Michigan Math. J. 47 (2000)

Periodic Orbits and Homoclinic Loops for Surface Homeomorphisms

Morris W. Hirsch

0. Introduction

Poincaré invented homoclinic orbits, conjectured their existence in the planar three-body problem, and despaired of understanding their complexity. Research by Birkhoff, Cartwright and Littlewood, and Levinson revealed that near transverse homoclinic points there are robust periodic points. A pinnacle of this line of research, and the basis for much of modern dynamical theory, is Smale's "horseshoe" theorem [22]. For a diffeomorphism f of a manifold of any dimension, it states that every neighborhood of a transverse homoclinic point meets a structurally stable, hyperbolic compact invariant set K on which some iterate f^k is topologically conjugate to the shift map on the Cantor set $2^{\mathbf{Z}}$.

Similar results have been obtained under weakenings of the transversality assumption, including work by Burns and Weiss [6], Churchill and Rod [7], Collins [8], Gavrilov and Šilnikov [13; 14], Guckenheimer and Holmes [15], Mischaikow [18], Mischaikow and Mrozek [19], Newhouse [20], and Rayskin [21].

Among many important consequences is the existence of hyperbolic periodic orbits in K of all periods kn, $n \geq 1$. Note, however, that k is not specified in the horseshoe theorem, and in most cases there is no way to estimate it (but see [19]). Collins [8] has shown that a differentiably transverse homoclinic point implies the existence of periodic points of all sufficiently high minimum periods; estimating such periods, however, requires detailed knowledge of the associated homoclinic tangle.

Although the horseshoe theorem guarantees infinitely many periodic orbits, it is insufficient for the existence of a second fixed point. For example, the toral diffeomorphism induced by the matrix $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right]$ has only one fixed point, even though transverse homoclinic points are dense.

It turns out that, for diffeomorphisms of the plane, even a nontransverse homoclinic point implies a second fixed point; in fact, there is a block of fixed points having index $+1$. But the proof of this (Hirsch [17]), based on Brouwer's plane translation theorem, gives no indication of the location of such a block.

In this paper we consider a saddle fixed point p for an orientation-preserving homeomorphism f of a surface X (definitions will be given in Section 1). Let p'

Received February 2, 2000. Revision received March 16, 2000.
This research was supported in part by the National Science Foundation.

be a homoclinic point associated to p, that is, a point different from p where the stable and unstable curves $W_s(p)$, $W_u(p)$ meet; no transversality or even crossing of these curves is assumed. Suppose $\Lambda = J_s \cup J_u$ is a homoclinic loop at p, where J_s and J_u are arcs in $W_s(p)$ and $W_u(p)$ respectively, having common endpoints p, p'. Assume there is a closed 2-cell, with interior V, whose boundary is the union of two arcs in $W_s(p)$ and $W_u(p)$ with endpoints p and p' in common but otherwise disjoint; such a 2-cell always exists when X is simply connected.

Our main result, stated more precisely in Theorem 1.2, is this:

> *If Λ is a Jordan curve bounding an open 2-cell V, then there exists $\rho \in \{1, 2\}$ such that the fixed point index of f^n in V is ρ for all $n \neq 0$, and ρ depends only on the geometry of V.*

An immediate consequence is that, for every $n \geq 2$, every map sufficiently close to f^n has a block of fixed points in V of index ρ.

Theorem 1.5 is a similar result for homoclinic loops that are homotopically trivial, but not necessarily Jordan curves.

The main theorem is stated in Section 1, and several applications are derived. Section 2 contains the proof of the main theorem.

1. The Main Result and Applications

We use $\mathbf{Z}$, $\mathbf{N}$, and $\mathbf{N}_+$ to denote the integers, natural numbers, and positive natural numbers. All maps are assumed to be continuous; $\approx$ denotes homeomorphism.

For any map g, the maps g^n ($n \geq 1$) are defined recursively by $g^1 = g$ and $g^{n+1}(x) = g(g^n(x))$ provided $g^n(x)$ is in the domain of g. By X we denote a connected, oriented surface with metric d, and $f : X \to X$ is an orientation-preserving injective map. We call f a diffeomorphism when f and f^{-1} are C^1 (continuously differentiable).

