# Chapter 2

# Homoclinic Points and the Horseshoe Map - A Reprise

As mentioned in the introduction, a central contribution of this thesis is the identification and visualization of *homoclinic tangles* and *horseshoes* in turbulent geophysical flows. This chapter provides an overview of the theoretical development of homoclinic and horseshoe dynamics. To this end, this chapter will not provide new results, but will rather furnish the mathematical setting and context for results that follow in later chapters.

Beyond simply providing background information, a review of this mathematical foundation will aid in identifying and appreciating the structures that will be subsequently demonstrated for turbulent flows in the ocean and atmosphere. Emphasis will be placed on the contributions of Poincaré and Smale; although, as will be seen, there are many participants, each of whose contributions are interwoven to produce a delightful panorama of mathematics. The narrative of the mathematical results will be historical in nature – the historical development of the ideas is too rich a story not to tell, and the appreciation of which heightens understanding of the main results.

A further reason for providing this introductory survey of homoclinic and horseshoe dynamics is to ensure that this thesis is self-contained and accessible to a broad audience of oceanographers and atmospheric scientists. For more mathematical details, the reader is directed to the texts of [Guckenheimer 1983] and [Wiggins 2003]. An excellent mathematical overview is provided in the survey of [Holmes 1990], and

the discussion here follows a similar progression of ideas. A detailed mathematical account from a historical perspective is provided by [Barrow-Green 1996] (also see the less technical account of [Szpiro 2007]). Of course, greatest appreciation for the beauty of these ideas is best obtained by studying the primary texts of [Poincaré 1899] and [Smale 1967].

This chapter will also help to provide definitions and notation that will be used uniformly throughout the thesis.

## 2.1 Poincaré and homoclinic points

The mathematical contribution of Poincaré that will concern us most in the present treatise is the discovery of homoclinic points and asymptotic solutions – a theory he developed during his study of the restricted three-body problem. Poincaré used a geometric and qualitative approach. In the introduction to his original memoir, he instructs his readers that [Barrow-Green 1996]:

> These theorems have been given in a geometric form which has to my eyes the advantage of making clearer the origin of my ideas ...

Poincaré's approach signaled the birth of geometric mechanics, and his initial exposition would be extended by the topological approach to dynamics of Birkhoff and Smale.

First we demonstrate the method Poincaré used to reduce a two-degree of freedom Hamiltonian system to a time-dependent Hamiltonian system with a single degree of freedom. The procedure uses two main ingredients: (1) the Hamiltonian integral, and (2) the replacement of time by a time-like coordinate. The particulars of the restricted three-body problem addressed by Poincaré are not essential to our purpose, and consequently we shall omit details and consider a general Hamiltonian system (see [Szebehely 1967] for a detailed treatment of the restricted three-body problem).

Consider the time-independent Hamiltonian of a system with two degrees of freedom:

$$H(q_1, q_2, p_1, p_2)\,. \tag{2.1}$$

Hamilton's equations of motion are

$$\dot{q}_1 = \frac{\partial H}{\partial p_1} \qquad\qquad \dot{p}_1 = -\frac{\partial H}{\partial q_2} \tag{2.2}$$

$$\dot{q}_2 = \frac{\partial H}{\partial p_2} \qquad\qquad \dot{p}_2 = -\frac{\partial H}{\partial q_2}\,. \tag{2.3}$$

The derivative of the Hamiltonian with respect to time is:

$$\frac{dH}{dt} = \frac{\partial H}{\partial q_1}\dot{q}_1 + \frac{\partial H}{\partial q_2}\dot{q}_2 + \frac{\partial H}{\partial p_1}\dot{p}_1 + \frac{\partial H}{\partial p_2}\dot{p}_2 = 0\,, \tag{2.4}$$

so that we may immediately conclude that the Hamiltonian is constant:

$$H(q_1, q_2, p_1, p_2) = h = \text{constant}\,. \tag{2.5}$$

The availability of this integral relationship implies that we can solve for one of the coordinates as a function of the remaining three coordinates and the constant $h$. That is,

$$p_1 = p_1(q_1, q_2, p_2; h)\,. \tag{2.6}$$

Our next step is to replace time by another of the coordinates, $q_1$. In this way, the coordinate $q_1$ becomes an independent time-like variable. In doing so, we lose the true time parameterization of the trajectories and retain only the geometry of the solution – the time parameterization can always be reconstructed after solving the reduced system.

