

ROTATION NUMBERS AND INSTABILITY SETS

JOHN FRANKS

ABSTRACT. Translation and rotation numbers have played an interesting and important role in the qualitative description of various dynamical systems. In this exposition we are especially interested in applications which lead to proofs of periodic motions in various kinds of dynamics on the annulus. The applications include billiards and geodesic flows.

Going beyond this simple qualitative invariant in the study of the dynamics of area preserving annulus maps, G.D. Birkhoff was led to the concept of “regions of instability” for twist maps. We discuss the closely related notion of instability sets for a generic area preserving surface diffeomorphism and develop their properties.

1. INTRODUCTION

The concepts of translation number and rotation number were introduced by Henri Poincaré and have had a rich history. We describe several of many manifestations of these concepts as they apply to the qualitative description of various dynamical systems. We are especially interested in applications which lead to proofs of periodic motions in various kinds of area preserving dynamics on the annulus and S^2 . Applications include billiards and geodesic flows. This is the content of sections 2 to 5.

In pursuit of the qualitative analysis of such dynamics, G.D. Birkhoff was led to the concept of “regions of instability” for twist maps on an annulus. We discuss the closely related notion of instability sets for a generic area preserving surface diffeomorphism and describe some of their properties.

Guiding this is the deep theory of twist maps of the annulus where we have a large body of knowledge, thanks to the pioneering work of G.D. Birkhoff and subsequent work by Arnold, Kolmogorov, Herman, Mather, and Moser. However, the definition of an annular twist map depends on a distinguished choice of co-ordinates on the annulus. In a local setting, e.g. in the neighborhood of a generic fixed point, one can always obtain these co-ordinates, but they are generally unavailable in the larger global context. It would be nice to have the same or analogous results in this global context.

In sections 6, 7 and 8 we describe a residual class of C^r area preserving surface diffeomorphisms for $r \geq 16$ and describe some of their qualitative properties.

In §9 we discuss some of the intriguing open questions related to instability sets.

Received by the editors December 31, 2002.

2000 *Mathematics Subject Classification*. Primary 37E45.

Supported in part by NSF grant DMS0099640. This article is the written version of an invited address at the January 2002 AMS meeting in San Diego, California.

2. CIRCLE HOMEOMORPHISMS

The simplest instance of rotation number is its use as an invariant for orientation preserving circle homeomorphisms. Define the circle \mathbb{T}^1 to be the quotient \mathbb{R}/\mathbb{Z} . As a quotient \mathbb{T}^1 inherits both a group structure and a topology. There is a natural projection $\pi : \mathbb{R} \rightarrow \mathbb{T}^1$ defined by $\pi(x) = x + \mathbb{Z}$ which is both a (universal) covering map and a group homomorphism.

Definition 2.1. Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism of the circle. A homeomorphism $F : \mathbb{R} \rightarrow \mathbb{R}$ is a *lift* of f provided

$$f(\pi(t)) = \pi(F(t)).$$

If F is a lift of f it is not difficult to see that $G : \mathbb{R} \rightarrow \mathbb{R}$ is also a lift of f if and only if $F(t) = G(t) + n$ for some $n \in \mathbb{Z}$.

Definition 2.2. Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism and let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a lift. The *translation number*¹ of a point $x \in \mathbb{R}$ under F is

$$\tau(x, F) = \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n}.$$

Notice that the value of $\tau(x, F)$ is the average amount the point x is translated under iteration by F . As shown below, its *mod* 1 reduction depends only on f and measures the average rate at which a point is moved around the circle by each iteration of f . We summarize the elementary properties of the translation number of a lift F in the following proposition (see section 11.1 of [14] for a proof).

Proposition 2.3. *Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism and let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a lift.*

- *The limit $\tau(x, F)$ exists for all $x \in \mathbb{R}$.*
- *$\tau(x, F^n) = n\tau(x, F)$ for all $n \in \mathbb{Z}$.*
- *$\tau(x, F + m) = \tau(x, F) + m$ for $m \in \mathbb{Z}$.*
- *$\tau(F) := \tau(x, F)$ is independent of x .*

More striking is the fact that if $\tau(F)$ is rational, then there are periodic orbits.

Theorem 2.4. *Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism and let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a lift.*

- *$\tau(F) = 0$ if and only if F has a fixed point.*
- *f has a fixed point if and only if $\tau(F) \in \mathbb{Z}$.*
- *f has a periodic point of period q if and only if $\tau(F) = p/q \in \mathbb{Q}$, where p/q is in lowest terms.*

If F and G are two different lifts of $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$, then they differ by a constant integer; i.e. there is an $m \in \mathbb{Z}$ such that $F(t) = G(t) + m$ for all $t \in \mathbb{R}$. From this and the proposition above it follows that $\tau(x, F) = \tau(x, G) + m$. As a consequence if we reduce $\tau(x, F)$ modulo 1, or, more precisely, consider its projection in $\mathbb{T}^1 := \mathbb{R}/\mathbb{Z}$, the result is independent of the choice of the lift F and depends only on the original homeomorphisms f .

¹It is common to see this quantity referred to as the *rotation number*; however, the term translation number, while less commonly used, is more accurately descriptive. More importantly, there is a need to distinguish the translation number from its projection in \mathbb{T}^1 , which is properly called the rotation number (see Definition (2.5)).

Definition 2.5. Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism and let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a lift. The *rotation number* of f is the element of \mathbb{T}^1 defined by

$$\rho(f) = \pi(\tau(x, F))$$

where $\pi : \mathbb{R} \rightarrow \mathbb{T}^1$ is the natural projection, F is any lift of f and x is any point of \mathbb{R} .

From Proposition 2.3 we immediately obtain the following:

Proposition 2.6. *Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism.*

- $\rho(f^n) = n\rho(f)$ for all $n \in \mathbb{Z}$.
- f has a periodic point of period q if and only if $\rho(f) = (p/q + \mathbb{Z}) \in \mathbb{Q}/\mathbb{Z}$ where p/q is in lowest terms.

For the many important dynamical systems which preserve a measure there is often an important relation between the measure and the rotation numbers. In the case of circle homomorphisms this relation has an especially simple form which we will now describe.

