Pruning theory and Thurston's classification of surface homeomorphisms

André de Carvalho* and Toby Hall†

February 2000, revised January 2001

Abstract

Two dynamical deformation theories are presented — one for surface homeomorphisms, called *pruning*, and another for graph endomorphisms, called *kneading* — both giving conditions under which all of the dynamics in an open set can be destroyed, while leaving the dynamics unchanged elsewhere. The theories are related to each other and to Thurston's classification of surface homeomorphisms up to isotopy.

1 Introduction

In this paper, new techniques for studying the dynamics of families of surface homeomorphisms are introduced. Two dynamical deformation theories are presented: *pruning theory*, applicable to surface homeomorphisms, which is motivated by the work of Cvitanović, Gunaratne, and Procaccia [8]; and *kneading theory*, applicable to graph endomorphisms, which is a generalization of some of the topological aspects of Milnor and Thurston's work on endomorphisms of the interval [17]. Both theories give conditions under which all of the dynamics in an open set can be destroyed, while leaving the dynamics unchanged elsewhere. The two theories are related to each other and to Thurston's classification theorem for isotopy classes of surface homeomorphisms [21].

One of the main underlying ideas used here was introduced by Williams [22]: by collapsing segments of stable manifolds of a surface homeomorphism one obtains a graph endomorphism; the surface homeomorphism can be recovered by taking the inverse limit. Many researchers have found this interplay between 1- and 2-dimensional systems fruitful. The emphasis here lies on using the correspondence to obtain results about the dynamics of families of surface homeomorphisms.

*Institute for Mathematical Sciences, State University of New York at Stony Brook, NY 11794-3660, USA, andre@math.sunysb.edu

†Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK, T.Hall@liverpool.ac.uk

The paper introduces a 2-dimensional version of Milnor and Thurston's kneading theory for endomorphisms of the interval [17], and a generalization of the same theory for graph endomorphisms. These are then used to shed new light on the Bestvina-Handel algorithmic proof [5] of Thurston's classification theorem [21]. The approach adopted clarifies the mechanisms involved in their algorithm and improves some of its aspects. Like its 1-dimensional counterpart, pruning theory enables one to give symbolic descriptions of the dynamics of all of the maps in a family — the *pruning family* — in terms of a fixed underlying subshift of finite type.

The techniques presented here — specifically, the explicit relationship between 1- and 2-dimensional dynamics — have given rise to a conjecture giving a complete picture of how Boyland's forcing relation [6] organizes the periodic orbits of Smale's horseshoe. Under this conjecture, the set of pseudo-Anosov horseshoe braid types is partitioned into totally ordered subsets which organize the mechanism of horseshoe creation. A forthcoming paper will provide a statement of the conjecture together with proofs of some special cases.

The original motivation for this work came from the *Pruning Front Conjecture* of Cvitanović, Gunaratne, and Procaccia [8], which states that the dynamics of any map in the Hénon family can be viewed as what remains from the dynamics of Smale's horseshoe after some orbits (namely those that fall into the *pruning front*) are destroyed. More will be said about the Pruning Front Conjecture after a more detailed description of the constructions and results contained in this paper.

Pruning theory makes precise, and more general, a concept which was described intuitively in [8] in the particular case of Smale's horseshoe map [19]. The initial surface homeomorphism $F\colon S \to S$ is taken to be one whose dynamics is well understood, in the sense that all of the non-trivial dynamics is contained in an F-invariant *thick graph* $\mathbb{G}$, which has the property that $F\colon \mathbb{G} \to \mathbb{G}$ induces a graph endomorphism $f\colon G \to G$. The horseshoe is a very simple thick graph map (whose induced graph endomorphism is a unimodal map of the interval), and is a good example to keep in mind when studying the more general theory. In Section 2, conditions are given under which all of the dynamics in a topological disk D, called a *pruning disk*, can be destroyed by an isotopy supported in D (and an arbitrarily small region in $S \setminus \mathbb{G}$). The first main goal of the paper is described by the following statement about thick graph maps F with positive topological entropy, which is contained in Theorem 2.3 and the definition preceding Example 2.

The Pruning Family: *An uncountable family $\mathcal{P}(F)$ of homeomorphisms of S is constructed. The dynamics of maps in $\mathcal{P}(F)$ can be understood as the dynamics of F less that which is* pruned away *from a sequence of pruning disks.*

Section 3 introduces *generalized kneading theory.* Its main purpose is to make it possible to understand pruning of a thick graph homeomorphism $F\colon S \to S$ on the level of the induced graph endomorphism $f\colon G \to G$. (Other authors have made use of the relationship between the dynamics of 1- and 2-dimensional systems in an analogous way, notably Williams [22], Thurston [21], Barge-Martin [3], Bestvina-Handel [5], and Franks-Misiurewicz [10].) When $F: S \to S$ is the horseshoe, the simplest possible prunings project to kneadings of unimodal maps of the type studied by Milnor and Thurston [17]. The term *kneading* was chosen because this theory generalizes the *topological* aspects of the Milnor-Thurston theory of endomorphisms of the interval to other 1-dimensional spaces (although the original motivation of the action of kneading dough is no longer applicable). The combinatorial and analytic aspects of their theory, however, are not developed here: in particular, symbolic sequences play only a minor rôle in this paper. Baillif [2] and Alves and Sousa Ramos [1] have, independently and using different techniques, developed the notion of kneading determinant in this more general context, and have proved an analogue of the Milnor-Thurston formula relating the kneading determinant and the ζ-function for tree endomorphisms.

More precisely, given a graph endomorphism $f: G \to G$, conditions are given on open subsets K of G, called *kneading sets*, for which it is possible to modify both G and f to obtain a graph endomorphism $f_K: G_K \to G_K$ whose dynamics can be understood as being the dynamics of the original endomorphism f less the dynamics which is contained in K. If $F: \mathbb{G} \to \mathbb{G}$ projects to $f: G \to G$, then each pruning of F induces a corresponding kneading of f. The converse is not true in general, since thick graphs have more structure than graphs (for example, the cyclic order of edges around a vertex). Generalized kneading theory may have some independent interest in the study of graph endomorphisms: the approach taken in this paper, however, is to place additional conditions on kneading sets which ensure that they do indeed lift to pruning disks. The practical problem of finding pruning disks for a thick graph map can then be approached by finding suitable kneading sets for the corresponding graph endomorphism. In summary, the second main goal of this paper is described by the following statement about graph endomorphisms f with positive topological entropy:

The Kneading Family: *An uncountable family $\mathcal{K}(f)$ of endomorphisms of one-dimensional spaces can be constructed. The dynamics of maps in $\mathcal{K}(f)$ can be understood as the dynamics of f less that which is* pulled away *from a sequence of kneading sets.*

The pruning family and the kneading family are related by the following statement, which expresses the interplay between the 1- and 2-dimensional systems.

Both Families: *If $f: G \to G$ is induced by a thick graph map $F: \mathbb{G} \to \mathbb{G}$, the kneading family $\mathcal{K}(f)$ and the pruning family $\mathcal{P}(f)$ are in one-to-one correspondence (subject to obvious compatibility conditions between the 1- and 2-dimensional structures): collapsing segments of stable sets takes $\mathcal{P}(f)$ to $\mathcal{K}(f)$ and the inverse limit takes $\mathcal{K}(f)$ back to $\mathcal{P}(f)$.*

In the body of the paper these two statements are not presented as theorems: there are some technical problems in doing so which would lead too far afield. Since, as was mentioned before, the emphasis here is on drawing conclusions about 2-dimensional dynamics from the interplay between the two families, these technicalities are best avoided. This is discussed further in Section 3. The main results of the paper which express the relationship between the two families are Theorems 3.11 and 3.12.

The following is another precise statement relating the pruning family with 1-dimensional dynamics. It follows easily from the constructions described in the paper, as explained in the remark following Example 1.

The Pruning Family is Big: *The pruning family of Smale's horseshoe map F contains a subfamily whose dynamics mimics that of a* full[1] *family of unimodal maps of the interval. In particular, $\mathcal{P}(F)$ contains uncountably many distinct dynamical systems.*

In Sections 4 and 5, the Bestvina-Handel algorithm for finding train tracks of surface homeomorphisms [5] is rewritten in the language of pruning theory. Other train track algorithms have been presented by Franks and Misiurewicz [10] and by Los [16]. There are certain advantages, both practical and theoretical, in recasting the Bestvina-Handel algorithm in terms of pruning. On the practical side, pruning is a more general operation than folding, which is the principal move of the Bestvina-Handel algorithm, and long sequences of foldings can often be replaced by a single pruning, with corresponding efficiency benefits. An example is given in Section 4.3 in which several foldings (8 in one computer implementation [12]) can be replaced with a single pruning. On the theoretical side, pruning makes it possible to carry out the algorithm in such a way that new dynamics is never created at any stage: this makes it possible to follow dynamics through an isotopy from the starting map to a Thurston representative of its isotopy class. This is achieved by replacing Bestvina and Handel's operation of *performing a valence two isotopy* (which, although it never increases topological entropy, may create new dynamics while destroying other dynamics) with a new operation, described in terms of pruning, called *cleaning a valence two vertex*. Cleaning is, in general, a more complicated procedure than performing a valence two isotopy, and if the algorithm is being applied simply to calculate a train track then it is more sensible to use valence two isotopies. However, the fact that the algorithm can be carried out using no non-trivial isotopies other than prunings has the following important corollary, which is the third main result in the paper (see Theorem 4.3):

Pruning and Thurston Classification: *For any pseudo-Anosov map isotopic to F relative to some finite F-invariant set, there is a map in the pruning family $\mathcal{P}(F)$ with the same (non-trivial) dynamics.*

[1] A family of unimodal maps is said to be *full* if it presents all possible kneading sequences. The famous example is the quadratic family $f_\mu(x) = \mu x(1-x)$, for $0 < \mu \le 4$.

The maps in the pruning family usually have many wandering domains: in the case of a pseudo-Anosov isotopy class, collapsing these domains to points defines a semi-conjugacy to the appropriate pseudo-Anosov map. In fact, the maps in the pruning family have invariant laminations rather than invariant foliations.

Together with the results above, this provides evidence that $\mathcal{P}(F)$ may indeed contain enough topological models to be considered a *full* family of surface homeomorphisms. On the one hand, it contains analogues of full families of 1-dimensional endomorphisms of the interval. On the other hand, $\mathcal{P}(F)$ contains representatives of all of the Thurston maps that is possible for it to contain.

A treatment of pruning in a much more general context, where it is not assumed that F is a thick graph map, is given in [7]. The more restricted approach here, besides being more accessible, permits, as a consequence of the underlying one-dimensional structure, a much more precise understanding of the space $\mathcal{P}(F)$, and a more direct approach to the Bestvina-Handel algorithm.

As has already been mentioned, an important motivation for this work was the Pruning Front Conjecture (PFC) of Cvitanović, Gunaratne, and Procaccia [8]. It attempts to relate the dynamics of maps in the Hénon family of diffeomorphisms of the plane,

$$H_{a,b}(x,y) = (a - by - x^2, x),$$

to the family $\mathcal{P}(F)$, where F is Smale's horseshoe: it states that for every pair (a,b) with $0 < b < 1$, the Hénon map $H_{a,b}$ is semi-conjugate to some element of $\mathcal{P}(F)$. Thus every Hénon map can be viewed, up to semi-conjugacy, as a full horseshoe from which the dynamics in a given region (the *pruning front*) has been destroyed. A proof of this conjecture would provide a full understanding, on the topological level, of the dynamics of maps in the Hénon family. In [8], PFC is also stated for the Lozi family

$$L_{a,b}(x,y) = (a - by - |x|, x).$$

Using methods different from the ones presented here, Ishii [15] has proved PFC for the Lozi family.

The Hénon family reduces to the standard quadratic family $x \mapsto a - x^2$ when $b = 0$. A number of authors have explored the relationship between these two families, stressing similarities (notably in the work of Benedicks and Carleson [4]), differences (Holmes and Whitley [13], Holmes and Williams [14]), or both (van Strien [20], Hall [11]). In the pruning family, however, each map has a natural underlying one-dimensional structure: it can be collapsed to a tree (or dendrite) endomorphism in a way which preserves the topological dynamics of the original map: indeed, the 2-dimensional map can be recovered simply by taking the inverse limit of the 1-dimensional map. Thus the Pruning Front Conjecture, if true, provides a link between Hénon maps and 1-dimensional dynamics for *any* value of the parameters with $b \in (0,1)$.

A number of the definitions and constructions in this paper, while intuitively straightforward, involve a distracting amount of technical detail in order to deal with all of the different cases which may arise. The reader is encouraged to refer to the examples and figures whenever necessary for clarification.

Acknowledgements: Both authors are grateful for hospitality and funding from the other's institution, to Philip Boyland, who made many helpful comments on both early and late drafts of this paper, and to the two referees. The first author is also grateful to the Department of Mathematics of the University of California at Berkeley, where part of this work was carried out; to Viviane Baladi and the *Fonds National Suisse pour la Recherche Scientifique* who provided hospitality and funding at the University of Geneva; and the EPSRC for visiting fellowship grant number GR/M29702.

2 Pruning theory

2.1 Preliminary definitions

Pruning theory describes a way in which the dynamics of surface homeomorphisms can be destroyed in a controlled manner by isotopies relative to a given compact invariant set. This section contains a brief summary of some standard definitions which will be used in the remainder of the paper.

Let X be a topological space, and $f\colon X \to X$ a continuous map. Dynamical systems theory is concerned with asymptotic properties of the family $\{f^n\}$ of *iterates* of f, where f^n denotes the n-fold composition $f \circ \cdots \circ f$ of f with itself. If f is a homeomorphism, then n can be either positive or negative (with $f^{-n} = (f^{-1})^n$).

Non-trivial dynamics requires some form of recurrence. In this paper, points of X will be considered to be dynamically non-trivial if they are non-wandering.

Definitions: A connected open subset U of X is a *wandering domain* if $f^n(U) \cap U = \emptyset$ for all $n > 0$. The *non-wandering set* $\Omega(f)$ of f is the set of all points x which are not contained in any wandering domain.

$\Omega(f)$ is a closed subset of X with $f(\Omega(f)) \subseteq \Omega(f)$. If f is a homeomorphism then $f(\Omega(f)) = \Omega(f) = \Omega(f^{-1})$. A *period* n (or just *periodic*) point x of f is a fixed point of f^n which is not fixed by f^k for $1 \le k < n$: clearly all periodic points are in $\Omega(f)$.

Now suppose that $f\colon X \to X$ is a homeomorphism. Then f is said to be *supported* on a subset U of X if f is the identity on $X \setminus U$. A second homeomorphism $g\colon X \to X$ is *isotopic to* f if there is a continuous map $F\colon X \times [0,1] \to X$ such that each map $f_t\colon X \to X$ defined by $x \mapsto F(x,t)$ is a homeomorphism, with $f_0 = f$ and $f_1 = g$. The isotopy is said to be *supported* on a subset U of X if the homeomorphisms f_t are all equal on $X \setminus U$, and is said to be *relative* to U if it is supported on $X \setminus U$.

