $width(D_\mu) \sim height(D_\mu)^{1/2}$.

Thus, we have that the first return map F_μ on the rectangle D_μ is conjugate to a small perturbation of the Henon map

$$H_{\bar{\mu},b}(x, y) = (y, \bar{\mu} + y^2 - bx)$$

on the rectangle Q with $b = \lambda_\mu^n \sigma_\mu^n$ and n large.

We claim that, for small $|b|$, and $\bar{\mu} \in \left[\frac{1}{64}, \frac{1}{16}\right]$, the map $H_{\bar{\mu},b}$ has a unique attracting fixed point $z_0 = z_0(\bar{\mu}, b)$ such that $W^s(z_0)$ contains the rectangle $R = [-1, 1] \times \left[-\frac{7}{8}, \frac{7}{8}\right]$.

To see this, first consider the singular family $H_{\bar{\mu},0}$. The calculations reduce to analogous ones in the one-dimensional logistic family $y \to \bar{\mu} + y^2$. One easily computes that, for $\mu \in \left[\frac{1}{64}, \frac{1}{16}\right]$, $H_{\bar{\mu},0}$ has the attracting fixed point

$$z_0(\bar{\mu}, 0) = \left(\frac{1 - \sqrt{1 - 4\bar{\mu}}}{2}, \frac{1 - \sqrt{1 - 4\bar{\mu}}}{2}\right),$$

and its stable manifold contains R. Now, for small $|b|$, the fixed point $z_0(\bar{\mu}, 0)$ continues to a nearby one $z_0(\bar{\mu}, b)$ with the required properties. Here we use the fact that stable manifolds of hyperbolic fixed points depend continuously on compact sets even for singular maps. This follows from the proofs of the stable manifold theorem in [**15**] or [**33**].

Now, return to the map F_μ. Figure 2 shows a neighborhood of p in the linearizing coordinates ψ_μ. In the domain of the linearizing coordinates ψ_μ, enlarge the rectangle D_μ horizontally to a rectangle $\tilde{D}_\mu$ which projects vertically onto all of γ^s and has D_μ as a full-height subrectangle. Thus, $\tilde{D}_\mu$ will contain both D_μ and a full-height curve γ^u

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

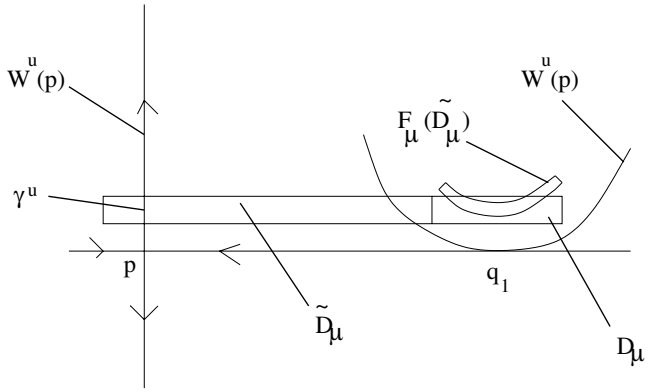


FIGURE 2. $\tilde{D}_\mu$ is a rectangle containing D_μ which is wide enough to project onto a line segment in $W^s(p)$ which contains both p and q_1; $F_\mu(\tilde{D}_\mu)$ is near q_1 and is a slight thickening of the curve $F_\mu(\gamma^u)$.

which is contained in $W^u(p_\mu)$. Since $\lambda_\mu \sigma_\mu < 1$, for n large, the image $F_\mu(\tilde{D}_\mu)$ is a slight thickening of $F_\mu(\gamma^u)$ which is much narrower than the width of D_μ as in Figure 2.

This implies that the stable manifold of the attracting fixed point of F_μ will meet $F_\mu(\gamma^u)$, proving Lemma 2.2. □

We can now complete the proof of Theorem 1.4.

For each positive integer n, let $\mathcal{U}_n$ be the subset of $\mathcal{D}^r(M)$ such that if $f \in \mathcal{U}_n$, then all periodic points of f of period less than or equal to n are hyperbolic with a characteristic exponent sum different from zero. As is well known, arguments as in the proof of the Kupka–Smale theorem [**34**, Theorem 1.1, p. 450] give that $\mathcal{U}_n$ is a dense open subset of $\mathcal{D}^r(M)$.