Definition 2.7. Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism and let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a lift. We define the *displacement function* $\phi : \mathbb{T}^1 \rightarrow \mathbb{R}$ of F by $\phi(x + \mathbb{Z}) = F(x) - x$.

Note that if x' is another choice for x (i.e. $x + \mathbb{Z} = x' + \mathbb{Z}$), then $x' = x + m$ for some integer m so $F(x) - x = F(x') - x'$ and ϕ is well defined. For a point $(x + \mathbb{Z}) \in \mathbb{T}^1$ the value of ϕ can be thought of as the amount this point is displaced around the circle. Having chosen a lift F , we have determined this as a well defined element of \mathbb{R} , not just an element of \mathbb{R} modulo 1.

Theorem 2.8. *Let $f : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ be an orientation preserving homeomorphism which preserves a probability measure μ , and let F be a lift of f and suppose $\phi : \mathbb{T}^1 \rightarrow \mathbb{R}$ is the displacement function of F . Then*

$$\tau(F) = \int_{\mathbb{T}^1} \phi \, d\mu.$$

Notice that the left hand side of this equality is independent of μ and hence the right hand side is the same for all choices of μ !

Proof. Here we use z to denote a point of \mathbb{T}^1 instead of the more cumbersome $x + \mathbb{Z}$. According to the Birkhoff Ergodic Theorem (see [14]), the function $\hat{\phi}$ defined by

$$\hat{\phi}(z) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(f^i(z))$$

is integrable and satisfies

$$\int_{\mathbb{T}^1} \phi \, d\mu = \int_{\mathbb{T}^1} \hat{\phi} \, d\mu.$$

But if $z = \pi(x)$, then $f^i(z) = \pi(F^i(x))$ and $\phi(f^i(z)) = F(F^i(x)) - F^i(x) = F^{i+1}(x) - F^i(x)$. It follows that the telescoping sum

$$\sum_{i=0}^{n-1} \phi(f^i(z)) = \sum_{i=0}^{n-1} (F^{i+1}(x) - F^i(x)) = F^n(x) - x.$$

Hence

$$\hat{\phi}(z) = \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n} = \tau(F).$$

So $\hat{\phi}$ is the constant function with value $\tau(F)$ and the result follows. \square

3. ANNULUS HOMEOMORPHISMS

The annulus $\mathbb{A} = \mathbb{T}^1 \times I$, where $I = [0, 1]$, provides much richer ground for application of the concept of rotation numbers.

Before exploring this, however, we motivate our interest in the annulus by describing one of the many mechanical systems which give rise to annulus maps. Poincaré's original interest in rotation numbers was related to the restricted three body problem, but a system much easier to describe is provided by billiards on a strictly convex table with smooth boundary.

The ball is idealized as a point which moves on the table with uniform motion and bounces off the edge in such a way that the angle of incidence equals the angle of reflection. The annulus associated with this dynamical system is $\mathbb{T}^1 \times [0, \pi]$ where the circle \mathbb{T}^1 parametrizes the edge of the table and $\theta \in [0, \pi]$ is the angle between the tangent to the table edge and the direction at which the ball leaves the edge at a bounce. If x is a point of contact and $\theta \in (0, \pi)$ is an angle of reflection, then the *billiard map* $F : \mathbb{A} \rightarrow \mathbb{A}$ is defined by $F(x, \theta) = (x', \theta')$, where x' is the next point of contact with the boundary and θ' is the next angle of reflection. Notice that by defining $F(x, 0) = x$ and $F(x, \pi) = x$ we can extend F to a homeomorphism of all of \mathbb{A} .

To understand the billiard dynamics qualitatively (for example the existence of periodic orbits) it suffices to understand this billiard map. A crucial aid in this analysis comes from the fact that the map preserves an area.

Theorem 3.1 (Birkhoff). *If we define an area on \mathbb{A} using the element of area $\sin \theta \, dx \, d\theta$, then this area is preserved by the billiard map $F : \mathbb{A} \rightarrow \mathbb{A}$.*

As in the case of a circle we can define the notions of translation number and rotation number.

Definition 3.2. Let $\tilde{\mathbb{A}} = \mathbb{R} \times I$. A homeomorphism $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ is a *lift* of f provided

$$f(\tilde{\pi}(x, y)) = \tilde{\pi}(F(x, y)),$$

where $\tilde{\pi} : \tilde{\mathbb{A}} \rightarrow \mathbb{A}$ is given by $\tilde{\pi}(x, y) = (\pi(x), y)$.

If F is a lift of f , it is not difficult to see that $G : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ is also a lift of f if and only if

$$F(x, y) = G(x, y) + (m, 0)$$

for some $m \in \mathbb{Z}$.

Definition 3.3. Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism which preserves boundary components and let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift. The *translation number* of a point $w = (x, y) \in \tilde{\mathbb{A}}$ under F is

$$\tau(w, F) = \lim_{n \rightarrow \infty} \frac{p_1(F^n(w) - w)}{n},$$

where $p_1 : \tilde{\mathbb{A}} \rightarrow \mathbb{R}$ is the projection on the first co-ordinate.

We would like an analogue of Proposition 2.3; i.e. we would like the properties of the translation number for circle homeomorphisms to hold for annulus homeomorphisms. Unfortunately, much less is true for the annulus. In particular $\tau(w, F)$ may fail to exist for many $w \in \tilde{\mathbb{A}}$. Moreover, when it does exist, it is only rarely independent of w .

Proposition 3.4. *Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism which preserves boundary components and let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift.*

- *If $\tau(w, F)$ exists, then $\tau(w, F^n) = n\tau(w, F)$ for all $n \in \mathbb{Z}$.*
- *If $\tau(w, F)$ exists, then $\tau(w, F + (m, 0)) = \tau(w, F) + m$ for $m \in \mathbb{Z}$.*
- *If f preserves a finite Borel measure μ and $\tilde{\mu}$ is the lift of this measure to $\tilde{\mathbb{A}}$, then $\tau(w, F)$ exists for all w except a set of $\tilde{\mu}$ measure zero.*

The proofs of the first two parts of this proposition are straightforward. The final part is a consequence of the Birkhoff ergodic theorem, and we give a proof below in Theorem 3.8.