2.2 Thick graph maps

Following the approach of [5, 10], the majority of the surface homeomorphisms in this paper will be assumed to be of a particular type, called *thick graph maps*. This means that all of their interesting dynamics is contained in a subsurface called a *thick graph*: intuitively, as the name suggests, this is a graph in which each point has been thickened up, either to a disk or an arc according as the point is a vertex or a regular point of the graph. The homeomorphisms are assumed to act on the thick graph in such a way that they induce endomorphisms of the underlying graph. The justification for taking this approach is that every surface homeomorphism is isotopic to a thick graph map.

Definitions: A *thick graph* is a triple $(S, \mathbb{G}, A)$, where S is a closed orientable surface endowed with a fixed metric compatible with its topology, $\mathbb{G}$ is a compact subsurface of S (with boundary) which is partitioned into compact *decomposition elements*, and A is a compact subset of $\mathbb{G}$, such that

a) Each decomposition element of $\mathbb{G}$ is either a *leaf* homeomorphic to $[0, 1]$, or a *junction* homeomorphic to D^2.

b) The boundary in $\mathbb{G}$ of each junction is a finite number of disjoint arcs: if there are k such arcs, then the disk is called a *k-junction*.

c) The set $\bigcup_{k\neq 2}\{k\text{-junctions}\}$ is finite.

d) Each decomposition element which is not in the accumulation of a sequence of distinct 2-junctions is contained in a chart as depicted in Figure 1.

e) Each component of $S \setminus \mathbb{G}$ is an open disk.

f) $A \cap \partial J = \emptyset$ for each junction J.

If $(S, \mathbb{G}, A)$ is a thick graph, let $\sim$ be the equivalence relation on $\mathbb{G}$ given by $x \sim y$ if and only if x and y lie in the same decomposition element. Then $G = \mathbb{G}/\!\sim$ is a graph, whose vertices (which may have valence 2) correspond to the junctions of $\mathbb{G}$: the canonical projection will be denoted $p\colon \mathbb{G} \to G$. The vertex set of G will be denoted V, and the union of the junctions of $\mathbb{G}$ will be denoted $\mathbb{V}$: thus $\mathbb{V} = p^{-1}(V)$. The components of $\mathbb{G} \setminus \mathbb{V}$ are called *strips*: each strip is therefore homeomorphic to $(0, 1) \times [0, 1]$.

Remarks:

a) A should be thought of as a set of punctures of S: all homeomorphisms considered will restrict to homeomorphisms of A, and the isotopies constructed later will be relative to this set. It is assumed for the sake of simplicity that S is a closed surface, but there is no difficulty in extending the theory to surfaces

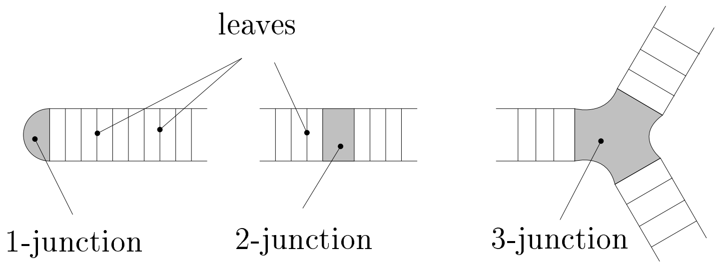

Figure 1: Charts in a thick graph

with boundary: in this case, each boundary component should be treated like a point of A. When A is empty or is not explicitly relevant (as is often the case in sections 2 and 3) it will usually be suppressed, and the thick graph referred to as $(S, \mathbb{G})$.

b) There are two cyclic orders on the set of strips incident to a given junction J, given by the order in which they are encountered as one moves around $\partial_S J$ in a chosen direction (strictly, these orders are on the set of initial segments of strips at the junction, since a strip may begin and end at the same junction). A subset U of the strips at J is said to be *connected in the cyclic order at J* if the strips of U are consecutive in one (and hence both) of these orders. A set of edges incident to a vertex v of G is said to be *connected in the cyclic order at v* if the corresponding set of strips is connected in the cyclic order at $p^{-1}(v)$.

c) Where there is no potential ambiguity, and where it is convenient to do so, $\mathbb{G}$ itself will be referred to as a thick graph.

d) In order to capture all the complexity of families of maps like the Hénon family it is necessary to allow a thick graph to have infinitely many strips (i.e., the graph G to have infinitely many edges). Otherwise, the discussion would be restricted to maps with Markov partitions. As far as this paper is concerned, we only need to consider the case where there are infinitely many 2-junctions: a discussion of the situation in which this becomes necessary can be found in Section 5. On a first reading, it is probably helpful to concentrate on the case in which $\mathbb{G}$ has finitely many junctions.

If $(S, \mathbb{G}, A)$ is a thick graph and $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ is a homeomorphism (i.e. a homeomorphism $S \to S$ with $F(\mathbb{G}) \subseteq \mathbb{G}$ and $F(A) \subseteq A$) under which the image of each decomposition element of $\mathbb{G}$ is contained in a decomposition element, then $F|_{\mathbb{G}}$ induces a unique graph endomorphism $f\colon G \to G$ such that $p \circ F|_{\mathbb{G}} = f \circ p$.

Definitions: A *thick graph map of* $(S, \mathbb{G}, A)$ is an orientation-preserving homeomorphism $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ such that:

a) $F(\mathbb{G}) \subset \operatorname{Int} \mathbb{G}$.

b) If γ is a decomposition element of $\mathbb{G}$, then $F(\gamma)$ is contained in a decomposition element, and $\operatorname{diam}(F^n(\gamma)) \to 0$ as $n \to \infty$. If γ is a junction, then $F(\gamma)$ is disjoint from the boundary of the junction which contains it.

c) The induced graph endomorphism $f\colon G \to G$ is piecewise monotone (that is, there is a finite subset L of V such that $f^{-1}(x) \cap U$ is connected for each $x \in G$ and each component U of $G \setminus L$); and is strictly monotone away from the preimages of vertices (that is, every $x \in G \setminus f^{-1}(V)$ has a neighbourhood on which f restricts to an embedding).

d) $F(A) = A$ (i.e. $F|_A\colon A \to A$ is a homeomorphism).

e) For each component U of $S \setminus \mathbb{G}$ there is some (least) integer $n_U > 0$ such that $F^{n_U}(U) \cap U \neq \emptyset$. Moreover, U contains a period n_U point p_U of F, which is a source whose immediate basin contains U: that is, $F^{-kn_U}(x) \to p_U$ as $k \to \infty$ for all $x \in U$.

A homeomorphism $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ which satisfies conditions a) to d) is called a *pre-thick graph map of* $(S, \mathbb{G}, A)$.

Remarks:

a) The dynamics of a thick graph map in $S \setminus \mathbb{G}$ is easily understood and uninteresting. In the case of finitely many junctions, the dynamics in $\mathbb{G}$ can be understood by standard Markov partition techniques.

b) The reason for defining pre-thick graph maps is that condition e) is lost at intermediate stages of some constructions. The full thick graph map structure is always recovered at the end.

c) If $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ is a thick graph map, then so are all of its forward iterates F^n ($n \geq 1$), and the graph endomorphism induced by F^n is $f^n\colon G \to G$.

In Sections 2.3 and 2.4, a method for isotoping away the dynamics of a thick graph map in a controlled manner is described. The isotopies which are used to accomplish this are called *prunings*: all of the dynamics in the interior of a given closed disk (and hence in the F-orbit of the interior) is destroyed by an isotopy which is supported in the union of the disk and an arbitrarily small region which lies outside of $\mathbb{G}$. It is also necessary to perform some 'adjusting' operations, which have no effect on the dynamics of F. There are two types of adjusting operations, namely *thickening* a decomposition element and *splitting* a junction, which are described below. Thickening involves small isotopies of F (and, in the case where the decomposition element being thickened is a junction, does not actually change the decomposition element), while splitting modifies $\mathbb{G}$ and its decomposition elements but leaves F unchanged.

Definition: Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a pre-thick graph map, and let γ be a decomposition element of $\mathbb{G}$. The following procedure is called *thickening* γ:

If γ is a junction, then postcompose F with a homeomorphism h supported in a small neighbourhood N in $\mathbb{G}$ of $\partial_{\mathbb{G}}\gamma$ with $N \cap A = \emptyset$, which satisfies $h(\partial_{\mathbb{G}}\gamma) \subset \text{Int}\,(\gamma)$, and which maps each decomposition element into a decomposition element.

If γ is a leaf then let $\gamma_0 = \gamma$, and for $i \geq 1$ let γ_i be the decomposition element which contains $F^i(\gamma)$. If γ_I is a junction for some (least) $I \geq 1$, then thicken each of the leaves $\gamma_0, \ldots, \gamma_{I-1}$ into 2-junctions (in such a way that any points of A, which before lay in some γ_i, now lie in the interiors of the corresponding junctions), and extend F over each junction so that $F(\gamma_i) \subset \text{Int}_S(\gamma_{i+1})$.

If no γ_I is a junction and there are only finitely many distinct leaves γ_i, then let I be least such that there is some (least) $J > 0$ with $\gamma_I = \gamma_{I+J}$. Thicken up each of the leaves $\gamma_0, \ldots, \gamma_{I+J-1}$ into 2-junctions, and extend F as in the previous case.

If the γ_i are all distinct leaves, then thicken up each of them with decreasing thicknesses tending to zero (that is, if the 2-junctions are regarded as rectangles, then the sides of the rectangles which lie on $\partial\mathbb{G}$ are chosen to be shorter and shorter with lengths whose sum converges), in such a way that F can be extended with $F(\gamma_i) \subset \text{Int}_S(\gamma_{i+1})$ for each i. Notice that in this case $\text{Int}\,(\gamma_0)$ becomes a wandering domain. (This is analogous to the construction of a Denjoy circle homeomorphism.)

In each of the cases where γ is a leaf, complete the process by postcomposing with a homeomorphism h as in the case where γ is a junction.

Thickening a decomposition element is a dynamically trivial operation which does not create or destroy any dynamics. For convenience of notation, a pre-thick graph map which is obtained from $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ by thickening a finite number of decomposition elements will usually also be denoted $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$. Observe that it is isotopic to the original pre-thick graph map relative to A.

There are two ways of splitting a junction J of $\mathbb{G}$. In each case, it is assumed that J is an isolated n-junction of $\mathbb{G}$, and that no strip of $\mathbb{G}$ has image contained entirely in J. Let Σ be the union of J and the strips which are adjacent to it.

Type a) splitting: Suppose that the oriented strips emanating from J can be partitioned into two non-empty sets, $U_1 = \{s_1, \ldots, s_k\}$ and $U_2 = \{s_{k+1}, \ldots, s_n\}$, each connected in the cyclic order at J, with the property that each component of $F(\mathbb{G}) \cap \Sigma$ is contained either in $J \cup \bigcup U_1$ or in $J \cup \bigcup U_2$ (note that it is possible for some s_i to be the same strip as s_j ($j \neq i$) with the opposite orientation). Let $\mathbb{G}'$ be the thick graph obtained by choosing two disjoint arcs α_1 and α_2 in J with endpoints in $\partial\mathbb{G}$ which separate the two types of components of $F(\mathbb{G}) \cap \Sigma$, and excising the subset of J between α_1 and α_2. Then $\mathbb{G}'$ inherits the obvious thick graph structure from $\mathbb{G}$ (in which there are now two junctions J_1 and J_2 corresponding to J), and

$F\colon (S, \mathbb{G}', A) \to (S, \mathbb{G}', A)$ is a pre-thick graph map (Figure 2). Notice that $A \subseteq F(\mathbb{G})$ (since $F(A) = A$), and hence is disjoint from the excised subset.

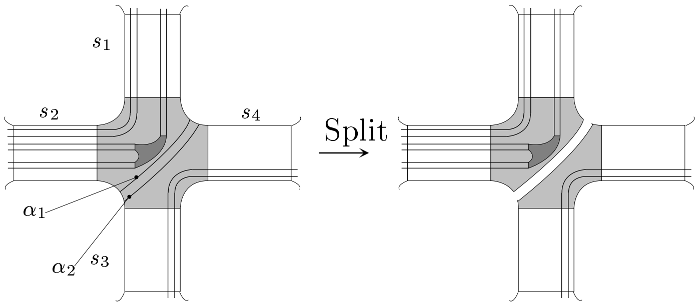

Figure 2: Type a) splitting

Type b) splitting: Let s be one of the oriented strips emanating from J whose other end is adjacent to a junction K (possibly equal to J), and suppose that the other strips emanating from J can be partitioned into two non-empty sets $U_1 = \{s_1, \ldots, s_k\}$ and $U_2 = \{s_{k+1}, \ldots, s_{n-1}\}$, each connected in the cyclic order of strips around J, with the property that each component of $F(\mathbb{G}) \cap \Sigma$ is contained either in $J \cup s \cup \bigcup U_1$ or in $J \cup s \cup \bigcup U_2$. Let α_1 and α_2 be disjoint arcs in $J \cup \bar{s}$, each having one endpoint in K and one in $J \cap \partial\mathbb{G}$, which separate the two types of components of $F(\mathbb{G}) \cap \Sigma$. Let $\mathbb{G}'$ be obtained by excising the subset of $J \cup s$ between α_1 and α_2. Then $\mathbb{G}'$ inherits the obvious thick graph structure from $\mathbb{G}$ (in which J has been split into two junctions J_1 and J_2, and the number of strips adjacent to K has increased by one if $K \neq J$), and $F\colon (S, \mathbb{G}', A) \to (S, \mathbb{G}', A)$ is a pre-thick graph map (Figure 3). Notice that $A \subseteq F(\mathbb{G})$ is disjoint from the excised subset.

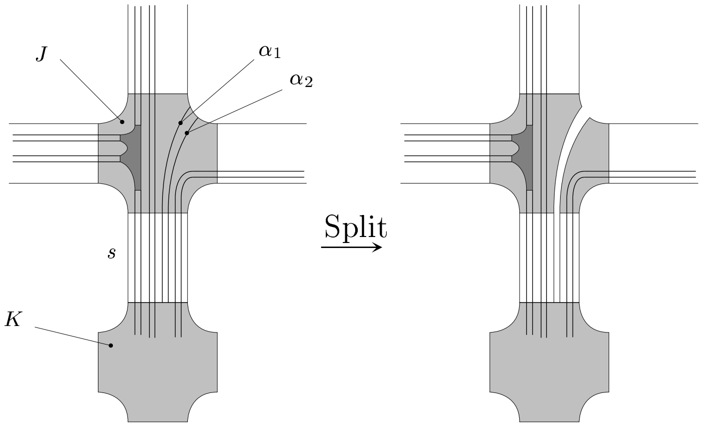

Figure 3: Type b) splitting

2.3 1-Pruning Disks

Definitions: Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a pre-thick graph map. A 1-*pruning disk for* F is a closed topological disk $D \subset S$, whose boundary is the union of two arcs C and E which intersect only at their endpoints, satisfying:

a) $F(D) \subset \operatorname{Int} \mathbb{G}$,

b) $E \subset \partial\mathbb{G}$.

c) D and $\mathbb{G}$ are *on the same side of* E: that is, there is a neighbourhood N of E in S such that $D \cap N \subseteq \mathbb{G}$.

d) $F(C)$ is contained in a single decomposition element, denoted γ_D.

e) $A \cap \operatorname{Int}(D) = \emptyset$.

