



Centrum voor Wiskunde en Informatica

**REPORT**RAPPORT

**MAS**

Modelling, Analysis and Simulation



*Modelling, Analysis and Simulation*

Dynamics of surface diffeomorphisms relative to  
homoclinic and heteroclinic orbits

Pieter Collins

**REPORT MAS-R0311 SEPTEMBER 30, 2003**

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).

CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

**Modelling, Analysis and Simulation (MAS)**

Information Systems (INS)

Copyright © 2003, Stichting Centrum voor Wiskunde en Informatica

P.O. Box 94079, 1090 GB Amsterdam (NL)

Kruislaan 413, 1098 SJ Amsterdam (NL)

Telephone +31 20 592 9333

Telefax +31 20 592 4199

ISSN 1386-3703

# Dynamics of surface diffeomorphisms relative to homoclinic and heteroclinic orbits

## ABSTRACT

We show how to obtain information about the dynamics of a two-dimensional discrete-time system from its homoclinic and heteroclinic orbits. The results obtained are based on the theory of trellises, which comprise finite-length subsets of the stable and unstable manifolds of a collection of saddle periodic orbits. For any collection of homoclinic or heteroclinic orbits, we show how to associate a canonical trellis type which describes the orbits. Given a trellis type, we show how to compute a graph representative which gives a combinatorial invariant of the trellis type. The orbits of the graph representative represent the dynamics forced by the homoclinic/heteroclinic orbits in the sense that every orbit of the graph representative is globally shadowed by some orbit of the system, and periodic, homoclinic/heteroclinic orbits of the graph representative are shadowed by similar orbits.

*2000 Mathematics Subject Classification:* 37E30 37B10 37C27 37E25

*Keywords and Phrases:* Surface diffeomorphism, homoclinic orbit, tangle, trellis, symbolic dynamics

*Note:* This work was carried out at the University of Liverpool, and-partially funded by Leverhulme Special Research Fellowship-SRF/4/9900172. The final version was completed at the Centrum voor-Wiskunde en Informatica.

# Dynamics of surface diffeomorphisms relative to homoclinic and heteroclinic orbits

Pieter Collins \*

Centrum voor Wiskunde en Informatia

P.O. Box 94079

1090 GB Amsterdam

The Netherlands

pieter.collins@cwi.nl

September 11, 2003

## Abstract

We show how to obtain information about the dynamics of a two-dimensional discrete-time system from its homoclinic and heteroclinic orbits. The results obtained are based on the theory of *trellises*, which comprise finite-length subsets of the stable and unstable manifolds of a collection of saddle periodic orbits. For any collection of homoclinic or heteroclinic orbits, we show how to associate a canonical *trellis type* which describes the orbits. Given a trellis type, we show how to compute a *graph representative* which gives a combinatorial invariant of the trellis type. The orbits of the graph representative represent the dynamics forced by the homoclinic/heteroclinic orbits in the sense that every orbit of the graph representative is *globally shadowed* by some orbit of the system, and periodic, homoclinic/heteroclinic orbits of the graph representative are shadowed by similar orbits.

Mathematics subject classification: Primary: 37E30. Secondary: 37B10, 37C27, 37E25.

## 1 Introduction

The importance of homoclinic orbits in dynamical systems theory was first realised by Poincaré [Poi99], who showed that the presence of a homoclinic tangle in the three-body problem was enough to show that the system was non-integrable. More work on homoclinic tangles was later undertaken by Birkhoff [Bir49], but modern interest was inspired by Smale, who showed that any system with a transverse homoclinic point must have a horseshoe in some iterate [Sma63]. In this paper, we extend the local analysis of Smale to obtain a global understanding of the dynamics for diffeomorphisms in two dimensions. The analysis depends on constructing a *graph representative* of the system,

---

\*This work was partially funded by Leverhulme Special Research Fellowship SRF/4/9900172

from which we obtain information about the dynamics in terms of symbolic dynamics relative to a Markov partition and global shadowing.

The use of Markov partitions to give symbolic dynamics is a fundamental tool in dynamical systems theory. For uniformly-hyperbolic systems, the dynamics is described completely by the symbolics, but for non-uniformly hyperbolic systems, the Markov partition typically only gives a lower bound for the dynamics. From this, we can often deduce information on the forcing relation, which describes when the presence of a particular geometric structure implies the existence of another structure. Sharkovskii's theorem for interval maps is a classic example of a forcing result: a total ordering  $\succ$  of the natural numbers is given such that the presence of a periodic orbit of period  $p$  forces periodic orbits of all periods  $q$  with  $p \succ q$ . Sharkovskii's theorem can be strengthened to consider the ordering of points on the orbit, and this gives a complete (though less elegant) description of the forcing relation for periodic orbits of interval maps. Interval maps can be studied by elementary methods, but obtaining symbolic dynamics for more complicated classes of system typically requires the use of some topological index theory, such as the fixed point index [Fri83, Jia93] or the Conley index [Con78].

The theory of isotopy classes of surface homeomorphisms is known as *Nielsen-Thurston theory* [CB88], and extends in a straightforward way to a theory of isotopy classes relative to periodic orbits. Here, the period is no longer a useful characterisation of the orbit; instead, we need to consider the *braid type* which is the conjugacy class of isotopy classes relative to the orbit in question. Just as for periodic orbits of interval maps, orbits of different periods may have different braid types. For each braid type there is a canonical representative such that a braid type  $BT_1$  forces  $BT_2$  if and only if the *Thurston minimal representative* of  $BT_1$  contains an orbit of type  $BT_2$ . The forcing relation on braid types is a partial order [Boy92]. A Thurston minimal representative has the minimum number of periodic points of any period [BK82, JG93, Boy99], and minimal topological entropy in the isotopy class. The most interesting case is when the canonical representative is pseudo-Anosov. A pseudo-Anosov map preserves a pair of transverse foliations, the leaves of the stable foliation being uniformly contracted and the unstable foliation being uniformly expanded. The orbits of a pseudo-Anosov map persist under isotopy [Fat90] and are globally shadowed by those of any other homeomorphism in the isotopy class [Han85]. The dynamics of a pseudo-Anosov map therefore have similar properties to those of uniformly hyperbolic diffeomorphisms, except that the persistence of the dynamics holds for arbitrarily large perturbations. Many of the above-mentioned results can be found in the survey article [Boy94].

Thurston's proof of the classification theorem [Thu76, Soc79] was non-constructive, and thus of little use in actually determining the forcing relation. The main breakthrough was an algorithm of Bestvina and Handel [BH92] for computing train tracks for automorphisms of free groups, which was adapted to the case of surface homeomorphisms by Franks and Misiurewicz [FM93] and Bestvina and Handel [BH95]. An alternative algorithm for the case of homeomorphisms of the disc was given by Los [Los93]. A description of the moves of this algorithm with an emphasis on implementation details was given by Keil [Kei97]. Even so, the global properties of the forcing relation are still only partially understood and this is an active area of research, especially in the case of the orbits of the Smale

horseshoe map [dCH02].

A more challenging problem is to consider the dynamics of surface diffeomorphisms relative to homoclinic and heteroclinic orbits of tangles. The standard results of Poincaré and Smale for transverse homoclinic points have since been extended to topologically transverse intersections which are not geometrically transverse [BW95, Ray98], and a great deal is known about bifurcations near a homoclinic tangency [New80, MV93]. The global dynamics and the forcing relation have been previously considered by McRobie and Thompson [MT94], who carried out a detailed study of the *spiral 3-shoe*, by Rom-Kedar [RK94], who obtained results on entropy bounds and homoclinic forcing using a *topological approximation method*, and by Easton [Eas98]. Algorithms for computing dynamics forced by homoclinic orbits were recently given independently by Handel [Han99] and Hulme [Hul00]. Homoclinic tangles have been used by Christiansen and Politi [CP96] to compute generating partitions for symbolic dynamics. Orbits of the Hénon map have also been obtained using continuation methods by Sterling, Dullin and Meiss [SRD99].

The aim of this paper is to give a complete and general theoretical basis for describing and computing the dynamics forced by a homoclinic or heteroclinic tangle. The results are a completion of the theory begun in [Col99], and are related to the pruning theory of de Carvalho [dC99]. Instead of considering an entire tangle, which consists of immersed curves of infinite length, we only consider a subset consisting of curves of finite length. We call such a figure a *trellis*, after Birkhoff [Bir49] and Easton [Eas86]. Aside from being easier to work with, trellises have the advantage that there exist a number of algorithms to compute them to a very high degree of accuracy, such as those of Simo [Sim89] and Krauskopf and Osinga [KO98]. We therefore obtain results which can be directly applied to physically-relevant systems.

In Section 2 we give a definition of trellises, which constitute the main object of study, and introduce terminology to describe important features of their geometry. We also introduce the related concepts of biasymptotic orbits and controlled graphs, which also play an important role in this paper. Finally, we discuss reducibility of surface diffeomorphisms and graph maps in the context of homoclinic and heteroclinic dynamics.

In Section 3, we show how to relate a collection of homoclinic and/or heteroclinic orbits to a canonical trellis type forcing the same dynamics, thus unifying the viewpoints of homoclinic/heteroclinic orbits and tangles. This result can be applied to classify homoclinic orbits by braid type. We then show how to extend trellises without increasing the associated dynamics, and give results concerning the relationship between extensions and reducibility.

In Section 4, we show how to associate a graph representative to a given trellis map, in much the same way that a pseudo-Anosov braid type gives rise to a train track. We give an algorithm modelled on the Bestvina-Handel algorithm to compute the graph representative. Following an idea of Los [Los96], we replace the valence-2 homotopy by a *valence-3 homotopy* which does not require a global knowledge of the map. Unfortunately, the entropy may increase during this move, so we give a new analysis based on dynamical zeta functions to show that the algorithm terminates. We also unify the process of absorbing into the peripheral subgraph with the main algorithm, which also simplifies matters

somewhat. As with the Bestvina-Handel algorithm, reductions are obtained and can be dealt with accordingly. Finally, we give a simplified algorithm for braid types based on our analysis of the algorithm for trellises. The graph representative we obtain is unique (unlike a train track) and gives a convenient way of specifying the trellis. From this, we also obtain a canonical graph representative for a homoclinic orbit.

In Section 5, we discuss how the dynamics of the graph representative model the dynamics of the surface diffeomorphism. The existence of periodic orbits was given in [Col99]; the existence of other orbits follows by taking limits. In particular, these results show that the topological entropy of the graph representative is a lower bound for the topological entropy of the surface diffeomorphism. Orbits are related by *global shadowing*, a relation introduced by Katok, though here we give a new definition more in keeping with the spirit of our topological analysis. Periodic orbits which globally shadow each other have the same braid type, so we obtain information about the braid types of the orbits of the system. Related results on braid types and strong Nielsen equivalence can be found in [BH99]. Finally, we show by a careful construction that homoclinic and heteroclinic orbits of the graph map are shadowed by homoclinic and heteroclinic orbits of the surface diffeomorphism.

Many of the results contained in this paper are based on the Nielsen theory of self-maps of a topological pair, as developed in [Col01] based on standard Nielsen fixed-point theory [Bro71, Jia83]. The *Nielsen equivalence relation* for maps of pairs gives a relation between periodic points of the same (not necessarily least) period, and a Nielsen class equivalence class is *essential* if it has non-zero Lefschetz index. Essential Nielsen classes can be *continued* under homotopy, and hence a homotopy between two maps induces a *correspondence* between their essential Nielsen classes. The *Nielsen entropy*  $h_{niel}$  is given by the growth rate of the number of essential Nielsen classes, and is a lower estimate for the topological entropy. In the relative case, each Nielsen class is either Nielsen related to the invariant set, or has a well defined *itinerary* in terms of the components of the complement of the invariant set which is preserved by homotopy.

Another key ingredient is the study of the intersections of homotopy and isotopy classes of embedded curves. We are particularly concerned with curves that intersect each other a minimum possible number of times, and we call an intersection of two curves *essential* if it cannot be removed by a homotopy. The results we need are intuitively obvious “folk theorems,” and are used without proof.

## 2 Trellises, biasymptotic orbits and controlled graphs

In this section we define the various geometric structures which constitute the main objects of study. We introduce the notion of trellis and trellis type, and relate these to the well-known notion of homoclinic/heteroclinic orbit and tangle. We also give a number of definitions which allow us to describe some of the basic properties of trellises. We also review the definitions of graphs and thick graphs (or fibered surfaces) that we will use. We discuss reducibility of trellis mapping classes, and explain how this differs from reducibility of mapping classes relative to periodic orbits.

## 2.1 Tangles and Trellises

A tangle is the figure formed by the stable and unstable manifolds of a collection of periodic saddle orbits.

**Definition 2.1 (Tangle)** Let  $f$  be a diffeomorphism of a surface  $M$  with a finite invariant set  $P$  of hyperbolic saddle points. Let  $W^S(f; P)$  be the stable set of  $P$ , and  $W^U(f; P)$  the unstable set of  $P$ . The pair  $W = (W^U(f; P), W^S(f; P))$  is a *tangle* for  $f$ . If  $P$  consists of a single point, then  $W$  is a *homoclinic* tangle, otherwise  $W$  is a *heteroclinic* tangle.

We use the notation  $W^{U/S}$  in statements which hold for both stable and unstable manifolds. If the diffeomorphism  $f$  under consideration is clear from the context, we abbreviate  $W^{U/S}(f; P)$  to  $W^{U/S}(P)$ , and if the periodic point set  $P$  is also clear, we simply write  $W^{U/S}$ .

If  $p$  is a point of  $P$ , then the stable and unstable curves passing through  $p$  are denoted  $W^S(f; p)$  and  $W^U(f; p)$  respectively. We use the obvious notation for intervals of  $W^{U/S}$ : The closed interval in  $W^{U/S}$  with endpoints  $a$  and  $b$  is denoted  $W^{U/S}[a, b]$  and the open interval is denoted  $W^{U/S}(a, b)$ . Half-open intervals are denoted  $W^{U/S}[a, b)$  and  $W^{U/S}(a, b]$ . An open interval is a *branch* it is component of  $W^{U/S}(P) \setminus P$ . In general, a branch will be an immersed curve with an initial point  $p \in P$ , but no final point. We shall denote a branch at  $p$  containing a point  $b$  as  $W^{U/S}(p, b, \infty)$ .

An *intersection point* of a tangle  $W$  is a point in  $W^U \cap W^S$ , and is a point of a homoclinic or heteroclinic orbit. An intersection point  $q \in W^U(p_1) \cap W^U(p_2)$  is a *primary intersection point* or *pip* if  $W^U(p_1, q)$  and  $W^S(q, p_2)$  are disjoint [Wig91]. An intersection point  $q$  is *transverse* or *tangential* according to whether  $W^U$  and  $W^S$  cross transversely or tangentially at  $q$ .

There are a number of difficulties in working with tangles directly. These are due to the fact that if any branch intersects another, then it will be an immersed, rather than embedded, curve. The geometry of such a curve is extremely complicated, as is the global topology of the tangle. Further, since branches may have infinite length, tangles are impossible to compute directly. Instead, we work with subsets of tangles known as *trellises*, for which the stable and unstable curves have finite lengths.

**Definition 2.2 (Trellis)** Let  $f$  be a diffeomorphism of a surface  $M$  with a finite invariant set  $P$  of hyperbolic saddle points. A *trellis* for  $f$  is a pair  $T = (T^U, T^S)$ , where  $T^U$  and  $T^S$  are subsets of  $W^U(f; P)$  and  $W^S(f; P)$  respectively such that:

1.  $T^U$  and  $T^S$  both consist of finitely many compact intervals with non-empty interiors,
2.  $f(T^U) \supset T^U$  and  $f(T^S) \subset T^S$ .

We denote the set of periodic points of  $T$  by  $T^P$ , and the set of intersections of  $T^U$  and  $T^S$  by  $T^V$ . Note that  $T^U$  and  $T^S$  are sets of curves, but  $T^P$  and  $T^V$  are sets of points.

We write  $(f; T)$  to denote the *trellis mapping pair* consisting of a diffeomorphism  $f$  and a trellis  $T$  for  $f$ . The restriction of  $f$  to  $f^{-1}(T^U) \cup T^S$  has image  $T^U \cup f(T^S)$ , and may



be called a (*restricted*) *trellis map*. In this paper, however, we shall always be concerned with maps defined on the whole of the surface  $M$ .

There are two important relations between pairs  $(f; T)$ , namely *isotopy* and *conjugacy*. Note that in the definition of isotopy, we take isotopies through arbitrary diffeomorphisms; in particular,  $f_t$  need not fix the boundary of  $M$  pointwise.

**Definition 2.3 (Isotopy and conjugacy)** Let  $f_0$  and  $f_1$  be diffeomorphisms with trellises  $T_0$  and  $T_1$  respectively. Then  $(f_0; T_0)$  and  $(f_1; T_1)$  are *isotopic* if  $T_0 = T_1 = T$ , and there is an isotopy  $f_t$  from  $f_0$  to  $f_1$  such that for each  $t$ ,  $T$  is a trellis for  $f_t$ , and we say that the diffeomorphisms  $f_0$  and  $f_1$  are *isotopic relative to  $T$* . We say that  $(f_0; T_0)$  and  $(f_1; T_1)$  are *conjugate* if there is a homeomorphism  $h$  such that  $h \circ f_0 = f_1 \circ h$  and  $h(T_0^{U/S}) = T_1^{U/S}$ .

These relations allow us to define the equivalence classes of trellis map which will be our primary object of study.

**Definition 2.4 (Trellis mapping class and trellis type)** The *trellis mapping class*  $([f]; T)$  is the set of all trellis mapping pairs  $(\tilde{f}, T)$  which are isotopic to  $(f; T)$ . The *trellis type*  $[f; T]$  is the set of all trellis mapping pairs  $(\tilde{f}, \tilde{T})$  which are conjugate to a map in  $([f]; T)$ . If  $[f_0; T_0] = [f_1; T_1]$ , we consider  $(f_0; T_0)$  and  $(f_1; T_1)$  to be equivalent.

The most important subsets of  $T^{U/S}$  are intervals with endpoints in  $T^U \cap T^S$ , especially those which have no intersections in their interior.

**Definition 2.5 (Branch, arc, segment and end)** A *branch* of a trellis  $T$  is the closure of the intersection of  $T^{U/S}$  with a branch of the tangle  $W^{U/S}$ . An *arc* of  $T^{U/S}$  is a closed subinterval of  $T^{U/S}$  with endpoints in  $T^U \cap T^S$ . A *segment* of  $T^{U/S}$  is an arc of  $T^{U/S}$  with no topologically transverse intersection points in its interior. An *end interval* of a trellis is a subintervals of  $T^{U/S}$  which does not lie in any segment.