Definition 3.5. Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism and let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift. The *rotation number* of f is the element of \mathbb{T}^1 defined by

$$\rho(f) = \pi(\tau(x, F))$$

where $\pi : \mathbb{R} \rightarrow \mathbb{T}^1$ is the natural projection, F is any lift of f and x is any point of \mathbb{R} .

From Proposition 3.4 we immediately obtain the following:

Proposition 3.6. *Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism which preserves boundary components and let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift.*

- *If $\tau(w, F)$ exists, then $\rho(\tilde{\pi}(w), f)$ exists and is well defined, i.e. independent of the lift F .*
- *If $\rho(x, f)$ exists, $\rho(w, f^n) = n\rho(x, f)$ for all $n \in \mathbb{Z}$.*
- *If f preserves a finite Borel measure μ , then $\rho(x, f)$ exists for all $x \in \mathbb{A}$ except a set of μ measure zero.*

It is not hard to see that periodic orbits have rational translation numbers. But the issue of using translation numbers or rotation numbers to prove the existence of periodic points is much more subtle than in the case of the circle and that is the subject of §4 below.

The last item in Proposition 3.6 makes it very important to have a measure preserving annulus homeomorphism, and it is in this setting that we are able to provide the most interesting results.

As in the case of \mathbb{T}^1 it is useful to define the displacement function for a lift F .

Definition 3.7. Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism, preserving boundary components, and let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift. We define the *displacement function* $\phi : \mathbb{A} \rightarrow \mathbb{R}$ of F by $\phi(\tilde{\pi}(w)) = p_1(F(w) - w)$ for $w \in \tilde{\mathbb{A}}$, where $p_1 : \tilde{\mathbb{A}} \rightarrow \mathbb{R}$ is the projection on the first co-ordinate.

Note that if w' satisfies $\tilde{\pi}(w) = \tilde{\pi}(w')$, then $w' = w + (m, 0)$ for some integer m , so $F(w) - w = F(w') - w'$ and ϕ is well defined.

Theorem 3.8. *Let $f : \mathbb{A} \rightarrow \mathbb{A}$ be an orientation preserving homeomorphism, preserving boundary components, and preserving a probability measure μ , and let F*

be a lift of f . Then the rotation number $\rho(x, F)$ exists for μ -almost all x and is a μ integrable function. If $\phi : \mathbb{A} \rightarrow \mathbb{R}$ is the displacement function of F ,

$$\int_{\mathbb{T}^1} \rho(x, F) d\mu = \int_{\mathbb{T}^1} \phi d\mu.$$

The value of these integrals will be denoted by $\tau_\mu(F)$ and is called the mean translation number of F .

Proof. Again we appeal to the Birkhoff Ergodic Theorem (see [14]). As in the proof of Theorem 2.8, if $x = \bar{\pi}(w)$, then the telescoping sum

$$\sum_{i=0}^{n-1} \phi(f^i(x)) = \sum_{i=0}^{n-1} p_1(F^{i+1}(w) - F^i(w)) = p_1(F^n(w) - w).$$

So

$$\tau(w, F) = \lim_{n \rightarrow \infty} \frac{p_1(F^n(w) - w)}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(f^i(x)).$$

Hence, according to the ergodic theorem, $\tau(w, F)$ exists for μ -almost all x and is μ integrable. Moreover it satisfies

$$\tau_\mu(F) = \int_{\mathbb{T}^1} \tau d\mu = \int_{\mathbb{T}^1} \phi d\mu.$$

□

4. THE POINCARÉ-BIRKHOFF THEOREM

In this section we describe several theorems analogous to Theorem 2.4, in the sense that from a hypothesis on a translation number one is able to conclude the existence of a periodic orbit. The most famous of these is the Poincaré-Birkhoff Theorem. This was conjectured by Poincaré, and he was able to prove some special cases. His motivation was to prove the existence of some periodic motions of the restricted three-body problem (this is the Newtonian motion of three point masses, one of which has zero mass).

G.D. Birkhoff succeeded in proving the result Poincaré conjectured in [1]. This was perhaps the first of Birkhoff's many great contributions to bring him world recognition in the mathematical community.

Theorem 4.1 (G.D. Birkhoff). *Suppose $f : \mathbb{A} \rightarrow \mathbb{A}$ is an area preserving annulus homeomorphism which preserves orientation and boundary components. Let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift and let $F_0(x) = F(x, 0)$, $F_1(x) = F(x, 1)$. If $p/q \in \mathbb{Q}$ and*

$$\tau(F_0) \leq \frac{p}{q} \leq \tau(F_1),$$

then F has at least two points with translation number p/q which are lifts of periodic points of f with distinct orbits of period q and rotation number $p/q + \mathbb{Z}$.

In addition to the application which Poincaré had in mind, this result can be applied to billiards on a table with smooth convex boundary as described above. In fact, if $f : \mathbb{A} \rightarrow \mathbb{A}$ is the billiard map, there is a lift $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ with the property that for all $x \in \mathbb{R}$, $F_0(x) = x$, and $F_1(x) = x + 1$. It follows that $\tau(F_0) = 0$ and $\tau(F_1) = 1$, so according to the theorem for every $p/q \in [0, 1]$ there are two periodic orbits with translation number p/q and rotation number $p/q + \mathbb{Z}$. This is true for any billiard table as long as it has a smooth strictly convex boundary. (In fact,

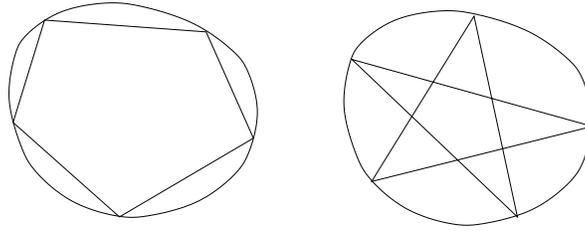


FIGURE 1. Periodic billiard orbits with $\rho(x, f) = \frac{1}{5} + \mathbb{Z}$ and $\rho(x, f) = \frac{2}{5} + \mathbb{Z}$

strict convexity is not really necessary; simple convexity and smoothness of the boundary suffices.)