Using the identities:

$$\dot{q}_2 = \frac{dq_2}{dq_1}\dot{q}_1\,, \qquad \dot{q}_2 = \frac{\partial H}{\partial p_2}\,, \quad \text{and} \quad \dot{q}_1 = \frac{\partial H}{\partial p_1}\,, \tag{2.7}$$

we can write

$$\frac{dq_2}{dq_1} = \frac{\frac{\partial H}{\partial p_2}}{\frac{\partial H}{\partial p_1}}\,, \tag{2.8}$$

so that time has been eliminated in favor of the new independent variable $q_1$.

Likewise,

$$\frac{dp_2}{dq_1} = -\frac{\frac{\partial H}{\partial q_2}}{\frac{\partial H}{\partial p_1}}\,. \tag{2.9}$$

When evaluating the partial derivatives of $H$ in the right-hand sides of (2.8) and (2.9), we use the expression for $p_1$ in (2.6) to write the result entirely in terms of the dependent variables $q_2$ and $p_2$, the independent variable $q_1$, and the constant $h$. Thus, the original Hamiltonian system has been reduced to

$$\frac{dq_2}{dq_1} = f(q_2, p_2; q_1; h) \tag{2.10}$$

$$\frac{dp_2}{dq_1} = g(q_2, p_2; q_1; h)\,, \tag{2.11}$$

a $q_1$-dependent (think, "time"-dependent), *single* degree of freedom system. What is more, the new system is in fact Hamiltonian, with the $q_1$-dependent Hamiltonian,

$$\bar{H} := -p_1(q_2, p_2; q_1; h)\,. \tag{2.12}$$

After solving this reduced Hamiltonian system, the full solution can be reproduced by chasing back up through the definitions.

In the restricted three-body problem, Poincaré chose an appropriate change of variables so that the $q_1$ coordinate corresponded to an angle variable with period $2\pi$. Hence, the problem was reduced to a single degree of freedom Hamiltonian system

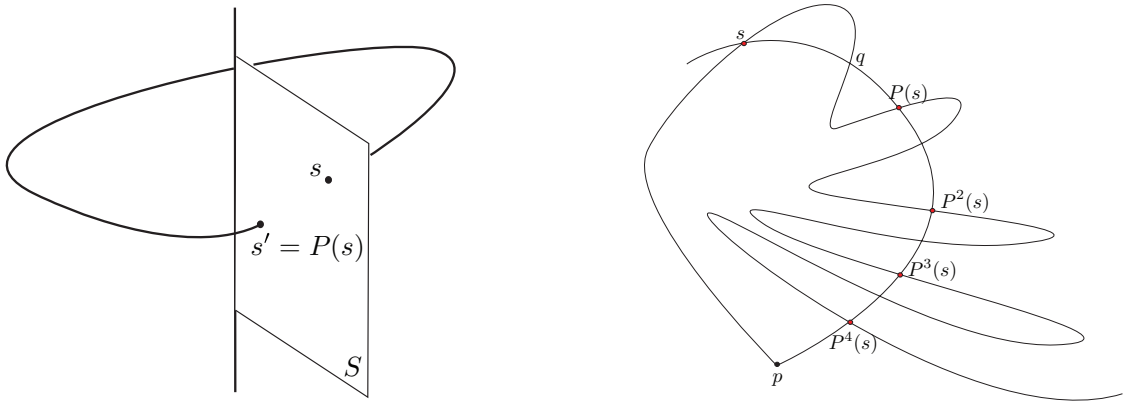



Figure 2.1: (a) The flow of the reduced Hamiltonian, $\bar{H}$, induces a discrete map, $P : S \to S$, on the Poincaré section, $S$. (b) Intersections of the stable and unstable manifolds of the fixed point $p$ on the Poincaré section $S$.

with periodic forcing. Many of the examples that we shall encounter in later chapters will be of this form.

The $q_1$-dependence of the reduced Hamiltonian implies that when viewed in the $(q_2, p_2)$-plane, trajectories may intersect, and consequently precludes a qualitative geometric analysis. Poincaré's insight was to reduce the system even further to remove the intersections and provide a clearer picture.

In the three-dimensional $(q_2, p_2; q_1)$-space, he introduced a two-dimensional cross-sectional surface, $S$, (now called a *Poincaré section*) defined by the plane $q_1 = 0$, such that all trajectories pass transversally through the surface. The transverse intersection of a trajectory with $S$ defines a point $s \in S$. Since the system is periodic in $q_1$, the trajectory returns to the surface $S$ at a point $s'$ after time $2\pi$. In this way, a study of the flow of the dynamical system under the reduced Hamiltonian $\bar{H}$, reduces to a study of a discrete *Poincaré map*, $P : S \to S$, as illustrated in Figure 2.1(a).

