PERGAMON

Nonlinear Analysis 41 (2000) 259–276

Nonlinear Analysis

www.elsevier.nl/locate/na

Homoclinic tangles for noninvertible maps

Evelyn Sander

Department of Mathematical Sciences, George Mason University, MS 3F2, 4400 University Drive, Fairfax, VA 22030, USA

Received 2 March 1998; accepted 17 July 1998

Keywords: Homoclinic tangles; Noninvertible maps; Transverse homoclinic point

1. Introduction

Not all naturally occurring iterated systems have the property that time is reversible. In which case, the dynamics must be modelled by a noninvertible map rather than by a diffeomorphism. While one-dimensional theory describes dynamics of noninvertible maps, higher-dimensional theory has focussed on diffeomorphisms, though there are many higher-dimensional examples of natural systems modelled by noninvertible maps. For example, noninvertible maps occur in the study of population dynamics [2], time one maps of delay equations [25], control theory algorithms [1, 5, 6], neural networks [16], and iterated difference methods [14]. In order to use dynamical systems theory in applications for which maps are noninvertible, it is important to know how the theory differs from the diffeomorphism case.

One of the principal indications of chaotic dynamics in deterministic systems is the existence of transverse homoclinic orbits. This paper examines the existence of homoclinic tangles and the associated dynamic behavior for noninvertible maps. In particular, we give examples of maps which have transverse intersection of stable and unstable manifolds but no horseshoe. Furthermore, we show that this is a codimension two phenomenon.

Papers by Steinlein and Walther [24,25] and Lerman and Shil'nikov [13] both establish conditions for a transverse homoclinic point of a noninvertible map to be contained in a horseshoe. This is discussed further in Section 4. The work of Laura Gardini [10] addressed transverse homoclinic orbits for planar maps. However, she missed the subtle distinction for noninvertible maps between a transverse homoclinic point and

E-mail address: sander@math.gmu.edu (E. Sander)

PII: S0362-546X(98)00277-6

transversality of every point on a homoclinic orbit. Thus she incorrectly concluded that a transverse homoclinic point implies the existence of a horseshoe. To our knowledge, no work has addressed bifurcations resulting from homoclinic orbits for noninvertible maps.

The paper is structured as follows: Section 2 contains definitions and background. It also contains a description of new phenomena that can occur for global invariant manifolds of noninvertible maps. Section 3 contains examples of maps with transverse homoclinic points with no nearby horseshoe. Section 4 states some conditions under which a homoclinic orbit has nearby chaotic behavior and shows that these conditions are generic. Section 5 describes the bifurcation picture for a two-parameter family containing such a no-horseshoe transverse homoclinic point, and shows that two-parameter families generically contain a map with such behavior.

2. Definitions and background

This section contains definitions and theorems regarding hyperbolic points, stable and unstable manifolds, homoclinic orbits, transversality, hyperbolicity, and shadowing.

Definition 2.1 (*Hyperbolic fixed point*). A fixed point p of a smooth map f is hyperbolic if Df_p has no eigenvalues on the unit circle. This does not exclude eigenvalue zero.

Definition 2.2 (*Stable and unstable manifolds*). For a map f with fixed point p, the stable and unstable manifolds $W^s(p)$ and $W^u(p)$ are the points with forward (resp. backward) orbits converging to p. Precisely:

$W^s(p)=\{z\in Z$: there exists an infinite forward orbit $\{z_k\}$ through z such that $z_k\to p$ as $k\to\infty\}$.

$W^u(p)=\{z\in Z$: there exists an infinite backward orbit $\{z_k\}$ through z such that $z_k\to p$ as $k\to-\infty\}$.

Definition 2.3 (*Local stable and unstable manifolds*). The local stable and unstable manifolds $W^s_U(p)$ and $W^u_U(p)$ are the points in the global stable and unstable manifolds which converge to p inside U, where U is some small neighborhood of p.

Note that $W^s(p)=\bigcup_{n\geq 0} f^{-n}(W^s_U(p))$ and $W^u(p)=\bigcup_{n\geq 0} f^{n}(W^u_U(p))$. For a proof of the following theorem, see [20] or [17].

Theorem 2.4 (Stable manifold theorem). *If f is a C^r map on R^n, and f has a hyperbolic fixed point p, then there exists a U such that $W^s_U(p)$ and $W^u_U(p)$ are graphs of C^r functions.*

2.1. New phenomena for global invariant manifolds

For diffeomorphisms, a corollary of the stable manifold theorem says that the global stable and unstable manifolds also have smoothness properties: namely, they are one-to-one immersed submanifolds of the appropriate dimensions. This does not hold for noninvertible maps. The global stable and unstable manifolds of smooth noninvertible maps are in general not even smooth. Here are some examples.

Example 2.5. The stable manifold can be disconnected. For example, consider the map $f : R^1 \to R^1$, where f is the quadratic map $f(x) = \mu x(1-x)$, $\mu > 4$. Then 0 is a fixed point with the stable manifold being the end points of a middle-third Cantor set, as these points eventually map onto 0. See for example [4].

Of course such disconnectedness can also occur in higher dimensions. For an example from adaptive control, see [1].

In contrast to the above example, the unstable manifold is always connected, since the local unstable manifold is connected, and the image of a connected set under a continuous map is connected.

Example 2.6. The stable manifold can increase in dimension. For example, let f be a map on R^2 such that for a hyperbolic fixed point, there is a one-dimensional curve γ of local stable manifold of f in an open set V; and f maps an open set U into V in such a way that the image of U is exactly γ. Then clearly there is a two-dimensional region contained in the stable manifold, with U a subset of this region.

Similarly, the unstable manifold can decrease in dimension. See [25] for a delay equation example.

Example 2.7. The stable and unstable manifolds can fail to be smooth.

If there is a neighborhood not near the fixed point such that the unstable manifold is of the form constant + $\{x = y\}$, and the map f is in this neighborhood of the form constant + $(x, y) \mapsto$ constant + (x^2, y^3), we get a cusp.

Similarly, if on a neighborhood V, W^s is of the form $\{(x, x^3)\}$, and $f : U \to V$ is of the form $(x, y) \mapsto (x, y^2)$, then in U, W^s has a cusp.

Example 2.8. The unstable manifold can have self-intersections.