A 1-pruning disk D is *non-trivial* if $D \cap \Omega(F) \neq \emptyset$ and $F(D)$ intersects more than one decomposition element of $\mathbb{G}$ (and hence intersects uncountably many decomposition elements). The arcs C and E will be referred to as the C-*side* and E-*side* of D. When D is a 1-pruning disk, the symbols C and E will always be taken to denote its C-side and E-side.

If D is a 1-pruning disk for F, then a D-*arc* is an arc α in $F^{-1}(\mathbb{G})$ which has the same endpoints as E and which is disjoint from $\mathbb{G}$ away from its endpoints. An *enlargement of* D is a closed disk Δ whose boundary is the union of C and a D-arc.

Lemma 2.1 *A* 1*-pruning disk* D *for* $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ *contains all of the decomposition elements of* $\mathbb{G}$ *which it intersects, with the possible exception of those whose image is contained in* γ_D *if* γ_D *is a junction.*

Proof. Since the decomposition elements are connected, and since $E \subseteq \partial\mathbb{G}$, any decomposition element γ which intersects D but is not contained in D must intersect C. Since $F(C) \subseteq \gamma_D$, it follows that $F(\gamma) \cap \gamma_D \neq \emptyset$, and hence $F(\gamma) \subseteq \gamma_D$. If γ_D is not a junction, then the local strict monotonicity of $f\colon G \to G$ implies that $\gamma \subseteq C$. ∎

Given a 1-pruning disk D, it is possible to modify F by a preisotopy which only changes the dynamics in $\bigcup_{n=-\infty}^{\infty} F^n(\operatorname{Int}(D))$, in such a way as to make every point of $\operatorname{Int}(D)$ wandering, with the possible exception of a single periodic point.

Lemma 2.2 *Let* $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ *be a pre-thick graph map,* D *be a* 1*-pruning disk for* F*, and* Δ *be an enlargement of* D*. Then there exists a pre-thick graph map* $F_D\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$*, which is obtained from* F *by thickening the decomposition element* γ_D *followed by an isotopy supported in* Δ*, such that* $\Omega(F_D) \cap D$ *is either empty or is a single periodic point.*

Proof. First thicken γ_D: this ensures that $F(C) \subseteq \text{Int}\,(\gamma_D)$. Let $M\colon S \to S$ be a homeomorphism supported in Δ (and hence isotopic to the identity relative to A) with the property that $M(D)$ is contained in a small enough neighbourhood of C that $F(M(D)) \subseteq \gamma_D$, and let $F_D = F \circ M$. Since F has only been modified on decomposition elements which intersect Int (D), and Lemma 2.1 guarantees that all such decomposition elements are mapped into γ_D by F_D, it follows that $F_D\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ is still a pre-thick graph map, provided that the diameter of $F_D^n(\gamma_D)$ tends to zero as $n \to \infty$. This may not be the case a priori if some image $F_D^k(\gamma_D)$ is contained in D: however, it can be ensured by taking the initial thickening of γ_D to involve postcomposition by a homeomorphism h with $h(\gamma_D)$ sufficiently small.

Either $F_D^n(\gamma_D) \subseteq \gamma_D$ for some n, in which case γ_D contains a periodic point of F_D whose orbit attracts every point of D; or $F_D^n(\gamma_D)$ is disjoint from γ_D for all $n > 0$, in which case every point of D is wandering under F_D. ∎

Definition: F_D is said to have been obtained from F by *pruning D away.*

Remarks:

a) F_D is isotopic to F relative to A.

b) D is still a pruning disk for F_D, but it is trivial since $F_D(D)$ is contained in the decomposition element γ_D.

c) If F is a thick graph map, rather than a pre-thick graph map, then so is F_D.

Example 1 One of the simplest interesting examples of a thick graph map is Smale's horseshoe diffeomorphism [19]. Good detailed treatments of the horseshoe can be found in many texts on dynamical systems (e.g. [9, 18]). In this example $S = S^2 = \mathbb{R}^2 \cup \{\infty\}$ and $\mathbb{G}$ consists of a single strip $Q = (0,1) \times [0,1]$, foliated by leaves $\{x\} \times [0,1]$, together with two 1-junctions J_0 and J_1, which are semicircles of diameter 1 centred on $(0, 1/2)$ and $(1, 1/2)$ respectively. The action of $F|_{\mathbb{G}}$ is to stretch Q uniformly in the horizontal direction by a factor λ (which must be greater than 2), to contract Q uniformly in the vertical direction by a factor $1/\lambda$, to bend the deformed Q into the shape of a horseshoe, and to map it into $\mathbb{G}$ as shown in Figure 4, in such a way that not only is each decomposition element mapped into a decomposition element, but also the foliation of Q by horizontal lines $(0,1) \times \{y\}$ is preserved. Finally, the two 1-junctions J_0 and J_1 are contracted uniformly by a factor $1/\lambda$. Outside of $\mathbb{G}$, F has a fixed point at ∞ with $F^{-n}(x) \to \infty$ as $n \to \infty$ for all $x \in S^2 \setminus \mathbb{G}$.

The non-wandering set $\Omega(F)$ of F is the union of ∞, a fixed point in J_0, and a Cantor set

$$\Lambda = \{x \in Q\colon F^n(x) \in Q \text{ for all } n \in \mathbb{Z}\}.$$

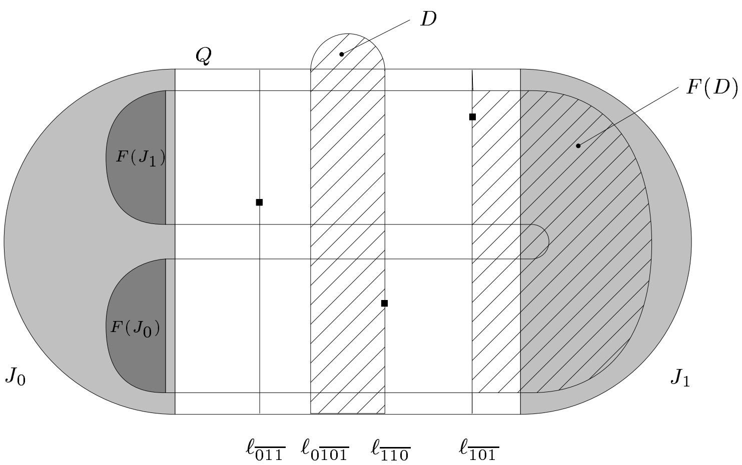

Figure 4: A 1-pruning disk in the horseshoe

The dynamics of F on Λ can be understood symbolically: let $p\colon Q \to (0,1)$ be projection onto the first coordinate, and define a map $h\colon \Lambda \to \Sigma = \{0,1\}^{\mathbb{Z}}$ by

$$h(x)_i = \begin{cases} 0 & \text{if } \ p(F^i(x)) < 1/2 \\ 1 & \text{if } \ p(F^i(x)) > 1/2. \end{cases}$$

Then h is a homeomorphism, and conjugates the action of F on Λ to that of the *shift map* $\sigma\colon \Sigma \to \Sigma$, defined by $\sigma(s)_i = s_{i+1}$: that is, $F|_\Lambda = h^{-1} \circ \sigma \circ h$. Thus every point x of Λ has a *code* $h(x) \in \Sigma$: it is often convenient to refer to x by its code, without explicitly invoking the homeomorphism h. In particular, periodic points of F correspond one-to-one with periodic codes of the same period. When writing elements $s \in \Sigma$, it is common to juxtapose a point between s_{-1} and s_0 to indicate the origin of the sequence, and to use an overbar to denote infinite repetition of a given word either at the beginning or the end of s. When s is periodic, and hence can be written in the form $\overline{w}$ for some word w, then the origin of the sequence is taken to be at the first letter of w, i.e. $\overline{w} = \ldots www \cdot www \ldots$.

The horizontal and vertical ordering of points of Λ can be determined using the *unimodal order* $\preceq$ on $\Sigma_+ = \{0,1\}^{\mathbb{N}}$, which is defined as follows: if $s = s_0 s_1 \ldots$ and $t = t_0 t_1 \ldots$ are distinct elements of Σ_+, and $k \in \mathbb{N}$ is least such that $s_k \neq t_k$, then $s \prec t$ if and only if $\sum_{i=0}^{k} s_i$ is even. If $s, t \in \Sigma_+$, then $s \preceq t$ if and only if either $s = t$ or $s \prec t$. Now if $(x_1, y_1), (x_2, y_2) \in \Lambda$ with $h(x_1, y_1) = \ldots s_{-2} s_{-1} \cdot s_0 s_1 \ldots$ and $h(x_2, y_2) = \ldots t_{-2} t_{-1} \cdot t_0 t_1 \ldots$, then

$$\begin{aligned} x_1 < x_2 \quad &\iff \quad s_0 s_1 s_2 \ldots \prec t_0 t_1 t_2 \ldots \ \text{ and} \\ y_1 < y_2 \quad &\iff \quad s_{-1} s_{-2} \ldots \prec t_{-1} t_{-2} \ldots \ . \end{aligned}$$

In particular, leaves which contain elements of Λ can be specified by an element of Σ_+: the leaf whose horizontal coordinate is given by $s \in \Sigma_+$ will be written ℓ_s.

Let $\ell_{\overline{101}}$, $\ell_{\overline{011}}$, and $\ell_{\overline{110}}$ be the leaves containing the three points $\overline{101}$, $\overline{011}$, and $\overline{110}$, which constitute a period 3 orbit of F, and consider the disk D whose E side is an arc along the lower boundary of Q joining $\ell_{0\overline{101}}$ to $\ell_{\overline{110}}$, and whose C side is contained in $F^{-1}(\ell_{\overline{101}})$ (see Figure 4). This is clearly a 1-pruning disk for $F\colon (S^2, \mathbb{G}) \to (S^2, \mathbb{G})$, with $F(C) \subseteq \gamma_D = \ell_{\overline{101}}$. The first step in pruning D away is to thicken γ_D, as shown in Figure 5, in order to make room for the pruning isotopy. Since $F^3(\gamma_D) \subseteq \gamma_D$, this requires thickening each of the leaves $\ell_{\overline{101}}$, $\ell_{\overline{011}}$, and $\ell_{\overline{110}}$ into 2-junctions: these 2-junctions will be denoted γ_{101}, γ_{011}, and γ_{110} respectively. Let α be a D-arc yielding an enlargement Δ of D, and modify F by a preisotopy supported in Δ to the pruned map F_D which satisfies $F_D(D) \subseteq \gamma_D$, as shown in Figure 6 (recall that $\gamma_D = \gamma_{101}$). Observe that $\Omega(F_D) \cap D$ consists of a single period 3 point.

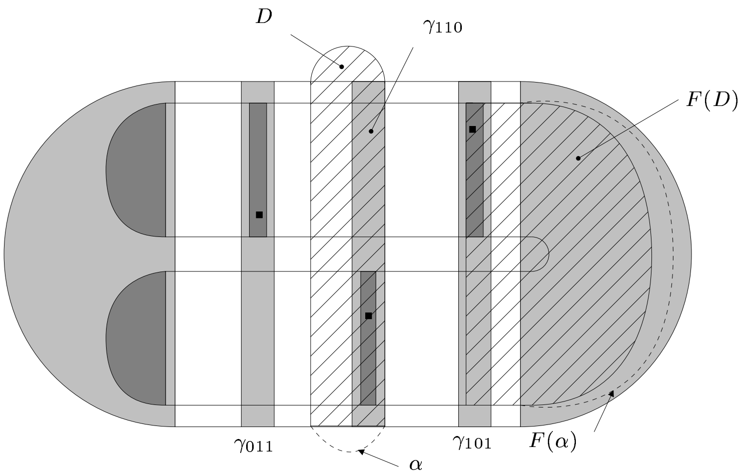

Figure 5: Thickening γ_D

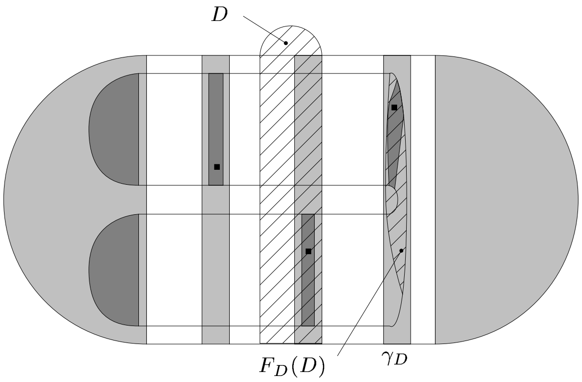

Figure 6: The pruned homeomorphism F_D

On the symbolic level, the dynamics of F_D is described by the subshift of $\sigma\colon \Sigma \to \Sigma$ in which all semi-infinite words $w \succ \overline{101}$ are excluded.

Remark: A similar pruning could be carried out up to any leaf in the horseshoe. By the standard theory of unimodal maps (compare with Example 4 below), it follows that homeomorphisms with uncountably many distinct topological entropies can be obtained from the horseshoe by 1-prunings.

2.4 n-Pruning Disks

Pruning away a 1-pruning disk is quite a simple operation: a more interesting situation occurs when D is a 1-pruning disk for an iterate F^n of F. Under an additional assumption, the *pruning condition*, it will be shown that the thick graph $\mathbb{G}$ can be modified so as to turn such a disk D into a 1-pruning disk for F^{n-1}. By repeating this process $n-1$ times, D becomes a 1-pruning disk for F, which can then be pruned away.

Definitions: Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a pre-thick graph map, and suppose that D is a 1-pruning disk for F^n where $n > 1$. Then D is said to be an *n-pruning disk for F* if the *pruning condition* is satisfied: that $F^k(D)$ does not intersect E away from its endpoints for $1 \leq k < n$. An n-pruning disk for F is *non-trivial* if it is a non-trivial 1-pruning disk for F^n.

From now on, all pruning disks will be assumed non-trivial unless otherwise stated. The main theorem of this section is the following:

Theorem 2.3 *Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a thick graph map, D be an n-pruning disk for F, and Δ be an enlargement of D. Then there exists a thick graph $(S, \mathbb{G}_D, A)$ and a thick graph map $F_D\colon (S, \mathbb{G}_D, A) \to (S, \mathbb{G}_D, A)$, which is obtained from F by thickening a finite number of decomposition elements, performing an isotopy supported in Δ, and carrying out some type a) splittings, such that $\Omega(F_D) \cap D$ is either empty or is a single periodic point.*

Proof. Let γ_D be the decomposition element of $\mathbb{G}$ which contains $F^n(C)$. The construction starts by thickening γ_D (so in particular it is a junction).