The *orbit* of x is the set $\gamma(x) = \{f^i(x) : i \in \mathbf{Z}\}$. The fixed point set of f is denoted by $\mathrm{Fix}(f)$. We call $q \in \mathrm{Fix}(f)$ *smooth* if it belongs to a coordinate chart in which f is represented by a C^1 map; such a chart is *smooth* for q. If f is C^1, then of course all fixed points are smooth. But in many constructions some fixed points of a nonsmooth map are smooth, as when a diffeomorphism of the plane is extended to the 2-sphere.

Let $q \in \mathrm{Fix}(f)$ be smooth. We call q *simple* if 1 is not an eigenvalue of the linear operator df_q, *hyperbolic* if no eigenvalue lies on the unit circle $S^1 \subset \mathbf{C}$, a *sink* if eigenvalues are inside S^1, a *source* if they are outside, and *elliptic* if the eigenvalues are on S^1 but different from 1.

A fixed point p is a *saddle* if (a) it is not in the boundary of X, (b) there is a chart at p in which f is locally represented as a linear map $\begin{bmatrix} \mu & 0 \\ 0 & \lambda \end{bmatrix}$, and (c) either $\mu > 1 > \lambda > 0$, making p a *direct* saddle, or $\mu < -1 < \lambda < 0$, defining a *twisted saddle.* Such a chart is *diagonalizing.* By the Hartman–Grobman linearization theorem (Hartman [16]), for p to be a saddle it is sufficient that there be a smooth chart at p in which df_p has eigenvalues μ, λ as just described.

An *n-periodic* point for f means a fixed point z for f^n, $n \geq 1$. When n is the minimum period, $\gamma(z)$ is an *n-orbit*. An n-periodic point is simple, hyperbolic, and so forth when it has the corresponding property as a fixed point for f^n.

The *stable curve* $W_s = W_s(p)$ of a saddle fixed point p is the connected component of p in the set of x for which there is a convergent sequence $x_k \to p$ in X with $x_0 = x$ and $f(x_k) = x_{k+1}$. The *unstable curve* W_u at p is defined as the stable curve for f^{-1}. Note that W_s and W_u are mapped homeomorphically onto themselves by f. Owing to the linearization assumption, there are bijective maps $\zeta_u, \zeta_s : \mathbf{R} \to W_s$ taking 0 to p, called *parameterizations* of W_u, W_s, respectively. The images of $[0, \infty)$ and $(-\infty, 0]$ are the four *branches* at p.

A *homoclinic point* for p is any point $p' \in W_s \cap W_u \setminus \{p\}$, in which case the *homoclinic loop* Λ defined by p' is the closed path formed by the two arcs $J_s \subset W_s$ and $J_u \subset W_u$ having common endpoints p and p'. There corresponds an element $[\Lambda]$ of the fundamental group of X at p, determined by first traversing Λ from p to p' in J_u and then from p' to p in J_s. If $[\Lambda]$ is the unit element then Λ is an *inessential homoclinic loop.* The loop Λ is *simple* if $J_u \cap J_s = \{p, p'\}$, in which case Λ is homeomorphic to the unit circle. Every homoclinic loop contains a simple homoclinic loop.

Suppose Λ is a simple homoclinic loop in X bounding a closed 2-cell $D \subset X$. The corresponding open 2-cell $V = D \setminus \partial D$ is a *homoclinic cell*. We call V a *positive* cell if some diagonalizing chart takes p to the origin $0 \in \mathbf{R}^2$ and a neighborhood of p in D onto a neighborhood of 0 in the first quadrant. In the contrary case, D is a *negative* cell: there is a diagonalizing chart taking a neighborhood of p in D onto a neighborhood of the origin in the complement of the open first quadrant (see Figure 1). Thus, when seen through a diagonalizing chart, a positive cell appears convex near p while a negative region appears concave.

(a) (b)

Figure 1 Homoclinic cells: (a) positive, (b) negative

Let $U \subset X$ be an open set such that $U \cap \mathrm{Fix}(f)$ is compact. The *fixed point index of f in U* is denoted by $I(f, U) \in \mathbf{Z}$; if it is nonzero, there exists a fixed point in U (Dold [9]). When U is a coordinate chart identified with an open set in

$\mathbf{R}^2$, we can calculate $I(f, U)$ as follows. Let $M \subset U$ be a compact surface with boundary whose interior contains $\mathrm{Fix}(f) \cap U$. Then $I(f, U)$ is the degree of the map

$$\partial M \to S^1, \quad x \mapsto \frac{x - f(x)}{\|x - f(x)\|},$$

where ∂M and S^1 inherit their orientations from $\mathbf{R}^2$. If ∂M is replaced by any oriented Jordan curve Γ on which f has no fixed points, then the same formula defines the *index of f along* Γ.