We will not give an example here, as this concept is well known; all snap-back repellers [15] are by definition the case in which a full-dimensional unstable manifold intersects itself. A more visually clear variation of this occurs when a less than full-dimensional unstable manifold intersects itself. See examples in [2, 22, 25].

Many numerical examples of global stable and unstable manifolds for specific maps have been computed. See for example [12]. [7] gives a more detailed overview of new phenomena for global manifolds of noninvertible maps.

2.2. Homoclinic tangles

Definition 2.9 (*Homoclinic point*)**.** If a smooth map has a hyperbolic fixed point p with global stable and unstable manifolds W^s and W^u, then a homoclinic point is a point in $W^s \cap W^u \setminus \{p\}$.

Definition 2.10 (*Orbit*)**.** A sequence $\{z_k\}$ is called an orbit for f if $f(z_{k-1})=z_k$.

Definition 2.11 (*Homoclinic orbit*)**.** A bi-infinite orbit $\{z_k\}$ is called a homoclinic orbit to the fixed point p, if $\lim_{k\to\infty} z_k = \lim_{k\to-\infty} z_k = p$.

Through an arbitrary point in a noninvertible map, there may be multiple orbits or no bi-infinite orbit. However, from the definitions of stable and unstable manifolds, through each homoclinic point, there is a homoclinic orbit.

Definition 2.12 (*Transverse homoclinic point*)**.** Assume a map has a hyperbolic fixed point with stable and unstable manifolds W^s and W^u. A transverse homoclinic point is a point at which W^s and W^u intersect transversally.

For a diffeomorphism, the homoclinic orbit through a transverse homoclinic point only contains transverse homoclinic points. Thus a simple geometric condition at just one point gives information about the entire orbit. Further, let U be a neighborhood containing a hyperbolic fixed point p and a transverse homoclinic point q. Then Smale's theorem says that U contains a compact hyperbolic invariant set K, and there is some n such that on K, f^n is topologically conjugate to the shift map on two symbols. In other words, the existence of a transverse homoclinic point implies chaotic dynamical behavior. Further, transverse intersection of manifolds is stable under perturbation. In contrast, it will be shown that for noninvertible maps, the mere existence of a transverse homoclinic point implies nothing about the dynamics. Nor is transverse intersection even stable under perturbation. This is a codimension two phenomenon.

The key observation of Smale's proof is that there exists a hyperbolic structure on the closure of the orbit of q. (The closure of the orbit of q is equal to the orbit union p.) Precisely, setting $E_x^s = T_xW^s$ and $E_x^u = T_xW^u$, it is possible to show that the homoclinic orbit has the proper expanding and contracting behavior. Once this is verified, the proofs of the two statements in the above theorem follow easily by prescribing pseudoorbits and using the shadowing lemma. This is the key to why the theorem does not hold in the noninvertible case.

2.3. Hyperbolicity and shadowing for noninvertible maps

The following is a discussion of the concepts of hyperbolicity and shadowing for noninvertible maps. These definitions and theorems follow [22], which gives these theorems in the more general class of smooth multivalued noninvertible maps. Also see [24] and [25] for another approach to hyperbolicity and shadowing for noninvertible maps. The results in this section follow from either treatment.

Definition 2.13 (*Stable and unstable cones*)**.** Given $\alpha > 0$, and a splitting of the tangent space at each point $T_zR^n = E_z^s \times E_z^u$, then the stable and unstable α cones are defined by

$$C_z^s = \{(v_s, v_u) \in E_z^s \times E_z^u\colon\ |v_u| \le \alpha|v_s|\}$$

$$C_z^u = \{(v_s, v_u) \in E_z^s \times E_z^u\colon\ |v_s| \le \alpha|v_u|\}.$$

Definition 2.14 (*Cone condition*). Let f be a smooth map and K a compact set. Then f satisfies the cone condition on K if there is some continuous splitting and a continuous metric, and a uniform $\lambda < 1$ such that for all $(z, w) \in K \times K$, and $w = f(z)$, vectors in the unstable λ cone at z map to vectors in the unstable λ cone at w under Df_z, and these vectors are backwards λ-contracting. In other words, if a vector v in the unstable cone maps to a vector v', then $|v| < \lambda |v'|$.

Similarly, vectors in the stable λ cone at w only come from the stable λ cone at z under Df_z, and are λ-contracting.

The following definition of hyperbolicity is in terms of stable and unstable cones. See [18] for such a theory for diffeomorphisms.

Definition 2.15 (*Hyperbolicity*). A compact set K is said to be hyperbolic for f if K satisfies the cone condition for f. The subspaces E_z^s and E_z^u mentioned in the definition above are called the stable and unstable subspaces, respectively.

As for diffeomorphisms, hyperbolicity implies shadowing.

Definition 2.16 (*Pseudo-orbit*). A sequence $\{z_i\}_{i \in I}$ is called a δ-pseudo-orbit for map f when $\text{dist}(f(z_i), z_{i+1}) < \delta$ whenever $i, i+1 \in I$.

Definition 2.17 (*Shadow*). An orbit of F $\{w_i\}_{i \in I}$ ε-shadows a sequence $\{y_i\}_{i \in I}$ when for all $i \in I$, $\text{dist}(w_i, y_i) < \varepsilon$.

Theorem 2.18 (Shadowing (Sander [22], Steinlein and Walther [24,25])). *If K is a hyperbolic set for map f, then for any $\varepsilon > 0$, there is a $\delta > 0$ such that any δ-pseudo-orbit in $B_\delta(K)$ is ε-shadowed by an orbit of f. If ε is small enough and the pseudo-orbit is bi-infinite, then its ε-shadow is unique. Further, if the pseudo-orbit is periodic, so is its shadow.*

One is tempted to think that since hyperbolicity and shadowing are the major tools needed to show homoclinic tangles for diffeomorphisms, the previous theorem should be sufficient in the noninvertible case as well. However, the story of homoclinic tangles for maps is not so idyllic. First of all, the concept of a transverse homoclinic point is not always well defined in this situation. Specifically, transverse intersection is defined only when the stable and unstable manifolds are smooth at the point of intersection. This is not true in general for noninvertible maps. This problem is avoidable, but a more severe one remains: a transverse homoclinic point does not imply chaotic behavior.

3. Transverse crossings without chaos

This section contains examples of noninvertible maps with transverse homoclinic points, such that nearby, there is no hyperbolic structure. The different examples illustrate different ways in which hyperbolicity can fail. The first example illustrates a codimension two phenomenon, whereas the others show higher codimension behavior.