Poincaré referred to the locations of forward iterations of the map as *consequents*, and to the locations of iterations of the inverse map, $P^{-1} : S \to S$, as *antecedents*.

If $\bar{H}$ is smooth, then $P$ will be an orientation-preserving map. Furthermore, since the flow under $\bar{H}$ is Hamiltonian and preserves volume, the map $P$ will preserve area.

A *fixed point* of the map $P : S \to S$ is a point $p \in S$ such that $P(p) = p$. Clearly, a fixed point on $S$ corresponds to a periodic trajectory in $(q_2, p_2; q_1)$-space. Just as with fixed points in continuous dynamical systems, we can investigate the *stability* of $p$ under iterations of the map. To do so, we first compute the *linearization* of the map P at $p$:

$$s \to DP(p) \cdot s\,, \tag{2.13}$$

where $DP(p)$ is represented by a 2-by-2 matrix. Thus, the linearization is a linear map from $S$ to itself. If the two eigenvalues of $DP(p)$, denoted $\lambda_s$ and $\lambda_u$, are such that $|\lambda_s| < 1 < |\lambda_u|$, then the fixed point is said to be *hyperbolic*. The eigenvectors associated with the eigenvalues of the linearization of a hyperbolic fixed point define linear subspaces, $E^s$ and $E^u$, that are invariant under the flow of the linearization map, $DP(p)$. The stable/unstable manifold theorem for maps states that for the fully nonlinear map $P : S \to S$, in a neighborhood of $p$, there exist local invariant manifolds $W^s_{loc}$ and $W^u_{loc}$ that are tangent to the linear subspaces $E^s$ and $E^u$ at $p$, respectively; and are as smooth as the map $P$. Finally, iterating points on each of these local manifolds reveals the global stable and unstable manifolds:

$$W^s = \bigcup_{k=0}^{\infty} P^{-k}(W^s_{loc}(p)) \quad \text{and} \quad W^u = \bigcup_{k=0}^{\infty} P^{k}(W^u_{loc}(p))\,. \tag{2.14}$$

Successive consequents of a point $s$ on the stable manifold will asymptote to $p$ along $W^s$, while successive antecedents of a point $u$ on the unstable manifold will asymptote to $p$ along $W^u$.

It should be made clear that these asymptotic manifolds, associated with the fixed point $p$, cannot self-intersect since this would violate the smoothness condition of the mapping $P$. Nevertheless, the two asymptotic manifolds *may intersect each other*. In his memoir, Poincaré refers to these points of transverse intersection as *doubly asymptotic* points, referring to the manner in which they asymptote to the same fixed point in both forward and backward time. Later, in the third volume of *Method Nouvelles* [Poincaré 1899], he referred to these points as *homoclinic points*, and also

introduced the concept of *heteroclinic points* — points whose consequents asymptote toward a fixed point $p_+$, but whose antecedents asymptote toward a different fixed point $p_-$.

The remaining discussion in this chapter is essentially an attempt to mathematically describe the implications resulting from the fact that these asymptotic manifolds intersect. A first consequence that is trivial to derive, yet has dizzying implications, is the fact that a single intersection implies infinitely many intersections. Starting with a homoclinic point $s \in W^s \bigcap W^u$, we realize that by definition, $P(s)$ also lies on both the stable and unstable manifolds. Repeated applications of the map reveal an infinite number of intersections.

An important point is that since the map $P$ is orientation preserving, the unstable manifold will have to double-back and intersect at an intermediate point $q$, before intersecting again at the point $P(s)$ as shown in Figure 2.1(b). Geometrically, we begin to envision the unstable manifold doubling back repeatedly across the stable manifold as the points of intersection asymptote toward the periodic point. In a similar way, when we consider antecedents, the stable manifold doubles back across the unstable manifold as they asymptote toward the fixed point in backward time.

The intersections and doubling back of the manifolds define closed regions in the section $S$, called *lobes*. As the iterated map takes successive consequents closer and closer to the fixed point, the distance between intersections that define the base of the lobe becomes arbitrarily small. Consideration of the fact that the map is area-preserving implies that the lobe must become longer and longer after each iteration while never intersecting neighboring lobes. Hence, the manifold that outlines the lobes oscillates wildly as it doubles back and forth across the other manifold. Lobes from each of the manifolds will inevitably intersect one another creating even more homoclinic points yielding a complicated picture that is referred to today as a *homoclinic tangle*. It is noteworthy that the map need not be area-preserving in order for the tangle to form – the linear rate of compression and expansion along the stable and unstable linear subspaces associated with the hyperbolic fixed point are sufficient

to produce the tangle.