Since $F^n(D)$ is contained in the interior of $\mathbb{G}$, it can be decomposed into disjoint arcs and disks which are the components of the intersections of decomposition elements of $\mathbb{G}$ with $F^n(D)$. This decomposition of $F^n(D)$ can then be pulled back to a decomposition of $F^{n-1}(D)$. Let $\mathbb{G}_1 = \mathbb{G} \cup F^{n-1}(D)$, and define a decomposition of $\mathbb{G}_1$ by *amalgamating* the decompositions of $\mathbb{G}$ and $F^{n-1}(D)$: that is, the decomposition element containing a point p of $\mathbb{G}_1$ is the smallest union of decomposition elements of $\mathbb{G}$ and $F^{n-1}(D)$ which contains p and all of the decomposition elements (of either surface) which it intersects. Because γ_D is a junction, the decomposition element containing $F^{n-1}(C)$ is a disk, and hence this decomposition gives $(S, \mathbb{G}_1, A)$ the structure of a thick graph (see Figure 10 for an example). Moreover, decomposition elements are mapped into decomposition elements by construction, and so $F: (S, \mathbb{G}_1, A) \to (S, \mathbb{G}_1, A)$ is a pre-thick graph map (the circumstances in which $F: (S, \mathbb{G}_1, A) \to (S, \mathbb{G}_1, A)$ may not be a thick graph map are discussed below). Finally, thicken the junction which contains $F^{n-1}(C)$.

The purpose of this construction is that D is now an $(n-1)$-pruning disk for $F: (S, \mathbb{G}_1, A) \to (S, \mathbb{G}_1, A)$. To see this, observe that:

a) $F^{n-1}(D)$ doesn't intersect E away from its endpoints by the pruning condition, so $E \subseteq \partial \mathbb{G}_1$.

b) $F^{n-1}(D) \subseteq \mathrm{Int}\,(\mathbb{G}_1)$, since $F^{n-1}(E) \subseteq \mathrm{Int}\,(\mathbb{G})$, and the final step of the construction was to thicken the junction which contains $F^{n-1}(C)$.

c) $F^{n-1}(C)$ is contained in this single junction.

d) The pruning condition holds since D is unchanged and F has only been modified by thickenings.

Let $F: (S, \mathbb{G}_{n-1}, A) \to (S, \mathbb{G}_{n-1}, A)$ be the pre-thick graph map obtained by repeating the above procedure $n-1$ times. Then D is a 1-pruning disk for F, which has been obtained from the original thick graph map by thickening a finite number of decomposition elements. Let $F_D: (S, \mathbb{G}_{n-1}, A) \to (S, \mathbb{G}_{n-1}, A)$ be the pre-thick graph map obtained by pruning D away.

F_D will not be a thick graph map if some iterates $F^k(D)$ $(1 \le k < n)$ of D intersect $\mathbb{G}$ in a disconnected set (notice that this implies that D must also intersect $\mathbb{G}$ in a disconnected set). In this case the above construction alters the topology of $\mathbb{G}$, creating additional components of $S \setminus \mathbb{G}$. However, it is possible to split the thick graph $\mathbb{G}_{n-1}$ to obtain a thick graph $\mathbb{G}_D$ with the same topology as $\mathbb{G}$, so that $F_D: (S, \mathbb{G}_D, A) \to (S, \mathbb{G}_D, A)$ is a thick graph map. Figure 7 illustrates this process in a simple example, where D is a 2-pruning disk which intersects $\mathbb{G}$ in two components. It is assumed that these two components are mapped by F into strips s and t. Following the arrows around Figure 7, the first diagram shows D intersecting two strips of $\mathbb{G}$, and the second shows the image $F(D)$, intersecting s and t, and the images of the two strips which D intersects: it is assumed that the image of some other strip also passes through t but is disjoint from $F(D)$. The induced decomposition of $F(D)$ is also marked on this diagram: the back-hatched regions and the dotted lines indicate how this decomposition will be amalgamated with that of $\mathbb{G}$. The result of this amalgamation is shown in the third diagram: two new junctions have been created, J (which contains $F(C)$), and J_1, which is connected to J by a strip u which is foliated by the leaves of $F(D)$ amalgamated with those of t. The strips s and t are each divided by J into two new strips: the new strips which come from s are denoted s_1 and s_2, and those from t are denoted t_1 and t_2.

Let $\Sigma = J \cup s_1 \cup s_2 \cup t_1 \cup t_2 \cup u$. It is clear that all of the components of $F(\mathbb{G}) \cap s$ correspond to components of $F(\mathbb{G}_1) \cap (\Sigma \cup J_1)$ which are disjoint from $t_1 \cup t_2$; and

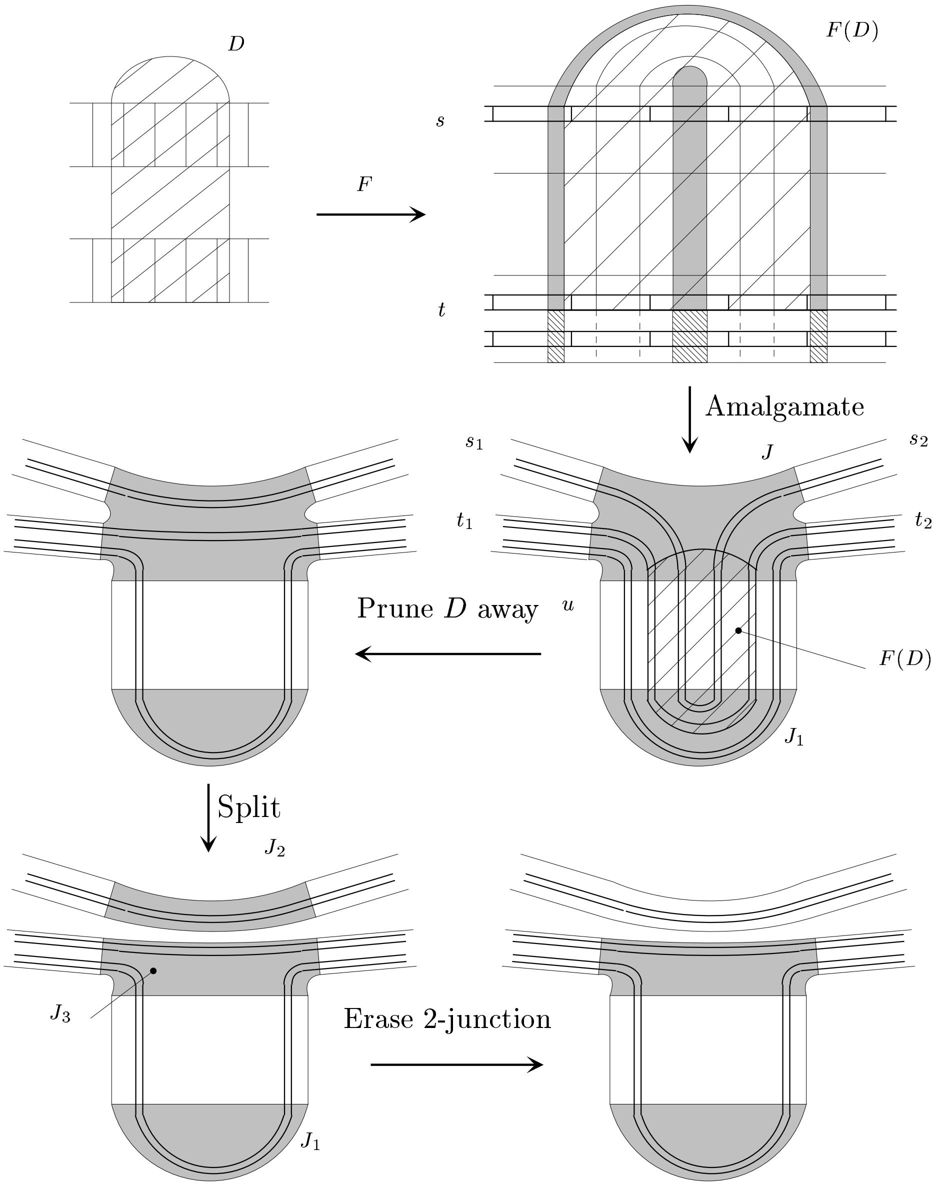

Figure 7: Splitting to restore a thick graph map

similarly, the components of $F(\mathbb{G}) \cap t$ correspond to components of $F(\mathbb{G}_1) \cap (\Sigma \cup J_1)$ which are disjoint from $s_1 \cup s_2$. After pruning D away (which only affects the preimage of $F(D)$), the components of $F(\mathbb{G}) \cap s$ correspond to components of $F_D(\mathbb{G}_1) \cap \Sigma$ which are contained in $s_1 \cup s_2 \cup J$, while the components of $F(\mathbb{G}) \cap t$ correspond to components of $F_D(\mathbb{G}_1) \cap \Sigma$ which are contained in $t_1 \cup t_2 \cup u \cup J$ (see the fourth diagram of Figure 7). Thus J can be split as shown in the fifth diagram. For an n-pruning, this procedure must be carried out successively for $F(D), F^2(D), \ldots, F^{n-1}(D)$.

The topology of the resulting thick graph $\mathbb{G}_D$ is the same as that of $\mathbb{G}$, and condition e) in the definition of a thick graph map can be achieved by composing with an isotopy supported in $S \setminus \mathbb{G}$ to ensure that each complementary component of $\mathbb{G}$ is a subset of the immediate basin of the single periodic point which it contains. ∎

Definition: It is said that $F_D\colon (S, \mathbb{G}_D, A) \to (S, \mathbb{G}_D, A)$ has been obtained from $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ by *pruning D away.*

Remark: In the example of Figure 7, s and t are strips (containing no junctions). After pruning and splitting, three junctions were created: the 1-junction J_1, the 3-junction J_3, and the 2-junction J_2. Since s was a strip, the 2-junction J_2 between s_1 and s_2 cannot carry any backtracking or contain the image of a junction. Thus it can be decomposed into leaves, as shown in the sixth diagram of Figure 7, joining the two strips s_1 and s_2 back into a single strip s.

It is easy to see that this is true in general: if $F^k(D)$ intersects $\mathbb{G}$ in j components, and J is the junction which arises from thickening $F^k(C)$, then, after pruning D away, J splits into j junctions. If one of these junctions corresponds to a strip in $\mathbb{G}$, then it is a 2-junction which can be decomposed into leaves, restoring the original strip of $\mathbb{G}$. The significance of this observation will become apparent in Section 4.3: the important point to be noticed is that no new 2-junctions are created.

The family of all prunings of a thick graph map is now defined. By the remark at the end of the previous subsection, it typically contains uncountably many distinct dynamical systems (see also Section 3). It will be shown later (Section 4) that it also contains the family of *Thurston maps* in the isotopy class of F relative to finite F-invariant sets. The dynamics of each map in the family can be seen as the dynamics of F less that which has been pruned away.

Definition: Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a thick graph map. Define the *pruning family* $\mathcal{P}(F)$ of F to be the set of all homeomorphisms which can be obtained from F by a finite sequence, or as the limit of an infinite sequence, of prunings, cleanings[2], and thickenings.

[2]A *cleaning* consists of a possibly infinite convergent sequence of prunings. A description is given in Section 5.

Example 2 Let $F\colon (S^2, \mathbb{G}) \to (S^2, \mathbb{G})$ be the horseshoe map. $\mathbb{G}$ and its image under F^2 are shown in Figure 8, together with a 2-pruning disk D and its images under F and F^2. The marked point which lies on C, $F(C)$, and $F^2(C)$ is the fixed point $\overline{1}$. It is clear that D is a 1-pruning disk for F^2 and, since $F(D)$ only intersects E at an endpoint, the pruning condition is satisfied and hence D is a 2-pruning disk for F.

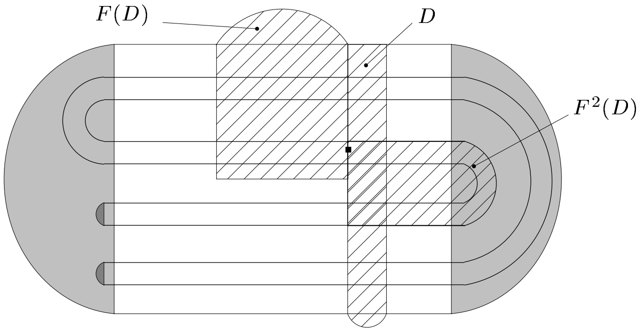

Figure 8: A 2-pruning disk in the horseshoe

The first step in pruning D away is to thicken the leaf $\gamma_D = \ell_{\overline{1}}$, to pull back the decomposition of $F^2(D)$ to $F(D)$ (Figure 9), and then to amalgamate the decompositions of $\mathbb{G}$ and $F(D)$ to yield a new thick graph $\mathbb{G}_1$ (Figure 10). The dotted lines in Figure 9 denote the boundaries of the 2-junction which is obtained by thickening $\ell_{\overline{1}}$. The disk D is a 1-pruning disk for $F\colon (S^2, \mathbb{G}_1) \to (S^2, \mathbb{G}_1)$, which can be pruned away to give the pruned map $F_D\colon (S^2, \mathbb{G}_D) \to (S^2, \mathbb{G}_D)$ (see Figure 11, in which $\mathbb{G}_1$ has been drawn in a different configuration).

On the symbolic level, the dynamics of F_D is described by the subshift of finite type of $\sigma\colon \Sigma \to \Sigma$ in which the words 1010 and 1110 are prohibited (it can easily be seen that the points $x \in \Lambda$ whose orbits intersect Int (D) are precisely those for which $h(x)$ contains one of these words).