We start with an orientation-preserving diffeomorphism $f: R^2 \to R^2$ with a transverse homoclinic orbit. Assume that the upper branch of the unstable manifold contains the homoclinic orbit, and that the lower branch of the unstable manifold converges to some other fixed point far away from p. Thus near a point q in the homoclinic orbit, all points on one side of the stable manifold never return near q under f. Namely, these points converge to the lower branch of the unstable manifold, and thus never return near q. Using this fact, we perturb the diffeomorphism in a neighborhood of q, without changing the map outside this neighborhood. We do so in such a way that all the points in the neighborhood converge to the never-returning branch of the unstable manifold. In other words, this new map has no recurrent behavior near q. See the examples in Ch. 5 of Palis and Takens [19] for a similar kind of alteration of a map in a neighborhood to study the change in stable and unstable manifolds.

3.1. Pips, lobes, and transport

The terminology introduced here for diffeomorphisms of R^2 follows [26]. First, the idea that r is a primary intersection point of W^s and W^u.

Definition 3.1 (*Primary intersection point, or pip*). Assume f is a diffeomorphism with hyperbolic fixed point p. Homoclinic point r is a primary intersection point, or pip, when the segments of W^s and W^u joining p to r intersect only at p and r.

Definition 3.2 (*Lobe*). Let r_1 and r_2 be two adjacent pips. Precisely, there are no pips between them along the segment joining r_1 to r_2 either along W^s or along W^u. The lobe is a region bounded by the segments joining r_1 to r_2 along W^s and along W^u.

It is possible to bound a region using the segments of W^s and W^u joining p to r. This union of segments is called a *pseudoseparatrix*. Assuming that there is only pip r_1 between r and $f^{-1}(r)$, it is possible to use the r, r_1 and $r_1, f^{-1}(r)$ lobes to classify the movement of points in and out of the pseudoseparatrix. Refer to Fig. 1.

Lemma 3.3 (Turnstile lobes). *If pip r_1 is the only one between r and $f^{-1}(r)$, then the r, r_1 lobe contains all points entering the interior of the pseudoseparatrix in one iterate, and the $r_1, f^{-1}(r)$ lobe contains all the points leaving the interior of the pseudoseparatrix in one iterate.*

For a proof, see [26].

3.2. An example

Using the notation developed above, assume that r is a pip, and that there is one pip r_1 between r and $f^{-1}(r)$. Thus the r, r_1 lobe contains all points entering the interior of the pseudoseparatrix made with r in one iterate. Let $q = f(r_1)$. By the above lemma, and the fact that points cannot map across the W^u, in a neighborhood of q, points outside the pseudoseparatrix never return near q under forward images of f. See Fig. 1.

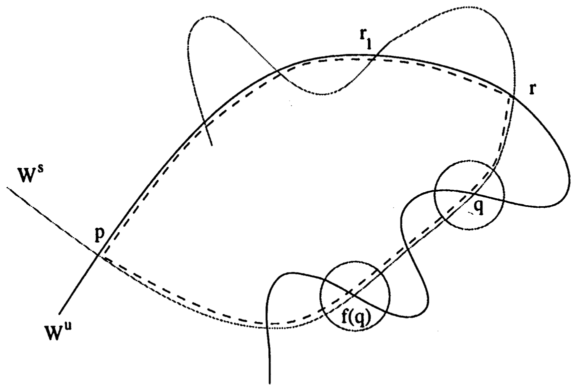

Fig. 1. Transverse homoclinic orbit of an orientation preserving diffeomorphism. The dotted line marks the pseudoseparatrix. The regions near q and $f(q)$ mark the perturbed domain and range, respectively.

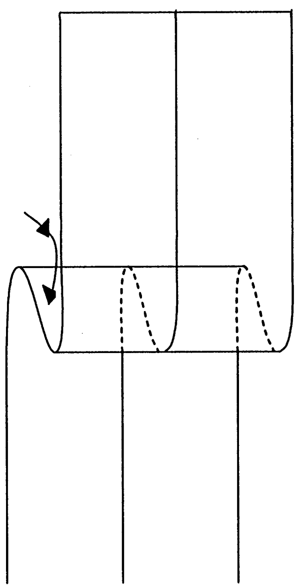

Fig. 2. A cubic map folds the plane over itself.

Choose a neighborhood of q. Smoothly perturb f in this neighborhood to get a new map g. Do this in such a way that outside the neighborhood, the map is equal to f, and inside a slightly smaller neighborhood, all points not on $W^s(f)$ map to the outside of the pseudoseparatrix, and all points on $W^s(f)$ map onto $W^s(f)$. This can be done, for example, by composing the original diffeomorphism f with a cubic map on a neighborhood of q. See Fig. 2.

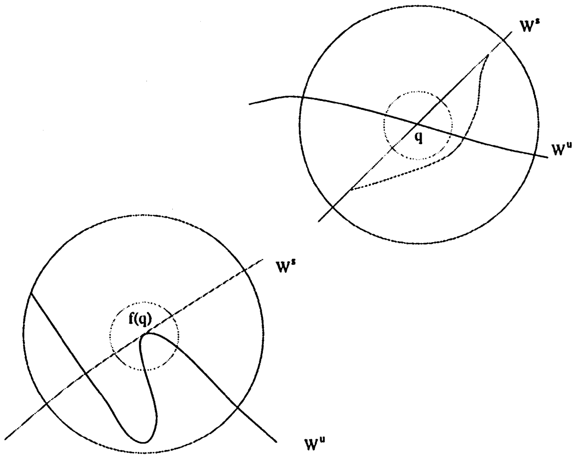

Fig. 3. A look at the change of stable and unstable manifolds near q and $f(q)$. Note that there is an extra curve added to W^s.

In a neighborhood of q, the unstable manifold for g is the same as the unstable manifold for f, since unstable manifolds are defined by images, and the image near q remains fixed. By mapping points on $W^s(f)$ to other points on $W^s(f)$, we assure that the stable manifold is locally near q the same as before the perturbation. Away from q, but in the perturbed neighborhood, there is actually an extra portion of $W^s(g)$ formed as the preimage under the cubic map, but this does not affect the behavior sufficiently near q. See Fig. 3. $W^s(g)$ and $W^u(g)$ still intersect transversally at q. Since all points near q map outside the pseudoseparatrix under g, and since outside the neighborhood of q, the map g is a diffeomorphism, these points converge to the lower branch of the unstable manifold. By our assumption, this lower branch is in the basin of attraction of some other fixed point; therefore no point near q ever returns near q. Thus there is no chaotic behavior near the homoclinic orbit containing q. In fact, by specifying the perturbation more carefully, it is possible to construct a noninvertible map with no recurrent behavior near the orbit of any transverse homoclinic point. This involves doing the same construction, but including the entire lobe between q and $f(r)$ in the set in the perturbed neighborhood, while taking care that the extra preimage of W^s does not intersect W^u.