Let $B \subset X$ be a *block* of fixed points; that is, B is compact and relatively open in $\mathrm{Fix}(f)$. There exists an open neighborhood $U_0 \subset X$ such that $B = \mathrm{Fix}(f) \cap U_0$. The number

$$\mathrm{Ind}(f, B) = \mathrm{Ind}\big(f\big|_{U_0}\big) \in \mathbf{Z},$$

called the *index of f at B*, is independent of the choice of U_0. When p is an isolated fixed point we set $\mathrm{Ind}(f, \{p\}) = I(f, p)$, called the *index of f at p*. A direct saddle has index -1. Twisted saddles, sources, sinks, and elliptic fixed points have index $+1$.

The following assumptions are in force throughout the rest of this article.

HYPOTHESIS 1.1.

(i) $f : X \approx X$ is an orientation-preserving homeomorphism of a surface X.
(ii) $p \in X \setminus \partial X$ is a direct saddle fixed point for f.
(iii) $V \subset X$ is an open 2-cell bounded by a simple homoclinic loop Λ at p.

To V we assign the number

$$\rho = \rho(V) = \begin{cases} 1 & \text{if } V \text{ is a positive region,} \\ 2 & \text{if } V \text{ is a negative region.} \end{cases}$$

For each $n \in \mathbf{N}_+$ we define an open set $V_n \subset V$,

$$V_n = V_n(f) = \{x \in V : f^i(x) \in V,\ i = 1, \dots, n-1\}.$$

Thus, $\mathrm{Fix}(f^n) \cap V_n$ is the union of the n-periodic orbits in V.

The following theorem is our fundamental result.

THEOREM 1.2. *Under Hypothesis 1.1,* $\mathrm{Fix}(f^n) \cap V_n$ *is a block of fixed points for* f^n *of index* $\rho(V)$ *for all* $n \geq 1$.

Before giving the proof of Theorem 1.2 in Section 2, we present several consequences. Hypothesis 1.1 is always assumed.

Homeomorphisms of the Sphere

Assume that $g : S^2 \approx S^2$ is an orientation-preserving homeomorphism having a simple homoclinic loop Λ at a direct saddle.

THEOREM 1.3. *The fixed point index of g in one of the two complementary components of* Λ *is* 1, *and the index in the other is* 2.

Proof. This follows from Theorem 1.2, because one complementary component of Λ has positive type and the other has negative type. □

The persistence of blocks having nonzero index implies the following.

COROLLARY 1.4. *Every map $S^2 \to S^2$ sufficiently close to g has at least three fixed points.*

A homoclinic loop constrains fixed point indices. Suppose, for example, that there are exactly three fixed points: a direct saddle and two other fixed points with respective indices 5 and -2. Then the saddle does not admit a homoclinic point.

Inessential Homoclinic Loops and Nielsen Classes

Fixed points a, b are in the same *Nielsen class* provided they are endpoints of a path that is homotopic to its composition with f, keeping endpoints fixed. Equivalently, f is covered by a map in a universal covering space having fixed points over a and b.

When X is compact, every Nielsen class is a block of fixed points, and the *Nielsen number* of f is the number of Nielsen classes having nonzero index. This number, a homotopy invariant of f, is a lower bound for the number of fixed points for any map homotopic to f.

THEOREM 1.5. *Assume Hypothesis 1.1, and let p belong to an inessential homoclinic loop. Then its Nielsen class contains a block of positive index, and such a block must contain a fixed point $q \neq p$. When the Nielsen class of p is finite, q can be chosen with positive index.*

Thus, in the presence of an inessential homoclinic loop, the number of fixed points exceeds the Nielsen number. Theorem 1.10 is a similar result for Lefschetz numbers.

COROLLARY 1.6. *If a direct saddle p is the only member of its Nielsen class, then p does not belong to an inessential homoclinic loop.*

Proof of Theorem 1.5. We assume that X is not simply connected, otherwise using Theorem 1.2. Since X is orientable, there is a universal covering space $\pi : \mathbf{R}^2 \to X$.