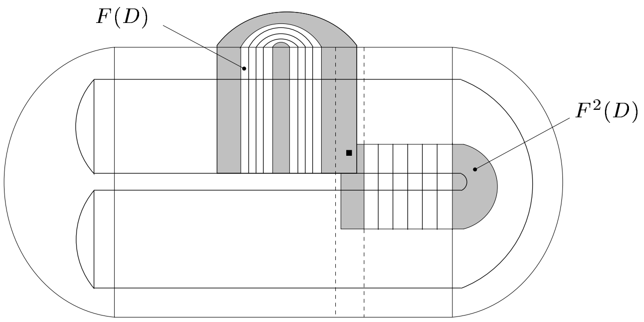

Figure 9: The induced decomposition of $F(D)$

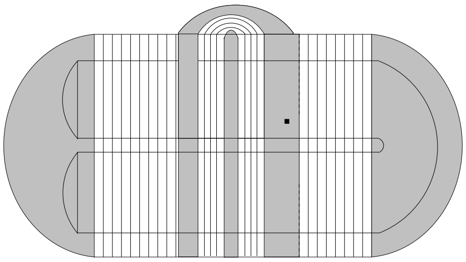

Figure 10: Amalgamating $\mathbb{G}$ and $F(D)$

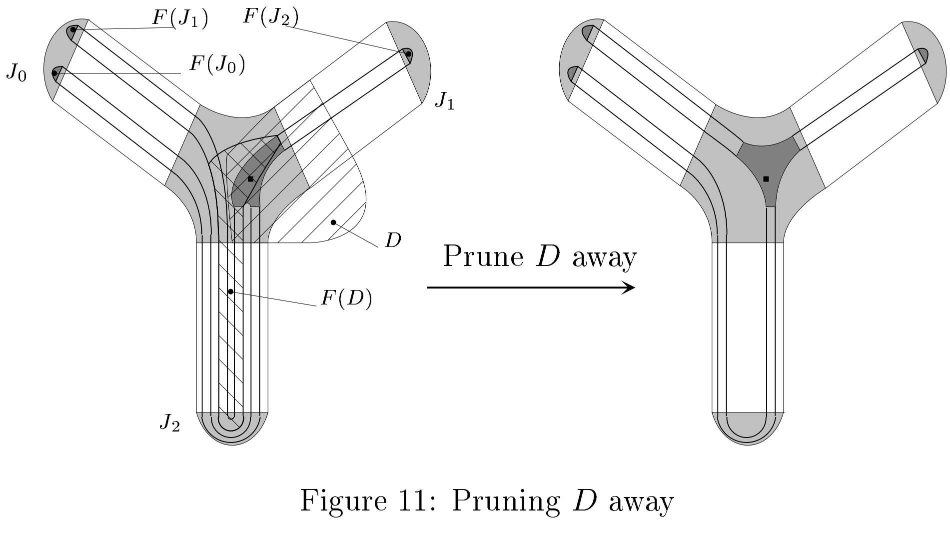

Figure 11: Pruning D away

The fact that $F^n(D) \subseteq \operatorname{Int}\mathbb{G}$ makes it possible to define decompositions on the disks $F^k(D)$ for $0 \le k \le n$: these decompositions will be important in Section 3.

Definition: Let D be an n-pruning disk for $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$. The *induced decomposition* on $F^n(D)$ is defined by taking as decomposition elements the connected components of the intersections of decomposition elements of $\mathbb{G}$ with $F^n(D)$. The *induced decomposition* on $F^k(D)$ for $0 \le k < n$ is defined by pulling back the induced decomposition of $F^n(D)$ using F^{n-k}.

Lemma 2.4 *Suppose that D is an n-pruning disk for $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$. Then for each k with $0 \le k < n$, D contains all of the decomposition elements of $F^k(D)$ which it intersects, with the possible exception of those whose image under F^n is contained in γ_D, if γ_D is a junction.*

Proof. Suppose conversely that γ is a decomposition element of $F^k(D)$ which intersects D but is not contained in D. Since γ is connected, and by the pruning condition doesn't intersect E away from its endpoints, it follows that γ intersects C, and hence $F^n(\gamma)$ intersects γ_D. Now $F^{n-k}(\gamma)$ is contained in a single decomposition element of $\mathbb{G}$, and hence $F^n(\gamma)$ is contained in γ_D. If γ_D is not a junction, then this is prohibited by the local strict monotonicity of $f\colon G \to G$. ∎

In fact, it is possible to remove the exception clause of this lemma without significant loss of generality, by restricting the n-pruning disk D.

Definition: An n-pruning disk for $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ is said to be *trimmed* if $F^n(C)$ is a connected component of $F^n(D) \cap \gamma_D$.

In particular, D is always trimmed if γ_D is a leaf. If γ_D is a junction, then it is possible to replace D with a finite union of trimmed n-pruning disks D_i such that the dynamics of F_D is the same as that of the homeomorphism $F_{D_1,\ldots,D_k}$ obtained by pruning each of the trimmed disks D_i in turn. To do this, let Σ be the component of $F^n(D) \cap \gamma_D$ which contains $F^n(C)$, and let D_i be the components of $F^{-n}(\overline{F^n(D) \setminus \Sigma})$. Then each D_i is clearly a trimmed n-pruning disk (with C-side a component of $(\overline{F^n(D) \setminus \Sigma}) \setminus (F^n(D) \setminus \Sigma))$, and $\Omega(F) \cap (D \setminus \bigcup D_i)$ is either empty, or consists of a single periodic point whose orbit attracts γ_D. In either case, the non-wandering sets of F_D and $F_{D_1,\ldots,D_k}$ are equal.

If D is a trimmed n-pruning disk, then Lemma 2.4 can be replaced by

Lemma 2.5 *Suppose that D is a trimmed n-pruning disk for $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$. Then for each k with $0 \le k < n$, D contains all of the decomposition elements of $F^k(D)$ which it intersects. Moreover, each such decomposition element is contained either in the interior or in the boundary of D.*

The proof follows that of Lemma 2.4.

3 Kneading Sets

In this section a deformation theory for graph endomorphisms, called *kneading theory*, is developed: it generalizes the topological aspects of Milnor and Thurston's kneading theory for endomorphisms of the interval. The theory is presented in such a way as to draw out the parallels with pruning theory: although it applies to arbitrary piecewise monotone graph endomorphisms, the emphasis is on graph endomorphisms which are induced by thick graph maps. The main results are Theorem 3.11, which describes how n-pruning disks descend to n-kneading sets, and conversely how n-kneading sets which are compatible with the two-dimensional structure can be lifted to n-pruning disks; and Theorem 3.12, which states that pruning away an n-pruning disk is equivalent to *pulling away* the corresponding n-kneading set.

3.1 1-Kneading sets

The results in this subsection are quite simple: its main purpose is to establish notation and introduce some important definitions.

Let G be a (compact) graph, perhaps having infinitely many vertices of valence 2, and $f: G \to G$ be a piecewise monotone graph endomorphism.

Definitions: A 1-*kneading set K for f* is an open subset K of G with finitely many components such that $f|_{\overline{K}}$ factors through a tree, with all the points of ∂K factoring through the same point of the tree: that is, there exist a tree T, piecewise monotone maps $\psi: \overline{K} \twoheadrightarrow T$ and $\varphi: T \to G$, and a point $t \in T$, such that $f|_{\overline{K}} = \varphi \circ \psi$ and $\psi(\partial K) = \{t\}$. (The double-headed arrow in $\psi: \overline{K} \twoheadrightarrow T$ means that ψ is onto). A 1-kneading set K for f is *non-trivial* if $f(\overline{K})$ contains more than one point of $\Omega(f)$.

Whenever K is a 1-kneading set, the symbols T, ψ, φ, and t will be interpreted according to this definition.

Examples 3 Figure 12 shows examples of some simple 1-kneading sets which illustrate various aspects of the definition. In a), a disconnected 1-kneading set $K = a \cup b \cup c \cup d$ is shown on the left. Its image under f is shown to the right: the dotted lines represent the subset of G which contains the image, and the image arcs have been 'spread out' for clarity. The tree T is an interval, and the maps $\psi: \overline{K} \twoheadrightarrow T$ and $\varphi: T \to G$ (which in this case is an embedding) are shown.

The definition of 1-kneading sets is contrived so that they can be made trivial by a homotopy of f supported in the kneading set. In this example it is clear that such a homotopy can be performed by dragging the image of K across the depicted edge of G adjacent to $f(p)$. The components $K_1 = a \cup b \cup c$ and $K_2 = d$ of K are also 1-kneading sets, which can be made trivial independently by appropriate homotopies. However, if $f: G \to G$ is induced by a thick graph map, and the way in which $f(K)$ has been spread out corresponds to the ordering of the corresponding image strips, then the dynamics in K_1 cannot be destroyed by an isotopy of the thick graph map without affecting the dynamics in K_2: this is because K_1 is not *innermost* (see later definition).

Example b) illustrates that φ need not be an embedding. Four points x, y, z, and w in K are marked, with the property that $f(x) = f(y) = f(z) = f(w)$, while $\psi(x) = \psi(w) \neq \psi(y) = \psi(z)$.

Example c) is one in which the tree T cannot be chosen to be an interval.

If K is a 1-kneading set for $f : G \to G$ then a new graph endomorphism $f_K: G_K \to G_K$ can be constructed by defining G_K to be G with new valence two vertices at all points of the orbit of $\varphi(t)$ which are not already vertices, and

$$f_K(x) = \begin{cases} f(x) & \text{if } x \notin \overline{K} \\ \varphi(t) & \text{if } x \in \overline{K}. \end{cases}$$

For notational convenience, G_K will be written simply as G.

Definition: f_K is said to be obtained from f by *pulling K away.*

Example 4 Consider the graph endomorphism $f: G \to G$ induced by the horseshoe map described in Example 1: G is an interval $[0, 1]$ (with valence 1 vertices at 0 and 1), and the graph endomorphism $f: G \to G$ is depicted in Figure 13. The non-wandering set $\Omega(f)$ of f is the union of the fixed point 0 and a Cantor set

$$\Lambda = \{x \in G : f^n(x) \in (0, 1) \text{ for all } n \in \mathbb{N}\},$$

which is the projection of the corresponding Cantor set in the horseshoe. As in that case, the dynamics of $f|_\Lambda$ can be understood symbolically: define $h: \Lambda \to \Sigma_+$ by

$$h(x)_i = \begin{cases} 0 & \text{if} \quad f^i(x) < 1/2 \\ 1 & \text{if} \quad f^i(x) > 1/2. \end{cases}$$

Then h is a homeomorphism, which conjugates $f|_\Lambda$ to the shift map $\sigma: \Sigma_+ \to \Sigma_+$: that is, $f|_\Lambda = h^{-1} \circ \sigma \circ h$. The usual order of points of Λ as a subset of the interval is reflected by the unimodal order on Σ_+: if $x, y \in \Lambda$ then $x < y$ if and only if $h(x) \prec h(y)$.

Let $K = (0\overline{101}, \overline{110})$, $T = [\overline{101}, 1] \subseteq G$, $t = \overline{101}$, $\psi = f|_{\overline{K}}: \overline{K} \twoheadrightarrow T$, and $\varphi: T \to G$ be the inclusion. Then $f|_{\overline{K}} = \varphi \circ \psi$ and $\psi(\partial K) = \{t\}$, so K is a 1-kneading set for f. Pulling K away yields a graph endomorphism $f_K: G \to G$ with $f_K(x) = \overline{101}$ for all $x \in \overline{K}$ (see Figure 14). Notice that, by the definition of pulling K away, the three points $\overline{011}$, $\overline{110}$, and $\overline{101}$ are now valence 2 vertices of G.

Comparing this example with Example 1, observe that $K = \operatorname{Int}(p(D \cap \mathbb{G}))$, and that $f_K: G \to G$ is the graph endomorphism induced by $F_D: (S^2, \mathbb{G}) \to (S^2, \mathbb{G})$. The aim of the following definitions and results is to show that a 1-pruning disk always projects to a 1-kneading set in this way, and to give conditions on 1-kneading sets which imply that they lift to 1-pruning disks.

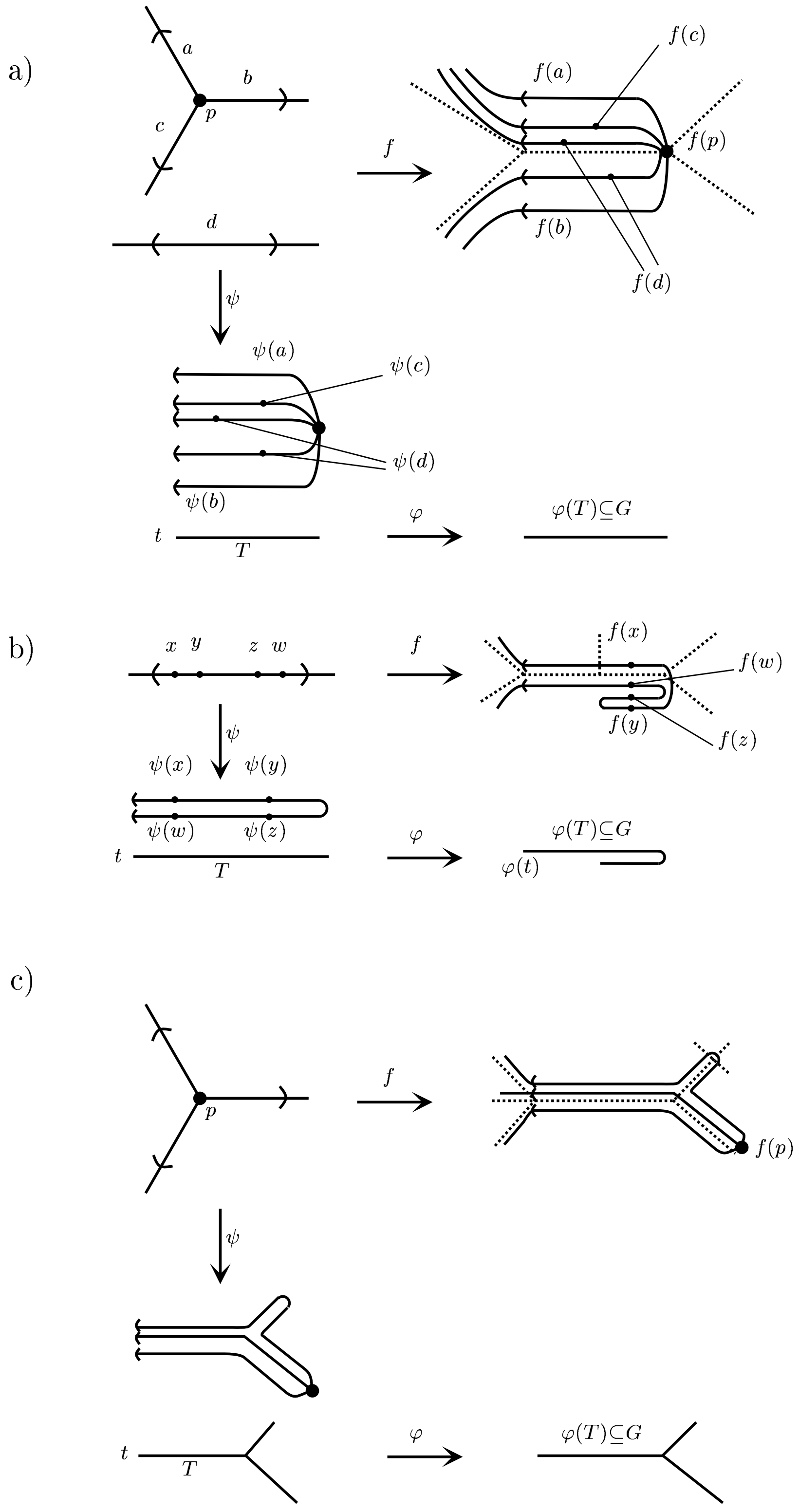

Figure 12: Examples of 1-kneading sets

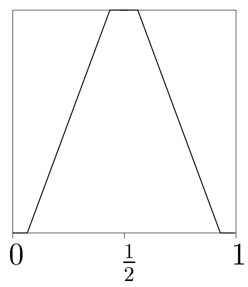

Figure 13: The graph endomorphism induced by the horseshoe

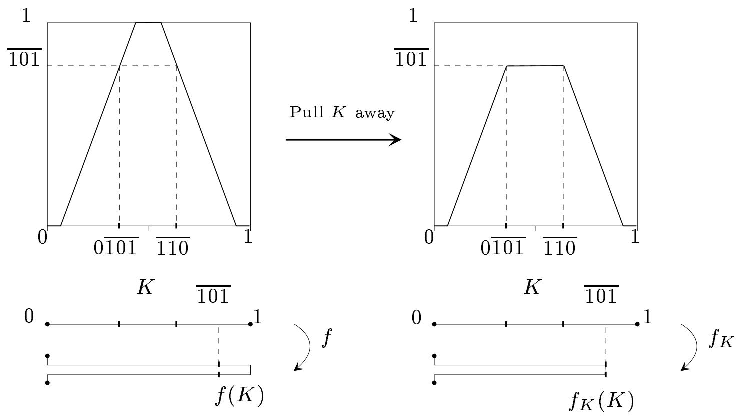

Figure 14: The graph endomorphism obtained by pulling K away

Definition: Let G be a graph, and v be a valence one vertex of G. A *thick graph above G truncated at v* is a thick graph $(S, \mathbb{G}, A)$ with projection $p\colon \mathbb{G} \to G$ in which the 1-junction $p^{-1}(v)$ is replaced by an arc. This arc is considered to be a leaf in the thick graph decomposition.