The closure of the above homoclinic orbit cannot be a hyperbolic set, as there is no shadowing. Namely, for all δ, there are δ-pseudo-orbits through q, p, and q again. However, no true orbit shadows, since near q, there is no recurrence.

3.3. Analysis of the example

Unlike diffeomorphisms, the homoclinic orbit above has both transversal homoclinic points and nontransversal homoclinic points. This occurs because the derivative map is singular on the tangent space to the unstable manifold, and in addition, the unstable manifold is tangent to the stable manifold.

The fact that a transverse homoclinic orbit for a diffeomorphism has a hyperbolic structure rests on the fact that the tangent direction to the unstable manifold at a point in the orbit is transverse to the stable manifold and eventually expanding. The unstable direction can then be chosen as the unstable subspace at each point in the orbit. This proof breaks down in the current example, since at a point in the orbit, the unstable manifold becomes tangent to the stable manifold, though previously in the orbit, the intersection was transverse; this can only happen because derivative is singular in the unstable direction.

3.4. Another example

As above, start with an orientation preserving diffeomorphism with a transverse homoclinic orbit. Assume that near a homoclinic point q, there are coordinates such that the unstable manifold is of the form $q+\{x=0\}$, and such that the vertical lines form the unstable foliation. Also assume that near $f(q)$, there are coordinates such that the stable manifold can be written as $f(q)+\{y=0\}$, and horizontal lines form the stable foliation. Further, assume that the mapping between the two neighborhoods takes vertical lines to vertical lines.

Perturb again in the same neighborhood as before in such a way that locally $q+(x,y)\mapsto f(q)+(x^2y^2,y^3)$. Thus the unstable manifold still maps to a vertical line above $f(q)$, but all other vertical lines map to curves with cusps. One of these curves must be the unstable manifold to points which converge to the fixed point after going an additional time around near the original homoclinic orbit. Thus the smoothness of generalized stable and unstable manifolds is no longer preserved. Thus the stable manifold theorem for hyperbolic sets implies that the closure of the homoclinic orbit of the new map is not a hyperbolic set.

For the original unperturbed diffeomorphism, the maximal invariant set of a neighborhood of $p\cup q$ for some iterate f^n has dynamics conjugate to the subshift on two symbols. For example, there is a point s near q such that $f^{-n}(s)$ is near q, but $f^{nj}(s)$ is near p for all $j\neq 0,-1$. Again we reparametrize so that the primary branches of W^u and W^s correspond to the x- and y-axes respectively, and the unstable leaves are vertical lines. Note that the point s lies on the intersection of one of these vertical lines with the x-axis. If we had perturbed the map in such a way that $q+(x,y)\mapsto f(q)+(x,y^3-x^2y)$. Then for the perturbed map, all unstable leaves remain vertical lines, and the stable and unstable manifolds remain the x- and y-axes. Every line other than $x=0$ has three intersections with the x-axis. Thus for example the pseudo-orbit corresponding to the point s above now has three distinct shadows. Thus the uniqueness of shadows does not hold, implying again that the closure of the homoclinic orbit is not a hyperbolic set.

In the above two examples, the images of the stable and unstable manifolds intersect transversally. However, the derivative of the map in the direction of the unstable manifold is singular at the homoclinic point. When there is a singularity in the direction of the tangent plane to the unstable manifold, the closure of the homoclinic orbit is not a hyperbolic set.

4. Genericity

As discussed in the previous section, a necessary condition for the closure of a homoclinic orbit to be a hyperbolic set is that there are no singularities in the map along the unstable manifold. As stated below, Steinlein and Walther [24] and Lerman and Shil'nikov [13] independently showed that this is a sufficient condition as well. In this section, we use the noninvertible version of the Kupka–Smale theorem to show that this condition is generic.

Theorem 4.1 (Homoclinic tangles for maps). *Let p be a hyperbolic fixed point for a map f, and let $\{z_k\}$ be a homoclinic orbit to p. Then there exists a sufficiently large N such that for all $m<-N$ and $n>N$, $z_m \in W^u_{loc}$, and $z_n \in W^s_{loc}$. If $Df^{n-m}_{z_m}$ is injective on $T_{z_m}W^u$ and maps $T_{z_m}W^u$ to a subspace transversal to $T_{z_n}W^s$, then the closure of $\{z_k\}$ is hyperbolic.*

Replacing f with f^n, this result also applies to periodic orbits.

Theorem 4.2 (Shub [23]; *Kupka–Smale for noninvertible maps*). *Let M be a compact manifold. Then maps with the following conditions form a residual set in the space of C^r maps on M in the C^r topology*:

1. *All periodic points of f are hyperbolic.*
2. *If p is a hyperbolic periodic point of f, then $W^s(p)$, the global stable manifold to p, is an injective immersed submanifold of M.*
3. *If p and q are periodic points of periods j and k for f, the map $f^{nj}|_{W^u_{loc}(p)}$ is transverse to $W^s(q)$ for all $n \geq 0$.*

This is due to [23]. The theorem and proof are essentially the same as for the diffeomorphism case. The changes come from the fact that it is not possible to talk about transversal intersection of the global stable and unstable manifolds, these not being well defined. Shub first shows that generically global stable manifolds are immersed submanifolds. He then considers the transversality of the global stable manifold with

$$f^{nj}|_{W^u_{loc}}, \quad n>0,$$

the positive iterates of the map f, restricted to each local unstable manifold. By restricting in this way, he is able to avoid working with the global unstable manifolds, which are not in general smooth. The proof that generically these intersect transversally is only a slight modification of the proof for diffeomorphisms that generically the global stable and unstable manifolds intersect transversally.