Note that points  $q_1$  and  $q_2$  lie in the same branch of  $T^{U/S}$  if and only if  $f^{\mp n}(q_1)$  and  $f^{\mp n}(q_2)$  lie in the same component of  $T^{U/S} \setminus P$  for some  $n$ . Unlike the case of a tangle, a branch of a trellis need not be connected, and contains its endpoints. A branch is *trivial* if it does not intersect any other branch. An *end intersection* of a trellis is an intersection which does not lie in two segments of  $T^{U/S}$  (so lies in at least one end interval of  $T^{U/S}$ ).

A surface diffeomorphism  $f$  with a periodic saddle orbit has infinitely many trellises, which are partially ordered by inclusion. Taking a smaller trellis gives a *subtrellis*, and a larger trellis a *supertrellis*.

**Definition 2.6 (Subtrellis and supertrellis)** Let  $f$  be a diffeomorphism with a tangle  $W = (W^U, W^S)$ , and trellises  $T$  and  $\hat{T}$ . Then  $T$  is a *subtrellis* of  $\hat{T}$  if  $T^{U/S} \subset \hat{T}^{U/S}$ , and  $T$  is a *subtrellis* of  $W$  if  $T^{U/S} \subset W^{U/S}$ . We say  $\hat{T}$  is a *supertrellis* of  $T$  and  $\widehat{W}$  is a *supertangle* of  $T$ .

Of particular importance are those supertrellises which can be obtained by iterating segments or branches.

**Definition 2.7 (Iterate and extension)** Let  $T$  be a trellis for a diffeomorphism  $f$ . Then a trellis  $\widehat{T}$  is an  $f$ -iterate of  $f$  if there exist positive integers  $n_u$  and  $n_s$  such that  $\widehat{T}^U = f^{n_u}(T^U)$  and  $\widehat{T}^S = f^{-n_s}(T^S)$ . A trellis  $\widehat{T}$  is an  $f$ -extension of  $T$  if there exists  $n$  such that

$$T^U \subset \widehat{T}^U \subset f^n(T^U) \quad \text{and} \quad T^S \subset \widehat{T}^S \subset f^{-n}(T^S).$$

An iterate/extension is a *stable iterate/extension* if  $\widehat{T}^U = T^U$  and an *unstable iterate/extension* if  $\widehat{T}^S = T^S$ . In a similar way, we say the trellis mapping class  $([f]; \widehat{T})$  is an *iterate* or *extension* of  $([f]; T)$  if  $(\widehat{f}; T) \in ([f]; T)$  and  $\widehat{T}$  is an  $\widehat{f}$ -iterate or extension of  $T$ .

A trellis  $\widehat{T}$  is therefore an extension of  $T$  if it is a subtrellis of some iterate of  $T$ . The main difference between extensions and supertrellises is that the periodic point sets  $\widehat{T}^P$  and  $T^P$  are equal for an extension, but  $T^P$  may be a strict subset of  $\widehat{T}^P$  for a supertrellis. This difference makes the analysis of supertrellises slightly more complicated than that of extensions.

The most important subsets of the surface  $M$  are those bounded by  $T^U$  and  $T^S$ .

**Definition 2.8 (Regions)** An *open region* of  $T$  is a component of  $M \setminus (T^U \cup T^S)$ . A *closed region* is the closure of an open region, and hence includes the stable and unstable boundary segments.

We will usually denote regions by the letter  $R$ . There are two special types of region which play an important role later.

**Definition 2.9 (Bigon and rectangle)** A *bigon* is a region which is a topological disc bounded by one stable and one unstable segment with internal angles less than  $\pi$ . A *rectangle* is a region which is a topological disc bounded by two stable and two unstable segments with internal angles less than  $\pi$ .

Bigons play a similar role in the trellis theory as do critical points in the kneading theory, and may be called *critical* regions.

The most important dynamical invariant of a trellis type is its *entropy*.

**Definition 2.10 (Entropy)** The *entropy* of a trellis type  $[f; T]$ , denoted  $h_{top}[f; T]$  is the infimum of the topological entropies of diffeomorphisms in  $[f; T]$ ; that is

$$h_{top}[f; T] = \inf\{h_{top}(\widehat{f}) : \widehat{f} \in [f; T]\}.$$

If this infimum is not a minimum, we sometimes write  $h_{top}[f; T] = \inf\{h_{top}(\widehat{f})\} + \epsilon$ .

We now illustrate the concepts introduced so far with one of the most important trellis types, the *Smale horseshoe trellis*.

**Example 2.11 (The Smale horseshoe trellis)**

The Smale horseshoe trellis is formed by the stable and unstable manifolds of the direct saddle fixed point (i.e. the saddle point with positive eigenvalues) of the Smale horseshoe

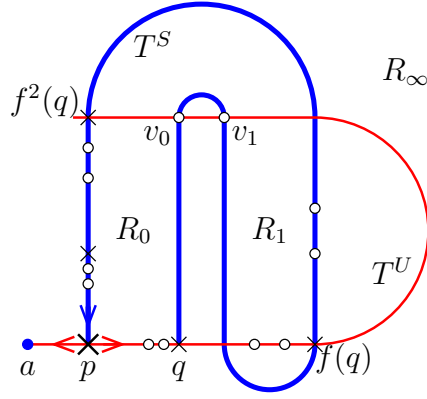


Figure 1: The Smale horseshoe trellis.

map, and is depicted in figure 1. The stable and unstable sets are subsets of the stable and unstable manifolds of the saddle fixed point  $p$ . The branches of this trellis are connected and the all intersection points are transverse. The points  $q$ ,  $f(q)$  and  $f^2(q)$  are primary intersection points on a single homoclinic orbit. The trellis type is completely determined by the geometry of  $T$  and the primary homoclinic orbit. Points on the orbits of the intersection points  $v_0$  and  $v_1$  are shown in white dots. These orbits are the *forcing orbits* for the Smale horseshoe trellis type; it is essentially impossible to remove any intersection points of the Smale horseshoe trellis by an isotopy without first removing  $v_0$  and  $v_1$  in a homoclinic bifurcation. Throughout this paper we follow the convention of showing forcing orbits as which dots, and primary intersection points as crosses. There are eight regions, an unbounded region  $R_\infty$ , three bigons and four rectangles. Under the Smale horseshoe map, there is a Cantor set of nonwandering points contained in the (closed) rectangular regions  $R_0$  and  $R_1$ , including a fixed point in  $R_1$ . All other points are wandering except for an attracting fixed point at the end of one unstable branch of  $T^U(p)$  in  $R_\infty$ . The topological entropy of the Smale horseshoe map is  $\log 2$ . We shall see later that any diffeomorphism with this trellis type must have topological entropy at least  $\log 2$ .

Notice that one of the unstable branches of the Smale horseshoe trellis ends in an attracting fixed point  $a$ , and one of the stable branches is non-existent. Both of these are therefore trivial branches.

Given a trellis, we can obtain a canonical map of pairs by *cutting* along the unstable curve. A formal description of cutting is given in [Col99], here we give an intuitive definition.

**Definition 2.12 (Cutting)** Let  $T$  be a trellis for a diffeomorphism  $f$  of some surface  $M$ . The topological pair obtained by cutting along the unstable curve is denoted  $\mathcal{C}T = (\mathcal{C}_{T^U}M, \mathcal{C}_{T^U}T^S)$ . The diffeomorphism  $f$  lifts for a map  $\mathcal{C}f$  on  $\mathcal{C}T$ .

Notice that the pair  $\mathcal{C}T = (\mathcal{C}_{T^U}M, \mathcal{C}_{T^U}T^S)$  contains the pair  $(M \setminus T^U, T^S \setminus T^U)$  as an open subset which is invariant under  $\mathcal{C}f$ . Indeed,  $\mathcal{C}T$  can be regarded as a natural compactification of  $M \setminus T^U$ . This compactness is important in the rigorous application of the Nielsen periodic point theory. However, the homotopy properties of  $(\mathcal{C}_{T^U}M, \mathcal{C}_{T^U}T^S)$  and  $(M \setminus T^U, T^S \setminus T^U)$  are essentially the same.

We finally give two subclasses of trellis which are important for technical reasons. We consider *proper trellises* in order that the stable set be well-behaved after cutting, and *well-formed trellises* in order that the isotopy classes of curves with endpoints in  $T^S$  are well behaved under iteration.

**Definition 2.13 (Proper and well-formed trellises)** We say a trellis  $T = (T^U, T^S)$  is *proper* if  $\partial T^S \subset T^U$  and  $\partial T^U \cap T^S = \emptyset$ . We say a trellis type  $[f; T]$  is *well-formed* if every component of  $T^U \cup f(T^S)$  contains a point of  $T^P$ .

In other words, a trellis is proper if the endpoints of intervals in  $T^S$  lie in  $T^U$ , but the endpoints of intervals in  $T^U$  do not lie in  $T^S$ . Note that considering proper trellises is no great loss of generality; an unstable curve with an endpoint in  $T^S$  can be extended by an arbitrarily short piece of curve, and a stable curve with an endpoint not in  $T^U$  can be shrunk by removing its end. Further, since the ends of a trellis give no useful information about the dynamics, both these procedures are harmless and do not affect dynamical computations in any way. Henceforth, all trellises will be taken to be proper trellises unless otherwise stated. Restricting to well-formed trellises will be important when considering the graph representative and gives a necessary condition for the topological entropy of the graph representative to be an optimal entropy bound for the trellis mapping class.

## 2.2 Biasymptotic orbits

By a *biasymptotic orbit*, we mean a homoclinic or heteroclinic orbit to an invariant set of periodic points. Just as we can consider isotopy classes of surface diffeomorphisms relative to periodic orbits and trellises, we can also consider isotopy classes relative to biasymptotic orbits.

**Definition 2.14 (Biasymptotic mapping class and biasymptotic type)** Let  $f$  be a surface diffeomorphism, and  $X$  a closed, invariant set consisting of a set of periodic saddle orbits  $X^P$  and a set  $X^V$  of biasymptotic orbits to  $X^P$ . We write  $(f; X)$  if  $f$  is a diffeomorphism with a set  $X$  which is the closure of a set of biasymptotic orbits. The isotopy class of  $f$  relative to  $X$  is called the *biasymptotic mapping class*  $([f]; X)$ . The conjugacy class of  $([f]; X)$  is the *biasymptotic type*  $[f; X]$ .

Methods for computing the dynamics forced by a biasymptotic type have been given by Handel [Han99] and Hulme [Hul00]. Here we show how to relate a biasymptotic type to a trellis type class, and hence derive a new method for computing the dynamics.

Since  $X^P$  is a finite set of saddle orbits, we can compute the stable and unstable manifolds  $W^{U/S}(f; X^P)$ . Since the points of  $X^V$  are biasymptotic to  $X^P$ , they are all intersection points of  $W^U$  and  $W^S$ . However,  $W^U$  and  $W^S$  may have extra intersections which are not in  $X$ , and indeed, typically have infinitely many such intersections. Different representatives of a biasymptotic mapping class will have different tangles.

**Definition 2.15 (Compatible tangle/trellis)** Let  $([f]; X)$  be a biasymptotic mapping class. Then if  $\hat{f}$  is any diffeomorphism isotopic to  $f$  relative to  $X$ , the tangle  $\widehat{W} =$

$W(\widehat{f}; X^P)$  is a *compatible tangle* for  $([f]; X)$ , and any trellis  $T$  with  $T^{U/S} \subset W^{U/S}$  is a *compatible trellis*.

There are many different trellis types compatible with a biasymptotic type  $([f]; X)$ . To restrict the possible trellises, we consider only trellises with connected branches, and specify the end intersections of each branch as a point of  $X$ . This gives rise to the notion of a *Birkhoff signature*.

**Definition 2.16 (Birkhoff signature)** A *Birkhoff signature* for a biasymptotic mapping class  $([f]; X)$  is a pair  $\mathcal{B} = (\mathcal{B}^U, \mathcal{B}^S)$  of subsets of  $X$  such that  $\mathcal{B}^U$  contains one point of  $X$  from each branch of  $W^U(f; X^P)$  and  $\mathcal{B}^S$  contain one point of  $X$  from each branch of  $W^S(f; X^P)$ . If  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are two signatures, we say  $\mathcal{B}_1 \leq \mathcal{B}_2$  if the trellis  $T_1$  for  $f$  compatible with  $([f]; X)$  and with end intersections in  $\mathcal{B}_1$  is a subtrellis of the compatible trellis  $T_2$  with end intersections in  $\mathcal{B}_2$ .

Passing from a trellis to a set of biasymptotic orbits is simpler; we take  $V$  to be a subset of the set of intersections of  $T$ , and  $X = \text{cl}(\bigcup_{n=-\infty}^{\infty} f^n(V))$ . An important question is whether  $X$  and  $T$  force the same dynamics. This will be answered in Section 3.

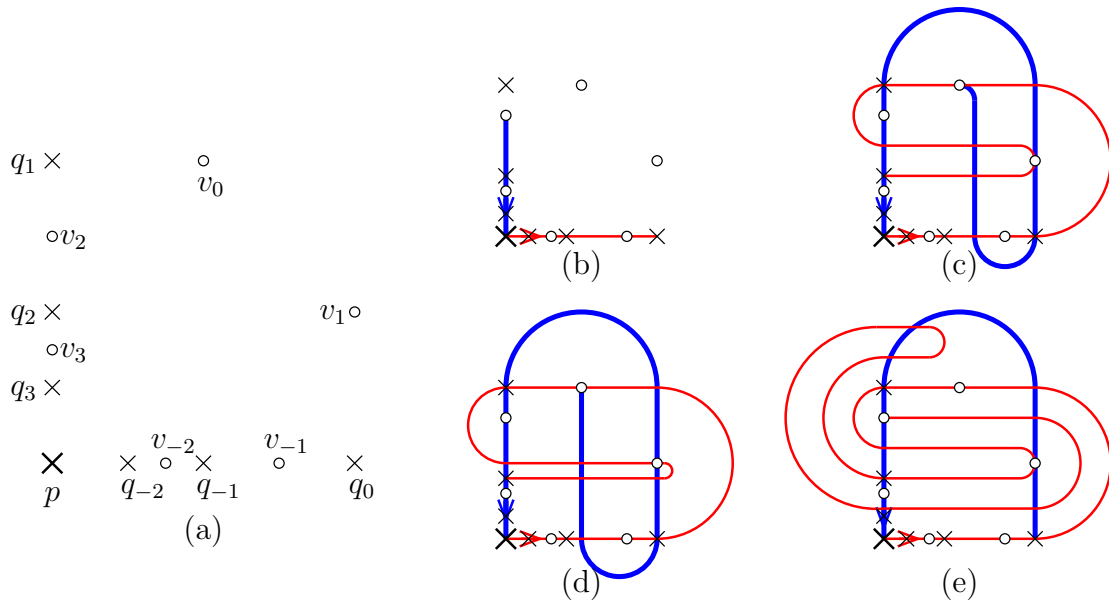


Figure 2: (a) Orbits  $\{q_i\}$  and  $\{v_i\}$  of the Smale horseshoe map. (b-e) Trellises compatible with  $\{q_i\}$  and  $\{v_i\}$ .

**Example 2.17 (Compatible trellises)**

In figure 2 we show four trellises compatible with the biasymptotic mapping class  $([f]; X)$ , where

$$X = \{p\} \cup \{q_i\} \cup \{v_i\}$$

The orbits  $q_i$  and  $v_i$  shown in Figure 2(a) are orbits of the Smale horseshoe map homoclinic to the saddle point  $p$ . Both these orbits lie in the same branch of  $W^U(p)$  and  $W^S(p)$ , so a signature consists of a pair of points from either orbit. The trellis in Figure 2(b) has

signature  $(\{q_0\}, \{v_2\})$  and no intersections other than  $p$ . The trellises in Figure 2(c) and Figure 2(d) have signature  $(\{q_2\}, \{v_0\})$ . However, the trellis in Figure 2(c) has a tangency at  $v_1$ , whereas the trellis in Figure 2(d) has a transverse intersection with negative orientation. The trellis in Figure 2(e) has signature  $(\{v_2\}, \{q_0\})$ , but has a pair of transverse intersections which is not forced by  $X$ .

## 2.3 Graph maps

Our main tool for computing and describing the dynamics forced by a trellis mapping class or trellis type is to relate the trellis map to a graph map. We will mostly use the same terminology and notation for graphs as in [BH95], though our terminology for thick graphs is closer to that of [FM93].

A *graph*  $G$  is a one-dimensional CW-complex with vertices  $V(G)$  and edges  $E(G)$ . An *edge-path* is a list  $e_1 \dots e_n$  of oriented edges of  $G$  such that  $\iota(\bar{e}_i) = \iota(e_{i+1})$  for  $1 \leq i < n$ , and an *edge loop* is a cyclically-ordered list of edges. The *trivial* edge-path contains no edges and is denoted  $\cdot$ . An edge-path  $e_1 \dots e_n$  *back-tracks* if  $e_{i+1} = \bar{e}_i$  for some  $i$ , otherwise it is *tight*. A *graph map*  $g$  is a self-map of  $G$  taking a vertex to a vertex, and edge  $e$  to an edge-path  $e_1 \dots e_k$  such that  $\iota(e_1) = g(\iota(e))$  for all directed edges  $e$ . The *derivative* map  $\partial g$  takes oriented edges to oriented edges or  $\cdot$ , with  $\partial g(e_i) = e_j$  if  $g(e_i) = e_j \dots$  and  $\partial g(e_i) = \cdot$  if  $g(e_i) = \cdot$ .

We will always consider a graph embedded in a surface by an embedding  $i$ . This induces a natural cyclic order  $\triangleleft$  on the oriented edges starting at each vertex. A pair of edges  $(e_1, e_2)$  is a *turn* in  $G$  at vertex  $v$  if  $v = \iota(e_1) = \iota(e_2)$  and  $e_1 \triangleleft e_2$ , so  $e_2$  immediately follows  $e_1$  in the cyclic order at  $v$ . An edge-loop  $\pi = \dots p_1 p_2 \dots p_n, \dots$  is *peripheral* in  $G$  if  $(p_{i+1}, \bar{p}_i)$  is a turn in  $G$  for all  $i$ .

A graph embedded in a surface can be “thickened” to obtain a *thick graph* (or *fibred surface*) with *thick vertices* and *thick edges*, the latter foliated by stable and unstable leaves. The inverse of this thickening is achieved by collapsing thick vertices to vertices and thick edges to edges. A *thick graph map* is an injective map of a thick graph which preserves the stable and unstable foliations, is a strict contraction on the set of thick vertices and on the stable leaves, and is a strict expansion on the unstable leaves. A graph map is *embeddable* if it can be obtained from a thick graph map by collapsing the stable leaves.

The *peripheral subgraph*  $P$  of an embeddable graph map  $g$  is a maximal invariant subset of  $G$  consisting of simple peripheral loops. Edges of  $P$  are called *peripheral edges*. If  $g^n(e) \subset P$  for some  $n$ , then  $e$  is *pre-peripheral*. The set of pre-peripheral edges is denoted  $\text{Pre-}P$ , and contains  $P$ .