In his memoir, Poincaré declined to draw the homoclinic tangle, but his oft-quoted description of the situation is provided here [Poincaré 1899]:

> If one seeks to visualize the pattern formed by these two curves and their infinite number of intersections, each corresponding to a doubly symptotic solution, these intersections form a kind of lattice-work, a weave, a chain-link network of infinitely fine mesh; each of the two curves can never cross itself, but it must fold back on itself in a very complicated way so as to recross all the chain-links an infinite number of times.
>
> One will be struck by the complexity of this figure, which I am not even attempting to draw. Nothing can give us a better idea of the intricacy of the three-body problem, and of all the problems of dynamics in general, when there is no uniform integral and the Bohlin series[1] diverge.

The homoclinic tangle is considered the trademark of *chaos*. Although the map is both differentiable and deterministic, the effect of its repeated action is to tear apart nearby trajectories in such a way that the trajectories can have arbitrarily different outcomes and may have arbitrarily complicated behavior. For instance, Poincaré showed in *Method Nouvelle* Volume I [Poincaré 1899] that given any particular solution, there exists a periodic solution of a sufficiently large period that is arbitrarily close to the particular solution for an arbitrary long period of time. The full implications of the homoclinic point will be discussed later in the context of the Smale horseshoe map which provides an elegant description of the chaotic dynamics near the homoclinic point.

Amongst the mathematical community, homoclinic points were not thoroughly investigated after Poincaré, and became largely forgotten. Smale described the situation in this way [Smale 1998]:

> Unfortunately, the scientific community soon lost track of the important ideas surrounding the homoclinic points of Poincaré. In the conferences in differential equations and dynamics that I attended in the late 50's, there was no awareness of this work. Even Levinson never showed in his book, papers, or correspondence with me that he

[1] Bohlin had published a series method just before the submission of his memoir.

> was aware of homoclinic points. It is astounding how important scientific ideas can get lost, even when they are exposed by leading mathematicians of the preceeding decades.

This being said, there were several mathematicians who were influenced by Poincaré's work, and kept his ideas alive.

Soon after the death of Poincaré, Birkhoff, a young American mathematician, proved Poincaré's "Last Geometric Theorem," and continued to develop the theory of dynamics from this geometric perspective. Morse, a student of Birkhoff, developed the theory of *symbolic dynamics* in which the motion of a trajectory is abstracted as a sequence of labels, a key to the topological treatment of chaotic trajectories.

Hadamard, a French mathematician, greatly admired Poincaré's work and studied geodesics on curved surfaces using many of the tools Poincaré had developed. He used Poincaré's Recurrence Theorem and the Poincaré section to analyze geodesics. He showed that in the neighborhood of every bounded geodesic, there were infinite unbounded geodesics. Thus, small changes in the initial condition could lead to geodesics of entirely different character. Hadamard observed this *sensitivity to initial conditions* (a hallmark of chaos encoded in the homoclinic tangle), and realized that it was nonsensical to discuss stability of the solar system, for example, since outcomes were not continuous with respect to initial conditions [Barrow-Green 1996].

## 2.2 Smale and the horseshoe map

The work of Poincaré and Birkhoff continued with Smale who provided an elegant topological description of the chaotic nature of the trajectories in the homoclinic tangle. By abstracting the dynamics of the tangle as a differentiable map, he showed conjugacy between the chaotic trajectories envisioned by Poincaré, and a shift automorphism on a sequence of binary symbols. In essence, he showed that the trajectories in a homoclinic tangle are as chaotic as a repeated coin flipping experiment.

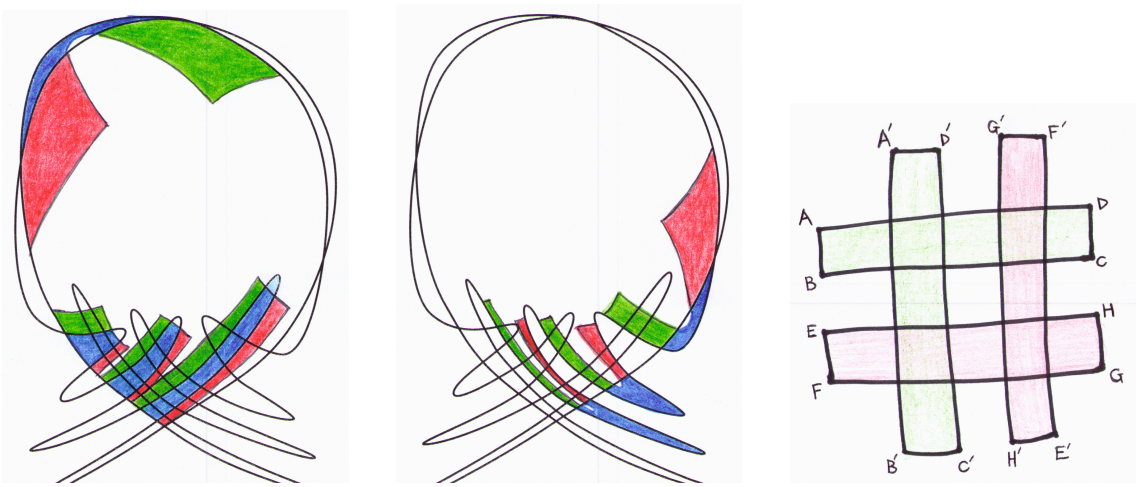
At the time of his discovery of the horseshoe map and its relationship to homoclinic solutions, Smale was a postdoctoral scholar at IMPA in Brazil [Smale 1998]. In