Notice that the conditions on the stable and unstable manifolds for a generically occurring map as described in the Kupka–Smale theorem are exactly the necessary and sufficient conditions for a homoclinic orbit to be a hyperbolic set. Thus generically, a map has the property that near every homoclinic orbit, there is chaotic behavior:

Corollary 4.3 (Homoclinic tangles are generic). *Within the set of C^r maps on a compact manifold, those with all homoclinic orbits being homoclinic tangles are generic.*

5. Two-parameter families

The first example in Section 3 is degenerate, since the stable manifold locally corresponds exactly to the set for which the derivative is singular. However, it illustrates the fact that only knowing that stable and unstable manifolds intersect transversally at a single point is not enough information to draw conclusions about the dynamics of a map near the homoclinic orbit. The loss of information occurs precisely when homoclinic orbits intersect the set for which the Jacobian is singular, which results in the stable and unstable manifolds becoming tangent. This section gives an analysis of the homoclinic bifurcation arising from crossing this set and shows that such a bifurcation is codimension two. We are interested in two sets in parameter space which determine the behavior of the homoclinic orbits: the *singular set* is the set of parameter values for which the determinant of the Jacobian is zero at a homoclinic point; the *tangency set* is the set of parameters for which at some homoclinic point, the image of the unstable manifold intersects the stable manifold tangentially. In a generic two-parameter family with a map for which one homoclinic orbit has both transversal and tangential intersections of stable and unstable manifolds, the singular set and the tangency set are tangent to each other.

The bifurcation is illustrated by the following example.

Example 5.1. Let $f_{(\mu,\delta)}$ be a two-parameter family of maps on R^2 with a saddle fixed point p. Choose open sets U' and V' not near p such that $f_{(\mu,\delta)}(U') \subset V'$ for all (μ,δ) near $(0,0)$. Assume there is a smooth portion of the stable manifold in V' and a smooth portion of the unstable manifold in U'.

Let the portion of the unstable manifold in U' be of the form $\{q + (cy + \delta, y)\}$, where $|c| < 1$ is a constant. Let the portion of the stable manifold in V' be of the form $\{f(q) + (x, x^2 + \mu)\}$. Let $f : U' \to V'$ be of the form $q + (x, y) \mapsto f(q) + (x, y^2)$. See Figs. 4–6.

For $\mu > 0$, $\delta = 0$ the stable manifold in U' consists of the curves $\{f(q) + (x, \sqrt{x^2 + \mu})\}$ and $\{f(q) + (x, -\sqrt{x^2 + \mu})\}$. Thus there are two transverse homoclinic points in U',

$$q + \left(c\sqrt{\frac{\mu}{1-c^2}}, \sqrt{\frac{\mu}{1-c^2}}\right) \quad \text{and} \quad q + \left(-c\sqrt{\frac{\mu}{1-c^2}}, -\sqrt{\frac{\mu}{1-c^2}}\right).$$

See Fig. 4. For $(\mu,\delta) = (0,0)$, the stable manifold in U' consists of the curves $\{q + (x,x)\}$ and $\{q + (x,-x)\}$. Again, these curves are transverse to the unstable manifold, so $q + (0,0)$ is a transverse homoclinic point. See Fig. 5. However, for $\mu < 0$, $\delta = 0$, the stable manifold in U' satisfies $\{q + (x, y) \colon y^2 = x^2 + \mu\}$, which does not intersect the line $\{q + (cy, y)\}$. Thus there are no longer any homoclinic points. See Fig. 6. A loss of stability occurs at the point $(x, y, \mu, \delta) = (q, 0, 0)$, though this is a transverse homoclinic point. This demonstrates that transverse homoclinic points are not stable under a perturbation of the map if the derivative is singular. Although this homoclinic orbit contains transversal intersections, it behaves much more like a homoclinic tangency. The behavior described took place on the μ-axis, but in fact qualitatively

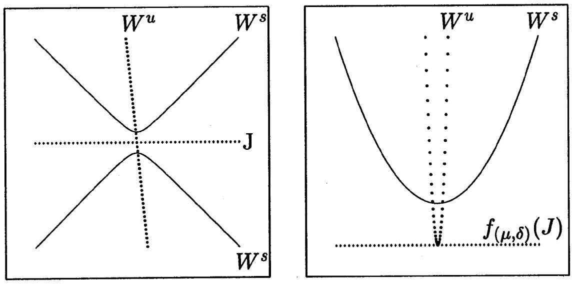

Fig. 4. The homoclinic bifurcation described in Example 5.1. This shows the $\mu > 0$, $\delta = 0$ case, in which there are two homoclinic points. The left picture shows the stable and unstable manifolds and the set for which the derivative is singular, denoted by J, near q within the region U. The right picture being the stable and unstable manifolds and the image of the singular set near $f_{(\mu,\delta)}(q)$ in the region V. At this parameter value, there are two preimages of the stable manifold in U which intersect the unstable manifold. Thus there are two homoclinic orbits.

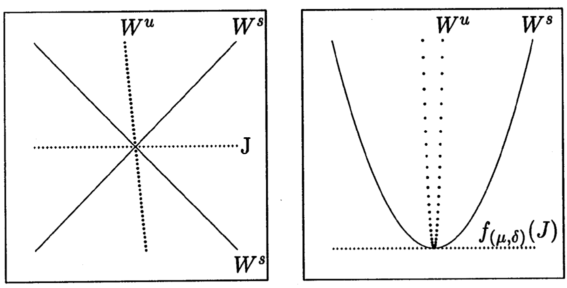

Fig. 5. Same as Fig 4, shown at the bifurcation point $\mu = \delta = 0$. In this case, there is exactly one homoclinic point.

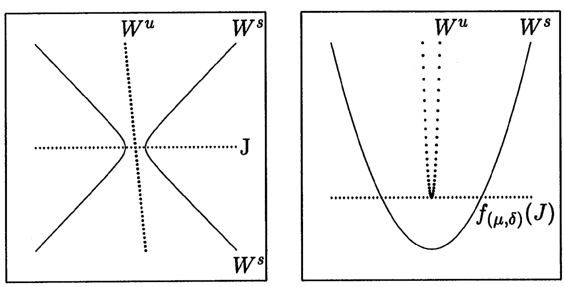

Fig. 6. Same as Fig 4, at $\mu < 0$, $\delta = 0$. There are no longer any homoclinic points.

similar behavior occurs along any curve through the origin in parameter space which is transverse to the δ-axis.