The *transition matrix* of a graph map  $g$  is the matrix  $A = (a_{ij})$  where  $a_{ij}$  is the number of times the edge  $e_j$  appears in the image path of edge  $e_i$ . The largest eigenvalue of  $A$  is the *growth rate*  $\lambda$  of  $g$ , and the logarithm of the growth rate gives the topological entropy of  $g$ .

We now relate a trellis mapping class to a graph map via the map  $\mathcal{C}f$  of the topological

pair  $\mathcal{CT}$  obtained by cutting along the unstable curves. The result is a map of a topological pair  $(G, W)$  where  $W$  is a finite subset of  $G$ .

**Definition 2.18 (Controlled graph map)** Let  $(G, W)$  be a topological pair where  $G$  is a graph and  $W$  is a finite subset of  $G$ . Then edges of  $G$  containing points of  $W$  are called *control edges*. A graph map  $g : (G, W) \rightarrow (G, W)$  is a *controlled graph map* if  $g(z)$  is a control edge whenever  $z$  is a control edge.

A vertex of  $G$  which is the endpoint of a control edge is called a *control vertex*. All other vertices are called *free vertices*, and edges which are not control edges are called *free edges*. Free edges which are neither peripheral nor pre-peripheral are called *expanding edges*.

A controlled graph is *proper* if it is connected and each free vertex has valence at least 3. A proper controlled graph has at most  $3\#(W) - 3\chi$  edges, and  $3\#(W) - 2\chi$  vertices, where  $\#(W)$  is the number of points of  $W$ , or equivalently, the number of control edges.

To relate maps of pairs on different spaces, we need the notion of *exact homotopy equivalence* defined in [Col01]. A graph map representing the topology of a trellis via exact homotopy equivalence is called *compatible* with the trellis.

**Definition 2.19 (Compatible graph map)** Let  $(G, W)$  be a topological pair where  $G$  is a graph and  $W$  is a finite subset of  $G$ . Then  $(G, W)$  is *compatible* with a transverse trellis  $T$  if  $(G, W)$  and  $(\mathcal{C}_{TV}M, \mathcal{C}_{TV}T^S)$  are exact homotopy equivalent by an embedding  $i : (G, W) \rightarrow (\mathcal{C}_{TV}M, \mathcal{C}_{TV}T^S)$ , and  $G$  crosses  $T^S$  transversely. A controlled graph map  $g$  of  $(G, W)$  is *compatible* with the trellis mapping class  $([f]; T)$  if the embedding  $i$  is an exact homotopy equivalence between  $g$  and  $\mathcal{C}f$ , and  $g$  and  $\mathcal{C}f$  have the same orientation at points of  $W$ .

Note that the inclusion  $i$  induced a bijection between the regions of  $\mathcal{CT}$  (and hence of  $T$ ) with the regions of  $(G, W)$ , and that all compatible graphs are exact homotopy equivalent. We restrict to transverse trellises since a trellis with tangencies may not have a compatible controlled graph; while it is always possible to find a topological pair  $(G, H)$  which is exact homotopy equivalent to  $(\mathcal{C}_{TV}M, \mathcal{C}_{TV}T^S)$  for which  $G$  is a graph, it may not be possible to take  $H$  to be a finite set of points.

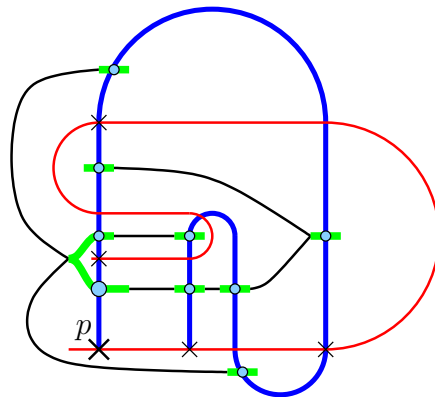


Figure 3: The controlled graph  $(G, W)$  is compatible with the trellis  $T$ .

The controlled graph shown in figure 3 is compatible with the trellis  $T$ . The control edges are shown as thick shaded lines.

## 2.4 Reducibility

A system is *reducible* if it can be split into simpler subsystems. In the Nielsen-Thurston theory, we say a homeomorphism  $f$  of a surface  $M$  is reducible if there is a collection  $C$  of simple closed curves such that every component of  $M \setminus C$  has negative Euler characteristic,  $f$  is isotopic to a homeomorphism  $\tilde{f}$  such that  $\tilde{f}(C) = C$ . Since the Euler characteristic of  $M \setminus C$  is the same as that of  $M$ , either  $M \setminus C$  consists of a single component with the same Euler characteristic but lower genus, or  $M \setminus C$  consists of components whose Euler characteristic is greater (less negative) than that of  $C$ .

For the case of a trellis mapping class, the situation is complicated somewhat. A trellis  $T$  on the sphere may consist of two separate pieces (i.e.  $T^U \cup T^S$  is not connected) in which case we may be able to simplify by considering each piece separately. More importantly, we may have a situation in which  $T$  has two periodic orbits,  $P_A$  and  $P_B$ , for which  $T^U(P_A) \cap T^S(P_B) = \emptyset$ . In this case, we may be able to isotope to obtain a diffeomorphism  $f$  which has an attractor  $A$  containing  $P_A$  and a repeller  $B$  containing  $P_B$ , and again we can consider these situations separately.

**Definition 2.20 (Reducibility of a trellis mapping class)** A trellis mapping class  $([f]; T)$  has an *invariant curve reduction* if there is a diffeomorphism  $\tilde{f}$  homotopic to  $f$  relative to  $T$  and a closed one-manifold  $C$  such that  $C$  is invariant under  $\tilde{f}$ , is disjoint from  $T^U \cup T^S$  and each component of the complement of  $C$  either contains a subtrellis of  $T$ , or has negative Euler characteristic. If the complement of  $C$  is disconnected, we have a *separating reduction*, otherwise we have a *non-separating reduction*. A trellis mapping class  $([f]; T)$  has an *attractor-repeller reduction* if there is a diffeomorphism  $\tilde{f}$  homotopic to  $f$  relative to  $T$  and a closed one-manifold  $C$  such that  $C$  divides  $M$  into subsets  $A$  and  $B$  such that  $\text{cl}(\tilde{f}(A)) \subset \text{int}(A)$ , and both  $A$  and  $B$  contain a point of  $T^P$

In both cases, the set  $C$  is called a set of *reducing curves*. It is clear that reducibility is a property of the trellis type, and not just the trellis mapping class. For most of this paper, we only consider irreducible trellis mapping classes.

### Example 2.21 (An attractor-repeller reduction)

A trellis type with an attractor-repeller reduction is shown in figure 4. A sub-basin  $A$  of an attractor is shaded. Notice that the unstable set  $T^U(p_1)$  lies in  $A$ , which means it is possible to find a diffeomorphism  $\tilde{f}$  for which  $W^U(\tilde{f}; p_1) \cap W^S(\tilde{f}; p_0) = \emptyset$ . The closure of the unstable manifold of  $p_1$  is then a non-trivial chaotic attractor, though it may have non-trivial sub-attractors.

The reducibility of a trellis mapping class  $([f]; T)$  can be seen from a compatible graph map  $(g; G, W)$ . If  $g$  has an invariant subgraph  $H$  which does not contain any control edges, then  $([f]; T)$  has an invariant curve reduction if either  $H$  has negative Euler characteristic, or is a union of non-peripheral simple closed curves. If both  $H$  and the complement of  $H$



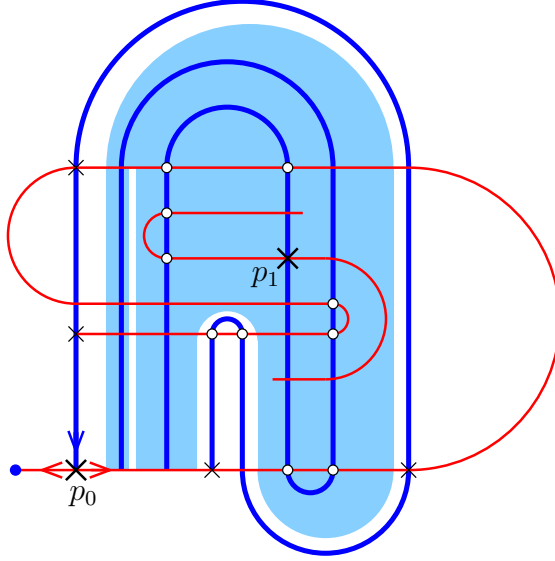


Figure 4: An trellis with an attractor-repeller reduction.

contain periodic control edges, and  $H$  does not deformation-retract onto its control edges, then  $([f]; T)$  has an attractor-repeller reduction.

### 3 Minimal Trellises

In this section, we consider how to construct a canonical trellis from a set of biasymptotic orbits to a trellis, and how to extend a trellis to a larger trellis. We shall usually consider extensions with a given Birkhoff signature. As previously remarked, this does not guarantee a unique trellis type. Instead, we define what it means for a trellis to be a *minimal* trellis compatible with a given biasymptotic mapping class. By carefully isotoping a given trellis mapping pair, we can construct a minimal compatible trellis. We then show that the type of a minimal compatible trellis is almost unique given its end points, the only ambiguities arising from the orientation (transverse or tangent) of the intersections.

We also consider minimal supertrellises of a given trellis mapping class. A minimal supertrellis is forced by the biasymptotic orbits of the intersections of the original trellis. Unlike a minimal trellis for a biasymptotic mapping class, the orientations of the intersections are uniquely determined.

#### 3.1 Pruning isotopies

If  $f_t$  is an isotopy of surface diffeomorphisms with a continuously-varying collection  $P_t$  of periodic saddle orbits, then the stable and unstable manifolds  $W^{U/S}(P_t; f_t)$  can also be continued under the isotopy. We can therefore restrict to a continuously-varying set of trellises  $T_t$  for  $f_t$ . The trellis type  $[f_t; T_t]$  changes on passing through a tangency, or if an end intersection is created or removed. The bifurcations occurring near tangencies of

dissipative saddle points are extremely complicated. In particular, it is known that homoclinic orbits must be both created and destroyed in a neighbourhood of the bifurcation. However, if we restrict to a trellis, we can isotope in such a way as to only remove intersection points. Such an isotopy is called a *pruning isotopy*, by analogy with the prunings of de Carvalho [dC99].

**Definition 3.1 (Pruning isotopy)** Let  $f_t$  be an isotopy and  $T_t$  be a trellis for  $f_t$ . Then  $(f_t; T_t)$  is a *pruning isotopy* if for every  $t' \in [0, 1]$ , there is an isotopy of embeddings  $T_{t'}^V \hookrightarrow T_t^V$  for  $t \leq t'$ .

Without loss of generality, we can restrict to pruning isotopies which preserve the unstable set  $T^U$ . A tangency can only be created during a pruning isotopy by collapsing certain bigons. These bigons are called an *inner bigons*

**Definition 3.2 (Inner bigon)** Let  $(f; T)$  be a trellis mapping pair. A bigon  $B$  is *inner* if  $B \cap \bigcup_{n=-\infty}^{\infty} f^n(T^V) = B^V$ , where  $B^V$  denotes the vertices of  $B$ .

Of course, an inner bigon containing a point of  $X$  cannot be removed by an isotopy relative to  $X$ .

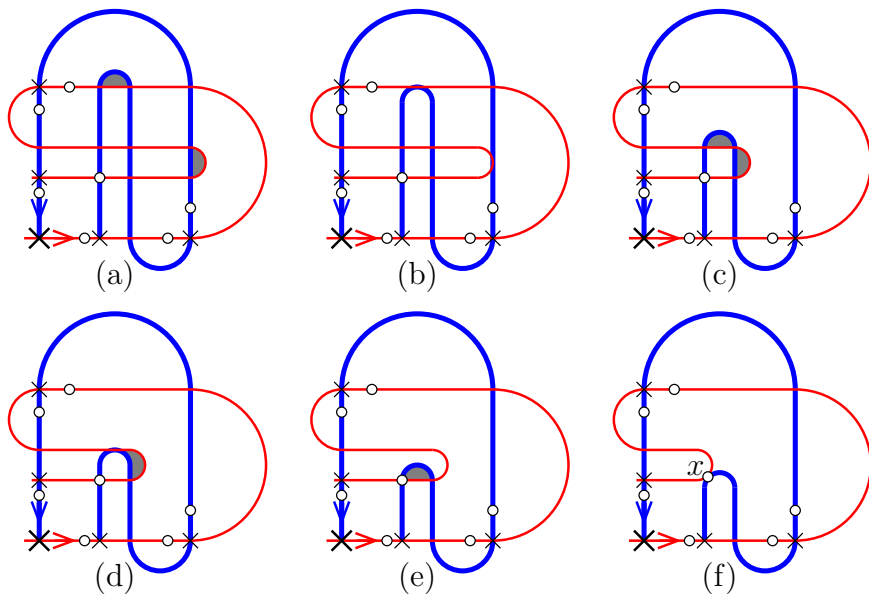


Figure 5: A pruning isotopy starting with the Smale horseshoe trellis (a) and ending with a minimal trellis compatible with the orbit of  $x$  (f).

An example of a pruning isotopy is shown in figure 5. The isotopy starts with the Smale horseshoe trellis, and ends with a trellis which is minimal with respect to the orbit of  $x$ . The inner bigons are shaded.

One of the fundamental properties of pruning is that it reduces the number of essential Nielsen classes. An immediate consequence is that the Nielsen entropy of the trellis mapping class is reduced during a pruning isotopy.

**Theorem 3.3 (Entropy decreases during pruning)** *Suppose  $(f_t; T_t)$  is a pruning isotopy. Then  $h_{niel}[f_0; T_0] \geq h_{niel}[f_1; T_1]$ . Further, any essential Nielsen class for  $[f_1; T_1]$  can be continued to an essential Nielsen class of  $[f_0; T_0]$ .*

PROOF: Without loss of generality, we can assume  $T^U$  is constant and that  $T_1^V \subset T_t^V$  for all  $t$ . We need only show that  $(f_0; T_0)$  has at least as many essential Nielsen classes of any period as  $f_1$ . Let  $\mathcal{C}F$  be the fat homotopy of  $\mathcal{C}(f)_t$ , so  $\mathcal{C}F(x, t) = (\mathcal{C}f_t(x), t)$  for all  $t$ . Note that  $\mathcal{C}F$  is defined on  $\mathcal{C}_{T^U}M \times I$ , and maps the set  $\bigcup_{t \in [0, 1]} \mathcal{C}(T^U(T_t^S)) \times \{t\}$  into itself.

Let  $(x_i)$  and  $(y_i)$  be periodic orbits of  $\mathcal{C}f_1$  in different Nielsen classes. Suppose  $(x_i, 1)$  and  $(y_i, 1)$  are Nielsen related for  $\mathcal{C}F$ .

Then there are curves  $\alpha_i(s)$  from  $x_i$  to  $y_i$  in  $\mathcal{C}_{T^U}M$  and parameters  $\tau(s)$  with  $\tau(0) = \tau(1) = 1$  such that  $(\alpha_i, \tau_i)$  is an exact curve from  $(I, J)$  for some  $J$  to  $(\mathcal{C}_{T^U}M \times I, \bigcup_{t \in [0, 1]} \mathcal{C}_T(T_t^S) \times \{t\})$ . If  $J$  contains parameter values other than 0 and 1, (that is, the curves  $(\alpha_i, \tau_i)$  have interior crossings with the stable manifolds), we can restrict attention to each piece with no interior crossings separately. Hence, we can assume  $J \subset \{0, 1\}$ .

Now since  $\mathcal{C}F \circ (\alpha_i \tau)$  is homotopic to  $(\alpha_{i+1}, \tau)$ , we have  $f_{\tau(\cdot)} \circ \alpha_i \sim \alpha_{i+1}$ , and hence  $f_1 \circ \alpha_i \sim \alpha_{i+1}$ . We need to show that the  $\alpha_i$  have no essential intersections with  $T_1^S$ . Since  $(\alpha_i(s))$  does not intersect  $\mathcal{C}_T(T_\tau(s)^S)$ , the curves  $\alpha_i$  have no essential intersections with arcs of  $T_1^S$  considered as cross-cuts relative to  $T_1^V$ , and so we can homotope  $\alpha_i$  to  $\beta_i$  such that  $\beta_i$  has no interior intersections with  $T_1^S$ .

Therefore, if  $(x_i)$  and  $(y_i)$  are periodic orbits of  $f_1$  which are Nielsen related for the fat homotopy of  $f_t$  on  $[0, 1]$ , then  $(x_i)$  and  $(y_i)$  are Nielsen related as periodic orbits of  $f_1$ . Hence any two Nielsen classes which are Nielsen separated for  $f_1$  can be continued for  $f_t$  with  $t < 1$  and remain Nielsen separated. Therefore the number of essential Nielsen classes increases as  $t$  decreases.  $\square$

We note that while the topological entropy appears to change continuously on passing through a tangency, the Nielsen entropy undergoes a discontinuous change.

## 3.2 Minimal compatible trellises

Given a Birkhoff signature, there are still many, indeed infinitely many, trellis types compatible with a biasymptotic type. Of these, we are most interested in trellises which have no intersections which can be removed by a pruning isotopy in the biasymptotic mapping class. Such a trellis is a *minimal* compatible trellis.

**Definition 3.4 (Minimal compatible trellis)** Let  $([f]; X)$  be a biasymptotic mapping class. Then a trellis mapping pair  $(\tilde{f}; \tilde{T})$  is a *minimal* compatible trellis mapping pair if  $\tilde{f}$  is isotopic to  $f$  relative to  $X$ , and

1. Every bigon of  $\tilde{T}$  either contains a point of  $X$  in the interior of one of its boundary segments, or both vertices are points of  $X$ .

2. Every tangency of  $\tilde{T}^U$  and  $\tilde{T}^S$  is a point of  $X$ .
3. The end intersections of  $\tilde{T}^U$  and  $\tilde{T}^S$  are points of  $X$ .

The trellis  $\tilde{T}$  is called a *minimal compatible trellis* for  $([f]; X)$ .

Clearly, a trellis with a minimal number of intersections in the biasymptotic mapping class is a minimal compatible trellis. We shall see later that a minimal compatible trellis is (essentially) unique.

The main result of this section is that for every set  $X$  of biasymptotic orbits, there is a minimal compatible trellis with any Birkhoff signature. The basic idea of the proof is to start with any compatible trellis mapping pair and try to remove intersection points by a pruning isotopy. If  $(f; T)$  is not minimal, then either  $T$  has a tangency  $v$  which is not

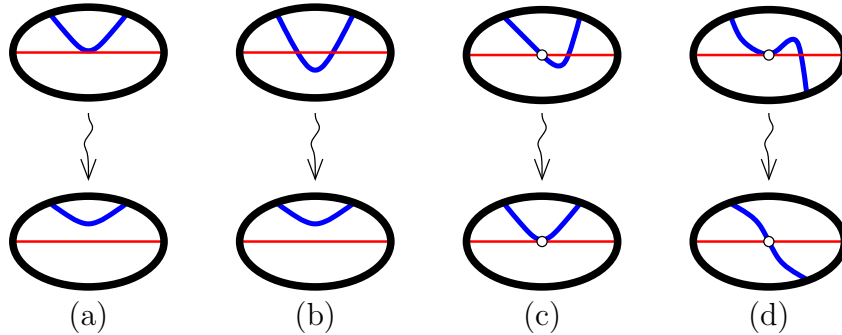


Figure 6: Isotopies removing intersections in  $K$ : (a) removing intersections from a bigon, (b) removing a tangency, (c) making a tangency at a point of  $X$  and (d) moving a transverse intersection to a point of  $X$ .

in  $X$ , or an inner bigon, and we can reduce the number of intersections in a disc  $K$  by performing an isotopy of  $T^S$  of the form shown in figure 6.