Choose $\tilde{p} \in \pi^{-1}(p)$, and let $\tilde{f} : \mathbf{R}^2 \to \mathbf{R}^2$ be the unique lift of f with a fixed point at $\tilde{p}$. Then $\tilde{p}$ is a direct saddle for $\tilde{f}$.

Let Γ be a null homotopic homoclinic loop at p. There is a unique homoclinic loop $\tilde{\Gamma} \subset \mathbf{R}^2$ for $\tilde{f}$ that contains $\tilde{p}$ and projects onto Γ under π. Let $\tilde{\Lambda} \subset \tilde{\Gamma}$ be a simple homoclinic loop at $\tilde{p}$. There is a unique open 2-cell $V \subset \mathbf{R}^2$ bounded by Λ, and $\bar{V}$ is a closed 2-cell. Applying Theorem 1.2, we choose a block $L \subset \mathrm{Fix}(\tilde{f}) \cap V$ such that

$$\mathrm{Ind}(\tilde{f}, L) = \sigma \in \{1, 2\}.$$

Notice that $\pi(L)$ lies in the Nielsen class of p.

Every fixed point $\tilde{z} \in \pi^{-1}(p)$ has index -1, since $\tilde{f}$ in a neighborhood of $\tilde{z}$ is conjugate to f in a neighborhood of $\pi(z)$. Because $\bar{V}$ is compact, $\pi^{-1}(p) \cap V$ is finite. Therefore $L \setminus \pi^{-1}(p)$ is nonempty, for otherwise L would be a nonempty finite subset of $\pi^{-1}(p)$ and thus have negative index.

It follows that $\pi(L \setminus \pi^{-1}(p))$ is a nonempty subset of $\mathrm{Fix}(f)$ disjoint from p that is contained in the Nielsen class of p. Suppose this class is finite; then $L \setminus \pi^{-1}(p)$ is finite. Let $L \cap \pi^{-1}(p)$ have cardinality ν, $1 \leq \nu < \infty$. Then

$$\begin{aligned}\mathrm{Ind}(\tilde{f}, L \setminus \pi^{-1}(p)) &= \mathrm{Ind}(\tilde{f}, L) - \mathrm{Ind}(\tilde{f}, L \cap (\pi^{-1}(p))) \\ &= \mathrm{Ind}(\tilde{f}, L) - \nu \,\mathrm{Ind}(f, p) \\ &= \sigma + \nu \geq 2.\end{aligned}$$

Hence there exists $\tilde{q} \in L \setminus \pi^{-1}(p)$ with $0 < \mathrm{Ind}(\tilde{f}, \tilde{q}) = \mathrm{Ind}(f, \pi(\tilde{q}))$, and $\pi(\tilde{q})$ is in the Nielsen class of p. This completes the proof of Theorem 1.5. □

Periodic Orbits in a Homoclinic Cell

The following theorem can be used to demonstrate the existence of infinitely many periodic orbits in situtations where the horseshoe theorem may not apply.

THEOREM 1.7. *Let $r \in \mathbf{N}$ be such that every 2^k-orbit $(0 \leq k \leq r)$ in the homoclinic cell V is hyperbolic. Then, either*:

(a) *V contains an attracting or repelling 2^k-orbit for some $k \in \{0, \ldots, r\}$; or else*
(b) *V contains a twisted saddle orbit of cardinality 2^k for every $k = 0, \ldots, r$.*

Proof. Suppose (a) does not hold. Fix $n = 2^k$ $(0 \leq k \leq r)$ and let $B \subset \mathrm{Fix}\big(f^n\big|_{V_n}\big)$ be a block having index $\rho \in \{1, 2\}$ (Theorem 1.2). Then some $q \in B$ has index 1 for f^n. Since (a) is ruled out, q is not a source or sink for f^n. The only other possibility for a hyperbolic, index-1 fixed point for f^n is a twisted saddle. This implies that n is the minimal period for q; thus, (b) holds. □

COROLLARY 1.8. *Assume that every periodic orbit in V whose cardinality is a power of 2 is a saddle. Then V contains a twisted saddle orbit of cardinality 2^k for every $k \in \mathbf{N}$.*

COROLLARY 1.9. *If f is C^1 and if $0 < \mathrm{Det}\, df_x < 1$ in a dense subset of V and if all periodic points in V are hyperbolic, then V contains either a periodic attractor or an orbit of cardinality 2^k for every $k \in \mathbf{N}$.*

It is interesting to compare these results to a theorem of Franks [10]. Specialized to an orientation-preserving diffeomorphism of the 2-sphere, it states:

> *If all periodic points are hyperbolic, and if at most one orbit whose cardinality is a power of 2 is repelling or attracting, then there are infinitely many periodic orbits.*

Corollary 1.8 makes no assumptions on orbits outside the homoclinic cell V, but it does not allow any attractors or repellors of cardinality 2^k in V. It gives sharper information than the conclusion of Franks's theorem on the periods and locations of periodic orbits.