Now consider the singular set and tangency set in parameter space, as defined in the introductory paragraph to this section. Homoclinic points occur at $\{(x,y,\mu,\delta)=(cs+t,s,(1-c^2)s^2-2cts-t^2,t)\}$. Note that this is a smooth surface in $R^2\times R^2$. The derivative is singular at $y=0$, so the singular set is $\{\mu=-\delta^2\}$. The tangency set is $\{\mu=-\delta^2/(1-c^2)\}$. Note that the tangency set is a dividing curve for parameters with two homoclinic orbits and those with none. The tangency set and the singular set intersect tangentially at $(0,0)$.

The following theorem says that generically, a codimension two bifurcation occurs in a way which was illustrated in the example above. The result is inherently noninvertible, as the bifurcation occurs as a result of the loss of the multiple preimages of the stable manifold.

Theorem 5.2 (Homoclinic singularity bifurcation). *Let M be a compact smooth m-dimensional manifold. There is an open set in the space of smooth (at least C^2) two-parameter families on M, consisting of $f_{(\mu,\delta)}:M\times R^2\to M$ with the following properties*:

1. *f has a hyperbolic fixed point p_0 at parameter value (μ_0,δ_0). The stable and unstable manifolds to this fixed point are denoted by W^s and W^u.*
2. *At (μ_0,δ_0), there is a point $q_0\in W^u$ such that $f_{(\mu_0,\delta_0)}$ is singular at q_0, and W^u is tangent to W^s at $f(q_0)$. (In addition, the crossing of these singular and tangency sets must be nondegenerate, as made precise in the course of the proof.)*

For an open dense set of f satisfying 1 and 2, the following conclusions can be drawn regarding the homoclinic points and nearby dynamics:

The set of homoclinic points near (q_0,μ_0,δ_0) form a smooth two-dimensional surface in $M\times R^2$. The tangency and singularity sets are smooth curves in R^2. Generically, the two curves intersect in a quadratic tangency at the bifurcation point.

The tangency curve divides the parameter space into parameters for which there are locally two homoclinic points and those for which there are no homoclinic orbits. The singularity curve corresponds to parameters for which the homoclinic orbits lie on the essentially noninvertible part of the map. Parameters on the singularity curve but not on the tangency curve do not have a singularity in the direction of the unstable manifold; the closures of corresponding homoclinic orbits are hyperbolic sets.

Proof. Assume that W^u is k-dimensional, where $0<k<m$. The case $k=m$ needs modifications, as discussed at the end. The case $k=0$ is not meaningful for maps.

Let U be a small neighborhood of q_0 and V a small neighborhood of $f_{(\mu_0,\delta_0)}(q_0)$. Assume that in U, the unstable manifold is a smooth embedded k-dimensional submanifold, and that in V the stable manifold is a smooth $(m-k)$-dimensional embedded submanifold. In other words, we assume that the set for which the derivative is singular only intersects the homoclinic orbit at one point. Call these portions $\tilde{W}^u$ and $\tilde{W}^s$. The assumption is dense in C^r, since the homoclinic orbit is zero-dimensional. It is also open in C^r, since the set where the derivative is singular changes continuously with the parameter. From now on we restrict to parameters in an open set containing (μ_0, δ_0) small enough that the condition still holds. On this set of parameter values, there is a smooth continuation $p_{(\mu,\delta)}$ of p_0, and a smooth continuation $\tilde{W}^u_{(\mu,\delta)}$ of $\tilde{W}^u$ in U and $\tilde{W}^s_{(\mu,\delta)}$ of $\tilde{W}^s$ in V. $R^m = R^{m-k} \times R^k$ by the standard splitting, and we write $z \in R^m$ in coordinates as (x, y).

Locally, we can instead consider a map $G : R^m \times R^2 \to R^m$ as follows. By the definition of a submanifold, if U and V are sufficiently small, there is a smooth family of diffeomorphisms $\phi_{(\mu,\delta)} : U \to R^m$ such that $\phi_{(\mu,\delta)}(\tilde{W}^u_{(\mu,\delta)})$ =the y-subspace (where $x = 0$). Likewise, there exists a smooth family of diffeomorphisms $\rho_{(\mu,\delta)} : V \to R^m$ such that $\rho_{(\mu,\delta)}(\tilde{W}^s_{(\mu,\delta)}) =$ the x-subspace (where $y = 0$). We can also prescribe $\phi_{(\mu_0,\delta_0)}(q_0) = (0, 0)$, and $\rho_{(\mu_0,\delta_0)}(f(q_0)) = (0, 0)$.

Define the smooth family $G : R^{m+2} \to R^m$ by $G(x, y, \mu, \delta) = \rho_{(\mu,\delta)} \circ f_{(\mu,\delta)} \circ \phi^{-1}_{(\mu,\delta)}(x, y)$. Thus point $(0, 0, \mu_0, \delta_0)$ corresponds to the point (q_0, μ_0, δ_0). For convenience, we assume $(\mu_0, \delta_0) = (0, 0)$. Let π_2 be the projection to the y-subspace. The proof of the theorem now reduces to a statement about smooth maps from R^{m+2} to R^m. Specifically, let $K \subset C^2(R^{m+2}, R^m)$ be the subspace such that:

1. $G(0, 0, 0, 0) = (0, 0)$. This corresponds to the bifurcation point.
2. $DG(0, 0, 0, 0)$ is singular. This corresponds to a homoclinic singularity.
3. $\partial/\partial y\ G(0, 0, 0, 0)$ is singular. This corresponds to a homoclinic tangency.

We show that for an open and dense set of diffeomorphisms within the set K (and therefore an open dense set of $f_{(\mu,\delta)}$ on M), the following statements hold on M cross parameter space:

A. The homoclinic points form a smooth two-dimensional surface.
B. The set of homoclinic tangencies form a smooth curve, which projects smoothly to the parameter plane.
C. The set of singular points form a smooth $(k + 1)$-dimensional surface projecting smoothly to both the parameter plane and to the surface of homoclinic points.
D. The set of parameters for which the map has a homoclinic tangency is a curve dividing the parameter space into a region in which there are two homoclinic points and a region for which there are no homoclinic points. Furthermore, the set of homoclinic tangencies and the set of homoclinic singularities intersect in a quadratic tangency in parameter space.
E. The existence of the bifurcation point described above is stable under perturbation in C^2.