**Theorem 3.5 (Existence and uniqueness of minimal compatible trellis)** *Suppose  $([f]; X)$  is a biasymptotic mapping class and  $\mathcal{B} = (\mathcal{B}^U, \mathcal{B}^S)$  is a signature in  $X$ . Then there is a minimal  $(\tilde{f}; \tilde{T})$  compatible with  $([f]; X)$  with signature  $\mathcal{B}$ . Further, if  $(\tilde{f}_0; \tilde{T}_0)$  and  $(\tilde{f}_1; \tilde{T}_1)$  are minimal trellis mapping pairs compatible with  $([f]; X)$  and have the same type (transverse or tangential) of intersection at each point of  $X$  then  $[\tilde{f}_0; \tilde{T}_0] = [\tilde{f}_1; \tilde{T}_1]$ . In particular, if  $([f]; X)$  has a minimal transverse compatible trellis, then there is a unique minimal transverse compatible trellis type.*

**PROOF OF THEOREM 3.5 (EXISTENCE):** Suppose  $\tilde{f}$  is a diffeomorphism which agrees with  $f$  except on a compact set  $K$ . Suppose further that there exist positive integers  $n_u$  and  $n_s$  such that  $f^{n_u}(K) \cap T^U = f^{n_s}(K) \cap T^S = \emptyset$  and the sets  $f^n(K)$  are disjoint for  $-n_s \leq n \leq n_u$ . Then  $\tilde{f}^n(K) = f^n(K)$  for  $-n_s \leq n \leq n_u$ . Let  $\tilde{T}$  be the trellis for  $\tilde{f}$  with  $\tilde{T}^{U/S} = \tilde{f}^{\pm n_{u/s}}(f^{\mp n_{u/s}}(T^{U/S}))$ . Then  $\tilde{T}^U \setminus (\bigcup_{n=1}^{n_u} f^n(K)) = \tilde{T}^U \setminus (\bigcup_{n=1}^{n_u} f^n(K))$  and  $\tilde{T}^S \setminus (\bigcup_{n=0}^{n_s} f^{-n}(K)) = \tilde{T}^S \setminus (\bigcup_{n=0}^{n_s} f^{-n}(K))$ . We also have  $K \cap \tilde{T}^S = \tilde{f}^{-1}(f(T^S \cap K))$ , so writing  $\tilde{f} = f \circ h^{-1}$ , we see that  $h$  is supported in  $K$  and  $\tilde{T}^S \cap K = h(T^S \cap K)$ . Hence  $\tilde{T}^V \cap K = \tilde{T}^U \cap \tilde{T}^S \cap K = T^U \cap h(T^S) \cap K$ .

Now suppose we also have  $f^{n_u}(K) \cap T^S = f^{n_u}(K \cap T^S)$  and  $f^{-n_s}(K) \cap T^U = f^{-n_s}(K \cap T^U)$ , or equivalently,  $K \cap f^{n_u}(T^S) = K \cap T^S$  and  $K \cap f^{-n_s}(T^U) = K \cap T^U$ . Then also  $\tilde{f}^{n_u}(K) \cap T^S = \tilde{f}^{n_u}(K \cap T^S)$  and  $\tilde{f}^{-n_s}(K) \cap \tilde{T}^U = \tilde{f}^{-n_s}(K \cap \tilde{T}^U)$ . Hence  $\tilde{T}^V \cap f^n(K) = \tilde{T}^U \cap \tilde{T}^S \cap \tilde{f}^n(K) = \tilde{T}^U \cap \tilde{f}^n(\tilde{T}^S \cap (K)) \subset \tilde{f}^n(\tilde{T}^U \cap \tilde{T}^S \cap (K)) = \tilde{f}^n(\tilde{T}^V \cap (K))$  for  $1 \leq n \leq n_u$ , and a similar argument shows  $\tilde{T}^V \cap f^n(K) \subset \tilde{f}^n(\tilde{T}^V \cap (K))$  for  $-n_s \leq n \leq 0$ . Hence  $\tilde{T}^V = T^V$  outside  $\bigcup_{n=-n_s}^{n_u} f^n(K)$  and  $\tilde{f}^n(K) \cap \tilde{T}^V \subset \tilde{f}^n(K \cap \tilde{T}^V)$ . In particular, if  $K \cap \tilde{T}^V = \emptyset$ , then  $\tilde{T}^V = T^V$ .

We can always ensure the sets  $f^n(K)$  for  $-n_s \leq n \leq n_u$  are disjoint by choosing  $K$  to be a sufficiently small disc. Further, if  $K$  does not contain a point of  $\mathcal{B}$ , then  $K \cap f^{-n_u}(T^S) = K \cap T^S$  and  $K \cap f^{n_s}(T^U) = K \cap T^U$ . If  $K$  is a disc containing at most one point of  $X$ , then  $\tilde{f}$  is isotopic to  $f$  relative to  $X$ .

We now take isotopies to reduce the number of intersections. First, we can then isotope in a neighbourhood of any non-isolated intersection points to obtain a trellis  $T_1$  for a diffeomorphism  $f_1$  with finitely many intersections.

If  $(f_1, T_1)$  is minimal, then we are done. If  $v$  is a tangency of  $T_1$  which is not a point of  $X$ , we can isotope at  $v$  as shown in Figure 6(a) to obtain a trellis with fewer intersections. We obtain a pair  $(f_2, T_2)$  where  $T_2$  is a trellis with no tangencies except at points of  $X$ . If  $T_2$  has any bigons which contain no points of  $X$ , then we must be able to find a bigon  $B$  for which  $B \cap \bigcup_{n=-\infty}^{\infty} f_2^n(T_2^V) = B^V$ . Then we can find a disc  $K$  containing  $B$ , and isotope in  $K$  to remove all intersections in  $K$  as shown in Figure 6(b) with fewer intersections. We obtain a pair  $(f_3, T_3)$  for which the only bigons containing a point of  $X$  are as shown in Figure 6(c,d). A further isotopy again reduces the number of intersections.

After finitely many steps we therefore obtain a pair  $(\tilde{f}; \tilde{T})$  which is minimal □



Figure 7: Possibilities for bigons with both vertices in  $X$ .

We now show that a minimal trellis compatible with a given biasymptotic mapping class with a given signature is essentially unique. Unfortunately, there is one problem; if there are two points  $x_0$  and  $x_1$  of  $X$  which form the intersections of an inner bigon, the orientation of these intersections is unspecified by the minimality conditions, as shown in figure 7. This may result in different minimal compatible trellises.

As part of the proof, we use the following result, which is important in its own right, since it shows that a trellis type is specified by a trellis  $T$  and a biasymptotic mapping class containing the end intersections of  $T^U$  and  $T^S$ . This greatly simplifies the representation of a trellis mapping class required for computations, and justifies our custom of specifying a trellis mapping class by drawing the trellis and part of a biasymptotic orbit.

**Theorem 3.6** *Suppose  $(f_0; T)$  and  $(f_1; T)$  are compatible with the biasymptotic mapping class  $([f]; X)$ . Then the trellis mapping classes  $([f_0]; T)$  and  $([f_1]; T)$ , are equal.*

PROOF: Let  $h = f_0^{-1} \circ f_1$ . Then  $h$  is isotopic to  $id$  relative to  $X$ . Since  $T^U$  consists of cross-cuts to  $X$ , there is an isotopy  $\tilde{h}_t$  such that  $\tilde{h}_0 = id$ ,  $\tilde{h}_1 = h$  and  $\tilde{h}_t(T^U) = T^U$  for all  $t$ . If we now cut along  $T^U$ , we find that  $h$  is isotopic to the identity by the isotopy  $\tilde{h}$ , and  $T^S$  is a set of cross-cuts. Therefore, by the isotopy extension theorem, there is an isotopy  $h_t$  such that  $h_0 = id$ ,  $h_1 = h$ , and for all values of  $t$ ,  $h_t(T^U) = T^U$  and  $h_t(T^S) = T^S$ . To show that  $f_t = f_0 \circ h_t$  gives the required isotopy from  $f_0$  to  $f_1$ , we need only check that  $f_t$  preserves  $T^U$  and  $T^S$ . We have  $f_t^{-1}(T^U) = h_t^{-1}(f_0^{-1}(T^U)) = f_0^{-1}(T^U)$  since  $f_0^{-1}(T^U) \subset T^U$ , and  $f_t(T^S) = f_0(h_t(T^S)) = f_0(T^S)$ , as required.  $\square$

We can now complete the proof of Theorem 3.5.

PROOF OF THEOREM 3.5 (UNIQUENESS): Since the curves  $T^U$  and  $T^S$  are mutually homotopic, have the same types of intersection at  $X$ , and have minimal intersections with respect to  $X$ , there is a homeomorphism  $h$  such that  $h(T_0) = T_1$  and  $h$  is isotopic to the identity relative to  $X$ . Then  $(h^{-1} \circ f_1 \circ h; h^{-1} \circ T_1) = (h^{-1} \circ f_1 \circ h; T_0)$ , and  $f_0$  and  $h^{-1} \circ f_1 \circ h$  are isotopic relative to  $X$ . Then by Theorem 3.6,  $(f_0; T_0)$  and  $(h^{-1} \circ f_1 \circ h; T_0)$  are in the same trellis mapping class, so  $(f_0; T_0)$  and  $(f_1; T_1)$  have the same type.  $\square$

Since minimal trellises are obtained by removing intersections by pruning, we immediately obtain:

**Corollary 3.7 (Minimal trellises have minimal numbers of intersections)** *Let  $[f; T]$  be a minimal trellis type for the biasymptotic mapping class  $([f]; X)$ . Then  $T$  has finitely many intersections, and if  $[\tilde{f}; \tilde{T}]$  is any other trellis type compatible with  $([f]; X)$  with the same signature, then  $\tilde{T}$  has more intersections than  $T$ .*

### 3.3 Minimal extensions and supertrellises

We have considered how to construct a from a set of biasymptotic orbits a trellis with a given signature. We can also extend trellises to longer trellises. Of particular interest are those which do not introduce any more intersections than necessary. Such trellises are called *minimal iterates* and *minimal extensions*.

**Definition 3.8 (Minimal iterate, extension and supertrellis)** Let  $([f]; T)$  be a trellis mapping class. A supertrellis  $(\tilde{f}; \tilde{T})$  of  $([f]; T)$  is a *natural minimal supertrellis* of  $([f]; T)$  if

1. Every bigon contains a point of  $\bigcup_{n=-\infty}^{\infty} \tilde{f}^n(T^V)$ .
2. Every tangency is a point of  $\bigcup_{n=-\infty}^{\infty} \tilde{f}^n(T^V)$ .
3. The ends intersections of  $\tilde{T}^U$  are points of  $f(T^S)$  and the end intersections of  $\tilde{T}^S$  are points of  $f^{-1}(T^U)$ .

A *minimal supertrellis* is any supertrellis which is a subtrellis of a natural minimal supertrellis. A minimal supertrellis which is an iterate or extension is a *minimal iterate* or *minimal extension*.

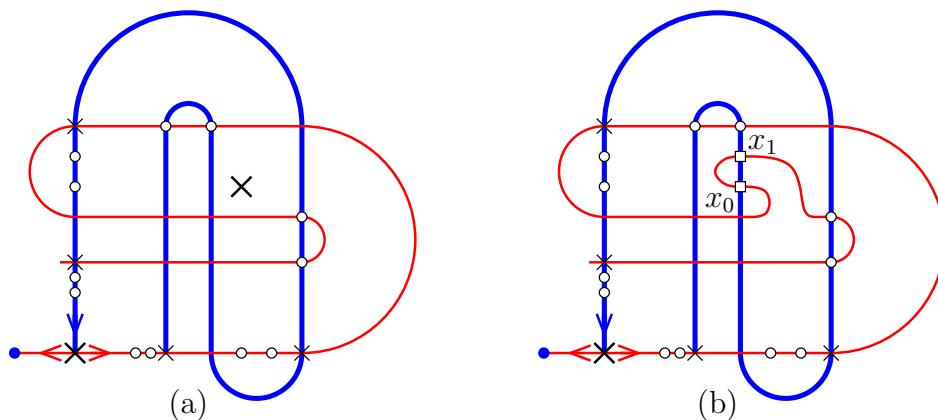


Figure 8: Extensions of the Smale horseshoe trellis (a) is a minimal extension, (b) is not.

The trellis shown in Figure 8(a) is a minimal extension of the Smale horseshoe trellis, but the trellis shown in Figure 8(b) is not, since the intersection points  $x_0$  and  $x_1$  can be removed. Note that if the end intersection condition is not satisfied, it is possible to find a supertrellis of  $([f]; T)$  for which every bigon contains a point of  $\bigcup_{n=-\infty}^{\infty} f^n(T^V)$ , but which has an end intersection whose braid type can be removed by isotopy. We will mostly be concerned with minimal extensions, since they have similar properties to minimal trellises compatible with biasymptotic mapping class.

**Theorem 3.9 (Existence and uniqueness of minimal extensions)** *Let  $([f]; T)$  be a trellis mapping class,  $X = \bigcup_{n \in \mathbb{Z}} f^n(T^V)$ ,  $\mathcal{B}$  the signature class of  $T$  in  $X$ , and  $\tilde{\mathcal{B}}$  a signature class in  $X$  with  $\tilde{\mathcal{B}} \geq \mathcal{B}$ . Then there is a unique trellis type  $[\tilde{f}; \tilde{T}]$  with signature  $\tilde{\mathcal{B}}$  such that  $(\tilde{f}; \tilde{T})$  is a minimal extension of  $([f]; T)$ . Further, if  $T$  is a transverse trellis, so is  $\tilde{T}$ .*

The proof is essentially the same as that of Theorem 3.5, and is omitted. Note that any inner bigon of  $\tilde{T}$  is an iterate of some inner bigon of  $T$ , so the intersection type is given by  $T$ .

Since minimal supertrellises have essentially the same forcing orbits as the original trellis, we expect the dynamics to be the same as that forced by the original trellis, or indeed, the dynamics forced by the biasymptotic orbits themselves. Minimal minimal supertrellises therefore provide a way of getting more information about the dynamics without increasing entropy.

**Theorem 3.10 (Nielsen entropy of minimal supertrellis)** *Let  $([f]; T)$  be a well-formed trellis mapping class. If  $(\tilde{f}; \tilde{T})$  is a minimal supertrellis of  $([f]; T)$ , then  $h_{niel}[\tilde{f}; \tilde{T}] = h_{niel}[f; T]$ .*

PROOF: Consider the case for which  $T^U = \widetilde{T}^U$ . Let  $(\alpha_i)$  be a family of relating curves for  $(\widetilde{f}; T)$ . Without loss of generality, we need only consider the case for which intersections of  $\alpha_i$  with  $T^S$  are at the endpoints. Further, we can homotope the  $\alpha_i$  so that  $\alpha_i$  has minimal intersections with  $\widetilde{T}^S$ .

Suppose there is an intersection of  $\alpha_i$  with  $\widetilde{T}^S$  that maps to a non-essential intersection of  $\widetilde{f} \circ \alpha_i$  with  $\widetilde{T}^S$ . Then there is a disc  $D$  bounded by an arc of  $\widetilde{f} \circ \alpha_i$  and a segment  $S$  of  $\widetilde{T}^S$  which is not in  $T^S$  such that  $D$  is disjoint from  $T^U$ . Further, since the intersection is a point of  $\widetilde{f}(\widetilde{T}^S)$  and the endpoints of  $\widetilde{T}^S$  are points of  $T^U$ , the segment  $S$  entirely lies in  $f(\widetilde{T}^S)$ , so  $\widetilde{f}^{-1}(S) \subset \widetilde{T}^S$ . Then  $\widetilde{f}^{-1}(D)$  is bounded by an arc of  $\alpha_i$  and an arc of  $\widetilde{T}^S$ . However,  $\alpha_i$  has minimal intersections with  $\widetilde{T}^S$ , so  $D$  contains a point of  $\widetilde{T}^U$ . Since the endpoints of  $T^U$  lie in  $T^S$ , the disc  $D$  must contain a bigon. However, any bigon bounded by  $T^U$  and an arc of  $\widetilde{T}^S$  must contain a point of  $T^S$ , a contradiction.

Hence any intersection of  $\alpha_i$  with  $\widetilde{T}^S$  maps to an essential intersection of  $\widetilde{f} \circ \alpha_i$  with  $\widetilde{T}^S$ , and hence is preserved under isotopy to  $\alpha_{i+1}$ . Therefore the number of intersections of  $\alpha_i$  increases as  $i$  increases, but since  $\alpha_n = \alpha_0$ , this number is constant, and  $\widetilde{f} \circ \alpha_i \sim \alpha_{i+1}$  taking  $\alpha_i$  as a relative maps on  $(I, \alpha_i^{-1}(T^S))$ .

The case for  $T^S = \widetilde{T}^S$  follows on reversing time and noticing that Nielsen classes for  $\widetilde{f}$  and  $\widetilde{f}^{-1}$  are identical except for finitely many classes related to  $T^P$ .  $\square$

### 3.4 Extensions of reducible trellises

We now examine the hypothesis of irreducibility in more detail. Recall that the main effect of reducibility is that the dynamics of a reducible trellis mapping class can be separated into simpler pieces. We now show that the converse is also true; the dynamics forced by an irreducible trellis essentially lies in one minimal component. To prove this, we need to carefully consider the intersections of branches of the trellis.

We consider a relationship based on *heteroclinic chains*.

**Definition 3.11 (Heteroclinic chains)** Let  $([f]; T)$  be a transverse trellis mapping class, and  $p_u, p_s \in T^P$ . We say there is a *heteroclinic chain* from  $p_u$  to  $p_s$  of length  $n$  if there are points  $p_u = p_0, p_1, \dots, p_n = p_s$  in  $T^P$  and an integer  $m$  such that for every  $\widehat{f}$  isotopic to  $f$  relative to  $T$  we have  $T^U(p_i) \cap \widehat{f}^{-m}(T^S)(p_{i+1}) \neq \emptyset$  for  $i = 0, 1, \dots, n-1$ . We write  $p_u \succ p_s$  if there is a heteroclinic chain from  $p_u$  to  $p_s$ . A subset  $C$  of  $T^P$  is *chain transitive* if  $p_1 \succ p_2$  for all  $p_1, p_2 \in C$ .