It is not trivial to construct diffeomorphisms of the disk or sphere, all of whose periodic orbits are saddles. However, examples are known that even have the Kupka–Smale property: stable and unstable curves of periodic points have only transverse intersections (Bowen and Franks [3]; Franks and Young [11]). Gambaudo and colleagues [12] have constructed real analytic Kupka–Smale examples on the disk.

Lefschetz Numbers

Let $\#Q$ denote the cardinality of a set Q.

Suppose that the surface X is a compact surface and that $h\colon X \to X$ is continuous. The *Lefschetz number* $\mathrm{Lef}(h)$, defined as the alternating sum of the traces of the induced endomorphisms of the singular homology groups $H_i(X)$, $i = 0, 1, 2$, is equal to $\mathrm{Ind}(h, X)$. Lefschetz proved that, when the fixed point set is finite, $\mathrm{Lef}(h)$ is the sum of the fixed point indices. When every fixed point has index $+1$, -1, or 0, this gives the useful estimate

$$\#(\mathrm{Fix}(f)) \ge |\mathrm{Lef}(h)|.$$

The following results show that, when fixed points are simple, homoclinic cells entail the existence of more fixed points than are counted by the Lefschetz number.

THEOREM 1.10. *Assume X is a compact surface,* $\mathrm{Fix}(f)$ *is finite, and every fixed point has index* $+1$, -1, *or* 0. *If f admits a homoclinic cell, then*

$$\#(\mathrm{Fix}(f)) \ge |\mathrm{Lef}(f) + 1 - \rho| + 1 + \rho \ge |\mathrm{Lef}(f)| + 2.$$

Proof. For any open set $A \subset X$, summing indices over fixed points $z \in A$ yields

$$\begin{aligned} |\mathrm{Ind}(f, A)| &= \left|\sum_z \mathrm{Ind}(f, z)\right| \le \sum_z |\mathrm{Ind}(f, z)| \\ &\le \#(\mathrm{Fix}(f) \cap A). \end{aligned}$$

Applying this to a homoclinic cell V, from Theorem 1.2 we have

$$\begin{aligned} \#(\mathrm{Fix}(f) \cap \bar{V}) &= 1 + \#(\mathrm{Fix}(f) \cap V) \ge 1 + |\mathrm{Ind}(f, V)| \\ &= 1 + \rho, \end{aligned}$$

because $\mathrm{Ind}(f, V) = \rho$. Also,

$$\begin{aligned} \#(\mathrm{Fix}(f) \cap (X \setminus \bar{V})) &\ge |\mathrm{Ind}(f, X \setminus \bar{V})| \\ &= |\mathrm{Ind}(f, X) - (\mathrm{Ind}(f, p) + \mathrm{Ind}(f, V))| \\ &= |\mathrm{Lef}(f) + 1 - \rho|, \end{aligned}$$

because $\mathrm{Ind}(f, p) = -1$. Therefore,

$$\begin{aligned}\#(\mathrm{Fix}(f)) &= \#(\mathrm{Fix}(f) \cap \bar{V}) + \#(\mathrm{Fix}(f) \cap (X \setminus \bar{V}))\\ &\geq (1+\rho) + |\mathrm{Lef}(f) + 1 - \rho|\\ &\geq |\mathrm{Lef}(f)| + 2. \qquad \square\end{aligned}$$

COROLLARY 1.11. *Assume X is a compact surface,* $\mathrm{Fix}(f)$ *is finite, and every fixed point has index* $+1$, -1, *or* 0. *If* $\#(\mathrm{Fix}(f)) \leq |\mathrm{Lef}(f)| + 1$, *then there are no homoclinic cells.*

2. Fixed Point Indices and Retractions

This section contains the proofs of Theorems 1.2 and 1.5. Hypothesis 1.1 continues to hold. Let $D \subset X$ denote the closure of the homoclinic cell V. Then D is a compact 2-cell whose boundary is the simple homoclinic loop Λ.