It will be shown in Theorem 3.1 below that for any trimmed 1-pruning disk D, the projection $\mathrm{Int}\,(p(D \cap \mathbb{G}))$ is a 1-kneading set for the induced graph endomorphism. The converse is not generally true: an additional compatibility condition between the 1- and 2-dimensional structures is required. This is the motivation for the following definition:

Definition: Let $F\colon (S, \mathbb{G}) \to (S, \mathbb{G})$ be a thick graph map with induced graph endomorphism $f\colon G \to G$. A 1-kneading set K for f is *F-compatible* if it is the projection of a 1-pruning disk D for F, i.e. if $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$.

The aims of the following discussion and results are: a), to show that (trimmed) 1-pruning disks and F-compatible 1-kneading sets are in one-to-one correspondence; and b), to derive a sufficient condition for a 1-kneading set to be F-compatible which can easily be checked: this means that, in applications, pruning disks can readily be found by searching for kneading sets which satisfy this condition. The essential compatibility requirement is that the kneading set must correspond to an 'innermost' pruning disk. Specifically, suppose that a pruning disk D is such that $D \cap \mathbb{G}$ has more than one component. Then any union of the corresponding components of the kneading set $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$ may also be a kneading set, but the only such unions which can be projections of pruning disks are those which correspond to collections of innermost components of $F(D \cap \mathbb{G})$ (see Figure 16 for an explanatory example).

Suppose, then, that $F\colon (S, \mathbb{G}) \to (S, \mathbb{G})$ is a thick graph map, and that D is a trimmed 1-pruning disk for F. Then $F(D) \subseteq \mathbb{G}$, and hence there is an induced decomposition on $F(D)$, where the decomposition elements are connected components of the intersections of decomposition elements of $\mathbb{G}$ with $F(D)$. Since $F(D)$ is a disk, this decomposition gives it the structure of a thick tree $\mathbb{T}$, which is truncated at the decomposition element $F(C) \subseteq \gamma_D$. Moreover, $F|_D$ can be factored as $F|_D = \Phi \circ \Psi$, where $\Psi = F|_D\colon D \to \mathbb{T}$ and $\Phi\colon \mathbb{T} \to \mathbb{G}$ is inclusion. By construction, Ψ and Φ are thick graph maps (i.e. they send decomposition elements into decomposition elements), and $F(D \cap \mathbb{G}) = \Phi(\mathbb{T}) \cap F(\mathbb{G})$.

Notice that the induced decompositions on $D, F(D), \ldots, F^{n-1}(D)$ described in section 2 (obtained by pulling back the decomposition of $F^n(D)$) can be defined equivalently by pushing forward a decomposition of D which is obtained by identifying it with $\mathbb{T}$ using Ψ.

Given this structure, it is easy to see that D projects to a 1-kneading set for the induced graph endomorphism $f\colon G \to G$: the converse is immediate from the definition.

Theorem 3.1 *Let $F: (S, \mathbb{G}) \to (S, \mathbb{G})$ be a pre-thick graph map with induced graph endomorphism $f: G \to G$.*

a) Let D be a trimmed 1-pruning disk for F. Then $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$ is an F-compatible 1-kneading set for f, and the graph endomorphism induced by $F_D: (S, \mathbb{G}) \to (S, \mathbb{G})$ is $f_K: G \to G$.

b) Let K be an F-compatible 1-kneading set for f. Then there exists a 1-pruning disk D for F such that $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$.

Proof.
Part a): Using the construction before the statement of the theorem, let T be the tree obtained by collapsing decomposition elements of $\mathbb{T}$, and let $\pi: \mathbb{T} \to T$ be the canonical projection. Define $\psi = \pi \circ \Psi \circ (p|_{D \cap \mathbb{G}})^{-1}: \overline{K} \to T$, which is well-defined since if $x, y \in D \cap \mathbb{G}$ with $p(x) = p(y)$, then they lie in the same decomposition element of $\mathbb{G}$, and hence $\Psi(x) = F(x)$ and $\Psi(y) = F(y)$ lie in the same component of the intersection of the image decomposition element with $F(D)$. Similarly, $\varphi = p \circ \Phi \circ \pi^{-1}$ is well defined, since if $x, y \in \mathbb{T}$ with $\pi(x) = \pi(y)$, then $\Phi(x) = x$ and $\Phi(y) = y$ lie in the same decomposition element of $\mathbb{T}$, and hence of $\mathbb{G}$. Now $f|_{\overline{K}} = \varphi \circ \psi$, and $\psi(\partial K) = \pi(\Psi(C \cap \mathbb{G})) = \pi(\gamma_D \cap \Psi(C))$ is a point t. Thus K is a 1-kneading set for $f: G \to G$: that it is F-compatible follows immediately from the definition, since $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$.

Finally, observe that F_D is equal to F on $\mathbb{G} \setminus D$, and $F_D(D) \subseteq \gamma_D$. Thus it induces a graph endomorphism $G \to G$ which is equal to f on $G \setminus \overline{K}$, and which maps all points of $\overline{K}$ to $p(\gamma_D) = \varphi(t)$. This graph endomorphism is exactly f_K.
Part b): is just a restatement of the definition of F-compatibility. ∎

Notice that the definitions have been set up precisely so that

$$\begin{array}{ccccc} D \cap \mathbb{G} & \xrightarrow{\Psi} & \mathbb{T} & \xrightarrow{\Phi} & \mathbb{G} \\ {\scriptstyle p}\downarrow & & {\scriptstyle \pi}\downarrow & & {\scriptstyle p}\downarrow \\ \overline{K} & \xrightarrow{\psi} & T & \xrightarrow{\varphi} & G \end{array} \tag{1}$$

commutes.

The final result in this section gives an easily verifiable condition for a 1-kneading set K to be F-compatible. As stated earlier, the issue is that the image of a 1-pruning disk must be 'innermost' in $F(\mathbb{G})$, but no corresponding notion exists for abstract graph endomorphisms.

Suppose, then, that K is a 1-kneading set for $f: G \to G$, the graph endomorphism induced by $F: (S, \mathbb{G}) \to (S, \mathbb{G})$. Then $\overline{p^{-1}(K)}$ is a union of subdisks of $\mathbb{G}$, each of which has boundary composed of arcs of $\partial\mathbb{G}$, and the same number of arcs contained in decomposition elements of $\mathbb{G}$ (if any such decomposition element is a junction, then the corresponding arc is contained in the boundary of the junction in $\mathbb{G}$). Each of the arcs contained in a decomposition element has image contained in the decomposition element $\gamma = p^{-1}(\varphi(t))$. Let $B \subseteq \gamma$ be the union of these images.

Definition: K is said to be *innermost* if

a) There is an arc α in γ such that $F(\mathbb{G}) \cap \alpha = B$.

b) There is a neighbourhood N of ∂K in $\overline{K}$ such that $f(N)$ is (the image of) an arc with endpoint $\varphi(t)$.

(Condition b) just says that the images of all of the 'ends' of K leave $\varphi(t)$ in the same direction).

Lemma 3.2 *An innermost* 1*-kneading set* K *is* F*-compatible.*

Proof. Let α be an arc as in the definition of innermost, and choose a subarc if necessary so that the boundary of α is contained in B, and hence in $F(\partial\mathbb{G})$. By condition b), there is an arc β in $F(\partial\mathbb{G})$ with the same endpoints as α, such that the disk Δ bounded by α and β contains $F(\overline{p^{-1}(K)})$ (namely, the arc obtained by following $F(\partial\mathbb{G})$ from one endpoint of α to the other, in the direction given by the images of the 'ends' of K). By condition a), and the fact that β is contained in $F(\partial\mathbb{G})$, it follows that $\Delta \cap F(\mathbb{G}) = F(\overline{p^{-1}(K)})$, and hence $D = F^{-1}(\Delta)$ is a 1-pruning disk for F with $p(\mathrm{Int}\,(D \cap \mathbb{G})) = K$. ∎

3.2 n-Kneading Sets

As with 1-pruning, the process of pulling away a 1-kneading set for f is straightforward, and a more interesting situation arises when K is a 1-kneading set for some iterate f^n of f. In this section it is shown that, provided K satisfies an additional condition (the *kneading condition*), it is possible to carry out some identifications on G to yield a new graph endomorphism $f_1\colon G_1 \to G_1$, such that the subset of G_1 corresponding to K is a 1-kneading set for f_1^{n-1}, which also satisfies the kneading condition. Repeating the process $n-1$ times thus yields a 1-kneading set which can be pulled away. The main point of this construction, as expressed by Theorems 3.11 and 3.12, is that this iterative process corresponds exactly to that by which an n-pruning disk is pruned away.

Suppose that K is a 1-kneading set for $f^n\colon G \to G$. For each $k \in \{0, \ldots, n-1\}$, define an equivalence relation $\sim_k$ on G by $x \sim_k y$ if and only if $x = y$ or there exist $z, w \in \overline{K}$ with $f^k(z) = x$, $f^k(w) = y$, and $\psi(z) = \psi(w)$ (recall from Example 3 b) that $\psi(z) = \psi(w)$ is a stronger condition than $f^n(z) = f^n(w)$). Notice that if $x \neq y$ and $x \sim_k y$ then $x, y \in f^k(\overline{K})$.

Lemma 3.3 *If* $x \sim_k y$ *then* $f^{n-k}(x) = f^{n-k}(y)$.

Proof. If $z, w \in \overline{K}$ with $\psi(z) = \psi(w)$, then $f^n(z) = f^n(w)$, since $f^n|_{\overline{K}} = \varphi \circ \psi$. The result follows immediately from the definition. ∎

Lemma 3.4 *For each $k \in \{0, \dots, n-1\}$, each equivalence class of $\sim_k$ has only finitely many components.*

Proof. Every non-trivial equivalence class of $\sim_0$ is a set of the form $\psi^{-1}(x)$ for some $x \in T$, and hence has only finitely many components since ψ is piecewise monotone. The equivalence classes of $\sim_k$ are the images of those of $\sim_0$ under f^k. ∎

Definitions: A 1-kneading set K for f^n is said to be an *n-kneading set for f* if the *kneading condition* is satisfied: that if $x \in K$ and $x \sim_k y$, then $y \in K$ and $x \sim_0 y$. If $f\colon G \to G$ is induced by a thick graph map $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$, then an n-kneading set for f is said to be *F-compatible* if it is an F^n-compatible 1-kneading set for f^n (i.e., it is $p(\mathrm{Int}\,(D \cap \mathbb{G}))$ for some 1-pruning disk D for F^n), and *non-trivial* if it is a non-trivial 1-kneading set for f^n.

From now on, all n-kneading sets will be assumed to be non-trivial unless otherwise stated. Notice that when checking whether or not the kneading condition is satisfied, it can clearly be assumed that $x, y \in f^k(\overline{K})$ and $x \neq y$.

By Lemma 3.3, if $x \sim_{n-1} y$, then $f(x) = f(y)$ (it is important to note that the converse is not generally true, since φ need not be an embedding). Thus it is possible to take the quotient of $f\colon G \to G$ by $\sim_{n-1}$ yielding a graph endomorphism $f_1\colon G_1 \to G_1$. Write $q\colon G \to G_1$ for the canonical projection, and $K_1 = q(K)$.

The following simple lemma is the essential consequence of the kneading condition:

Lemma 3.5 $K = q^{-1}(K_1)$.

Proof. Suppose that there exists $y \in G \setminus K$ with $q(y) \in K_1$. Pick $x \in K$ with $q(x) = q(y)$. Then x and y lie in the same $\sim_{n-1}$-equivalence class: that is, $x \in K$ and $x \sim_{n-1} y$. It follows from the kneading condition that $y \in K$, which is a contradiction. ∎

The aim of the following results is to show that K_1 is an $(n-1)$-kneading set for $f_1\colon G_1 \to G_1$.

Lemma 3.6 *$f_1\colon G_1 \to G_1$ is a piecewise monotone graph endomorphism, and K_1 is an open subset of G_1.*

Proof. That G_1 is a graph and f_1 is piecewise monotone follows from Lemma 3.4, while the openness of K_1 in G_1 is immediate from Lemma 3.5 (and the definition of the quotient topology). ∎

Now define $\psi_1 = \psi \circ q|_{\overline{K}}^{-1}: \overline{K_1} \to T$. To check that ψ_1 is well defined, it is necessary to show that if $x, y \in \overline{K}$ with $q(x) = q(y)$, then $\psi(x) = \psi(y)$. If either x or y is in K, then it follows by the kneading condition that both lie in K and that $x \sim_0 y$: i.e. $\psi(x) = \psi(y)$. If both lie in ∂K, then $\psi(x) = \psi(y) = t$.