Clearly,  $\succ$  is a transitive relation, though it need not be reflexive or symmetric, and  $\succ$  induces a transitive relation on the chain transitive sets. If  $\succ$  is not reflexive, then there may be a chain-transitive set  $C$  such that  $W^U(C)$  does not intersect  $W^S(C)$ .

Note that although  $p_u \succ p_s$  implies  $W^U(\widehat{f}; p_u)$  intersects  $W^S(\widehat{f}; p_s)$  for any  $\widehat{f}$  isotopic to  $f$  relative to  $T$ , it is not necessarily true that there exists an  $n$  such that  $T^U(p_u)$  intersects  $\widehat{f}^{-n}(T^S(p_s))$  for any such  $\widehat{f}$ . In other words, the intersections may lie in arbitrarily high



iterates.

There is a strong relation between chain transitive components and reduction. If  $C$  is a reflexive chain transitive component, then  $C$  contains a homoclinic orbit, so induces some chaotic behaviour. A chain transitive component  $C$  which is not reflexive must consist of a single periodic point  $p$  with  $W^U(p) \cap W^S(p) = \emptyset$ , so there are no homoclinic orbits to  $p$ . If  $C_R$  and  $C_A$  are chain transitive components and  $C_R \succ C_A \not\succeq C_R$ , then we hope to find an attractor-repellor reduction with  $C_A$  in the attractor and  $C_R$  in the repellor. If  $C = f(C)$ , then the tangle of the orbits of  $C$  is connected, but if  $C = f^N(C)$  for some least  $N > 1$ , then the tangle of the orbits of  $C$  may have several components, which may form the basis for some invariant curve reduction.

**Theorem 3.12 (Intersections of irreducible tangles)** *Let  $([f]; T)$  be a well-formed irreducible transverse trellis mapping class, and let  $W$  be the infinite  $f$ -extension of  $T$ . Then if  $T^U[p_u, b_u]$  and  $T^S[p_s, b_s]$  are nontrivial branches of  $T$ , the branches  $W^U(p_u, b_u, \infty)$  and  $W^S(p_s, b_s, \infty)$  intersect transversely.*

PROOF: First, notice that if  $p_u = p_0, p_1, \dots, p_n = p_s$  is a heteroclinic chain, and  $W$  is the infinite  $f$ -extension of  $T$ , then by the Lambda lemma, since  $W^U(p_{i-1}, b_{i-1}, \infty)$  transversely intersects  $W^S(p_i)$  and  $W^U(p_i)$  transversely intersects  $W^S(p_{i+1}, b_{i+1}, \infty)$ , we see that  $W^U(p_{i-1}, b_{i-1}, \infty)$  transversely intersects  $W^S(p_{i+1}, b_{i+1}, \infty)$ . Induction on the length of the heteroclinic chain shows that  $W^U(p_u)$  transversely intersects  $W^S(p_s)$ .

Suppose  $p_u \succ p_s$ . Let  $\tilde{f}$  be a diffeomorphism such that  $\tilde{f}^n(T^U(p_u))$  transversely intersects  $T^S(p_s)$  and  $(\tilde{f}^{n+1}(T^U), T^S)$  is a minimal extension of  $T$ . Then  $\tilde{f}^{n+1}(T^U)(p_u)$  transversely intersects  $\tilde{f}(T^S(p_s))$ , so  $\tilde{f}^n(T^U(f(p_u)))$  transversely intersects  $T^S(f(p_s))$ , so  $f(p_u) \succ f(p_s)$ . Similarly, we can see that  $f^{-1}(p_u) \succ f^{-1}(p_s)$ , and hence  $f^n(p_u) \succ f^n(p_s)$  for any  $n \in \mathbb{Z}$ .

Now suppose  $C$  is a chain-transitive subset of  $T^P$ . Since the points of  $C$  are periodic and reflexive under  $\succ$ , there exists least  $N \in \mathbb{N}$  such that  $p \succ f^N(p)$  for some  $p \in C$ . Then  $f^{iN}(p) \succ f^{(i+1)N}(p)$ , for any  $i$ , and since  $p$  is periodic, we deduce  $p \succ f^{iN}(p) \succ p$  for any  $i$ . Note that the construction implies that  $N$  must divide the period of any point of  $C$ .

Now suppose  $q \in C$ , so  $p \succ q \succ p$ . Then for any  $i$ , we have  $p \succ f^{iN}(p) \succ f^{ni}(q) \succ f^{iN}(p) \succ p$ , so  $f^{iN}(q) \in C$ . Conversely, we can show that if  $f^{iN}(q) \in C$ , then  $q \in C$ . Therefore,  $C = f^N(C)$ .

The relationship  $\succ$  therefore defines a partial order on the chain transitive sets. Choose a set  $C$  which is minimal in this partial order, and let  $C_A = \bigcup_{n=1}^{\infty} f^n(C)$ .

Let  $([f_0]; T_0) = ([f]; T)$ , and for  $i = 1, \dots, \infty$ , let  $([f_i]; T_i)$  be a minimal extension of  $([f_{i-1}]; T_{i-1})$  with  $T_i = (f_i(T_{i-1}^U), T_{i-1}^S)$ . Let  $K_i$  be a set consisting of  $T_i^U(C_A)$ , and all regions of  $T_i$  which are topological discs or annuli and all of whose stable boundary segments lie in  $T^S(C_A)$ . Since  $T_i^U(C_A) \cap T_i^S(C_R) = \emptyset$ ,  $K_i \subset K_{i+1}$ . Let  $A_i$  be a closed neighbourhood of  $K_i$  which deformation retracts onto  $K_i$ , and such that  $A_i \subset A_{i+1}$ . Now,  $\chi(K_i \setminus T^U) \geq \chi(K_{i+1} \setminus T^U) \geq \chi(M)$ , so the Euler characteristics are a decreasing bounded sequence of integers, so have a limit, which occurs for some  $A_n$ . Then  $A_{n+1}$  deformation retracts onto  $K_n$ . Further,  $f(K_n) \subset K_{n+1} \subset A_{n+1}$ , so by taking  $A_n$  sufficiently small, we can ensure  $f(A_n) \subset A_{n+1}$ . Choose an isotopy  $h_t$  such that  $h_1(A_{n+1}) \subset A_n$  and  $h_t$  is fixed on  $K$ , and let  $\tilde{f}_t = h_t \circ f_n$ . Then  $\tilde{f}_1(A_n) = h_1(f_n(A_n)) \subset h_1(A_{n+1}) \subset A_n$ , so  $\tilde{f}_1 \in ([f]; T)$

and maps  $A_n$  into itself.

If  $C_A \neq C$ , then this gives an attractor-repellor reduction, a contradiction. If  $C_A = C$  and  $\chi(A_n) \neq \chi(M)$ , we can isotope  $f$  so that the boundary of  $A_N$  maps to itself, giving an invariant curve reduction, also a contradiction. Therefore  $C_A = C$  and there is no reduction. Since  $T^U(p_u, b_u)$  and  $T^S(p_s, b_s)$  are closed branches, there are periodic points  $\tilde{p}_u$  and  $\tilde{p}_s$  such that  $T^U(p_u, b_u)$  transversely intersects  $T^S(\tilde{p}_u)$  and  $T^U(\tilde{p}_s)$  transversely intersects  $T^S(p_s, b_s)$ . Then  $T^U(\tilde{p}_u)$  transversely intersects  $T^S(\tilde{p}_s)$ , and so  $T^U(p_u, b_u)$  transversely intersects  $T^S(p_s, b_s)$  as required.  $\square$

In other words, if  $([f]; T)$  is irreducible, then there exists  $n$  such that the  $n$ th image of every unstable branch under  $f$  intersects every stable branch. This is, essentially, the converse to the trivial observation that if  $([f]; T)$  has an attractor-repellor reduction, then there is a diffeomorphism  $\tilde{f}$  in  $([f]; T)$  and branches of  $W(\tilde{f}; T^P)$  which do not intersect.

## 4 Graph Representatives

Any connected compact surface with nonempty boundary is homotopy-equivalent to a one-dimensional space. Further, such spaces, and maps on them, are very easy to describe combinatorially. We can use these properties to provide a framework for representing surface diffeomorphisms and computing their dynamical properties. In classical Nielsen-Thurston theory, surface homeomorphisms are represented on one-dimensional spaces with a differentiable structure called *train tracks*. When representing trellises, we also need to take into account the topology of the stable and unstable curves, and hence obtain a controlled graph map.

There are many controlled graph maps compatible with a trellis type  $[f; T]$ . To use graph maps to describe the dynamics forced by  $[f; T]$ , we need to define a subclass of controlled graph maps which have minimal entropy in the exact homotopy class. These graph maps are called *efficient*, by analogy with Nielsen-Thurston theory. Of the efficient graph maps, we show in Section 4.3 that there is at most one which is *optimal*, giving a canonical *graph representative* for the trellis type.

Tight, efficient and optimal graph maps can be defined in terms of their actions on the turns of  $G$ . Some of following notions are used in the definitions, and all are useful in describing the algorithms.

**Definition 4.1 (Good, controlled, bad and inefficient turns)** Let  $g$  be a controlled graph map of  $(G, W)$ , and  $e_0 \triangleleft e_1$  a turn of  $G$ . If  $\partial g(e_0) \neq \partial g(e_1)$ , then  $e_0 \triangleleft e_1$  is a *good* turn. If  $\partial g(e_0) = \partial g(e_1)$  and at least one of  $e_0$  and  $e_1$  is a control edge, then  $e_0 \triangleleft e_1$  is a *controlled* turn, otherwise  $e_0 \triangleleft e_1$  is a *bad turn*. A *bad* turn  $e_0 \triangleleft e_1$  is *inefficient* if there is an edge  $e$  such that  $g^n(e) = \dots \bar{e}_0 e_1 \dots$

**Definition 4.2 (Efficient and optimal graph map)** Let  $g$  be a graph map of a controlled graph  $G$ . The controlled graph map  $g$  is *efficient* if there are no inefficient turns, and *optimal* if there are no bad turns, and every invariant forest contains a control edge.

We can now define what it means for a graph map to be a representative of  $[f; T]$ . As well as the (complete) graph representative, we also define the restricted graph representative which contains all the interesting dynamical information. and the topological graph representative is an invariant of the forcing orbits.

**Definition 4.3 (Graph representatives)** A controlled graph map  $(g; G, W)$  is a *graph representative* of an transverse trellis type  $[f; T]$  if  $g$  is an optimal graph map which is compatible with  $[f; T]$ . The *restricted graph representative* of  $[f; T]$  is the controlled graph map  $(g; \bar{G}, \bar{W})$ , where  $\bar{G} = \bigcap_{n=0}^{\infty} g^n(G)$ , and  $\bar{W} = \bar{G} \cap W$ . The *topological graph representative* of  $[f; T]$  is the topological conjugacy class of graph map obtained by collapsing all control edges of the restricted graph representative to points.

The following theorem asserts that every proper irreducible trellis type has a unique graph representative.

**Theorem 4.4 (Existence and uniqueness of graph representatives)** *Let  $[f; T]$  be a proper trellis type with no invariant curve reduction. Then  $[f; T]$  has a unique graph representative  $(g; G, W)$ . Further, if  $[f_0; T_0]$  and  $[f_1; T_1]$  are trellis types, the graph representatives  $(g_0; G_0, W_0)$  and  $(g_1; G_1, W_1)$  are homeomorphic if and only if  $[f_0; T_0] = [f_1; T_1]$ .*

We prove the existence of a graph representative in Section 4.2 by giving an algorithm to compute it, and uniqueness in Section 4.3. Since the maps  $\mathcal{C}f_0$  and  $\mathcal{C}f_1$  for different trellis types are not exact homotopy equivalent, it is trivial that different trellis types have different graph representatives. This means that the graph representative provides a convenient way of specifying a trellis type.

**Example 4.5 (Graph representatives)**

The graph representatives for the trellis type shown in figure 3 are shown in figure 9.

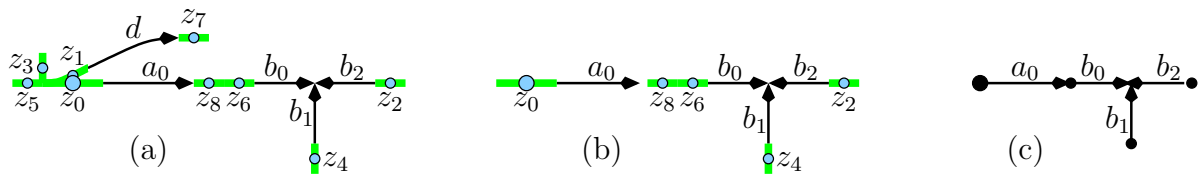


Figure 9: Graph representatives for the trellis type shown in figure 3. (a) is the graph representative, (b) the restricted graph representative and (c) the topological graph representative.

The control edges map

$$z_0, z_1, z_2 \mapsto z_0; \quad z_3 \mapsto z_1; \quad z_4 \mapsto z_2; \quad z_5 \mapsto z_3; \quad z_6, z_7, z_8 \mapsto z_4;$$

and the expanding edges map

$$a_0 \mapsto a_0 \bar{z}_8 z_6 b_0 \bar{b}_1; \quad b_0 \mapsto b_1; \quad b_1 \mapsto b_2; \quad b_2 \mapsto a_0 \bar{z}_8 z_6 b_0; \quad d \mapsto a_0 \bar{z}_8 z_6 b_0 \bar{b}_1.$$

The edges  $z_1, z_3, z_5, z_7$  and  $d$  do not lie in  $g^2(G)$  so are not part of the restricted graph representative as shown in Figure 3(b). Collapsing the control edges gives the topological graph representative as shown in Figure 3(c), with

$$a_0 \mapsto a_0 b_0 \bar{b}_1; \quad b_0 \mapsto b_1; \quad b_1 \mapsto b_2; \quad b_2 \mapsto a_0 b_0.$$

The transition matrix for this graph map is

$$\left( \begin{array}{c|cccc} & a & b_0 & b_1 & b_2 \\ \hline a & 1 & 1 & 1 & 0 \\ b_0 & 0 & 0 & 1 & 0 \\ b_1 & 0 & 0 & 0 & 1 \\ b_2 & 1 & 1 & 0 & 0 \end{array} \right)$$

and the growth rate is given by the largest root of the characteristic polynomial  $\lambda(\lambda^3 - \lambda^2 - 2) = 0$ . Numerically, the growth rate is 1.696, giving topological entropy  $h_{top}(g) \approx 0.528$ .

## 4.1 Graph moves and zeta functions

The moves we need for the algorithm are based on those of [BH95]. We use four basic moves, each of which is an exact homotopy equivalence and so preserves the Euler characteristic and number of control edges are unchanged. We define the moves at the graph level, but all moves can be performed on thick graphs.

**Edge tightening** Let  $e$  be an edge of  $G$  with  $g(e) = e_1 \dots e_i e_{i+1} \dots e_n$  and  $e_{i+1} = \bar{e}_i$ . Homotope  $g$  to  $g'$  with  $g'(e) = e_1 \dots e_{i-1} e_{i+2} e_n$ .

**Vertex homotopy** Let  $v$  be a vertex of  $G$  with incident edges  $e_1, \dots, e_k$  and let  $e_0$  be an edge with  $\iota(e_0) = g(v)$ . Homotope  $v$  across  $e_0$  to obtain a new map  $g'$  such that  $g'(e) = \bar{e}_0 g(e_i)$  unless  $g(e_i) = e_0 \epsilon$  in which case  $g'(e_i) = \epsilon$ .

**Edge collapsing** If  $e$  is an edge with  $g(e) = \cdot$ , collapse  $e$  to a point, and remove all occurrences of  $e$  in the image of  $g$ .

**Vertex splitting** Suppose  $v$  is a vertex of  $G$ , and edges  $a_1, \dots, a_j$  are adjacent in the cyclic ordering at  $v$ . Let  $G'$  be a graph formed by introducing a new vertex  $w$  and new edge  $e_0$  to  $G$  from  $v$  to  $w$  so that the edges at  $w$  are  $e_0$  and  $a_1, \dots, a_j$  as shown in figure 10. Let  $g'$  be the graph map obtained by taking  $g'(e_0) = \cdot$ , and by replacing every occurrence of an edge  $a_i$  for  $1 \leq i \leq j$  in an image  $g(e)$  with the edge-path  $e_0 a_i$ .

The moves of edge tightening and vertex homotopy do not affect the graph  $(G, W)$ , but change the map  $g$  to a new map  $g'$ . The moves of edge collapsing and vertex splitting are inverse operations and change a graph  $G$  to a graph  $G'$ .

The following moves can all be expressed in terms of the basic moves described above.

**Vertex tightening** A vertex homotopy across  $e_0$  is a tightening if  $\partial g(e_i) = e_0$  for all  $i = 1, \dots, n$ .

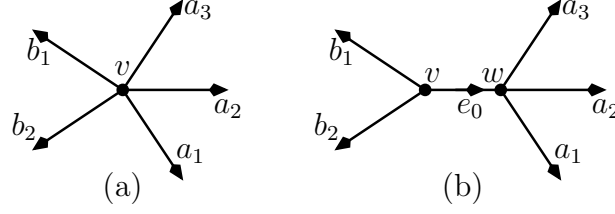


Figure 10: (a) Before and (b) after a vertex splitting.

**Collapsing an invariant forest** If  $H \subset G$  is an invariant forest of expanding edges, collapse every edge of  $H$  to a point.

**Tidying** Perform edge tightenings, vertex tightenings, and collapsing invariant forests until no further tightening or collapsing is possible, with the exception that we never collapse peripheral edges at this stage (even if they have a trivial image).

**Valence-3 homotopy** Suppose  $v$  is a valence-3 vertex with incident edges  $e_1, e_2$  and  $e_3$  such that  $g(e_i) = e_0 \epsilon_i$  for  $i = 1, 2$  and  $g(e_3) = \epsilon_3$ . Homotope  $v$  across  $e_0$  and to give  $g'(e_i) = \epsilon_i$  for  $i = 1, 2$  and  $g'(e_3) = \bar{e}_0 \epsilon_3$ .

**Folding** Suppose there are edges  $e_1 \triangleleft e_2$  at a vertex  $v$  with valence at least 4 such that  $g(e_1) = e_0 \epsilon_1$  and  $g(e_2) = e_0 \epsilon_2$ . First split off  $e_1$  and  $e_2$  to a new vertex  $w$  with new edge  $e_0$  from  $v$  to  $w$ . Then perform any possible edge tightenings to give a graph map  $g'$  with  $g'(e_0) = \cdot$ ,  $g'(e_1) = e'_0 \epsilon'_1$  and  $g'(e_2) = e'_0 \epsilon'_2$ . Finally, homotope at  $w$  across the edge-path  $e'_0$  to obtain a graph map  $g''$  with  $g''(e_0) = e_0$ ,  $g''(e_1) = \epsilon_1$  and  $g''(e_2) = \epsilon_2$ .