A *retraction* of a space Y onto a subset $Y_0 \subset Y$ is a map $Y \to Y_0$ fixing every point in Y_0.

LEMMA 2.1. *Assume we are given $n \in \mathbf{N}_+$ and a map $g\colon D \to D$ with the following properties*:

(i) *g coincides with f^n on a neighborhood of p in D; and*
(ii) $\mathrm{Fix}(g) = K \cup \{p\}$, *where $K \subset V$ is compact.*

Then $\mathrm{Ind}(g, K) = \rho(V)$.

Proof. Fix a coordinate chart in which p is the origin and f^n is represented by a linear map

$$T(x, y) = (\lambda x, \mu y), \quad 0 < \lambda < 1 < \mu.$$

We identify points near p with their images in $\mathbf{R}^2$ under this chart.

Consider the case that V is a positive homoclinic cell ($\rho = 1$). Then there is a compact disk neighborhood $N \subset \mathbf{R}^2$ centered at the origin and meeting D in only one of the four closed quadrants; to fix ideas, we assume it is the first quadrant Q_{I}. We take N so small that g coincides with T in $N \cap D$, $N \cap K = \emptyset$, and $N \cup D$ is a 2-cell.

Choose a retraction $s\colon N \to N \cap Q_{\mathrm{I}}$. We compute the fixed point index $\mathrm{Ind}(T \circ s, N)$. Let $\varepsilon > 0$ be so small that the disk D_ε of radius ε lies in N. Let S^1_ε denote the circle bounding D_ε. Since $T \circ s$ has the unique fixed point 0, the index equals the degree of the map

$$u\colon S^1_\varepsilon \to S^1, \quad z \mapsto \frac{z - T \circ s(z)}{\|z - T \circ s(z)\|}.$$

The retraction s sends any point $z \in N \setminus Q_{\mathrm{I}}$ to the unique point $s(z) \in \partial Q_{\mathrm{I}}$ such that z and $s(z)$ are the endpoints of line segment having slope 1; and s is the identity on $N \cap Q_{\mathrm{I}}$. A simple computation shows that u takes no values in the first quadrant of the unit circle and thus has degree 0. Thus $\mathrm{Ind}(T \circ s, N) = 0$.

Now consider the map $h\colon N \cup D \to D \subset N \cup D$ defined to be $T \circ s$ in N and g in D; this definition is consistent because s is a retraction and g coincides with T in $N \cap D$. Clearly

$$\mathrm{Fix}(h) = \{p\} \cup K \subset \mathrm{Int}(N \cup D).$$

Therefore,

$$\mathrm{Lef}(h) = \mathrm{Ind}(h, \mathrm{Int}(N \cup D)) = \mathrm{Ind}(h, p) + \mathrm{Ind}(h, K).$$

The Lefschetz number is 1 because $N \cup D$ is a compact 2-cell, and

$$\mathrm{Ind}(h, \mathrm{Int}(N \cup D)) = \mathrm{Ind}(T \circ s, N) = 0.$$

Hence

$$1 = \mathrm{Ind}(h, K) = \mathrm{Ind}(g, K)$$

as required.

When V is a negative homoclinic cell, we can assume that $N \cap D$ *excludes* the interior of the first quadrant. The retraction $s\colon N \to N \setminus \mathrm{Int}\, Q_{\mathrm{I}}$ is defined by sending $z \in N \cap Q_{\mathrm{I}}$ to the unique point of ∂Q_{I} such that z and $r(z)$ are the endpoints of line segment having slope 1; r is the identity on $N \setminus Q_{\mathrm{I}}$. The degree of u in this case is -1. Define h as before. An argument similar to the preceding shows that

$$1 = \mathrm{Ind}(h, K) = \mathrm{Ind}(h, \{p\}) + \mathrm{Ind}(h, K) = -1 + \mathrm{Ind}(g, K). \qquad \square$$

Let $J_u \subset W_u(p)$ and $J_s \subset W_s(p)$ denote the two compact arcs whose union is Λ; these arcs meet at their common endpoints, which are p and the homoclinic point $p' \neq p$, but nowhere else.

Our next goal is the following.