Figure 1 shows examples with two periodic billiard orbits on a convex table with rotation numbers $\rho(x, f) = \frac{1}{5} + \mathbb{Z}$ and $\rho(x, f) = \frac{2}{5} + \mathbb{Z}$.

The Poincaré-Birkhoff Theorem has a number of generalizations. One which is important for our purposes asserts that we do not need information about the behavior of F on the boundary of $\tilde{\mathbb{A}}$; it suffices to have any two points $w_0, w_1 \in \tilde{\mathbb{A}}$ with distinct translation numbers to obtain periodic lifts with all rational translation numbers in the interval $[\tau(w_0, F), \tau(w_1, F)]$.

Theorem 4.2 (Franks [3],[4]). *Suppose $f : \mathbb{A} \rightarrow \mathbb{A}$ is an area preserving annulus homeomorphism which preserves orientation and boundary components. Let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift. If $w_0, w_1 \in \tilde{\mathbb{A}}$ and $p/q \in \mathbb{Q}$ satisfy*

$$\tau(w_0, F) \leq \frac{p}{q} \leq \tau(w_1, F),$$

then F has at least two points with translation number p/q which are lifts of periodic points of f with distinct orbits of period q and rotation number $p/q + \mathbb{Z}$.

Of course, it is natural to ask about points with irrational translation number.

Theorem 4.3 (M. Handel [11]). *Suppose that the homeomorphism $f : \mathbb{A} \rightarrow \mathbb{A}$ preserves orientation and boundary components and F is a lift. Then the set $\{\tau(w, F) \mid w \in \tilde{\mathbb{A}}\}$ is closed.*

Notice that unlike the Poincaré-Birkhoff Theorem, no area preserving hypothesis is necessary for Handel’s result. However, when the area preserving hypothesis holds, we can conclude from the results above that the set of all translation numbers of a lift F forms a closed interval.

Finally we note that there is also a result relating rationality of the mean translation number to the existence of periodic points.

Theorem 4.4 (Franks [5]). *Suppose $f : \mathbb{A} \rightarrow \mathbb{A}$ is an annulus homeomorphism which preserves orientation and boundary components. Let $F : \tilde{\mathbb{A}} \rightarrow \tilde{\mathbb{A}}$ be a lift. If f preserves Lebesgue measure μ and*

$$\tau_\mu(F) = \frac{p}{q},$$

then F has at least two points with translation number p/q which are lifts of periodic points of f with distinct orbits of period q and rotation number $p/q + \mathbb{Z}$.

5. THE GEODESIC RETURN MAP ON S^2

In this section we discuss another application of translation and rotation numbers, namely the problem of the existence of closed geodesics on the two-sphere S^2 .

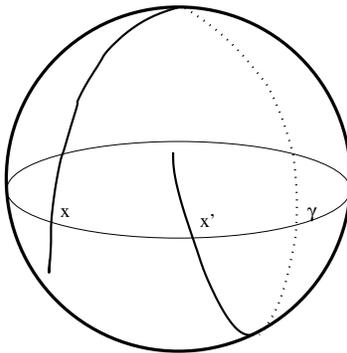


FIGURE 2. The geodesic return map

One of the standard ways to find closed geodesics is to try to minimize the length of a closed curve in its homotopy class. Of course, this cannot immediately be applied to S^2 since all closed curves are homotopic to curves of arbitrarily small length. Nevertheless a fascinating result of Birkhoff, using a mini-max argument, does succeed in obtaining at least one closed geodesic for S^2 provided with an arbitrary Riemannian metric. In fact the geodesic so obtained is *simple*, i.e. does not cross itself.

Theorem 5.1 (Birkhoff). *For any Riemannian metric on S^2 there is always at least one simple closed geodesic.*

In fact, Liusternik and Schnirrelman subsequently proved that there are at least three simple closed geodesics on such a sphere.

Birkhoff realized that, like the problem of periodic billiards, the existence of closed geodesics could often be reduced to the study of an annulus map.

Definition 5.2. The *geodesic return map* $F : \mathbb{A} \rightarrow \mathbb{A}$ for a simple closed geodesic $g \subset S^2$ is defined by

$$F(x, \theta) = (x', \theta'),$$

where x' is the second point where g is crossed by a geodesic which starts at x and forms angle θ with g . θ' is the new angle. (See Figure 2.)

Theorem 5.3 (Birkhoff). *The geodesic return map is well defined for a sphere when it has positive curvature. Whenever it is well defined it is an area preserving diffeomorphism.*

This motivates an interest in the existence of periodic orbits for area preserving annulus maps since there is a one-to-one correspondence between closed geodesics and periodic orbits of the geodesic return map. The following result shows, in particular, that whenever the geodesic return map is well defined for a Riemannian metric on S^2 , there are infinitely many distinct closed geodesics for that metric.

Theorem 5.4 (Franks [5]). *Suppose $f : \mathbb{A} \rightarrow \mathbb{A}$ is an area preserving annulus diffeomorphism which preserves orientation and boundary components. If f has at least one periodic point, then it has infinitely many.*

Theorem 5.5 (V. Bangert [1]). *When the geodesic return map is not well defined there are infinitely many closed geodesics on S^2 .*

Combining these results we get the following result which answers a long-standing question about the existence of closed geodesics on S^2 .

Theorem 5.6. *Every smooth Riemannian metric on S^2 has infinitely many closed geodesics.*

Shortly after the appearance of this result a rather different proof using Morse theory on the space of free loops on S^2 was provided by Nancy Hingston [13].

Theorem 5.4 can be strengthened in several ways. For example in [9] the following result is proved.

Theorem 5.7 (Franks and Handel, [9]). *Suppose*

$$f : \mathbb{A} \rightarrow \mathbb{A}$$

is an area preserving annulus homeomorphism which preserves orientation and boundary components. If f has at least one periodic point, then the periods of periodic points are unbounded.