For $f \in \mathcal{U}_n$, let $P_n^t(f)$ denote the periodic points p of f of period less than or equal to n such that:

(1) p is dissipative; and

(2) there is a sequence of diffeomorphisms $g_1, g_2, \ldots$ converging to f in $\mathcal{D}^r(M)$ such that, for each i, the h-closure of $p(g_i)$ contains a homoclinic tangency.

Let $\mathcal{V}_n$ be the set of f's in $\mathcal{U}_n$ such that $P_n^t(f)$ is empty. Then, $\mathcal{V}_n$ is an open subset of $\mathcal{U}_n$. Letting $\mathcal{F}_n = \mathcal{U}_n \setminus \mathcal{V}_n$, we have that $f \in \mathcal{F}_n$ implies that $P_n^t(f)$ is a finite non-empty set of periodic points. Label these as

$$P_n^t(f) = \{p_1(f), p_2(f), \ldots, p_{s_n}(f)\}$$

with $s_n = s_n(f)$.

For $m > 0$, let

$$\mathcal{F}_{n,m} = \{f \in \mathcal{F}_n : s_n(f) = m\},$$

and let

$$T_n = \{m : \mathcal{F}_{n,m} \neq \emptyset\}.$$

Since, the cardinality of T_n is less than the cardinality of the set of (hyperbolic) periodic points of f of period less than $n+1$, it is clearly a finite set of positive integers. Also, we

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

have the disjoint union

$$\mathcal{F}_n = \bigsqcup_{m \in T_n} \mathcal{F}_{n,m}.$$

For $m \in T_n$ and $1 \leq j \leq m$, let $\mathcal{F}_{n,m,j}$ be the set of f's in $\mathcal{F}_{n,m}$ such that the h-closure of $p_j(f)$ has persistent tangencies and a u-related sink.

By Theorem 1.1 and Lemma 2.2, each $\mathcal{F}_{n,m,j}$ is dense and open in $\mathcal{F}_{n,m}$. Hence, the set

$$\mathcal{E}_n \stackrel{\text{def}}{=} \bigsqcup_{m \in T_n} \left[\bigcap_{1 \leq j \leq m} \mathcal{F}_{n,m,j} \right]$$

is also dense and open in $\mathcal{F}_n$.

Let $\mathcal{G}_n = \mathcal{V}_n \bigcup \mathcal{E}_n$. This set is dense and open in $\mathcal{D}^r(M)$ for each n. If $f \in \mathcal{G}_n$, then any dissipative periodic point p of f with period less than or equal to n whose h-closure contains a tangency must have a u-related sink. So, a part of its unstable manifold tends toward a periodic sink. But, by Proposition 2.1, if such an h-closure carried an SRB measure ν, then $W^u(p)$ would be in the support of ν. This contradiction proves that no such h-closure can carry an SRB measure.

Now, set $\mathcal{U} = \bigcap_n \mathcal{G}_n$. This is the residual subset of $\mathcal{D}^r(M)$ required in Theorem 1.4. □

References

[1] M.-C. Arnaud. Création de connexions en topologie C^1. *Ergod. Th. & Dynam. Sys.* **21** (2001), 339–381.

[2] M. Benedicks and L. Carleson. The dynamics of the Henon map. *Ann. Math.* **133** (1991), 73–169.

[3] M. Benedicks and L.-S. Young. SBR measures for certain Henon maps. *Invent. Math.* **112** (1993), 541–576.

[4] C. Bonatti and S. Crovisier. Récurrence et généricite, to appear.

[5] C. Bonatti, L. Diaz and E. Pujals. A C^1 generic dichotomy for diffeomorphisms: weak forms of hyperbolicity or infinitely many sinks or sources. *Ann. Math*. to appear.

[6] M. Boyle and T. Downarowicz. The entropy theory of symbolic extensions. *Preprint*, http://www.math.umd.edu/∼mmb/papers.html.

[7] M. Boyle, D. Fiebig and U. Fiebig. Residual entropy, conditional entropy, and subshift covers. *Preprint*, http://www.math.umd.edu/∼mmb/papers.html.

[8] J. Buzzi. Intrinsic ergodicity for smooth interval map. *Israel J. Math.* **100** (1997), 125–161.

[9] T. Downarowicz. Entropy of a symbolic extension of a totally disconnected dynamical system. *Ergod. Th. & Dynam. Sys.* **21** (2001), 1051–1070.