the IMPA library, he browzed the collected works of Birkhoff[2] and started to work in the area of topological dynamics. After publishing a paper in which he claimed "chaos cannot exist", he received a letter from Levinson that caused him to rethink his conclusion. Levinson directed him to one of his papers that provided a counterexample to Smale's claims. Levinson's paper [Levinson 1949] provided an analysis of a simplified model for radar circuits that had been studied previously during the second world war by Cartwright and Littlewood. Cartwright and Littlewood investigated the anomalous behavior of radar circuits that had hampered military field operations. A study of the solutions of the differential equations governing the circuitry revealed "a rich variety of behaviour, some of it very bizarre" [Cartwright 1945]. Their analysis showed that the system possessed an infinite number of periodic motions of "a great variety of structures." They were in fact observing the complicated dynamics associated with a homoclinic tangle. In his study, Levinson provided a similar result although on a somewhat simpler system in an attempt to clarify their results.

Smale assimilated these results from Birkhoff, Levinson, Littlewood, and Cartwright, and developed the Smale horseshoe map to provide a complete description of these chaotic trajectories.

The horseshoe map is an abstraction of the flow near a homoclinic tangle. First, consider what happens to a rectangular patch on the surface of section $S$ in the vicinity of a hyperbolic fixed point under the iterated application of the map. In Figure 2.2, a rectangular patch, consisting of three strips colored red, blue, and green, is mapped forward around the tangle. (Figure 2.2(a) contains three iterates of the map, after which, Figure 2.2(b) illustrates an additional three iterates of the map.) In the immediate vicinity of the fixed point, the linearized flow dominates, and we see that the patch is squeezed in one direction and lengthened in the other. Under repeated action of the map, the patch is eventually brought again into the region near the fixed point. Thus for some positive integer $m$, the map $Q := P^m$ is a map from the patch into itself. A single iteration of the map, $Q$, yields an interesting outcome.

[2]Smale's PhD advisor was Bott, a student of Duffin, in turn a student of Bourgin, who was a student of Birkhoff.




(a) (b) (c)

Figure 2.2: The action of the homoclinic tangle on a rectangular patch is illustrated here. The patch consists of three stripes colored red, blue, and green. Figure (a) shows the first three iterates of the map, and Figure (b) shows the last three. The red and green strips are returned to the patch, while the blue strip is removed. The initial and final orientation of the red and green strips is shown in (c), where $A'$ is the consequent of $A$. Both strips return with a vertical orientation, and the red strip is turned upside down.

First we notice that the middle blue "third" of the patch has been removed entirely, while the red and green "thirds" are reorientated. The portion of the patch that remains inside the tangle will be mapped again under the action of the dynamics. The crux of this exercise is to carefully observe which points remain inside the tangle.

Carefully keeping track of the red and green strips during the mapping around the tangle, allows us to determine their orientation once they return to the patch. Figure 2.2(c) shows the initial horizontal orientation of the strips, overlaid with the final vertical orientation of the strips. Not only have the strips been squeezed and stretched – the red strip has been turned upside down!

The action of the homoclinic tangle just described is abstracted as the *horseshoe map* $Q : S \to \mathbb{R}^2$ acting on the square $[0,1] \times [0,1] \subset \mathbb{R}^2$ and is shown graphically in Figure 2.3. Again, we consider three horizontal strips of the square: the first labelled $A$, the third $C$, while the middle strip $B$ does not return to the square under the action of the map.