In fact this result is valid for more general surfaces. An area preserving diffeomorphism of a compact surface is called a Hamiltonian diffeomorphism if it is a commutator in the group of area preserving diffeomorphisms homotopic to the identity. In [9] it is shown that any Hamiltonian diffeomorphism of a surface of genus ≥ 1 must have periodic points of arbitrarily high period.

6. THE INSTABILITY SET

The purpose of this section is to delve more deeply into the investigation of the dynamics of typical area preserving diffeomorphisms of annuli and spheres.

A rather remarkable qualitative behavior is observed if one does graphical computer studies of even simple diffeomorphisms of this type. The *standard map family* $f : \mathbb{T}^1 \times \mathbb{R} \rightarrow \mathbb{T}^1 \times \mathbb{R}$ is given by

$$f(x, y) = (x + y + k \sin(2\pi x), y + k \sin(2\pi x)),$$

where k is a parameter. One checks easily that the determinant of the Jacobian for this diffeomorphism of the open cylinder is 1, so it is an area preserving diffeomorphism. If $k = 0$ this map exhibits rather simple dynamics, namely each circle $\mathbb{T}^1 \times y_0$ in the cylinder is invariant and the restriction of f to this circle is a rotation by y_0 .

If the parameter k is increased, the dynamics becomes much more complicated. For example Figure 3 shows a computer generated image of several thousand points on an orbit for the standard map with $k = 0.15$. Actually, $f(x, y)$ is periodic of period one, and what has been drawn is the region $0 \leq x \leq 2, -0.5 \leq y \leq 0.5$ in order to give a more informative picture.

The striking feature of this image is that the plotted points (a finite set) appear to be a connected set which is riddled with holes. The remarkable property of the image is that if another image were created by plotting points on the orbit of

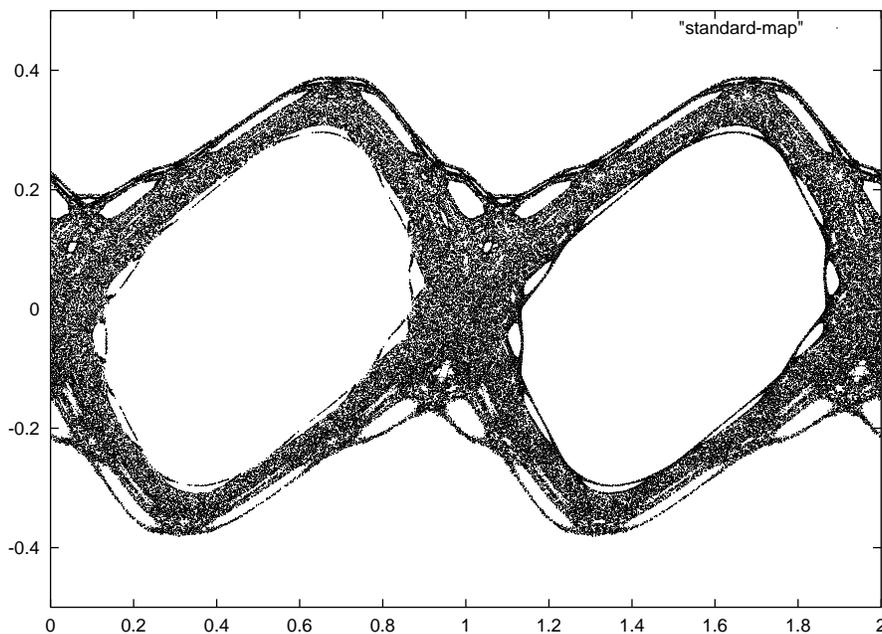


FIGURE 3. The standard map, $k = 0.15$

another point near this set (i.e. not in one of the holes), then the image obtained would be visually essentially identical. This is despite the fact that it consists of a disjoint set of points. As we will see, it is likely that there is a compact connected invariant set which these plots are approximating.

If we were to fill all the holes in this figure and take the interior of the resulting set, we would obtain an example of what Birkhoff called a “*region of instability*”. Our aim, however, is to investigate the black set in the figure, which we will define and refer to as an *instability set*.

Of particular importance in our investigation are the *stable and unstable sets* which we now define.

Definition 6.1. Suppose $f : X \rightarrow X$ is a homeomorphism and $P \subset X$ is a compact f -invariant set. The stable set $W^s(P)$ is defined by

$$W^s(P) := \{x \in X \mid \lim_{n \rightarrow \infty} \text{dist}(f^n(x), P) = 0\},$$

and the unstable set $W^u(P)$ is defined by

$$W^u(P) := \{x \in X \mid \lim_{n \rightarrow -\infty} \text{dist}(f^n(x), P) = 0\},$$

where $\text{dist}(x, P)$ is the distance from x to the set P .

If $f : S \rightarrow S$ is an area preserving diffeomorphism of a surface and $p \in S$ is a fixed point, then p is called *elliptic* if the eigenvalues of Df_p are complex conjugates of modulus 1 and are not equal to ± 1 . The point p is *hyperbolic* if the eigenvalues of Df_p are real and have modulus distinct from 1. Since the diffeomorphism is area preserving and orientation preserving the product of the eigenvalues = $\det(Df_p) = 1$.

If p is hyperbolic, then its stable set is an injectively immersed copy of \mathbb{R} , as is its unstable set, and these are usually referred to as the *stable and unstable manifold* of p . If p is a periodic point of period n and a hyperbolic fixed point of f^n , then $W^s(p)$ and $W^u(p)$ will denote the stable and unstable manifolds with respect to f^n .

We will say that an elliptic periodic point p of period n is *Moser stable* if it admits a fundamental system of neighborhoods which are closed disks D such that $f^n|_{\partial D}$ is minimal. A consequence of KAM theory is that this is the typical behavior for elliptic periodic points (see Theorem 6.3 below).

Definition 6.2. Suppose f is an area preserving diffeomorphism of an annular domain \mathbb{A}_0 and $\text{Per}_q(f)$ is the set of periodic points of f with period q . Assume that for all q

- i) the set $\text{Per}_q(f)$ is empty or finite and all periodic points are elliptic or hyperbolic;
- ii) every elliptic periodic point in $\text{Per}_q(f)$ is Moser stable, and the invariant, f minimal, Jordan curves bounding neighborhoods of an elliptic periodic point z have rotation numbers which are not constant in any neighborhood of z ; and
- iii) the intersection of any branches of hyperbolic periodic points are transverse and any two branches of the same periodic point have non-empty intersection.