Lemma 3.7 *$\psi_1: \overline{K_1} \to T$ is onto and piecewise monotone, and $\psi_1(\partial K_1) = \{t\}$.*

Proof. Surjectivity follows from that of ψ, and ψ_1 is piecewise monotone by Lemma 3.4 and the piecewise monotonicity of ψ. Since $K = q^{-1}(K_1)$ (Lemma 3.5), it follows that $\partial K_1 = q(\partial K)$, and hence $\psi_1(\partial K_1) = \psi(\partial K) = \{t\}$. ∎

Next define $\varphi_1 = q \circ f^{n-1} \circ \psi^{-1}: T \to G_1$. To check that φ_1 is well defined, it is necessary to show that if $x, y \in \overline{K}$ with $\psi(x) = \psi(y)$, then $q(f^{n-1}(x)) = q(f^{n-1}(y))$. However $\psi(x) = \psi(y)$ means exactly that $x \sim_0 y$, and hence $f^{n-1}(x) \sim_{n-1} f^{n-1}(y)$: that is, $q(f^{n-1}(x)) = q(f^{n-1}(y))$ as required.

Lemma 3.8 *$\varphi_1: T \to G_1$ is piecewise monotone, and $\varphi_1 \circ \psi_1 = f_1^{n-1}|_{\overline{K_1}}$.*

Proof. Piecewise monotonicity follows by Lemma 3.4 and the piecewise monotonicity of ψ and f. If $x \in \overline{K_1}$, then $\varphi_1 \circ \psi_1(x) = q \circ f^{n-1} \circ \psi^{-1} \circ \psi \circ q^{-1}(x) = (q \circ f \circ q^{-1})^{n-1}(x) = f_1^{n-1}(x)$. ∎

Theorem 3.9 *K_1 is an $(n-1)$-kneading set for $f_1: G_1 \to G_1$, with factor maps ψ_1 and φ_1.*

Proof. It follows from Lemmas 3.6, 3.7, and 3.8 that K_1 is a 1-kneading set for f_1^{n-1}: thus it only remains to verify the kneading condition: that if $x_1 \in K_1$ and if $x_1 \sim_k y_1$ for some $k \in \{0, \ldots, n-2\}$, then $y_1 \in K_1$ and $\psi_1(x_1) = \psi_1(y_1)$ (in this proof, elements of G_1 are denoted with a subscript 1, while elements of G have no subscript: the equivalence relations $\sim_k$ pertain either to $f: G \to G$ or to $f_1: G_1 \to G_1$, depending on which graph the elements which are being compared belong to).

Suppose then that $x_1 \in K_1$ and $x_1 \sim_k y_1$. Thus there exist $z_1, w_1 \in \overline{K_1}$ with $f_1^k(z_1) = x_1$, $f_1^k(w_1) = y_1$, and $\psi_1(z_1) = \psi_1(w_1)$. Pick $z, w \in \overline{K}$ with $q(z) = z_1$ and $q(w) = w_1$. Since $\psi_1 = \psi \circ q^{-1}$, it follows from Lemma 3.5 that $z \sim_0 w$, and hence $f^k(z) \sim_k f^k(w)$. Moreover, $q(f^k(z)) = f_1^k(q(z)) = f_1^k(z_1) = x_1$, and similarly $q(f^k(w)) = y_1$.

Now $f^k(z) \in K$ by Lemma 3.5, and so $f^k(w) \in K$ and $f^k(z) \sim_0 f^k(w)$ by the kneading property for the original n-kneading set for $f: G \to G$. Hence $y_1 = q(f^k(w)) \in K_1$, and $\psi_1(y_1) = \psi(f^k(w)) = \psi(f^k(z)) = \psi_1(x_1)$, so $x_1 \sim_0 y_1$ as required. ∎

Definition: Suppose that K is an n-kneading set for $f: G \to G$. Let $f_K: G_K \to G_K$ be the piecewise monotone graph endomorphism obtained by applying the above process $n-1$ times, and then pulling away the resulting 1-kneading set. It is said that f_K has been obtained from f by *pulling K away.*

Example 5 This example of a 2-kneading mirrors the 2-pruning described in Example 2. Let $f: G \to G$ be the graph endomorphism induced by the horseshoe, as described in Example 4, and let $K = (\overline{1}, 10\overline{1}) \subseteq G$. Let $T = [\overline{1}, 1] \subseteq G$ and $t = \overline{1} \in T$. Let $\psi = f^2|_{\overline{K}}: \overline{K} \twoheadrightarrow T$, and $\varphi: T \to G$ be inclusion. Then $f^2|_{\overline{K}} = \varphi \circ \psi$, and $\psi(\partial K) = \{t\}$, so K is a 1-kneading set for f^2. Since $f(\overline{K}) = [0\overline{1}, \overline{1}]$ is disjoint from K, the kneading condition is satisfied trivially, and hence K is a 2-kneading set for f.

The graph endomorphism $f_1: G_1 \to G_1$ (shown in Figure 15) is obtained by identifying each pair of points $x, y \in f(\overline{K})$ with $x \sim_1 y$. Since $\psi = f^2$, this is equivalent to $f(x) = f(y)$: in general, of course, this need not be the case. Notice that K_1 is a 1-kneading set for f_1. Pulling K away yields the graph endomorphism $f_K: G_K \to G_K$, also depicted in Figure 15.

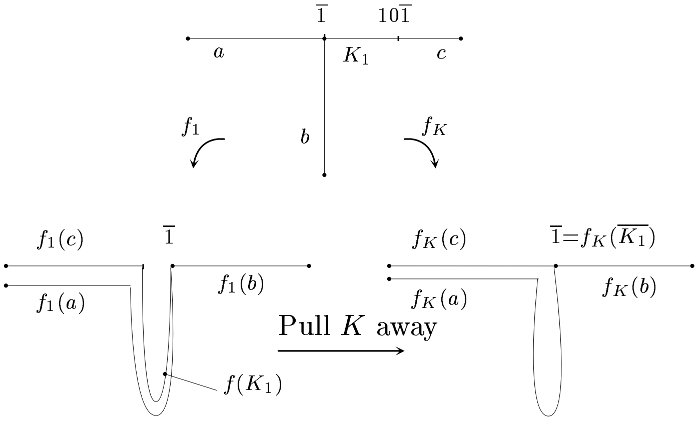

Figure 15: Pulling away a 2-kneading set

Comparing with Example 2, observe that $K = \operatorname{Int}(p(D \cap \mathbb{G}))$, and that $f_K: G_K \to G_K$ is the graph endomorphism induced by $F_D: (S^2, \mathbb{G}_D) \to (S^2, \mathbb{G}_D)$. The aim of the following discussion and results is to show that there is a one-to-one correspondence between (trimmed) n-pruning disks and F-compatible n-kneading sets, and that pruning an n-pruning disk is equivalent to pulling away the corresponding n-kneading set.

Suppose, then, that D is a trimmed 1-pruning disk for F^n. Recall that $F^n(D)$ can be given the structure of a thick tree $\mathbb{T}$, and that $F^n|_D$ factors as $F^n|_D = \Phi \circ \Psi$, where $\Psi: D \to \mathbb{T}$ and $\Phi: \mathbb{T} \to \mathbb{G}$ are thick graph maps. Define an equivalence relation $\approx_0$ on the decomposition elements of $\mathbb{G}$ by $\xi \approx_0 \eta$ if and only if either $\xi = \eta$, or both ξ and η intersect D and $\Psi(\xi)$ and $\Psi(\eta)$ are contained in the same decomposition element of $\mathbb{T}$. Then define equivalence relations $\approx_k$ on the set of decomposition elements of $\mathbb{G}$ for $0 < k < n$ by $\xi \approx_k \eta$ if and only if there exist decomposition elements ζ, ω with $\zeta \approx_0 \omega$ and $F^k(\zeta) \subseteq \xi$, $F^k(\omega) \subseteq \eta$. It is an immediate consequence of the commutativity of (1) that each $\approx_k$ corresponds to the equivalence relation $\sim_k$ defined by the 1-kneading set $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$ for f^n: that is, $x \sim_k y$ if and only if $p^{-1}(x) \approx_k p^{-1}(y)$.

Theorem 3.10 *Let $F\colon (S, \mathbb{G}) \to (S, \mathbb{G})$ be a pre-thick graph map, and $f\colon G \to G$ be the induced graph endomorphism. Let D be a trimmed 1-pruning disk for F^n, and $K = \mathrm{Int}\,(p(D \cap \mathbb{G}))$ the corresponding 1-kneading set for f^n. Then D is an n-pruning disk for F if and only if K is an n-kneading set for f.*

Proof. It is necessary to show that D satisfies the pruning condition if and only if K satisfies the kneading condition.

Suppose first that the pruning condition holds: i.e., $F^k(D)$ doesn't intersect E away from its endpoints for $0 < k < n$. Let $x \neq y$ and $k < n$ be such that $x \sim_k y$ and $x \in K$. It is necessary to show that $y \in K$ and $x \sim_0 y$.

Let $\xi = p^{-1}(x)$ and $\eta = p^{-1}(y)$. Then $\xi \subseteq \mathbb{G} \cap (D \setminus C)$, and ξ and η are distinct decomposition elements with $\xi \approx_k \eta$: that is, there are (distinct) decomposition elements ζ and ω such that $\zeta \approx_0 \omega$ and $F^k(\zeta) \subseteq \xi$, $F^k(\omega) \subseteq \eta$. Now $\zeta \approx_0 \omega$ means exactly that ζ and ω are decomposition elements of $\mathbb{G}$ which intersect D, and $\Psi(\zeta)$, $\Psi(\omega)$ are contained in the same decomposition element of $\mathbb{T}$. It follows that there is a decomposition element γ of $F^k(D)$ which contains both $F^k(\zeta)$ and $F^k(\omega)$, and hence intersects both ξ and η. Since $\xi \subseteq \mathbb{G} \cap (D \setminus C)$, it follows from Lemma 2.5 that $\gamma \subseteq \mathrm{Int}\,(D)$, and hence $y = p(\eta) \in K$. Moreover, ξ and η are contained in the same decomposition element of D, so that $\xi \approx_0 \eta$. It follows that $p(\xi) \sim_0 p(\eta)$, i.e. $x \sim_0 y$ as required.

For the converse, suppose that the kneading condition holds. If the pruning condition is violated, then there is some k with $0 < k < n$ and a leaf γ in the induced decomposition of $F^k(D)$ which intersects E away from its endpoints: let α be such a point of intersection. Since both endpoints of γ lie on $F^k(E) \subseteq \mathrm{Int}\,(\mathbb{G})$, it follows that γ intersects $\partial\mathbb{G}$ in at least one other point: choose such an intersection β so that the subarc ξ of γ between α and β only intersects $\mathbb{G}$ at its endpoints. Let $x = p(\alpha) \in K$ and $y = p(\beta)$. Since $p^{-1}(x)$ and $p^{-1}(y)$ are contained in the same decomposition element of $F^k(D)$, it follows that $p^{-1}(x) \approx_k p^{-1}(y)$, and hence that $x \sim_k y$. By the kneading condition, $y \in K$ and $x \sim_0 y$. Thus $\beta \in E$ and $p^{-1}(x) \approx_0 p^{-1}(y)$, so α and β lie in the same element η of the induced decomposition of D. Therefore $F^n(\xi \cup \eta)$ is contained in a single decomposition element of $\mathbb{G}$, which must be a junction since $\xi \cup \eta$ is homotopically non-trivial. It follows that η is a junction, whose intersection with E contains both α and β.

Since D is non-trivial, it must contain leaves: this implies that $F^k(E)$ is entirely contained in a junction contained in E, which is again impossible given the non-triviality of D.

∎

The following two theorems, which are the main results of this section, now follow easily from Theorems 3.1 and 3.10. For Theorem 3.11, recall that $F\colon (S, \mathbb{G}_1) \to (S, \mathbb{G}_1)$ is the pre-thick graph map obtained by amalgamating $\mathbb{G}$ with $F^{n-1}(D)$, and $f_1\colon G_1 \to G_1$ is the graph endomorphism obtained on factoring $f\colon G \to G$ by $\sim_{n-1}$.

Theorem 3.11 *Let $F\colon (S, \mathbb{G}) \to (S, \mathbb{G})$ be a pre-thick graph map with induced graph endomorphism $f\colon G \to G$.*

a) Let D be a trimmed n-pruning disk for F. Then $K = \operatorname{Int}(p(D \cap \mathbb{G}))$ is an F-compatible n-kneading set for f, and the graph endomorphism induced by $F\colon (S, \mathbb{G}_1) \to (S, \mathbb{G}_1)$ is $f_1\colon G_1 \to G_1$ (see the remark above regarding notation).

b) Let K be an F-compatible n-kneading set for f. Then there exists a 1-pruning disk D for F such that $K = \operatorname{Int}(p(D \cap \mathbb{G}))$.

Proof. The only part of the statement which is not an immediate consequence of Theorems 3.1 and 3.10 is that the graph endomorphism induced by $F\colon (S, \mathbb{G}_1) \to (S, \mathbb{G}_1)$ is $f_1\colon G_1 \to G_1$. This follows directly from the observation that two decomposition elements ξ and η of $\mathbb{G}$ are identified under amalgamation with $F^{n-1}(D)$ if and only if $\xi \approx_{n-1} \eta$. ∎

For the final result, recall that $F_D\colon (S, \mathbb{G}_{n-1}) \to (S, \mathbb{G}_{n-1})$ is the pre-thick graph map obtained by amalgamating $F^{n-1}(D), \ldots, F(D)$ with $\mathbb{G}$ and pruning away the resulting 1-pruning disk D; $\mathbb{G}_D$ is obtained from $\mathbb{G}_{n-1}$ by splitting in order to restore the structure of the complementary components.

Theorem 3.12 *Let $F\colon (S, \mathbb{G}) \to (S, \mathbb{G})$ be a thick graph map, $f\colon G \to G$ be the induced graph endomorphism, D be an n-pruning disk for F, and $K = \operatorname{Int}(p(D \cap \mathbb{G}))$ be the corresponding n-kneading set for f. Then the graph endomorphism induced by $F_D\colon (S, \mathbb{G}_{n-1}) \to (S, \mathbb{G}_{n-1})$ is $f_K\colon G_K \to G_K$.*

Proof. Immediate from Theorems 3.1 and 3.11 ∎

If it is necessary to split $\mathbb{G}_{n-1}$ to obtain $\mathbb{G}_D$, then the homotopy types of G_K and G are correspondingly different, but G_K can be split at a number of its vertices (corresponding to the splittings of $\mathbb{G}_{n-1}$) in order to restore the homotopy type of G.

Remark: At this point it would be natural to define the *kneading family* $\mathcal{K}(f)$. In order that it correspond to the pruning family $\mathcal{P}(f)$ defined in Section 2 using Theorem 3.11 above, however, it would be necessary to make precise the convergence of a sequence of kneadings, and hence of the sequence of spaces on which the maps are defined. Since a formal definition of $\mathcal{K}(f)$ is not required in what follows, it is unnecessary to consider the technical difficulties which this involves.

4 Pruning up to an invariant set

4.1 Basic definitions

Definitions: Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a thick graph map inducing $f\colon G \to G$. Then F is said to be *Markov* if

a) G has only finitely many vertices, and

b) $f\colon G \to G$ is onto, and does not collapse any edges to points.