PROPOSITION 2.2. *There is a retraction*

$$r\colon f(D) \cup D \to D$$

such that

$$r(f(D) \setminus D) \subset J_s. \tag{1}$$

Proof. We first prove

$$J_u \cap \mathrm{clos}(f(D) \setminus D) = \{p, p'\} \tag{2}$$

or, equivalently,

$$J_u \cap \mathrm{clos}(f(V) \setminus D) = \{p, p'\}.$$

Suppose (2) is false, so that there exists

$$b \in J_u \setminus \{p, p'\} \cap \mathrm{clos}(f(V) \setminus D).$$

Then $b = \lim_{i\to\infty} f(a_i)$ for some sequence $a_i \in V \setminus f^{-1}D$, and $b = f(a)$ by continuity. We claim that f maps a relatively open neighborhood $N_a \subset D$ of a onto a relatively open neighborhood $f(N_a) \subset f(D)$ of $f(a)$. This is because f maps the interior of D onto the interior of $f(D)$. The assumption that p is a direct saddle implies that f preserves orientation, and $f^{-1}\big|_{J_u}$ preserves orientation in J_u. From this it follows that N_a and $f(N_a)$ abut J_u from the same side. Consequently,

$f(N_a)$ contains a relatively open neighborhood $N_b \subset D$ of b. For sufficiently large i we have $a_i \in f^{-1}N_b$, and thus $a_i \in f^{-1}D$; this contradiction completes the proof of (2).

From equation (2) we see that

$$\operatorname{clos}(f(D) \setminus D) \cap D \subset J_s. \tag{3}$$

Note also that

$$f(D) \cup D = \operatorname{clos}(f(D) \setminus D) \cup D,$$
$$\operatorname{clos}(f(D) \setminus D) \cap D = \operatorname{clos}(f(D) \setminus D) \cap \partial D \subset J_s.$$

By Tietze's extension theorem, there is a retraction

$$r_0 \colon \operatorname{clos}(f(D) \setminus D) \cup J_s \to J_s;$$

r_0 agrees with the identity map of D on the intersection of their domains, which by (3) is J_s. Thus r_0 and the identity map of D fit together to give the desired retraction r. □

From now on, $r \colon f(D) \cup D \to D$ denotes a retraction as in Proposition 2.2.

LEMMA 2.3. *Let $n \in \mathbf{N}$. For every $q \in \operatorname{Fix}(f^n) \cap V_n$, there is a neighborhood $U \subset V_n$ of q such that $f^n\big|_U = (r \circ f)^n\big|_U$.*

Proof. The definition of V_n implies that $f^j(q) \in V_n \subset V$ for all $j \in \mathbf{N}$. Therefore, q has a neighborhood U such that $f^i(U) \subset V_n$ for $i = 0, \dots, n$. Assume inductively that $0 \le i < n$ and $f^i\big|_U = (r \circ f)^i\big|_U$; the case $i = 0$ is trivial. For $x \in U$ we have $(r \circ f)^i(x) = f^i(x)$, and both $f^i(x)$ and $f^{i+1}(x)$ are in V because $x \in V_n$. Hence

$$(r \circ f)^{i+1}(x) = (r \circ f)(f^i(x)) = r(f^{i+1}(x)) = f^{i+1}(x),$$

because r and f coincide on V. This completes the induction. □

LEMMA 2.4. $\operatorname{Fix}\big((r \circ f\big|_D)^n\big) = \{p\} \cup (\operatorname{Fix}(f^n) \cap V_n)$ *for all $n \ge 1$.*

Proof. Let $x \in D \setminus \{p\}$ be n-periodic for $r \circ f$. We first show $x \notin J_s$. We know that J_s is invariant under f, and $r\big|_{J_s}$ is the identity because $J_s \subset D$. Thus $r \circ f\big|_{J_s}$ coincides with $f\big|_{J_s}$, whose only periodic point is p. The foregoing implies that no point on the orbit x under $r \circ f$ lies in J_s. Thus no point y in this orbit maps outside D under f, for otherwise $(r \circ f)(y) \in J_s$ by equation (1). This proves $\gamma(x) \subset D$, and by induction we know that $(r \circ f)^k x = f^k x$ for all k. Since $J_u \setminus p$ contains no periodic points for f, the conclusion follows. □

Proof of Theorem 1.2

The set $B = \operatorname{Fix}(f^n) \cap V_n$ is open in $\operatorname{Fix}(f^n)$ because V_n is open. We prove that B is compact by showing it is closed in D. Since $\bar{B} \cap \partial D \subset \{p\}$, it suffices to prove that p is not a limit point of B. Clearly $p \notin B$, and p (being a saddle) has a neighborhood in which the only point of period n is p. Therefore B is a block.