Then f will be called *Moser generic*.

Combining the work of many authors, one can conclude the following:

Theorem 6.3. *If $\Sigma = S^2$ or \mathbb{A} and $r \geq 16$, then the set of Moser generic diffeomorphisms of Σ is a residual subset of the group $\text{Diff}_\omega^r(\Sigma)$ of C^r area preserving diffeomorphisms.*

We remark that the high smoothness appears to be required to have all elliptic points be Moser stable and the best published result of which we are aware is Douady [2], which requires $r \geq 16$. The genericity of the properties of hyperbolic periodic points of a Moser generic diffeomorphism follow from the Kupka-Smale theorem together with results of Mather [17], Pixton [18], and Robinson [20] (cf. Theorem 8.7 below).

The fact that the dynamics of the typical area preserving diffeomorphism is radically different from the dynamics of the typical dissipative diffeomorphism is amply illustrated in the following result of Mather.

Theorem 6.4 (Mather [17]). *If p is a hyperbolic fixed point of a Moser generic diffeomorphism of S^2 , then*

$$cl(W^s(p)) = cl(W^u(p)).$$

For a Moser generic diffeomorphism there are homoclinic points (i.e. points of $W^s(p) \cap W^u(p)$ in both components of $W^s(p) \setminus \{p\}$), but the reader should be cautioned that we do not know if the homoclinic points are dense in $W^s(p)$. In any case, this common closure of the stable and unstable manifold is a kind of organizing center for the dynamics. Figure 3 was constructed by plotting an orbit of the standard map, but had we picked a fixed point and plotted points on its stable manifold we would have obtained a visually indistinguishable picture. All this motivates the following definition.

Definition 6.5. Suppose $f : S^2 \rightarrow S^2$ is an area preserving diffeomorphism which is Moser generic and p is a hyperbolic point. The *instability set* Λ associated to p is defined by

$$\Lambda = cl(W^s(p)) = cl(W^u(p)).$$

The most fundamental open question concerning Moser generic diffeomorphisms would be answered by a proof of the following conjecture.

Conjecture 6.6. *If $f : S^2 \rightarrow S^2$ is an area preserving diffeomorphism which is Moser generic, then $\text{Per}(f)$ is dense in S^2 .*

Of course we would be satisfied with the proof that some other residual subset of $\text{Diff}_\omega^r(S^2)$ with r arbitrarily large has the property that all of its elements have a dense set of periodic points. This conjecture is intimately related with the question of whether there is a C^r “closing lemma” for area preserving C^r diffeomorphisms when $r \geq 2$. The C^r closing lemma asserts that by a C^r perturbation of a diffeomorphism a recurrent (or merely non-wandering) point can be made periodic. This is known in all dimensions even for volume preserving diffeomorphisms when $r = 1$ by a result of C. Pugh and R.C. Robinson [19], but is not known even on surfaces for $r \geq 2$. If Conjecture 6.6 is true, then the C^r closing lemma for area preserving C^r diffeomorphisms of S^2 holds when $r \geq 16$, because any C^r diffeomorphism could be C^r approximated by a Moser generic one with dense periodic points.

The strongest current result in the direction of Conjecture 6.6 is the following.

Theorem 6.7 (Franks and LeCalvez [10]). *Suppose $f : S^2 \rightarrow S^2$ is an area preserving diffeomorphism which is Moser generic. Then there are countably many instability sets, and their union is dense in S^2 . Equivalently, if $\text{Hyp}(f)$ denotes the set of hyperbolic periodic points, then both*

$$\bigcup_{p \in \text{Hyp}(f)} W^s(p) \quad \text{and} \quad \bigcup_{p \in \text{Hyp}(f)} W^u(p)$$

are dense in S^2 .

As a result of this, in order to prove Conjecture 6.6, it would suffice to show that the periodic points are dense in any instability set, or equivalently that the homoclinic points (i.e. points in $W^s(p) \cap W^u(p)$) are dense in the instability set associated to a hyperbolic periodic point p . At present this seems to be difficult, however. For example, we know of no diffeomorphism with a hyperbolic periodic point p and an open set U in which both $W^s(p)$ and $W^u(p)$ are dense but with no periodic points and no homoclinic points in U . Remarkably, however, we don’t know how to exclude this possibility even for Moser generic diffeomorphisms! It is difficult to imagine the structure of any such example, however.

7. PROPERTIES OF INSTABILITY SETS

From its definition the instability set associated to a hyperbolic periodic point p of a Moser generic diffeomorphism $f : S^2 \rightarrow S^2$ is a compact set invariant under f^n where n is the period of p . It has many other striking properties which we now summarize.

Theorem 7.1 (Franks and LeCalvez [10]). *Suppose $f : S^2 \rightarrow S^2$ is an area preserving diffeomorphism which is Moser generic and Λ is an f -invariant instability set. Then*

- All components of $S^2 \setminus \Lambda$ are open disks. For each such disk U there is an $n > 0$ depending on U such that $f^n(U) = U$.
- All periodic points in Λ are hyperbolic saddles.
- Λ is the closure of $W^s(q)$ for any $q \in \text{Per}(f) \cap \Lambda$. Also Λ is the closure of $W^u(q)$ for any $q \in \text{Per}(f) \cap \Lambda$.

Corollary 7.2. *If Λ_p and Λ_q are instability sets associated with periodic points p and q respectively, then either $\Lambda_p = \Lambda_q$ or $\Lambda_p \cap \Lambda_q = \emptyset$.*

We can also consider the rotation numbers of points in an instability set. However, rotation numbers are not defined for points on S^2 , so we must remove two fixed points to produce an (open) annulus.

Proposition 7.3. *Suppose Λ is an instability set and U is a component of $S^2 \setminus \Lambda$ satisfying $f^n(U) = U$. Then f^n has a fixed point in U .*

This is because $f^n : U \rightarrow U$ is an area preserving homeomorphism of an open disk. The Poincaré recurrence theorem asserts that almost all points of U are f^n recurrent, and then the Brouwer plane translation theorem implies that f^n has a fixed point.