[10] T. Downarowicz and S. Newhouse. Symbolic extensions in smooth dynamical systems, to appear, http://www.mth.msu.edu/∼sen/Papers.

[11] A. Fathi, M. Herman and J. C. Yoccoz. A proof of Pesin's stable manifold theorem. *Geometric Dynamics (Lecture Notes in Mathematics, 1007)*. Ed. J. Palis. Springer, Berlin, 1983, pp. 177–215.

[12] J. Graczyk and G. Swiatek. *The Real Fatou Conjecture (Annals of Mathematical Studies, AM-144)*. Princeton University Press, Princeton, NJ, 1998.

[13] B. Hasselblatt and A. Katok. *Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications, 54)*. Cambridge University Press, Cambridge, 1995.

[14] M. Hirsch and C. Pugh. Stable manifolds and hyperbolic sets. *Proc. AMS Symp. Pure Math.* **14**, 1970.

[15] M. C. Irwin. On the stable manifold theorem. *Bull. London Math. Soc.* **2** (1970), 196–198.

[16] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.* **51** (1980), 137–173.

[17] F. Ledrappier. Propriétés ergodiques des mesures de Sinai. *Inst. Hautes Études Sci. Publ. Math.* **59** (1984), 163–188.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press

[18] F. Ledrappier and L. S. Young. The metric entropy of diffeomorphisms. Part I: Characterization of measures satisfying Pesin's entropy formula. *Ann. Math.* **122**(3) (1985), 509–529.
[19] M. Lyubich. Amost every real quadratic map is either regular or stochastic. *Ann. Math.* **156**(1) (2002), 1–78.
[20] R. Mane. A proof of the C^1 stability conjecture. *Inst. Hautes Études Sci. Publ. Math.* **66** (1988), 161–210.
[21] M. Misiurewicz. Topological conditional entropy. *Studia Math.* **55** (1976), 175–200.
[22] L. Mora and M. Viana. Abundance of strange attractors. *Acta Mathematica*.
[23] J. Moser. *Stable and Random Motions in Dynamical Systems (Annals of Mathematical Studies)*. Princeton University Press, Princeton, NJ, 1973.
[24] S. Newhouse. Topological entropy and Hausdorff dimension for area preserving diffeomorphisms of surfaces. *Astèrisque* **51** (1978), 323–334.
[25] S. Newhouse. The abundance of wild hyperbolic sets and non-smooth stable sets for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 101–151.
[26] S. Newhouse. Lectures on dynamical systems. *Dynamical Systems, CIME Lectures, Bressanone, Italy, June 1978 (Progress in Mathematics, 81)*. Eds. J. Coates and S. Helgason. Birkhauser, Boston, MA, 1980, pp. 1–114.
[27] J. Palis. On the C^1 Ω-stability conjecture. *Inst. Hautes Études Sci. Publ. Math.* **66** (1988), 211–215.
[28] J. Palis. A global view of dynamics and a conjecture on the denseness of finitude of attractors. *Asterisque* **261** (2000), 339–351.
[29] J. Palis and F. Takens. *Hyperbolicity and Sensitive Chaotic Dynamics at Homoclinic Bifurcations (Cambridge Studies in Advanced Mathematics, 35)*. Cambridge University Press, Cambridge, 1993.
[30] J. Palis and M. Viana. Continuity of Hausdorff dimension and limit capacity for horseshoes. *Dynamical Systems (Lecture Notes in Mathematics, 1331)*. Springer, Berlin, 1988, pp. 150–160.
[31] C. Pugh and M. Shub. Ergodic attractors. *Trans. Amer. Math. Soc.* **312**(1) (1989), 1–54.
[32] E. Pujals and M. Samborino. Homoclinic tangencies and hyperbolicity for surface diffeomorphisms. *Ann. Math.* **151** (2000), 962–1023.
[33] J. Robbin. Stable manifolds of semi-hyperbolic fixed points. *Illinois J. Math.* **15** (1970), 595–609.
[34] C. Robinson. *Dynamical Systems, Stability, Symbolic Dynamics, and Chaos (Studies in Advanced Mathematics)*, 2nd edn. CRC Press, Baton Rouge, FL, 1999.
[35] D. Wang and L. S. Young. Analysis of a class of strange attractors. *Comm. Pure Applied Math*.

https://doi.org/10.1017/S0143385703000749 Published online by Cambridge University Press