Here are the details of steps A–E. The arguments apply for arbitrary dimensions, with explicit formulae for a saddle point in dimension two; for this case, we write the Taylor series expansion of G up to second order around $(0, 0, \mu_0, \delta_0)$, where for convenience, we let $(\mu_0, \delta_0) = (0, 0)$.

$$G = \begin{pmatrix} a^{10}_{00}\mu + a^{01}_{00}\delta + a^{00}_{10}x + a^{00}_{01}y + a^{10}_{10}x\mu \dots \\ b^{10}_{00}\mu + b^{01}_{00}\delta + b^{00}_{10}x + b^{00}_{01}y + b^{10}_{10}x\mu \dots \end{pmatrix}.$$

The fact that q_0 is a tangency between stable and unstable manifolds corresponds to $b^{00}_{01} = 0$, where this is the coefficient of y. The fact that $f(q_0)$ is a point of singularity corresponds to $a^{00}_{01}b^{00}_{10} = 0$.

Before proceeding, we describe the type of argument used in each case. To show that the set of diffeomorphisms with a certain condition is open, we just need to display a continuous map on C^2 so that the image of maps with the desired property is open. To show that a property is dense is more difficult; we construct (or describe the construction of) a specific arbitrarily small perturbation which has the desired property. To do so, we make use of the existence of a C^∞ bump function β such that near 0, $\beta \equiv 1$, away from 0, $\beta \equiv 0$, but for specified ε and k, $\|\varepsilon\beta\|_{C^k} < \varepsilon$.

A. Homoclinic points: Define $H_G : W \subset R^{k+2} \to R^k$ by $H_G(y,\mu,\delta) \equiv \pi_2 \circ G(0,y,\mu,\delta)$. The set $A = \{(y,\mu,\delta) : H_G(y,\mu,\delta) = 0\}$ is diffeomorphic to the homoclinic points of f near (q_0,μ_0,δ_0).

For the homoclinic set to be smooth on a neighborhood, we need $\mathrm{d}H_G(y,\mu,\delta)$ to be rank k at $(0,0,0)$. Here, d is used to denote the derivative with respect to y,μ,δ. Then, by the implicit function theorem, there is a neighborhood such that the homoclinic set is a smooth two-dimensional surface. That this is open is clear, since $G \mapsto \mathrm{d}H_G(0,0,0,0)$ is a continuous map, and the set of maps of rank k is an open set of the $k+2$ by k matrices.

To show that the set is dense, start with any $G \in K$ so that $\mathrm{d}H_G(0,0,0,0)$ is not rank k and choose a small ε. It is always possible to find an ε-small matrix

$$M = \begin{pmatrix} M_1 & M_3 & M_5 \\ M_2 & M_4 & M_6 \end{pmatrix}$$

so that

$$DG(0,0,0,0) + \begin{pmatrix} M_1 & M_3 \\ M_2 & M_4 \end{pmatrix}$$

is rank $n-1$, $\partial/\partial y\ H(0,0,0,0) + M_4$ is rank $k-1$, and $\mathrm{d}H + (M_4 M_6)$ is rank k. Let

$$P(x,y,\mu,\delta) = G(x,y,\mu,\delta) + \beta(G(x,y,\mu,\delta))M \begin{pmatrix} x \\ y \\ \mu \\ \delta \end{pmatrix}.$$

Then P is an ε perturbation of G which is in K, but has that set A is smooth on a neighborhood of $(0,0,0,0)$.

Explicitly for a two-dimensional saddle, we have

$$\mathrm{d}H = (0, b_{00}^{10}, b_{00}^{01}),$$

so we need either $b_{00}^{10} \neq 0$ or $b_{00}^{01} \neq 0$.

B. Tangencies: Define $T_G : W \subset R^{k+2} \to R$ by $T_G(y,\mu,\delta) \equiv det(\mathrm{d}/\mathrm{d}y(\pi_2 \circ G(0,y,\mu,\delta)))$.

$B=\{(y,\mu,\delta):T(y,\mu,\delta)=0\}$ is a smooth $(k+1)$-dimensional surface transverse to the homoclinic points on a neighborhood of $(0,0,0)$ as long as

$$\begin{pmatrix} \mathrm{d}H(0,0,0) \\ \nabla T(0,0,0) \end{pmatrix}$$

is rank $k+1$. ∇ denotes the derivative with respect to y,μ,δ, emphasizing the fact that this is a scalar function.

That the above occurs for an open set is not hard to see. That it is true for a dense set again follows from the fact that there is a C^k small perturbation of G such that the condition is satisfied. The argument is much the same as the argument in A.

For $m=2$:

$$\nabla T=(2b_{02}^{00},b_{01}^{10},b_{01}^{01}).$$

The intersection of A and B is diffeomorphic to the set of homoclinic tangencies.

C. Singularities: Similar to B, define $S(y,\mu,\delta)=\det(DG(x,y,\mu,\delta))|_{x=0}$. The derivative is taken with respect to x and y. Then $C=\{(y,\mu,\delta):S(y,\mu,\delta)=0\}$ is diffeomorphic to the set of singularities.

Using similar reasoning to the previous two arguments, for an open dense set of maps, there is a neighborhood such that the singularity set is smooth. In fact, there is an arbitrarily small perturbation Q of G in K such that

$$\begin{pmatrix} \mathrm{d}H_Q(0,0,0,0) \\ \nabla S_Q(0,0,0,0) \\ \nabla T_Q(0,0,0,0) \end{pmatrix}$$

is full rank: $k+2$, and $\partial/\partial y H_Q(0,0,0,0)$ is rank $k-1$. Therefore sets A, B, and C are mutually transverse on some neighborhood of the homoclinic bifurcation point. The last condition guarantees that these sets are transverse to the parameter plane.

Specifically for $m=2,\ k=1$:

$$\nabla S=\left(2a_{10}^{00}b_{02}^{00}-2b_{10}^{00}a_{02}^{00},a_{10}^{00}b_{01}^{10}-b_{10}^{00}a_{01}^{10},a_{10}^{00}b_{01}^{01}-b_{10}^{00}a_{01}^{01}\right).$$

D. Under the open dense conditions above, the tangency curve divides parameter space into a region with two homoclinic orbits and a region with no homoclinic orbits. This is because for the perturbation Q, H_Q is singular in one-parameter direction, but the second-order terms are positive.

The curves of homoclinic tangency and homoclinic singularity (resp. $A\cap B$ and $A\cap C$) project in the parameter plane to two curves intersecting at a quadratic tangency. This is because the homoclinic tangency projection bounds the set of images of homoclinic points (A), but the second-order terms are nonzero.