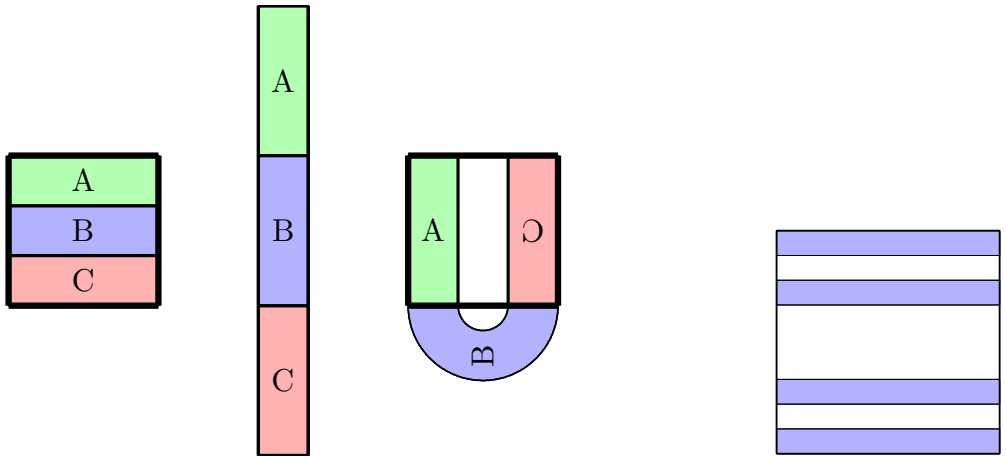



Figure 2.3: (a) Forward iteration of the Smale horseshoe map. (b) The set that remains in the square after two iterations of the *forward* map is colored blue.

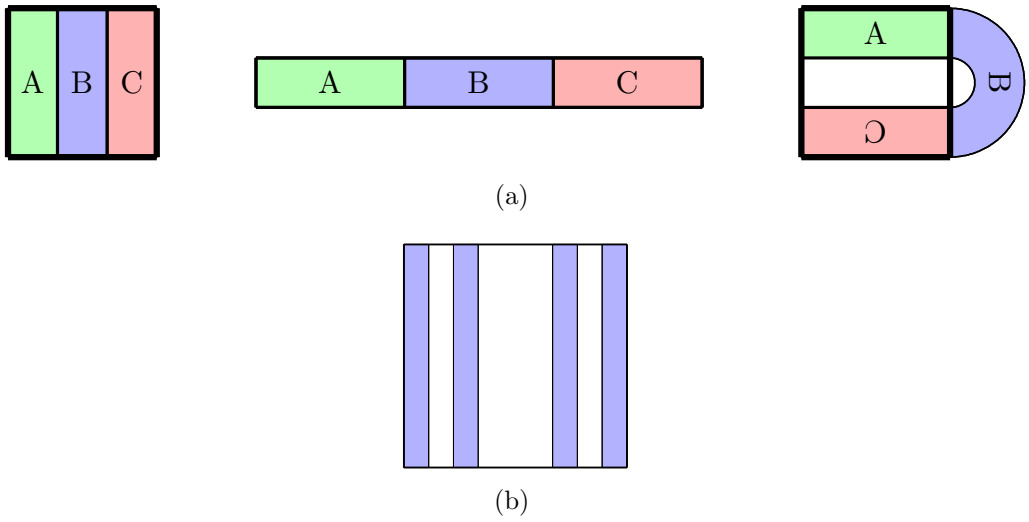
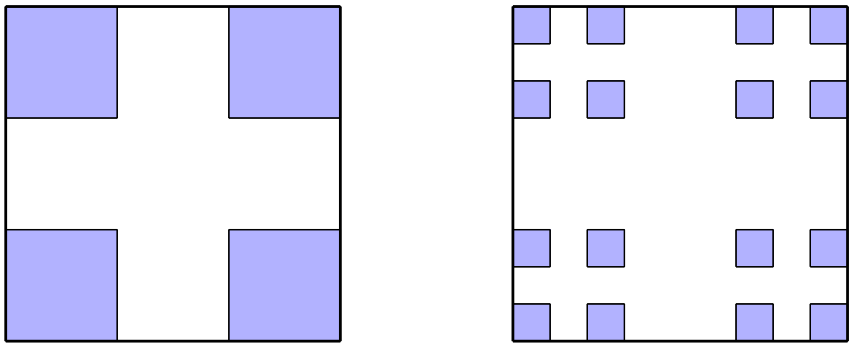



Figure 2.4: (a) Backward iteration of the Smale Horseshoe map. (b) The set that remains in the square after two iterations of the *backward* map is colored blue.



(a) The set that remains invariant under one iteration of both forward and backward maps.

(b) The invariant set after two iterations of both the forward and backward maps.

Figure 2.5: In the limit as the number of iterations of both the forward and backward maps approaches infinity, the invariant set approaches the Cantor set.