Suppose $f : S^2 \rightarrow S^2$ is a Moser generic diffeomorphism and that Λ is an f -invariant instability set. If U_0 and U_1 are components of the complement of Λ which are invariant under f^n , then we may remove a fixed point of f^n from each to obtain an open f^n -invariant annulus in which Λ is essential. Translation numbers and rotation numbers can be defined for an open annulus in precisely the same way as Definition 3.3 and Definition 3.5, or alternatively, we could “blow up” the two fixed points rather than deleting them and obtain a closed annulus.

We will denote by $\rho_\Lambda(f^n, U_0, U_1)$ the set of all rotation numbers of points in Λ measured in the annulus formed in this way. It is not difficult to see that all these rotation numbers do not depend on the choice of particular fixed points, but only on the disks U_0 and U_1 containing them.

Theorem 7.4 (Franks and LeCalvez [10]). *Suppose $f : S^2 \rightarrow S^2$ is a Moser generic diffeomorphism and that Λ is an f -invariant instability set. If U_0 and U_1 are any components of the complement of Λ which are invariant under f^n , then $\rho_\Lambda(f^n, U_0, U_1)$ is a non-trivial closed interval. Every rational number in $\rho_\Lambda(f^n, U_0, U_1)$ is the rotation number (with respect to U_0 and U_1) of a periodic point in Λ .*

In other words, for any two components of the complement of Λ there is an entire interval of rates at which points rotate around them!

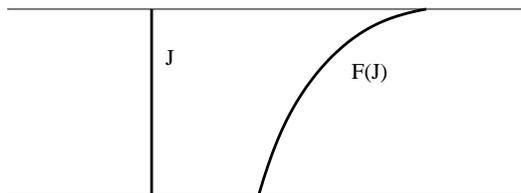
8. CONNECTIONS BETWEEN FRONTIER COMPONENTS

An important class of area preserving annulus diffeomorphisms, with numerous applications is the class of twist maps.

Definition 8.1. An area preserving diffeomorphism $f : \mathbb{A} \rightarrow \mathbb{A}$ is called a *twist map* if

$$\frac{\partial f_1}{\partial y} > 0$$

where $f(x, y) = (f_1(x, y), f_2(x, y)) \in \mathbb{T}^1 \times [0, 1]$.

FIGURE 4. The lift F of a twist map

Geometrically this means that a radial line segment

$$J_0 = \{(x, t) \in \mathbb{A} \mid 0 \leq t \leq 1\}$$

is twisted in a positive sense, or that in the lift a vertical line segment

$$J = \{(x, t) \in \tilde{\mathbb{A}} \mid 0 \leq t \leq 1\}$$

is bent in the positive direction (see Figure 4).

Theorem 8.2 (Birkhoff). *If $f : \mathbb{A} \rightarrow \mathbb{A}$ is a twist map, any essential component of the frontier of an invariant open set is the graph of a Lipschitz function; in particular it is a circle. Hence any essential component of the frontier of an instability set is topologically a Lipschitz circle.*

This is a very strong property which may be true more generally. However, it is unknown if for a twist map or a Moser generic diffeomorphism whether all components of the frontier must be topological circles. We do know that in the general (non-Moser generic, non-twist) case the frontier of an essential open invariant set can have components which are bad fractals.

Theorem 8.3 (Handel [12]). *There exists a C^∞ area preserving diffeomorphism of the annulus with an essential invariant continuum which is a pseudo-circle.*

Despite the fact that we do not know if essential components of the frontier of an instability set are topologically circles, it was shown by Mather using Caratheodory's theory of prime ends that there is always a well defined rotation number for any component of the frontier of an instability set of a Moser generic diffeomorphism. Indeed Mather proves the following important result.

Theorem 8.4 (Mather [17]). *Any component of the frontier of a Moser generic diffeomorphism contains no periodic points and has an irrational rotation number.*

Mather's result actually applies somewhat more generally than the Moser generic setting (see [17] for details).

For twist maps then, we have several interesting properties. If Λ is an instability set which is essential in the annulus, there are two distinguished components, X and Y , of the frontier of Λ , namely those which are also the frontier of the components of $\mathbb{A} \setminus \Lambda$ which contain the boundary components of \mathbb{A} . They are the only components of the frontier of Λ which are essential in \mathbb{A} . The beautiful result of Birkhoff, Theorem 8.2 above, asserts that these are embedded Lipschitz circles. Birkhoff also proved another striking result about these components. He showed that if U and V are neighborhoods of X and Y respectively, then there are $n, m > 0$ such that $f^n(U) \cap V \neq \emptyset$ and $f^m(V) \cap U \neq \emptyset$. This implies that one cannot dynamically separate X and Y in the sense that any open invariant set containing one of them

must contain the other in its closure. In particular there can be no invariant circles separating them.

This result was subsequently significantly strengthened by John Mather, who proved the following.

Theorem 8.5 (Mather [16]). *If $f : \mathbb{A} \rightarrow \mathbb{A}$ is a twist map and X and Y are the two essential components of the frontier of B , then*

$$W^s(X) \cap W^u(Y) \neq \emptyset \text{ and } W^u(X) \cap W^s(Y) \neq \emptyset.$$

It would be nice to know the analogous statement for other components of the frontier of an essential instability set. It would also be nice to know this with hypotheses more general than those for a twist map. For example, an instability set of an area preserving diffeomorphism of S^2 has no distinguished frontier components and no distinguished co-ordinates in which to measure twist. At least in the Moser generic case we have precisely this result.

Theorem 8.6 (Franks and LeCalvez [10]). *Suppose $f : S^2 \rightarrow S^2$ is Moser generic and Λ is an instability set. If X and Y are any two components of the frontier of Λ , then*

$$W^s(X) \cap W^u(Y) \neq \emptyset \text{ and } W^u(X) \cap W^s(Y) \neq \emptyset.$$

In some ways the components of the frontier of an instability set are like periodic points (for example, they are periodic). Pursuing this analogy we should compare Theorem 8.6 to the following result.