For $m=2,\ k=1$, a straightforward but tedious calculation shows that the tangent direction to the singular curve is equal to the tangent direction to the tangency curve, and they are both equal to $(-b_{00}^{10}/b_{00}^{01})$. An even more tedious straightforward calculation shows that the second derivatives to each of the curves at this point are given by large expressions in terms of at most second-order coefficients. For an open dense set, these two expressions are not equal.

E. The singularity point is stable under perturbation in C^2. This is due to the fact that the transversally intersecting sets A, B, and C are well defined for all C^2 diffeomorphisms. Thus under small perturbation, they must still intersect at a point of homoclinic singularity and tangency.

The case $k = m$: This is often referred to as a snap-back repeller, as mentioned in Example 2.8. The statements above still hold, except that the functions T and S are actually equal. Thus the bifurcation becomes codimension one bifurcation. The following statements still hold: A, C, the first sentence in the statement of D, and the fact that this holds in an open set of parameters. That the singularity curve generically divides the parameter space into two- and no-homoclinic orbit regions, is the main theorem of Batteli and Lazzari [3]. □

By looking at f^n rather than f, the above proof applies to periodic points.

Future plans are to consider what happens to Newhouse-type structures near the bifurcation described above. See [11, 19], and references contained therein.

Acknowledgements

A portion of this work was originally part of the author's Ph.D. dissertation under the guidance of Richard McGehee in the School of Mathematics at the University of Minnesota. I would also like to thank Konstantin Mischaikow, Dimitri Turaev, and Thomas Wanner for useful discussions and helpful suggestions.

References

[1] R. Adomaitis, I. Kevrekidis, Noninvertibility and the structure of basins of attraction in a model adaptive control system, J. Nonlinear Sci. 1 (1991) 95–105.

[2] D.G. Aronson, M.A. Chory, G.R. Hall, R.P. McGehee, Bifurcations from an invariant circle for two-parameter families of maps of the plane: a computer-assisted study, Communications in Mathematical Physics, vol. 83, Springer, Berlin, 1982, pp. 303–354.

[3] F. Batteli, C. Lazzari, On the bifurcation from critical homoclinic orbits in n-dimensional maps, Discrete Continuous Dyn. Systems 3 (2) (1997) 289–303.

[4] R. Devaney, An Introduction to Chaotic Dynamical Systems, Addison-Wesley, Reading, MA, 1989.

[5] C.E. Frouzakis, R.A. Adomaitis, I.G. Kevrekidis, M.P. Golden, B.E. Ydstie, The structure of basin boundaries in a simple adaptive control system, in: T. Bountis (Ed.), Chaotic Dynamics: Theory and Practice, Plenum Press, New York, 1992, pp. 195–210.

[6] C.E. Frouzakis, R.A. Adomaitis, I.G. Kevrekidis, On the dynamics and global stability characteristics of an adaptive control system, Comp. Chem. Eng. 1996.

[7] C.E. Frouzakis, L. Gardini, I.G. Kevrekidis, G. Millerioux, C. Mira, On some properties of invariant sets of two-dimensional noninvertible maps, Int. J. Bifur. Chaos Appl. Sci. Eng. 7 (6) (1997) 1167–1194.

[8] L. Gardini, Bifurcations and multistability in a class of two-dimensional endomorphisms, Nonlinear Anal. 28 (1) (1997) 61–85.

[9] L. Gardini, Homoclinic bifurcations in n-dimensional endomorphisms, due to expanding periodic points, Nonlinear Anal. 23 (8) (1994) 1039–1089.

[10] L. Gardini, Homoclinic orbits of saddles in two-dimensional endomorphisms, Iteration Theory (Batschuns, 1992), World Sci. Publishing, River Edge, NJ, 1996, pp. 81–99.
[11] S.V. Gonchenko, L.P. Shil'nikov, D.V. Turaev, On models with non-rough Poincaré homoclinic curves, Physica D 62 (1993) 1–14.
[12] I. Gumowski, C. Mira, Recurrences and Discrete Dynamic Systems, Springer, Berlin, 1980.
[13] L.M. Lerman, L.P. Shil'nikov, Homoclinic structures in infinite-dimensional systems, Sibirskii Matematicheskii Zhurnal 29 (3) (1988) 92–103.
[14] E.N. Lorenz, Computational chaos – a prelude to computational instability, Physica D 25 (1989) 299–317.
[15] F.R. Marotto, Snap-back repellers imply chaos in R^n, J. Math. Anal. Appl. 63 (1978) 199–223.
[16] R. Rico-Martínez, I.G. Kevrekidis, R.A. Adomaitis, Noninvertible Dynamics in Neural Network Models, Preprint.
[17] R. McGehee, E. Sander, A new proof of the stable manifold theorem, Z. angew. Math. Phys. 47 (1996) 497–513.
[18] S. Newhouse, J. Palis, Bifurcations of Morse–Smale dynamical systems, in: M.M. Peixoto (Ed.), Dynamical Systems, Sec. 3. Academic Press, New York, 1973, pp. 312–320.
[19] J. Palis, F. Takens, Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations, Cambridge Studies in Advanced Mathematics, vol. 35, Cambridge University Press, Cambridge, 1993.
[20] C. Robinson, Dynamical Systems, CRC Press, Boca Raton, 1995.
[21] E. Sander, Hyperbolic sets for noninvertible maps and relations, Ph.D. Thesis, University of Minnesota, June, 1996.
[22] E. Sander, Hyperbolic sets for noninvertible maps and relations, CDSNS96 Preprint 254, Georgia Tech., 1996, submitted.
[23] M. Shub, Endomorphisms of compact differentiable manifolds, Amer. J. Math. 91 (1969) 175–199.
[24] H. Steinlein, H.-O. Walther, Hyperbolic sets and shadowing for noninvertible maps, Advanced Topics in the Theory of Dynamical Systems, Academic Press, New York, 1989.
[25] H. Steinlein, H.O. Walther, Hyperbolic sets, transversal homoclinic trajectories, and symbolic dynamics for C^1-maps in banach spaces, J. Dyn. Differential Equations 2 (3) (1990) 325–365.
[26] S. Wiggins, Introduction to Applied Nonlinear Dynamical Systems and Chaos, Texts in Applied Mathematics 2, Springer, Berlin, 1990.