To prove $\mathrm{Ind}(f^n, B) = \rho$, let $r\colon f(D) \cup D \to D$ be a retraction as in Proposition 2.2. Lemmas 2.4 and 2.3 show that $\mathrm{Ind}(f^n, B) = \mathrm{Ind}\big((r \circ f\big|_V)^n, B\big)$. Now apply Lemma 2.1 to $g = \big(r \circ f\big|_D\big)^n$ to conclude that $\mathrm{Ind}\big((r \circ f\big|_V)^n, B\big) = \rho$. $\square$

References

[1] G. D. Birkhoff, *Proof of Poincaré's last geometric theorem,* Trans. Amer. Math. Soc. 14 (1913), 14–22.

[2] ———, *An extension of Poincaré's last geometric theorem,* Acta Math. 47 (1925), 297–311.

[3] R. Bowen and J. Franks, *The periodic points of maps of the disk and the interval,* Topology 15 (1976), 337–342.

[4] M. Brown, *The monotone union of open n-cells is an open n-cell,* Proc. Amer. Math. Soc. 12 (1961), 812–814.

[5] ———, *Homeomorphisms of two-dimensional manifolds,* Houston J. Math. 11 (1985), 455–469.

[6] K. Burns and H. Weiss, *A geometric criterion for positive topological entropy,* Comm. Math. Phys. 172 (1995), 95–118.

[7] R. Churchill and D. Rod, *Pathology in dynamical systems. III. Analytic Hamiltonians,* J. Differential Equations 37 (1980), 23–38.

[8] P. Collins, *Dynamics of surface maps with homoclinic and heteroclinic tangles,* Ph.D. dissertation, Univ. of California at Berkeley, 1999.

[9] A. Dold, *Fixed point index and fixed point theorem for Euclidean neighborhood retracts,* Topology 4 (1965), 1–8.

[10] J. Franks, *Some smooth maps with infinitely many hyperbolic periodic points,* Trans. Amer. Math. Soc. 226 (1977), 175–179.

[11] J. Franks and L.-S. Young, *A C^2 Kupka–Smale diffeomorphism of the disk with no sources or sinks,* Dynamical systems and turbulence (Warwick 1980), Lecture Notes in Math., 898, pp. 90–98, Springer-Verlag, Berlin, 1981.

[12] J.-M. Gambaudo, S. van Strien, and C. Tresser, *Hénon-like maps with strange attractors: There exist C^∞ Kupka–Smale diffeomorphisms on S^2 with neither sinks nor sources,* Nonlinearity 2 (1989), 287–304.

[13] N. Gavrilov and L. Šilnikov, *On three-dimensional dynamical systems close to systems with a structurally unstable homoclinic curve, I,* Math. USSR-Sb. 17 (1972), 467–485 (in Russian).

[14] ———, *On three-dimensional dynamical systems close to systems with a structurally unstable homoclinic curve, II,* Math. USSR-Sb. 19 (1973), 139–156 (in Russian).

[15] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields,* Springer-Verlag, New York, 1983.

[16] P. Hartman, *Ordinary differential equations,* Wiley, New York, 1964.

[17] M. Hirsch, *Fixed-point indices, homoclinic contacts, and dynamics of injective planar maps,* Michigan Math. J. 47 (2000), 101–108.

[18] K. Mischaikow, *Conley index theory,* Lecture Notes in Math., 1609, pp. 119–207, Springer-Verlag, Berlin, 1995.

[19] K. Mischaikow and M. Mrozek, *Chaos in the Lorenz equations: A computer-assisted proof,* Bull. Amer. Math. Soc. 32 (1995), 66–72.

[20] S. Newhouse, *Lectures on dynamical systems,* Prog. Math., 8, pp. 1–114, Birkhäuser, Boston, 1980.
[21] V. Rayskin, *Degenerate homoclinic crossing,* Ph.D. dissertation, Univ. of California at Berkeley, 1997.
[22] S. Smale, *Diffeomorphisms with many periodic points,* Differential and combinatorial topology (S. Cairns, ed.), pp. 63–80, Princeton Univ. Press, Princeton, NJ, 1965.

Mathematics Department
University of California at Berkeley
Berkeley, CA 94720-3840

hirsch@math.berkeley.edu