Theorem 8.7 (D. Pixton [18], C. Robinson [20]). *Suppose $f : S^2 \rightarrow S^2$ is Moser generic and Λ is an instability set. If $p, q \in \Lambda$ are periodic points, then*

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

9. QUESTIONS ABOUT MOSER GENERIC DIFFEOMORPHISMS

The study of Moser generic diffeomorphisms (or any other residual subset of C^r area preserving diffeomorphisms with $r \geq 2$) is a area with many more questions than answers. The results above as well as some additional results in [10] provide only a start in this investigation.

Here we list some open questions about instability sets. They are framed in the context of Moser generic diffeomorphisms, but answers would be very interesting for any large class of C^r diffeomorphisms with r larger than 1. It may be useful to view these questions after a further look at the computer image of Figure 3.

For all these questions we suppose $f : S^2 \rightarrow S^2$ is a C^r , $r \geq 2$, Moser generic diffeomorphism preserving Lebesgue measure and Λ is an instability set. Instability sets share some properties of hyperbolic sets (though they are never hyperbolic except in the case of an Anosov diffeomorphism on the torus). In particular it may be that instability sets are topologically transitive and have dense periodic points.

Question 9.1 (Transitivity). *Is there always a point in Λ whose orbit is dense in Λ ? Are the periodic points of Λ always dense in Λ ?*

An affirmative answer to the second question combined with Theorem 6.7 would, of course, imply Conjecture 6.6. However, it is conceivable that this conjecture could be true without the periodic points being dense in each instability set.

Question 9.2 (The frontier components). Are the components of the frontier of Λ topological circles? Is the restriction to such a component a minimal homeomorphism?

It is quite possible that even when a component of the frontier is a circle that the restriction of f is not minimal, despite the fact that it has irrational rotation number by the result of Mather cited above (Theorem 8.4).

Question 9.3 (The size of instability sets). Does every instability set Λ have positive measure? Can Λ have interior? Can the complement of Λ have finitely many components? Indeed, must the complement even be non-empty?

Of course, for diffeomorphisms on the torus \mathbb{T}^2 we have the example of linear Anosov diffeomorphisms. These are Moser generic and have a single instability set equal to all of \mathbb{T}^2 . This seems likely to be a phenomenon which cannot occur on S^2 .

Closely related is the following question which really displays the depth of our ignorance about Moser generic diffeomorphisms.

Question 9.4 (Elliptic Points). If $f : S^2 \rightarrow S^2$ is a Moser generic C^r , $r \geq 2$, diffeomorphism preserving Lebesgue measure, must it have infinitely many elliptic periodic points? Indeed must it have at least one elliptic periodic point?

REFERENCES

1. G.D. Birkhoff. Proof of Poincaré's Geometric Theorem, *Trans. Amer. Math. Soc.*, **14** (1913) 14–22.
2. R. Douady. Application du théorème des tores invariants, Thèse de troisième cycle, Univ. Paris 7, (1982).
3. J. Franks. Recurrence and fixed points of surface homeomorphisms, *Ergod. Th. Dynam. Sys.*, **8*** (1988), 99–107. MR **90d**:58124
4. J. Franks. Generalizations of the Poincaré-Birkhoff Theorem, *Annals of Math. (2)* 128:139–151, 1988. MR **89m**:54052
5. J. Franks. Geodesics on S^2 and periodic points of annulus homeomorphisms, *Inventiones Math.*, 108:403–418, 1992. MR **93f**:58192
6. J. Franks. Rotation numbers for Area Preserving Homeomorphisms of the Open Annulus, *Proceedings of the International Conference Dynamical Systems and Related Topics*, K. Shiraiwa, ed., World Scientific (1991), 123–128. MR **93e**:58153
7. J. Franks. Area Preserving Homeomorphisms of Open Surfaces of Genus Zero, *New York Jour. of Math.* 2:1–19, 1996. MR **97c**:58123
8. J. Franks. Rotation vectors and fixed points of area preserving surface diffeomorphisms. *Trans. Amer. Math. Soc.* 348:2637–2662, 1996. MR **96i**:58143
9. J. Franks and M. Handel. Periodic Points of Hamiltonian Surface Diffeomorphisms, preprint.
10. J. Franks and P. LeCalvez. Regions of instability for non-twist maps, to appear in *Ergodic Theory and Dynamical Systems*.
11. M. Handel. The rotation set of a homeomorphism of the annulus is closed, *Comm. Math. Phys.*, **127** (1990), no. 2, 339–349. MR **91a**:58102
12. M. Handel. A pathological area preserving C^∞ diffeomorphism of the plane, *Proc. A.M.S.*, **86** (1982), 163–168. MR **84f**:58040
13. N. Hingston. On the growth of the number of closed geodesics on the two-sphere, *Internat. Math. Res. Notices* 1993, no. 9, 253–262. MR **94m**:58044
14. A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, Cambridge, 1995. MR **96c**:58055
15. J. Mather. Topological proofs of some purely topological consequences of Caratheodory's Theory of prime ends, in *Selected Studies*, Th. M. Rassias, G. M. Rassias, eds., North-Holland (1982), 225–255. MR **84k**:57004

16. J. Mather. Variational construction of connecting orbits, *Ann. Inst. Fourier, Grenoble*, **43** 5 (1994), 1349–1386. MR **95c**:58075
17. J. Mather. Invariant subsets of area-preserving homeomorphisms of surfaces, *Advances in Math. Suppl. Studies*, **7B** (1981), 531–562. MR **84j**:58069
18. D. Pixton. Planar homoclinic points, *J. of Diff. Eq.*, **44** (1982), 365–382. MR **83h**:58077
19. Charles C. Pugh and C. Robinson. The C^1 closing lemma, including Hamiltonians, *Ergodic Theory Dynam. Systems*, **3** (1983), no. 2, 261–313. MR **85m**:58106
20. C. Robinson. Closing stable and unstable manifolds on the two sphere, *Proc. A.M.S.*, **41** (1973), 299–303. MR **47**:9674

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS 60208-2730
E-mail address: john@math.northwestern.edu