We say a graph map is *tight* if no edge or vertex tightenings are possible. A tight graph map is *tidy* if it has no invariant forest of expanding edges, and every valence-2 vertex is an endpoint of a control edge. Performing the tidying operation leave a nice graph map to use as a starting point for the more complicated moves of vertex homotopy and edge splitting.

We measure the complexity of a graph map by its *zeta functions*.

**Definition 4.6 (Zeta function)** Let  $g$  be a graph map of a graph  $G$  and  $H$  an invariant subgraph. The  $H$ -length of an edge path  $\epsilon$ , denoted  $|H|_\epsilon$  is the number of occurrences of an edge of  $H$  in  $\epsilon$ . The  $H$ -norm and the *control zeta function* of  $g$  are given respectively by

$$|g|_H = \sum_{e \in E(G)} |H|_{g(e)} \quad \text{and} \quad \zeta_{g;H}(t) = \sum_{n=0}^{\infty} |g^n|_H t^n.$$

It is straightforward to show that if every invariant subgraph of  $G$  contains an edge in  $H$ , then the logarithm of the asymptotic growth rate of  $|g^n|_H$  equals  $h_{top}(g)$ . If  $R$  is the radius of convergence of  $\zeta_{g;H}$ , then  $h_{top}(g) = \log(1/R)$ .

Zeta functions are ordered by the ordering  $\sum_{n=0}^{\infty} a_n t^n < \sum_{n=0}^{\infty} b_n t^n$  if for some  $k$  we have  $a_k < b_k$ , but  $a_i = b_i$  for all  $i < k$ . If  $k \leq N$ , we say  $\sum_{n=0}^{\infty} a_n t^n <_N \sum_{n=0}^{\infty} b_n t^n$ .

Equivalently,  $\zeta_1 < \zeta_2$  if there exists  $\epsilon > 0$  such that  $\zeta_1(t) < \zeta_2(t)$  for all  $t \in (0, \epsilon)$ , and  $\zeta_1 <_N \zeta_2$  if  $\zeta_1 + at^n < \zeta_2$  for all  $a \geq 0$  and  $n > N$ .

It is clear that edge tightening, vertex tightening, edge collapsing and vertex splitting do not increase the any graph norms, so the graph norms can only increase during a vertex homotopy. The following key lemma shows that valence-3 homotopies also reduce the graph norms.

**Lemma 4.7** *Let  $g$  be a graph map with a valence-3 vertex  $v$  which has incident edges  $e_1, e_2$  and  $e_3$  such that  $g(e_i) = e_0\epsilon_i$  for  $i = 1, 2$  and  $g(e_3) = \epsilon_3$ , and that  $g'$  is obtained from  $g$  by a homotopy at  $v$  across  $e_0$ . Suppose further that  $g^{n-2}(e_0)$  does not contain any edges of  $H$ , but  $g^{n-1}(e_0)$  does. Then*

1. For any  $k < n$  and any edge  $e$  of  $G$ ,  $|H|_{(g')^k(e)} = |H|_{g^k(e)}$ , and  $|(g')^k|_H = |g^k|_H$ .
2.  $|(g')^n|_H < |g^n|_H$

PROOF: Since  $g^k(e_0)$  does not contain any control edges for  $k < n - 1$ , so  $|g^k(e_0)|_H = 0$  for these values of  $k$ . For part 1, we use induction on  $k$ . The result is trivial for  $k = 0$ . Assume the result is true for some  $k < n$ . Then

$$\begin{aligned} |(g')^{k+1}(e_i)|_H &= |(g')^k(\epsilon_i)|_H = |g^k(\epsilon_i)|_H \text{ (by the inductive hypothesis)} \\ &= |g^k(e_0)g^k(\epsilon_i)|_H - |g^k(e_0)|_H = |g^k(e_0\epsilon_i)|_H - |g^k(e_0)|_H \\ &= |g^{k+1}(e_i)|_H - |g^k(e_0)|_H \text{ for } i = 1, 2 \\ |(g')^{k+1}(e_3)|_H &= |(g')^k(e_0\epsilon_3)|_H = |g^k(e_0\epsilon_3)|_H = |g^k(e_0)|_H + |g^k(\epsilon_3)|_H \\ &= |g^k(e_3)|_H + |g^k(e_0)|_H \\ |(g')^{k+1}(e_i)|_H &= |(g')^k(\epsilon_i)|_H = |g^k(\epsilon_i)|_H = |g^{k+1}(e_i)|_H \text{ otherwise.} \end{aligned}$$

Clearly then,  $|(g')^k|_H = \sum_{e \in G} |(g')^k(e)|_H = \sum_{e \in G} |g^k(e)|_H = |g^k|_H$  for  $k < n$ . Part 2 follows from the above calculations since  $|(g')^n|_H = |g^n|_H - |g^{n-1}(e_0)|_H < |g^n|_H$ .  $\square$

## 4.2 Computing a graph representative

We now give an algorithm for obtaining an optimal controlled graph map compatible with a given trellis mapping class. If the trellis mapping class has an invariant curve reduction, the algorithm instead finds this reduction, and we then compute a graph representative for the irreducible components. We can process the reduction by considering the invariant subgraph separately; see [Kei97] for details. Although the moves of the algorithm are based on the algorithm of Bestvina and Handel [BH95], but the method is quite different. Instead of reducing the topological entropy at each step, the algorithm reduces the zeta function, which gives a more precise measure of the growth based on the control and peripheral edges. This results in an algorithm where we need only consider the local action of the graph map.

In the case of a surface mapping class, there are no control edges, and no optimal graph map. We use a slightly different algorithm to find an efficient graph map, and consider

a zeta function based on the peripheral edges to show that the algorithm terminates. If there are no peripheral edges, we create temporary peripheral edges by puncturing at a periodic orbit.

**Algorithm 4.8** Suppose  $g$  is a controlled graph map. Perform the following moves, followed by tidying (noting that tidying does not affect the peripheral subgraph), until the graph map is optimal or there is an invariant subgraph of expanding edges which is not a forest.

1. Fold a bad turn at a vertex of valence greater than 3, or at a control vertex.
2. Homotope at a bad turn at a free vertex of valence 3.
3. If no other moves are possible, collapse any peripheral edges which have trivial image.

The proof of termination considers zeta functions of the graph map. We need to consider both the control zeta function  $\zeta_{g;Z}$  and the peripheral zeta function  $\zeta_{g;P}$ , and also the number of peripheral edges  $|P|$ . The standard procedure for dealing with the peripheral subgraph  $P$  is *absorbing*, which yields a graph map for which there is no invariant set which deformation-retracts onto  $P$  other than  $P$  itself. However, absorbing into the peripheral subgraph can be accomplished by the moves of Algorithm 4.8. A fundamental observation is that  $G$  has at most  $|G| = 2\#(Z) - 3\chi$  edges, where  $\chi$  is the Euler characteristic, and hence any decrease in a zeta function must occur in at most  $|G|$  coefficients.

**Theorem 4.9 (Computation of the graph representative)** *Algorithm 4.8 terminates at a reduction or an optimal graph map.*

PROOF: First note that, although it may be possible to successively decrease a zeta function arbitrarily many times, if there exists  $N$  such that all such moves decrease the  $n$ th coefficient for some  $n \leq N$ , then the process must terminate. In our case, we can take  $n = 2\#(Z) - 3\xi$ , the bound on the number of edges of  $G$ .

If there is an invariant subgraph which does not contain any control or peripheral edges and is not a forest,  $g$  is irreducible and the algorithm has found a reduction. Otherwise, every move apart from collapsing in the peripheral subgraph either reduces the control zeta function, increases the number of peripheral edges, or decreases the peripheral zeta function.

After collapsing peripheral edges in step 3, the link of the peripheral subgraph must be invariant under  $\partial g$ . If there is no reduction, every expanding edge in this subgraph must eventually map to a control edge, or else it would be contained in an invariant subgraph of free edges with negative Euler characteristic. We may perform homotopies across pre-peripheral edges disjoint from  $P$  which do not affect the control zeta function, but cannot introduce new peripheral edges without homotoping across an edge in  $\text{Lk}(P)$ . At this homotopy, the control zeta function decreases, so there is no infinite sequence of moves which does not decrease the control zeta function. Additionally, there are at most finitely

many moves which decrease the control zeta function. Hence the algorithm either finds a reduction, or terminates at an optimal graph map.  $\square$

The case of no control edges cannot be treated directly by Algorithm 4.8, since there is no optimal graph map in the homotopy class. The proof of Theorem 4.9 fails since after collapsing peripheral edges, all edges in  $\text{Lk}(P)$  map to cover the entire graph, which is itself an invariant subgraph of negative Euler characteristic which does not include any control edges. Instead we consider the entropy in addition to the peripheral zeta function in order to show that the algorithm finds an efficient graph representative.

**Algorithm 4.10** Suppose  $g$  is a graph map compatible with a surface mapping class, and  $g$  has nonempty peripheral subgraph  $P$ .

1. Perform Algorithm 4.8, stopping after collapsing peripheral edges in step 3.
2. If the graph map is efficient or has a reduction, the algorithm terminates. Otherwise, repeat step 1.
3. Perform Algorithm 4.8, checking after step 3 whether the graph map is efficient, at which point the algorithm terminates.

If there are no peripheral edges, we need to artificially introduce peripheral edges by puncturing at a periodic orbit.

**Algorithm 4.11** Suppose  $g$  is a map of a graph  $G$  with no peripheral curves. Perform the following moves until an efficient graph map or a reduction is found.

1. Let  $P$  be a periodic orbit of  $g$ . Puncture at points of  $P$  by introducing artificial peripheral loops at  $P$ .
2. Perform the Algorithm 4.10 for graph maps with peripheral loops.
3. Collapse the artificial peripheral loops at  $P$  and tidy. If the resulting graph map is efficient or has a reduction, the algorithm terminates. If the resulting graph map has peripheral loops, perform Algorithm 4.10. Otherwise, return to step 1

**Theorem 4.12** *Algorithms 4.10 and 4.11 terminate at a reduction or an efficient graph map.*

**PROOF:** We first consider Algorithm 4.10. If  $g$  is not efficient, there is a turn in the link of  $P$  which is inefficient. Folding this turn decreases entropy. Although other moves may increase entropy, the graph map obtained by folding and homotoping can also be obtained by the moves of the Bestvina-Handel algorithm, so the entropy must strictly decrease. Since the entropy is bounded above, and the number of edges of  $g$  is bounded by  $-3\chi$ , there are only finitely many possible values for the entropy. Hence Algorithm 4.10 terminates.

Algorithm 4.11 terminates for similar reasons, since the entropy is bounded, and must strictly decrease.  $\square$



We now give an example of computing an optimal graph map using Algorithm 4.8.

**Example 4.13 (Computing the graph representative)**

Consider the trellis type  $[f; T]$  shown in figure 11.

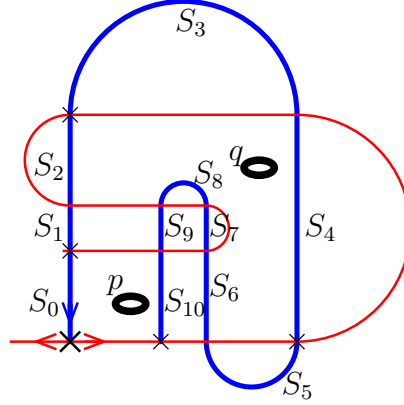


Figure 11: A trellis with a period-2 orbit.

There are eleven stable segments, and the surface has two punctures,  $p$  and  $q$ , which are permuted by the diffeomorphism  $f$ .

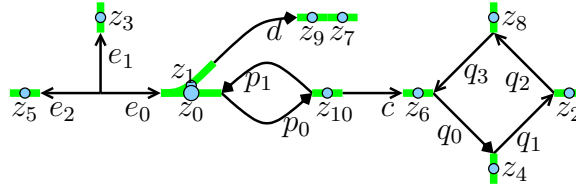


Figure 12: An initial compatible graph

An initial compatible graph is shown in figure 12. The edge-loop  $p_0p_1$  surrounds the puncture  $p$ , and the edge-loop  $q_0q_1q_2q_3$  surrounds the puncture  $q$ . The control edges map under  $g$  as follows:

$$z_0, z_1, z_2 \mapsto z_0; \quad z_3 \mapsto z_1; \quad z_4 \mapsto z_2; \quad z_5 \mapsto z_3; \quad z_6, z_7, z_8, z_9, z_{10} \mapsto z_4$$

We compute  $g$  on the expanding edges to obtain:

$$\begin{aligned} p_0 &\mapsto p_0 \bar{z}_{10} c z_6 q_0; & p_1 &\mapsto q_1 q_2 q_3 \bar{z}_6 \bar{c} z_{10} \bar{p}_0; \\ q_0 &\mapsto q_1; & q_1 &\mapsto q_2 q_3 \bar{z}_6 \bar{c} z_{10} p_1; & q_2 &\mapsto p_0 \bar{z}_{10} c z_6 \bar{q}_3 \bar{q}_2 \bar{q}_1; & q_3 &\mapsto \cdot; \\ c &\mapsto \cdot; & d &\mapsto p_0 \bar{z}_{10} c z_6 \bar{q}_3 \bar{q}_2 \bar{q}_1; & e_0 &\mapsto e_0; & e_1 &\mapsto e_0; & e_2 &\mapsto e_1. \end{aligned}$$

Note that the edges  $p_i$  and  $q_i$  are not peripheral edges, since they do not form an invariant set. However the edge-loop  $p_0p_1$  maps to the edge loop  $p_0 \bar{z}_{10} c z_6 q_0 q_1 q_2 q_3 \bar{z}_6 \bar{c} z_{10} \bar{p}_0$  which is freely homotopic to  $q_0 q_1 q_2 q_3$ .

The edges  $q_3, c, e_1, e_2$  and  $e_3$  form an invariant forest which does not contain any control edges. Each of these edges can be collapsed to a point to give the graph in figure 13. The

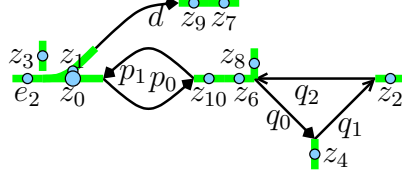


Figure 13: Graph obtained by collapsing the invariant forest.

expanding edges now map

$$\begin{aligned} p_0 &\mapsto p_0 \bar{z}_{10} z_6 q_0; & p_1 &\mapsto q_1 q_2 \bar{z}_6 z_{10} \bar{p}_0; \\ q_0 &\mapsto q_1; & q_1 &\mapsto q_2 \bar{z}_6 z_{10} p_1; & q_2, d &\mapsto p_0 \bar{z}_{10} z_6 \bar{q}_2 \bar{q}_1. \end{aligned}$$

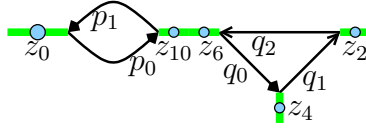


Figure 14: Restricted graph map.

The edge  $d$  has no preimage, so is not part of the restricted graph representative. We can remove  $d$ , and also  $z_1, z_3, z_5$  and  $z_8$  to obtain the graph map shown in figure 14. The control zeta function is  $\zeta_{g;Z}(t) = 5 + 13t + 25t^2 + \dots$ .

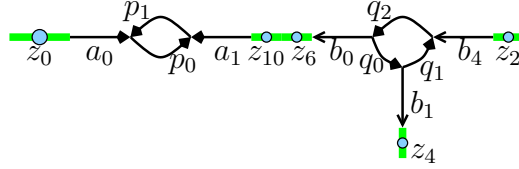


Figure 15: Graph obtained after folding

Performing five folding operations introduced edges  $a_0, a_1, b_0, b_1$  and  $b_4$ , as shown in figure 15. Before tightening, the graph map is

$$\begin{aligned} p_0 &\mapsto a_0 p_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0 q_0 b_1; & p_1 &\mapsto \bar{b}_1 q_1 \bar{b}_4 b_4 q_2 b_0 \bar{z}_6 z_{10} a_1 \bar{p}_0 \bar{a}_0; \\ q_0 &\mapsto \bar{b}_1 q_1 \bar{b}_4; & q_1 &\mapsto b_4 q_2 b_0 \bar{z}_6 z_{10} a_1 p_1 \bar{a}_0; & q_2 &\mapsto a_0 p_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0 \bar{q}_2 \bar{b}_4 \bar{q}_1 b_1. \end{aligned}$$

On performing tightenings and valence-3 homotopies, we obtain the graph map

$$\begin{aligned} p_0 &\mapsto q_0; & p_1 &\mapsto q_1 q_2; & q_0 &\mapsto \cdot; & q_1 &\mapsto q_2 b_0 \bar{z}_6 z_{10} a_1 p_1; & q_2 &\mapsto p_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0 \bar{q}_2; \\ a_0 &\mapsto a_0 p_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0; & a_1 &\mapsto \bar{b}_1; & b_0 &\mapsto \bar{q}_1 b_1; & b_1 &\mapsto \bar{b}_4; & b_4 &\mapsto a_0. \end{aligned}$$

Collapsing  $q_0$  and then  $p_0$  gives the graph map shown in figure 16, with

$$\begin{aligned} p_1 &\mapsto q_1 q_2; & q_1 &\mapsto q_2 b_0 \bar{z}_6 z_{10} a_1 p_1; & q_2 &\mapsto \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0 \bar{q}_2; \\ a_0 &\mapsto a_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0; & a_1 &\mapsto \bar{b}_1; & b_0 &\mapsto \bar{q}_1 b_1; & b_1 &\mapsto \bar{b}_4; & b_4 &\mapsto a_0. \end{aligned}$$

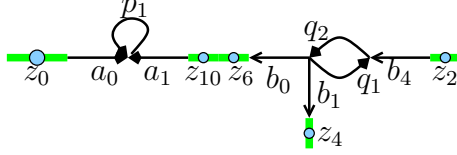


Figure 16: Graph obtained after collapsing  $q_0$  and  $p_0$ .

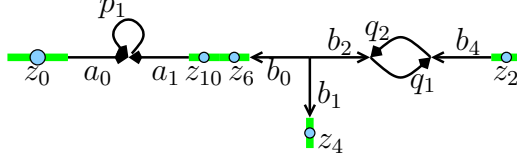


Figure 17: Graph obtained after folding  $q_1$  and  $q_2$ .

and control zeta function  $\zeta_Z(t) = 5 + 11t + 19t^2 + \dots$ .