We may also consider the backward iteration of the map (Figure 2.4). In this case, a vertical "third" of the patch is removed upon each iteration. In the limit as the number of iterations both forward and backward becomes infinite (Figure 2.5), the invariant set of points that remain, denoted $\Lambda := \bigcap_{k=-\infty}^{\infty} Q^k(S)$, is a *Cantor set*. The set $\Lambda$ contains uncountably infinite points, but has measure zero.

We assign to each point in $\Lambda$, a symbolic representation of its entire bi-infinite trajectory. We begin with the map $\varphi : \Lambda \to \{0, 1\}$ that assigns to a point, $q \in \Lambda$, a symbol, either 0 or 1, according to the rule:

$$\varphi(q) = 0 \quad \text{if} \quad q \in A\,, \qquad \text{and} \qquad \varphi(q) = 1 \quad \text{if} \quad q \in C\,. \tag{2.15}$$

Next, using this symbol assignment map $\varphi$, we define the map $\phi : \Lambda \to \Sigma$ that maps a point $q \in \Lambda$ to the space of all possible bi-infinite sequences of the symbols 0 and 1 according to:

$$\phi(q) = \left[\varphi(P^k(q))\right]_{k=-\infty}^{\infty}\,. \tag{2.16}$$

By this construction, we now have a representation for every trajectory under $P$ as a bi-infinite ordered sequence of symbols that describes precisely how the trajectory

$$\begin{array}{ccc} \Lambda & \xrightarrow{P} & \Lambda \\ \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \phi} \\ \Sigma & \xrightarrow{\sigma} & \Sigma \end{array} \tag{2.18}$$

Figure 2.6: The dynamics of $P$ on the invariant set $\Lambda$ is conjugate to a shift map on the space of bi-infinite sequences, $\Sigma$.

meanders back and forth between $A$ and $C$. The action of the map results in a simple shift on the sequence of symbols. That is,

$$\phi(P(q)) = \sigma(\phi(q))\,, \tag{2.17}$$

where $\sigma : \Sigma \to \Sigma$ simply shifts the list of symbols one space to the left. Furthermore, the fact that both the images and pre-images of $A$ lie in both $A$ and $C$ (and *vice versa*), implies that to every bi-infinite sequence of symbols, we may assign a trajectory in $\Lambda$. Thus, $\phi : \Lambda \to \Sigma$ is a homeomorphism (continuous, invertible mapping) between the set of trajectories in $\Lambda$, and the set of bi-infinite sequences, $\Sigma$. By providing this construction, Smale showed that the two sets, $\Lambda$ and $\Sigma$, are *topologically conjugate*. The commutation diagram that illustrates the conjugacy between the map $P : \Lambda \to \Lambda$ and $\sigma : \Sigma \to \Sigma$ through $\phi : \Lambda \to \Sigma$ is shown in Figure 2.6.

The upshot of this construction is that the morass of complicated trajectories in the homoclinic tangle envisioned by Poincaré is completely and elegantly described by a consideration of the shift map on the space of bi-infinite sequences of two symbols. Specifically, Smale was then able to use this conjugacy to describe the possible orbits in $\Lambda$, and by considering all possible sequences of symbols, conclude that $\Lambda$ has

1. a countable infinity of periodic orbits with arbitrarily high period,

2. an uncountable infinity of periodic orbits, and

3. a dense orbit.

The dense orbit is obtained by considering the concatenation of all possible finite

sequences. Moreover, given any finite trajectory, it is easy to see how repeating the sequence produces a periodic orbit that is arbitrarily close.

The (Poincaré)-Smale-Birkhoff Homoclinic Theorem is stated precisely as follows:

**Theorem:** (Smale-Birkhoff Homoclinic Theorem)
*Suppose $q$ is a transversal homoclinic point of $P \in Diff(S)$. Then there is a Cantor set $\Lambda \in S$, and $N \in \mathbb{Z}^+$ such that $P^N(\Lambda) = \Lambda$, and $P^N(\cdot)$ restricted to $\Lambda$ is topologically a shift automorphism.*

The value of $N$ is chosen large enough, precisely to ensure that the square patch in $S$ is returned by the map $P$ to $S$ after its journey around the homoclinic tangle.

In sum, the conjugacy of the horseshoe map with a shift automorphism provides a succint framework in which to understand homoclinic dynamics. For instance, we see the existence of sensitivity to initial conditions: there exist two sequences with identical histories, but with very different futures.