Folding  $q_1$  and  $\bar{q}_2$  to form a new edge  $b_2$  gives the graph map shown in figure 17, with

$$\begin{aligned} p_1 &\mapsto b_2 q_1 q_2 \bar{b}_2; & q_1 &\mapsto p_1; & q_2 &\mapsto \cdot; \\ a_0 &\mapsto a_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0; & a_1 &\mapsto \bar{b}_1; & b_0 &\mapsto \bar{q}_1 \bar{b}_2 b_1; & b_1 &\mapsto \bar{b}_4; & b_2 &\mapsto q_2 \bar{b}_2 b_0 \bar{z}_6 z_{10} a_1; & b_4 &\mapsto a_0. \end{aligned}$$

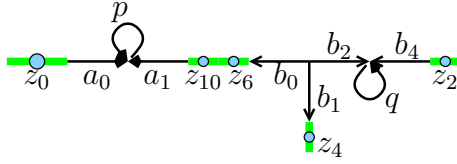


Figure 18: Graph obtained after collapsing  $q_2$ .

Since now  $q_2$  has trivial image, we can collapse  $q_2$  to obtain the graph map shown in figure 18, with

$$\begin{aligned} p_1 &\mapsto b_2 q_1 \bar{b}_2; & q_1 &\mapsto p_1; \\ a_0 &\mapsto a_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0; & a_1 &\mapsto \bar{b}_1; & b_0 &\mapsto \bar{q}_1 \bar{b}_2 b_1; & b_1 &\mapsto \bar{b}_4; & b_2 &\mapsto \bar{b}_2 b_0 \bar{z}_6 z_{10} a_1; & b_4 &\mapsto a_0. \end{aligned}$$

and control zeta function  $\zeta_Z(t) = 5 + 9t + 17t^2 + \dots$ .

Finally, we fold up  $p_1$  and  $q_1$  to give new edges  $a_2$  and  $b_3$  as shown in figure 19 The graph map is now

$$\begin{aligned} p_1 &\mapsto q_1; & q_1 &\mapsto p_1; \\ a_0 &\mapsto a_0 \bar{a}_1 \bar{z}_{10} z_6 \bar{b}_0; & a_1 &\mapsto \bar{b}_1; & a_2 &\mapsto b_3 \bar{b}_2 \\ b_0 &\mapsto \bar{b}_3 \bar{q}_1 b_3 \bar{b}_2 b_1; & b_1 &\mapsto \bar{b}_4; & b_2 &\mapsto \bar{b}_2 b_0 \bar{z}_6 z_{10} a_1; & b_3 &\mapsto a_2; & b_4 &\mapsto a_0. \end{aligned}$$

which is optimal. The control zeta function is  $\zeta_{g;Z}(t) = 5 + 9t + 15t^2 + \dots$ . The edges  $p_1$  and  $q_1$  are peripheral edges for the graph map.

Collapsing the control edges we obtain the topological graph representative as shown in figure 20.

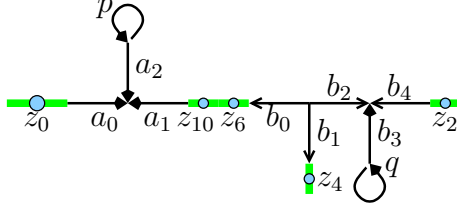


Figure 19: The optimal graph representative.

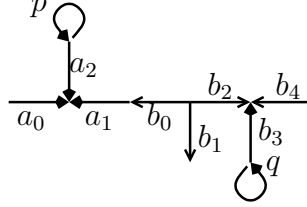


Figure 20: The optimal graph representative.

### 4.3 Uniqueness of graph representatives

It is clear that given an efficient graph map compatible with a trellis map, the trellis mapping type can be reconstructed from the graph map. The converse is also true; up to isomorphism there is only one efficient controlled graph map compatible with a given trellis mapping class.

**PROOF OF THEOREM 4.4 (UNIQUENESS):** We first give a verifiable criterion for an exact homotopy equivalence between two graphs to be homotopic to a homeomorphism. Suppose  $(G_1, W_1)$  and  $(G_2, W_2)$  are controlled graphs, and  $p$  is an exact homotopy equivalence between  $G_1$  and  $G_2$ . Suppose there are strictly positive real functions  $l_1$  and  $l_2$  on the edges of  $G_1$  and  $G_2$  such that whenever  $\alpha_1$  and  $\alpha_2$  are tight paths joining points of  $W$  with  $p \circ \alpha_1 \sim \alpha_2$  we have  $l_1(\alpha_1) = l_2(\alpha_2)$ . (We take the distance between a point of  $W$  and a control vertex to be half the length of the control edge.)

We claim  $p$  is homotopic to a homeomorphism from  $(G_1, W_1)$  to  $(G_2, W_2)$ . To show this, is sufficient to show that the universal covers  $(\tilde{G}_1, \tilde{W}_1)$  and  $(\tilde{G}_2, \tilde{W}_2)$  are homeomorphic. Since  $W_1$  and  $W_2$  are homeomorphic, and  $G_1$  and  $G_2$  are homotopy-equivalent, the sets  $\tilde{W}_1$  and  $\tilde{W}_2$  are homeomorphic, and we call them both  $\tilde{W}$ . The functions  $l_1$  and  $l_2$  lift naturally to  $\tilde{G}_1$  and  $\tilde{G}_2$ . For any four points  $\tilde{w}_1, \tilde{w}_2, \tilde{w}_3$  and  $\tilde{w}_4$  in  $\tilde{W}$ , there are only two possibilities for the span; either it contains a valence-4 vertex or two valence-3 vertices. Let  $\alpha_{ij}$  be the curve from  $w_i$  to  $w_j$ .

Suppose the span of  $\{w_1, w_2, w_3, w_4\}$  contains a valence-4 vertex  $v$  as shown in Figure 21(a) Then for  $i = 1, 2$ ,

$$l_i(\alpha_{12}) + l_i(\alpha_{34}) = l_i(\alpha_{13}) + l_i(\alpha_{24}) = l_i(\alpha_{14}) + l_i(\alpha_{23}) = l_i(\gamma_1) + l_i(\gamma_2) + l_i(\gamma_3) + l_i(\gamma_4)$$

However, if the span of  $G\{w_1, w_2, w_3, w_4\}$  contains a valence-3 vertices  $v_1$  and  $v_2$  as shown in Figure 21(b), we have

$$l_i(\alpha_{12}) + l_i(\alpha_{34}) = l_i(\gamma_1) + l_i(\gamma_2) + l_i(\gamma_3) + l_i(\gamma_4)$$

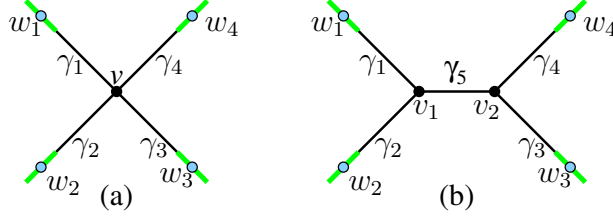


Figure 21: The span of four points of  $W$  in a simply connected region can either have a valence-4 vertex (a) or two valence-3 vertices (b).

but

$$l_i(\alpha_{13}) + l_i(\alpha_{24}) = l_i(\alpha_{14}) + l_i(\alpha_{23}) = l_i(\gamma_1) + l_i(\gamma_2) + l_i(\gamma_3) + l_i(\gamma_4) + 2l_i(\gamma_5),$$

so

$$l_i(\alpha_{12}) + l_i(\alpha_{34}) \leq l_i(\alpha_{13}) + l_i(\alpha_{24}) = l_i(\alpha_{14}) + l_i(\alpha_{23})$$

Hence the universal covers are homeomorphic by a homeomorphism  $\tilde{h}$  homotopic to  $\tilde{p}$ . Further, since the computations given above are equivariant, the homeomorphism  $\tilde{h}$  is equivariant, so projects to a homeomorphism  $h : (G_1, W_1) \rightarrow (G_2, W_2)$ . This completes the proof of the claim.

Let  $p : (g_1; G_1) \rightarrow (g_2; G_2)$  be an exact homotopy equivalence with homotopy inverse  $q$ . Thus  $q \circ p \sim id$ ,  $p \circ q \sim id$  and  $g_1 \sim q \circ g_2 \circ p$ , taking homotopies relative to  $W$ . It is clear that we can choose  $p$  and  $q$  to be mutual inverses on the control and peripheral edges, so merely need to consider the expanding edges. We therefore make no distinctions between control edges in  $G_1$  and  $G_2$ . For any curve  $\gamma : (I, J) \rightarrow (G, W)$  for some finite set  $J$ , we can find a curve homotopic to  $g \circ \gamma$  relative to  $J$  which minimises the number of intersections with  $W$ . It is easy to see that the exact homotopy class of such a curve is unique up to reparameterisation, giving a well-defined map  $g_{\min}$  on homotopy classes  $[\gamma]$  up to reparameterisation.

Let  $\gamma_1$  be a tight edge-path in  $G_1$ , and let  $\gamma_2$  be the tight edge-path in  $G_2$  homotopic to  $p(\gamma_1)$ . Consider  $\gamma_1$  and  $\gamma_2$  as exact curves  $(I, J_0) \rightarrow (G, W)$ , and take minimal iterates under  $g_1$  and  $g_2$ . Since  $G_1$  and  $G_2$  are graphs, these minimal iterates can each be represented by a single tight edge-path. Then  $\gamma_1 \sim q \circ p(\gamma_1) \sim q(\gamma_2)$ , so  $p((g_1)_{\min}[\gamma_1])$  is the same homotopy class as  $(g_2)_{\min}[\gamma_2]$ . Taking further iterates gives  $p((g_1)_{\min}^n[\gamma_1]) \sim (g_2)_{\min}^n[\gamma_2]$  for all  $n$ . Thus  $(g_1)_{\min}^n[\gamma_1]$  and  $(g_2)_{\min}^n[\gamma_2]$  contain the same number of control and peripheral edges for all  $n$ . Since this number must be positive for all edges for some  $n$ , we obtain a suitable function  $l$  on the expanding edges of  $G$ , and so the graphs  $G_1$  and  $G_2$  are homeomorphic by a graph map  $h$  homotopic to  $p$ . This homeomorphism gives the required topological conjugacy between  $g_1$  and  $g_2$ .  $\square$

## 5 Global Shadowing

In this section, we discuss ways of representing the dynamics of a chaotic map using the relation of *global shadowing*, which is an equivalence relation on orbits of a system which gives a condition under which two orbits can be considered to be close to each other over infinite time intervals. The classical definition was introduced by Katok, and is given in terms of lifts of the map to the universal cover.

As the concept is topological in spirit, it would be useful to have a purely topological classification. We will also be concerned with spaces which are topological pairs, and seek a definition which extends naturally to this setting. We would also like to talk about arbitrary sets of orbits as well as just pairs of orbits. As a frequently-recurring theme in this work is that of homotopy classes of curves, a definition in terms of this formalism, rather than that of universal covers will also be useful. The basic idea of our global shadowing relation is given below. It relies on the definition of sets of curves of *bounded lengths*.

**Definition 5.1 (Global shadowing)** Orbits  $(x_i)$  and  $(y_i)$  of a map  $f : (X, Y) \rightarrow (X, Y)$  are said to *globally shadow* each other if there are sets  $J_i \subset I$  such that  $J_i \subset J_{i+1}$  and exact curves  $\alpha_i : (I, J_i) \rightarrow (X, Y)$  from  $x_i$  to  $y_i$  such that  $f \circ \alpha_i$  is homotopic relative to endpoints to  $\alpha_{i+1}$  via curves  $(I, J_{i+1}) \rightarrow (X, Y)$ , and the curves  $\alpha_i$  have *bounded lengths*. The curves  $\alpha_i$  are called *relating curves*. Similarly,  $(x_i)$  and  $(y_i)$  *forwards shadow* each other if there are shadowing curves for  $i \geq 0$ , and *backwards shadow* each other if there are shadowing curves for  $i \leq 0$ .

There are many possible definitions of the term “bounded length,” and these mostly give identical definitions of global shadowing. In the setting of smooth curves on Riemannian manifolds, we can use the natural definition of the length of a smooth curve. In the definition of Katok, we consider the distance between the endpoints of lifts to a universal cover with equivariant metric.

In the non-relative case [Han85, Boy94], periodic orbits globally shadow each other if and only if they are Nielsen equivalent, and a homotopy between two maps induces a global shadowing relation between their orbits by considering the fat homotopy. Both these properties extend in a straightforward way to the relative case.

### 5.1 Shadowing by trellis orbits

It is important to know how the dynamics of the graph representative models that of the trellis map. The results stated in Theorem 5.3 are analogous to the shadowing results obtained by Handel [Han85] for surface diffeomorphisms relative to periodic orbits. We show that orbits of  $g$  are shadowed by orbits of  $f$ , and further, that periodic, asymptotic and biasymptotic orbits of  $g$  are shadowed by periodic, asymptotic and biasymptotic orbits of  $f$ . The results for periodic orbits were shown in [Col99], and the results for biasymptotic orbits are similar to those obtained by Rom-Kedar [RK94], who considered forward iterates of *strips* representing homotopy classes of curves parallel to  $T^U$ .

We assume throughout that  $(f; T)$  has graph representative  $(g; G, W)$ , and that  $(G, W)$  is embedded as a subset of  $(\mathcal{C}_{TV}M, \mathcal{C}_{TV}T^S)$  by an embedding  $i$ . We let  $\pi$  be a deformation retract  $(\mathcal{C}_{TV}M, \mathcal{C}_{TV}T^S) \rightarrow (G, W)$ , and  $f_t$  be a homotopy from  $\mathcal{C}f = f_0$  to  $f_1 = i \circ g \circ \pi$  with fat homotopy  $F$ . We first need to clarify what it means for an orbit of  $g$  to be backward asymptotic to the set  $Z$  of control edges.

**Definition 5.2 (Asymptotic graph orbits)** An orbit  $(y_i)$  of  $g$  is *forward asymptotic* to  $Z$  if  $y_j \in Z$  for some  $j$ . An orbit  $(y_i)$  of  $g$  is *backward asymptotic* to  $Z$  if there exists  $k \in \mathbb{Z}$  and exact curves  $\mu_i : (I, \{0, 1\}) \rightarrow (G, W)$  such that  $\mu_i$  begins in a periodic point of  $W$ ,  $\mu_i$  is homotopic in  $\mathcal{C}T$  to a curve in  $T^U$ ,  $g \circ \mu_i \supset \mu_{i+1}$  for  $i < k$  and  $y_i \in \mu_i$ .

Notice that if  $(y_i)$  is backward asymptotic to  $T^P$ , then the  $\alpha$ -limit set of  $(y_i)$  lies in the control set  $Z$ . The condition that the curves are homotopic to  $T^U$  is crucial; otherwise, we may have a periodic orbit of  $g$  which has  $\alpha$ -limit set in  $Z$ , but is not shadowed by an orbit of  $\mathcal{C}f$  in  $T^U$ .

**Theorem 5.3 (Shadowing by trellis orbits)** *Let  $(f; T)$  be a trellis mapping pair and  $(g; G, W)$  the graph representative of  $[f; T]$ . Then for any orbit  $(y_i)$  of  $g$  there is an orbit  $(x_i)$  of  $f$  which globally shadows  $(y_i)$  under the exact homotopy equivalence given by the inclusion  $(G, W) \rightarrow \mathcal{C}T$ . Further,*

1. *If  $(y_i)$  is periodic, then  $(x_i)$  can be chosen to be periodic.*
2. *If  $(y_i)$  is forward asymptotic to  $Z$  for some  $j$ , then  $(x_i)$  can be chosen in the stable set of  $T^P$  (so  $x_j \in T^S$  for some  $j$ ).*
3. *If  $(y_i)$  backward asymptotic to  $Z$ , then  $(x_i)$  can be chosen in the unstable set of  $T^P$ .*
4. *If  $(y_i)$  is both forwards and backwards asymptotic to  $Z$ , then  $(x_i)$  can be chosen to be biasymptotic to  $T^P$ .*

As an immediate corollary, we have the following result on the topological entropy. It can be shown that this entropy bound is sharp if  $T$  is well-formed, but otherwise need not be.

**Corollary 5.4 (Entropy of efficient graph maps)** *If  $g$  is an efficient controlled graph map compatible with  $(f; T)$ , then  $h_{top}(g) \leq h_{top}(f)$ .*

In [Col99], it is shown that periodic orbits of  $g$  are globally shadowed by an orbit of  $\mathcal{C}f$ . This proves Part 1 of Theorem 5.3, and forms the basis for all other shadowing results. To show that all orbits of  $g$  are shadowed by orbits of  $f$ , we need to find a uniform bound for the lengths of shadowing curves in  $g$ .

**Lemma 5.5** *If  $(x_i)$  and  $(y_i)$  are orbits of  $g$  which globally shadow each other, there are relating curves  $\alpha_i$  from  $x_i$  to  $y_i$  such that the  $\alpha_i$  lie in  $Z$ ,  $P$  or  $Pre-P$ .*

**PROOF:** Suppose the curves  $\alpha_i$  are tight and do not cross any control edges. Then  $\alpha_{i+1} = g \circ \alpha_i$  for all  $i$ , so the  $\alpha_i$  are eventually periodic or disjoint. Since  $g$  has no

invariant subgraphs apart from  $P$ , and  $g$  is expanding outside  $\text{Pre-}P$ , the  $\alpha_i$  can only be eventually periodic if they lie in  $\text{Pre-}P$ .

If the  $\alpha_i$  do cross  $Z$ , they must eventually do so in periodic control edges. We can split the curves into pieces that cross  $Z$  and pieces that do not. The pieces that do not must then lie in  $\text{Pre-}P$ .  $\square$

**PROOF OF THEOREM 5.3 (GLOBAL SHADOWING):** If  $(y_i)$  is periodic, the result follows from the relative Nielsen theory. If not, suppose there are periodic orbits  $(y_i^j)$  such that for every  $i$ ,  $\lim_{j \rightarrow \infty} y_i^j = y_i$ . We choose orbits  $(x_i^j)$  of  $f$  such that  $(x_i^j, 0)$  shadows  $(y_i^j, 1)$  under  $F$ . Since  $M$  is compact, by choosing a subsequence if necessary, we can assume the sequence  $x_0^j$  converges to a point  $x_0$  as  $j \rightarrow \infty$ , and this means that  $x_i^j$  converges to  $x_i = f_0^i(x_0)$  for all  $i > 0$ . Choosing further subsequences, we can assume further that  $x_i^j$  converges to a point  $x_i$  as  $j \rightarrow \infty$  for any  $i < 0$ .

We therefore have an orbit  $(x_i)$  such that  $x_i = \lim_{j \rightarrow \infty} x_i^j$  for all  $i$ . We now let  $\pi_i^j$  be the relating curves for  $(x_i^j, 0) \sim (y_i^j, 1)$ . Take an open cover  $\mathcal{U}$  of  $(\mathcal{C}_{T^U}M, \mathcal{C}_{T^U}T^S) \times I$  by sets  $U$  such that the components of  $U \setminus T^S$  are simply-connected. Let  $\beta_i^j$  and  $\gamma_i^j$  be curves from  $(x_i, 0)$  to  $(x_i^j, 0)$  and from  $(y_i, 1)$  to  $(y_i^j, 1)$  in a component of  $U$  and  $V$  respectively, which must exist for large enough  $j$ . Then  $f_0 \circ \beta_i^j \sim \beta_{i+1}^j$  and  $f_1 \circ \gamma_i^j \sim \gamma_{i+1}^j$  whenever  $j$  is sufficiently large.

Now let  $\alpha_i^j = \beta_i^j \cdot \pi_i^j \cdot \gamma_i^j$ . The curves  $\alpha_0^j$  have bounded lengths, since they are joins from three sets of curves which each have bounded lengths. Therefore, there must be finitely many homotopy classes  $[\alpha_0^j]$ , at least one of which must occur infinitely often. By taking an subsequence if necessary, we can assume  $[\alpha_0^j]$  has the same value for all  $j$ . Now,

$$F \circ \alpha_i^j = F \circ \beta_i^j \cdot F \circ \pi_i^j \cdot F \circ \gamma_i^j \sim \beta_{i+1}^j \cdot \pi_{i+1}^j \cdot \gamma_{i+1}^j = \alpha_{i+1}^j$$

and since this must be true whenever  $\alpha_i^j$  is defined, we have  $[\alpha_i^j]$  is independent of  $j$  for fixed  $i$ . By taking subsequences if necessary, we can also ensure this is the case for  $i$  negative (or this follows automatically if  $F$  has a homotopy inverse). Hence, we have curves  $\alpha_i$  from  $x_i$  to  $y_i$  of bounded lengths such that  $F \circ \alpha_i \sim \alpha_{i+1}$ .  $\square$

## 5.2 Shadowing by (un)stable and biasymptotic orbits

We now show that orbits of  $g$  which enter  $Z$  are shadowed by orbits of  $f$  in  $W^S(T^P)$ , that orbits backward asymptotic to  $Z$  are shadowed by orbits of  $f$  in  $W^U(T^P)$ , and that orbits of  $g$  which are biasymptotic to  $Z$  are shadowed by orbits of  $f$  in  $W^U(T^P) \cap W^S(T^P)$ . These results have no corresponding statements in Nielsen-Thurston theory, since we cannot guarantee the existence of periodic saddle points in this case. Note that due to the asymmetry in our treatment of stable and unstable curves, part (2) and part (3) are genuinely different statements and require different proofs. We first prove an initial result which shows that orbits of  $g$  are contained in curves with endpoints in  $W$  which map over each other under  $g$ . If  $\alpha_0 : (I, J) \rightarrow (G, W)$  is a tight exact curve in  $G$  with endpoints in  $W$ , then we write  $g(\alpha_0) \supset \alpha_1$  if  $g(\alpha)$  contains  $\alpha_1$  as a sub-curve.



**Lemma 5.6** *Let  $(y_i)$  be an orbit of  $g$ . Then there is a bi-infinite sequence of exact curves  $\mu_i : (I, \{0, 1\}) \rightarrow (G, W)$  such that  $y_i \in \mu_i(I)$  and  $g \circ \mu_i$  contains  $\mu_{i+1}$  as a sub-curve.*

PROOF: First suppose that every component of  $G \setminus W$  is homotopy-equivalent to a point or circle. Then every pair of points of  $W$  in the same region, there are two tight curves between them (for an homotopy circle) and one for a homotopy point. Let  $\Gamma$  be the set of all these curves. Let  $\Sigma_n, n \geq 0$  be the set of bi-infinite sequences of elements of  $\Gamma$  such that  $g(\gamma_i) \subset \gamma_{i+1}$  for  $i \geq -n$ . This is nonempty, since if  $y_i \in \gamma_i$ , then  $y_{i+1} \in g(\gamma_i)$ , and is compact since  $\Gamma$  is finite. Clearly  $\Sigma_{n+1} \subset \Sigma_n$  so the  $\Sigma_n$  form a sequence of compact nested sets, hence  $\Sigma = \bigcap_{n=0}^{\infty} \Sigma_n$  is nonempty. Then any sequence  $(\mu_i)$  is a sequence in  $\Sigma$ .

Now if  $g$  is irreducible optimal graph map, there exists  $N$  such that the complement of  $g^{-N}(W)$  consists of homotopy points and circles. We consider  $g$  as a self-map of  $(G, g^{-N}(W))$ , and let  $(\gamma_n)$  be a sequence of curves found by the previous arguments. Then  $g^N(\gamma_{i-N})$  is an exact curve  $(I, \{0, 1\}) \rightarrow (G, W)$  containing  $(y_i)$ , and maps across  $g^N(\gamma_{i+1-N})$ . Therefore taking  $\mu_i = g^N(\gamma_{i-N})$  gives the required bi-infinite sequence of curves.  $\square$

PROOF OF THEOREM 5.3 (STABLE ORBITS): Let  $(\mu_i)$  be a sequence of exact curves  $(I, \{0, 1\}) \rightarrow (G, W)$  such that  $g \circ \mu_i$  contains  $\mu_{i+1}$  as a sub-curve. and  $y_i \in \mu_i$ . Let  $S_i$  be the stable segment containing  $y_i$  for  $i \geq 0$ ; without loss of generality we assume  $y_{-1} \notin T^S$ .

We now use recursively construct a sequence of segments  $S_i$  of  $W^S(T^P)$  for  $i < 0$ . Suppose therefore for  $i > n$  (where  $n$  is negative) we have a stable segment  $S_i$  of  $f^{-n}(T^S)$  with an essential intersection with  $\mu_i$  at a point  $z_i$  and that  $f(S_i) \subset S_{i+1}$ . Then  $g \circ \mu_n$  has an essential intersection with  $S_{n+1}$  at a point  $z_{n+1}$ , so  $f \circ \mu_n$  has an essential intersection with  $S_{n+1}$ . Therefore, one of the segments  $S_n$  of  $f^{-1}(S_{n+1})$  is such that  $f \circ \mu_n$  contains an essential intersection with  $f(S_n)$  in the same intersection class as  $z_{n+1}$  under the homotopy  $f_t$ . Let  $f(z_n)$  be this essential intersection, so that  $z_n$  is an essential intersection of  $S_n$  with  $\mu_n$ , as shown in figure 22.

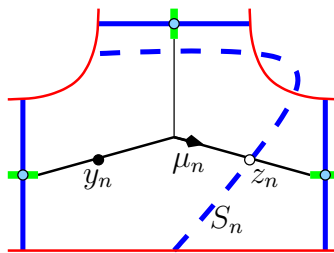


Figure 22: Construction of  $z_n$ .

The sequence of segments  $S_n$  of  $W^S(T^P)$  is such that  $f(S_n) \subset S_{n+1}$  for  $n < 0$ , so  $f^n(S_n) \subset f^{n+1}(S_{n+1}) \subset S_0$ . Hence the sets  $f^n(S_n)$  are a nested sequence of compact intervals, so contain a limit point  $x_0$ . We claim the orbit  $(x_i) = f^i(x_0)$  globally shadows  $(y_i)$  under the homotopy  $f_t$ .

We need to show that the orbits  $(x_n, 0)$  and  $(y_n, 1)$  globally shadow each other under  $F$ . Let  $\alpha_n$  be a curve from  $(x_n, 0)$  to  $(z_n, 0)$  in  $S_n \times \{0\}$ ,  $\beta_n$  a curve from  $(z_n, 0)$  to

$(g^{-1}(z_{n+1}), 1)$  in  $\mu_{n+1} \times I$  and  $\gamma_n$  the curve from  $(g^{-1}(z_{n+1}), 1)$  to  $(y_n, 1)$  in  $\mu_n(I) \times \{1\}$ . Then  $F \circ \alpha_n$  is a curve in  $T^S \times I$  and  $F \circ \gamma_n$  is a curve in  $\mu_{n+1} \times \{1\}$ . The curve  $F \circ \beta_n$  goes from  $(f(z_n), 0) \in S_{n+1} \times \{0\}$  to  $(z_{n+1}, 1) \in S_{n+1} \times \{0\}$ , and is homotopic to a curve whose path lies in  $S_{n+1} \times I$  since  $z_{n+1}$  is an essential intersection of  $S_n$  with  $\mu_n$ . Hence  $F \circ (\alpha_n \cdot \beta_n)$  is homotopic to a curve in  $S_{n+1} \times I$  from  $(x_{n+1}, 0)$  to  $(z_{n+1}, 1)$  and  $F \circ \gamma_n$  is homotopic to a curve from  $(z_{n+1}, 1)$  to  $(y_{n+1}, 1)$ . Since  $\beta_{n+1}$  is homotopic to a curve in  $\{z_{n+1}\} \times I$  joined with a curve in  $\mu_{n+1}(I) \times \{1\}$ , we have  $F \circ (\alpha_n \cdot \beta_n \cdot \gamma_n) \sim \alpha_{n+1} \cdot \beta_{n+1} \cdot \gamma_{n+1}$  as required.  $\square$

**PROOF OF THEOREM 5.3 (UNSTABLE ORBITS):** Let  $(y_i)$  be an orbit of  $g$  such that for  $i \leq 0$  there are exact curves  $\mu_i : (I, \{0, 1\}) \rightarrow (G, W)$  containing  $y_i$  such that  $g(\mu_i) \supset \mu_{i+1}$ ,  $\mu_i$  is unstable-parallel, has periodic initial vertex. For  $i \geq 0$ , let  $\sigma_i$  be a cross-cut in  $\mathcal{CT}$  which does not intersect  $T^S$ , but which crosses  $G$  exactly once in the edge containing  $y_i$ . Let  $U_0$  be the segment of  $T^U$  which is exact homotopic to  $\mu_i$ . Suppose for  $i \leq n$ , there is a segment  $U_i$  of  $W^U(T^P)$  such that  $U_i \supset U_{i-1}$  and  $U_i$  is exact homotopic to a tight curve  $\mu_i$  in  $G$  containing  $y_i$ . Then  $U_i$  has an essential intersection  $z_i$  with  $\sigma_i$ . Since  $y_{i+1} \subset g \circ \mu_i$ , there must be an essential intersection of  $g \circ \mu_i$  with  $\sigma_i$ , and this intersection lies in the same intersection class as an intersection  $z_{i+1}$  of  $f(U_i)$  with  $\sigma_{i+1}$ . Let  $U_{i+1}$  be the segments of  $W^U$  containing  $z_{i+1}$ , and let  $\mu_{i+1}$  be the tight curve in  $G$  which is exact homotopic to  $U_i$ , and so  $y_{i+1} \in \mu_{i+1}$ . Then the set  $U = \bigcap_{n=0}^{\infty} f^{-n}(U_n)$  is a nonempty subset of  $T^U$ .

We take  $x_0 \in U$ , and  $(x_i) = f^i(x_0)$  for  $i \neq 0$ . A similar analysis to that of the proof of Part 2 shows that  $(x_i)$  globally shadows  $(y_i)$ .  $\square$

**PROOF OF THEOREM 5.3 (BIASYMPTOTIC ORBITS):** Without loss of generality we can assume that  $y_k \in W$ , but  $y_{k-1} \notin Z$ . Take  $\sigma_i$  and  $U_i$  for  $i < k$  as in the proof of Part 3, and let  $\sigma_k$  to be the stable segment  $S_k$  containing  $y_k$ . Then  $z_k$  is a point in  $U_k \cap S_k$ . We take  $x_i = f^{i-k}(z_k)$ . Then a similar analysis to that of the proof of Part 2 shows that  $(x_i)$  globally shadows  $(y_i)$  and gives the construction of the stable segments  $S_i$ .  $\square$

## References

- [BH92] Mladen Bestvina and Michael Handel. Train tracks and automorphisms of free groups. *Ann. of Math. (2)*, 135(1):1–51, 1992.
- [BH95] Mladen Bestvina and Michael Handel. Train-tracks for surface homeomorphisms. *Topology*, 34(1):109–140, 1995.
- [BH99] Philip L. Boyland and Toby Hall. Isotopy stable dynamics relative to compact invariant sets. *Proc. London Math. Soc. (3)*, 79(3):673–693, 1999.
- [Bir49] Garrett Birkhoff. Théorie et applications des treillis. *Ann. Inst. H. Poincaré*, 11:227–240, 1949.

- [BK82] Joan S. Birman and Mark E. Kidwell. Fixed points of pseudo-anosov diffeomorphisms of surfaces. *Adv. in Math.*, 46(2):217–220, 1982.
- [Boy92] Philip Boyland. Rotation sets and monotone periodic orbits for annulus homeomorphisms. *Comment. Math. Helv.*, 67(2):203–213, 1992.
- [Boy94] Philip Boyland. Topological methods in surface dynamics. *Topology Appl.*, 58(3):223–299, 1994.
- [Boy99] Philip L. Boyland. Isotopy stability of dynamics on surfaces. In Marcy Barge and Krystyna Kuperberg, editors, *Geometry and Topology in Dynamics*, number 246 in Contemporary Mathematics, pages 17–45. American Mathematical Society, 1999.
- [Bro71] Robert Brown. *The Lefschetz Fixed Point Theorem*. Scott, Foresman and Company, 1971.
- [BW95] Keith Burns and Howard Weiss. A geometric criterion for positive topological entropy. *Comm. Math. Phys.*, 172(1):95–118, 1995.
- [CB88] A. Casson and S. Bleiler. *Automorphisms of surfaces, after Nielsen and Thurston*, volume 9 of *London Mathematical Society Student Texts*. Cambridge University Press, 1988.
- [Col99] Pieter Collins. Dynamics forced by surface trellises. In Marcy Barge and Krystyna Kuperberg, editors, *Geometry and Topology in Dynamics*, number 246 in Contemporary Mathematics. American Mathematical Society, 1999.
- [Col01] Pieter Collins. Relative periodic point theory. *Topology Appl.*, 115(1):97–114, 2001.
- [Con78] Charles Conley. *Isolated Invariant Sets and the Morse Index*, volume 38 of *CBMS Regional Conference Series in Mathematics*. AMS Publications, 1978.
- [CP96] Freddy Christiansen and Antonio Politi. Symbolic encoding in symplectic maps. *Nonlinearity*, 9(6):1623, 1996.
- [dC99] André de Carvalho. Pruning fronts and the formation of horseshoes. *Ergodic Theory Dynamical Systems*, 19(4):851–894, 1999.
- [dCH02] André de Carvalho and Toby Hall. The forcing relation for horseshoe braind types. *Experimental Math.*, 11(2):271–288, 2002.
- [Eas86] Robert Easton. Trellises formed by stable and unstable manifolds in the plane. *Trans. Amer. Math. Soc.*, 294(2):719–732, 1986.
- [Eas98] Robert W. Easton. *Geometric methods for discrete dynamical systems*. Number 50 in Oxford Engineering Science Series. Oxford University Press, 1998.
- [Fat90] Albert Fathi. Homotopical stability of pseudo-anosov diffeomorphisms. *Ergodic Theory Dynamical Systems*, 10(2):387–294, 1990.

- [FM93] John Franks and Michael Misiurewicz. Cycles for disk homeomorphisms and thick trees. In *Nielsen Theory and Dynamical Systems*, number 152 in Contemporary Mathematics, pages 69–139, 1993.
- [Fri83] David Fried. Periodic points and twisted coefficients. In *Geometric dynamics (Rio de Janeiro, 1981)*, volume 1007 of *Lecture Notes in Math.*, pages 261–293. Springer, Berlin-New York, 1983.
- [Han85] Michael Handel. Global shadowing of pseudo-Anosov homeomorphisms. *Ergodic Theory Dynamical Systems*, 5(3):373–377, 1985.
- [Han99] Michael Handel. A fixed-point theorem for planar homeomorphisms. *Topology*, 38(2):235–264, 1999.
- [Hul00] Helen Hulme. *PhD. Thesis*. PhD thesis, University of Liverpool, 2000.
- [JG93] Bo Ju Jiang and Jian Han Guo. Fixed points of surface diffeomorphisms. *Pacific J. Math.*, 160(1):67–89, 1993.
- [Jia83] Boju Jiang. *Lectures on Nielsen Fixed Point Theory*. Number 14 in Contemporary Mathematics. American Mathematical Society, 1983.
- [Jia93] Boju Jiang. Nielsen theory for periodic orbits and applications to dynamical systems. In Christopher McCord, editor, *Nielsen Theory and Dynamical Systems*, number 152 in Contemporary Mathematics, pages 183–202. American Mathematical Society, 1993.
- [Kei97] Ulrich Keil. *Construction Algorithmique D’un Collage Efficace Pour Les Homéomorphismes D’une Surface À Bord (Algorithmic Construction of an Efficient Splicing for Homeomorphisms of a Punctured Surface)*. PhD thesis, L’Université Paul Sabatier De Toulouse, 1997.
- [KO98] Bernd Krauskopf and Hinke Osinga. Growing 1d and quasi-2d unstable manifolds of maps. *J. Comput. Phys.*, 146(1):404–419, 1998.
- [Los93] Jérôme Los. Pseudo-Anosov maps and invariant train tracks in the disc: A finite algorithm. *Proc. London Math. Soc. (3)*, 66(2):400–430, 1993.
- [Los96] Jérôme Los. On the conjugacy problem for automorphisms of free groups. *Topology*, 35(3):779–806, 1996.
- [MT94] F. Allan McRobie and J. M. T. Thompson. Knot-types and bifurcations sequences of homoclinic and transient orbits of a single-degree-of-freedom driven oscillator. *Dynam. Stab. Sys.*, 9:223–251, 1994.
- [MV93] L. Mora and M. Viana. Abundance of strange attractors. *Acta Math.*, 171:1–71, 1993.
- [New80] Sheldon E. Newhouse. Lectures on dynamical systems. In *Dynamical Systems (C.I.M.E. Summer School, Bressanone, 1978)*, volume 8 of *Progress in Mathematics*, pages 1–114. Birkhäuser, 1980.

- [Poi99] Henri Poincaré. *Les méthodes nouvelles de la mécanique céleste*, 1892–1899.
- [Ray98] Victoria Rayskin. PhD thesis, University of California, Berkeley, 1998.
- [RK94] Vered Rom-Kedar. Homoclinic tangles—classification and applications. *Nonlinearity*, 7(2):441–473, 1994.
- [Sim89] C. Simó. On the analytical and numerical approximation of invariant manifolds. In D. Benest and C. Froeschlé, editors, *Les Méthodes Modernes de la Mécanique Céleste*, pages 285–329. Goutelas, 1989.
- [Sma63] Stephen Smale. A structurally stable differentiable homeomorphism with an infinite number of periodic points. In *Proceedings of the International Symposium on Nonlinear Vibrations*, volume 2, pages 365–366, 1963.
- [Soc79] Société Mathématique de France. *Travaux de Thurston sur les surfaces*, volume 66–67, 1979.
- [SRD99] D. Sterling, H. R. Dullin, and J. D. Meiss. Homoclinic bifurcations for the hnon map. *Phys. D*, 134(2):153–184, 1999.
- [Thu76] William P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. Preprint, 1976.
- [Wig91] Steven Wiggins. *Chaotic Transport in Dynamical Systems*. Introduction to Applied Mathematics. Springer-Verlag, 1991.