The lobes defined by the intersections of the stable and unstable manifolds, and their evolution under the action of the Poincaré map, reveal the transport mechanism of *lobe dynamics*, that is responsible for entraining and detraining lobes both into and out of the homoclinic tangle, as illustrated in Figure 2.7. This transport mechanism will feature prominently in the time-dependent flows we shall consider in later chapters.

Finally, in terms of our present study it is important to note that the framework provided here by Poincaré and Smale is for a system that exhibits time periodic forcing. In fact, it is precisely this fact that allows for abstraction of the dynamics as the repeated application of a differentiable map. Some work has been done for analogs of this theory in quasi-periodic flows [Beigie 1991]; however, no concrete theory is currently available for aperiodic turbulent flows.

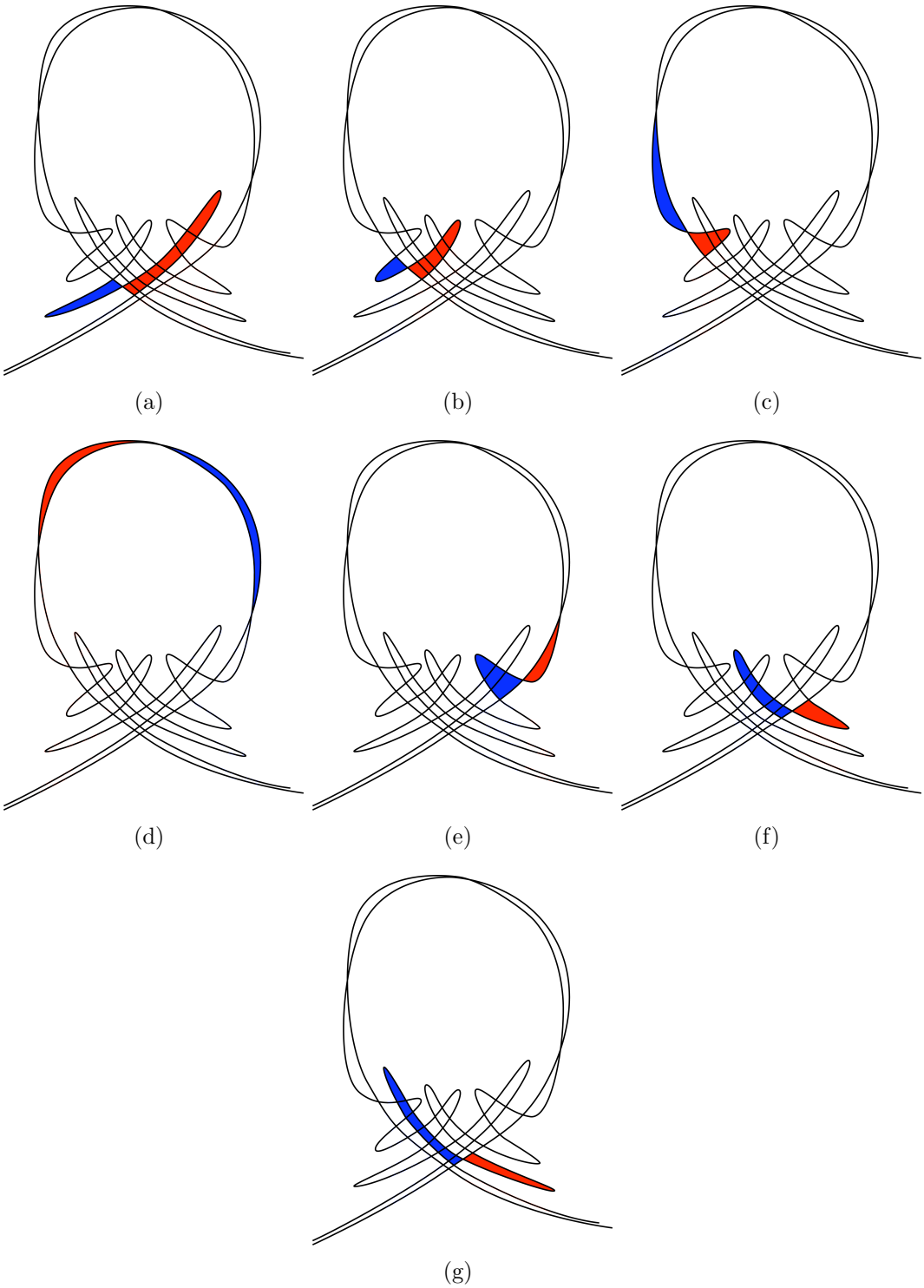



Figure 2.7: Transport via lobe dynamics in a homoclinic tangle. Under the action of the Poincaré map, the blue lobe is entrained into the vortex, while the red lobe is detrained.