Let $F\colon (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ be a Markov thick graph map, and write $E(G) = \{e_1, \ldots, e_r\}$ for the set of edges of the corresponding graph G, each endowed with a fixed orientation. The *set of oriented edges of* G is defined to be $OE(G) = \{e_1, \ldots, e_r, \bar{e}_1, \ldots, \bar{e}_r\}$, where $\bar{e}_i$ denotes the edge e_i with the opposite orientation. By convention, general elements of $E(G)$ will be denoted with lower case letters, and those of $OE(G)$ with upper case letters.

For each vertex v of G, define $\mathrm{Lk}(v) \subset OE(G)$ to be the set of oriented edges whose initial vertex is v. Since f doesn't collapse any edges and is strictly monotone away from the preimages of vertices, there is for each $n > 0$ a well-defined *derivative* map $Df^n\colon OE(G) \to OE(G)$ which maps each $E \in \mathrm{Lk}(v)$ to the element of $\mathrm{Lk}(f(v))$ which is the first oriented edge traversed by the image $f^n(E)$ (see [5]).

If $x \in G$ is not a vertex (so x lies in the interior of some edge e), then write $\mathrm{Lk}(x) = \{0, 1\}$. If $f^n(x)$ is a vertex v for some $n > 0$, then define $Df^n\colon \mathrm{Lk}(x) \to \mathrm{Lk}(v)$ by setting $Df^n(0)$ to be the first edge traversed by the nth iterate of the segment of e which starts at x, and $Df^n(1)$ to be the first edge traversed by the nth iterate of the segment of $\bar{e}$ which starts at x.

A point $x \in G$ is said to be *n-preturning (to the vertex v)* if $f^n(x) = v$, and Df^n is constant on $\mathrm{Lk}(x)$. It is said to be *n-turning (to v)* if in addition

a) $f^k(x)$ is not $(n-k)$-preturning for $1 \le k < n$, and

b) x is not k-preturning to v for $1 \le k < n$.

Notice that this does not preclude the possibility that x is k-turning to some other vertex for some $k < n$. If x is n-turning to v and $Df^n(\mathrm{Lk}(x)) = \{E\}$, then it is also said that x is *n-turning to* (v, E).

Lemma 4.1 *If x is n-turning, then $f(x)$ is a vertex.*

Proof. If x is itself a vertex, then there is nothing to prove. If both x and $f(x)$ were not vertices, then it would follow from the strict monotonicity of f away from the preimages of vertices that $f(x)$ would be $(n-1)$-preturning, contradicting the fact that x is n-turning. ∎

4.2 Finding pruning disks

Suppose that $F: (S, \mathbb{G}, A) \to (S, \mathbb{G}, A)$ is a Markov thick graph map, with induced graph endomorphism $f: G \to G$. The following definition is illustrated by Figure 16.

Definition: Suppose that x is n-turning to (v, L). If $J = p^{-1}(x)$ is a junction, let $E_1, \ldots, E_r$ be the elements of $\mathrm{Lk}(x)$ in their correct cyclic order (i.e. each pair $\{E_i, E_{i+1}\}$ and $\{E_r, E_1\}$ is connected in the cyclic order at x), and let $s_1, \ldots, s_r$ be the corresponding (oriented) strips in $\mathbb{G}$. Now $F(J) \subseteq p^{-1}(v)$, and the images of the initial segments of the strips s_i leave $p^{-1}(v)$ along the strip $p^{-1}(L)$. Amongst these initial segments there are two, connected in the cyclic order, whose images are outermost in $p^{-1}(L)$: suppose without loss of generality that these are s_1 and s_r. For each $y \in L$, define $\Delta(\cdot, y)$ to be the disk bounded by the smallest segment of $p^{-1}(y)$ which disconnects both of these images of initial segments, together with the arc of $F^n(\partial\mathbb{G} \cap (s_1 \cup s_r \cup J))$ which has the same endpoints and is contained in $p^{-1}(v) \cup p^{-1}(L)$. Let $D(\cdot, y) = F^{-n}(\Delta(\cdot, y))$, a disk with boundary the union of the arcs $E(\cdot, y) \subseteq \partial\mathbb{G}$ and $C(\cdot, y)$ with $F^n(C(\cdot, y)) \subseteq p^{-1}(y)$. (The reason for the $\cdot$ will become apparent shortly).

If $p^{-1}(x)$ is not a junction, then $\Delta(\cdot, y)$ and $D(\cdot, y)$ can be defined analogously, replacing the strips $s_1, \ldots, s_r$ with the two oriented strip segments emanating from the leaf $p^{-1}(x)$.

Lemma 4.2 *Suppose that x is n-turning to (v, L) and $A \cap p^{-1}(v) = \emptyset$. If $y \in L$ is sufficiently close to v (following the orientation of L), then $D(\cdot, y)$ is an n-pruning disk.*

Proof.

It is clear from the construction that $D(\cdot, y)$ is a 1-pruning disk for F^n provided y is close enough to v that $\Delta(\cdot, y)$ is disjoint from A. Thus it is only necessary to show that $D(\cdot, y)$ satisfies the pruning condition for y sufficiently close to v: that is, that $F^k(D(\cdot, y))$ does not intersect $E(\cdot, y)$ away from its endpoints for $1 \le k < n$. Equivalently (replacing k with $n-k$), it is necessary to show that $F^k(E(\cdot, y) \setminus \{\text{endpoints}\}) \cap \Delta(\cdot, y) = \emptyset$ for $1 \le k < n$ and y sufficiently close to v.

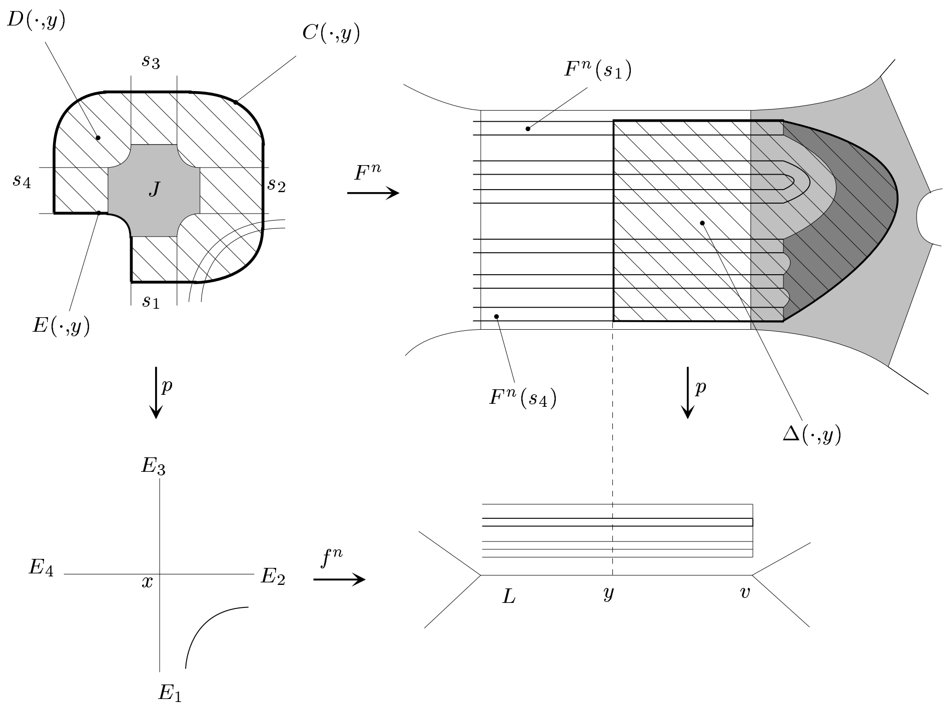

Figure 16: Opening out a pruning disk

By continuity, it is enough to show that $F^k(E) \cap \Delta = \emptyset$ for $1 \leq k < n$, where Δ is the limiting disk $\Delta(\cdot, v) \subseteq p^{-1}(v)$, and $E = E(\cdot, v)$.

Suppose, then, that $F^k(E) \cap \Delta \neq \emptyset$ for some k. Write $K = p^{-1}(v)$, $C = C(\cdot, v)$, $e = F^n(E)$, and $c = F^n(C)$ (so $\partial\Delta = c \cup e$). Notice first that $F^k(E) \subseteq \partial F^k(\mathbb{G})$, and $F^n(\partial\mathbb{G}) \subseteq \operatorname{Int}(F^k(\mathbb{G}))$, so $F^k(E) \cap e = \emptyset$. Second, $F^k(E)$ cannot be contained in Δ, since $\Delta \subseteq K$ and hence $F^{n-k}(\Delta)$ is contained in the interior of a junction, but $F^n(E)$ intersects ∂K. Third, the endpoints of $F^k(E)$ are disjoint from K, since otherwise the endpoints of $e = F^n(E)$ would lie in the interior of a junction. Finally, L is not contained in $p(F^k(E))$, since $F^n(E) \subseteq K$ and F does not collapse any edges to points. Thus $F^k(E)$ is contained in the disk $K \cup p^{-1}(L)$, intersects Δ, has endpoints disjoint from K, and p is injective on each component of $F^k(E) \setminus \Delta$: so $F^k(E) \setminus \Delta$ consists of exactly two components, each of which is contained in $p^{-1}(L)$.

It follows that $Df^k(E_1) = Df^k(E_r) = L$, and that $Df^{n-k}(L) = L$. Since F is orientation-preserving, this means that the relative positions of the initial segments of $F^k(s_1)$ and $F^k(s_r)$ in $p^{-1}(L)$ are the same as those of $F^n(s_1)$ and $F^n(s_r)$; and this, again using the fact that F is orientation-preserving, implies that $Df^k(s_i) = L$ for all i. Hence x is k-preturning to v, which is a contradiction. ∎

Definition: An n-turning point x to (v, L) is said to be a *prunable n-turning point* if $D(\cdot, y)$ is an n-pruning disk for all $y \in L$ sufficiently close to v.

Remark: If $A \cap p^{-1}(v) = \emptyset$, then Lemma 4.2 says that any n-turning point to v is prunable.

Suppose that the conditions of Lemma 4.2 are satisfied, so there exists a pruning disk $D(\cdot, y) = F^{-n}(\Delta(\cdot, y))$. It is clear that $\Delta(\cdot, y)$ is a disk for all $y \in L$, and hence that there is a well-defined disk $\Delta(\cdot, u_1)$, where u_1 is the terminal vertex of L. Whether or not it is possible to extend the disk beyond u_1 depends on the behaviour of the segments of $F^n(\partial\mathbb{G})$ which arise from the F^n-image of $E(\cdot, y)$. If both of them leave $p^{-1}(u_1)$ through the same strip, which is distinct from $p^{-1}(\overline{L})$, or if both leave through $p^{-1}(\overline{L})$ on the same side of $\Delta(\cdot, u_1)$, then it is said that $\Delta(\cdot, u_1)$ *can be extended through* u_1 (Figure 17). Otherwise (Figure 18) $\Delta(\cdot, u_1)$ *cannot be extended through* u_1.

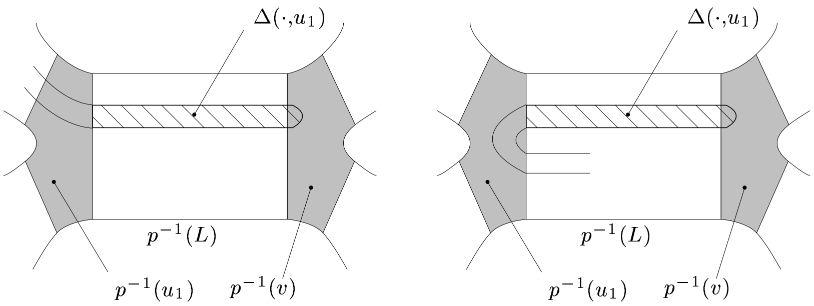

Figure 17: Extending Δ through a junction

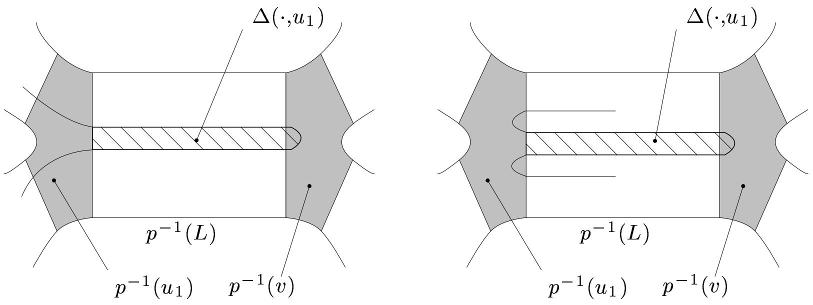

Figure 18: Δ cannot be extended

If $\Delta(\cdot, u_1)$ can be extended through u_1, and the segments of $F^n(\partial\mathbb{G})$ leave $p^{-1}(u_1)$ along an oriented strip $p^{-1}(L_1)$, then again, $\Delta(\cdot, u_1)$ can be extended all the way along $p^{-1}(L_1)$ to the junction $p^{-1}(u_2)$ at the other end of the strip. Write $\Delta(\cdot, u_1) = \Delta(L, u_1)$, and denote by $\Delta(L, y)$ the disk obtained by extending to the leaf $p^{-1}(y)$ for each $y \in L_1$.

Continuing in this way, there is some sequence $u_1, \ldots, u_{n-1}$ of vertices through which Δ can be extended, followed by a vertex u_n through which it cannot be extended: and a corresponding sequence $L, L_1, L_2, \ldots, L_{n-1}$ of oriented edges corresponding to the strips along which Δ is extended. Let α_k denote the edge-path $LL_1L_2 \cdots L_{k-1}$. The disk obtained by extending through $u_1, u_2, \ldots, u_k$ to the leaf $p^{-1}(y)$ in $p^{-1}(L_k)$ is denoted $\Delta(\alpha_k, y)$, and its preimage $F^{-n}(\Delta(\alpha_k, y))$ by $D(\alpha_k, y)$. The arc of $\partial D(\alpha_k, y)$ which lies on $\partial \mathbb{G}$ is denoted $E(\alpha_k, y)$, while the arc in $F^{-n}(p^{-1}(y))$ is denoted $C(\alpha_k, y)$.

In the above discussion, the only question is whether or not the disk $\Delta(\alpha_k, y)$ can be extended, not whether the disk $D(\alpha_k, y)$ remains a pruning disk. There are three reasons why it may become impossible to extend $D(\alpha_k, y)$ further while it remains an n-pruning disk:

a) That the pruning condition would be violated if $D(\alpha_k, y)$ were extended further. If this is the case then $F^j(C(\alpha_k, y))$ intersects $C(\alpha_k, y)$ for some $j < n$.

b) That the interior of the pruning disk would intersect A if $D(\alpha_k, y)$ were extended further. If this is the case then $A \cap C(\alpha_k, y) \neq \emptyset$.

c) That $k = n$ and $y = u_n$: that is, that $\Delta(\alpha_n, u_n)$ cannot be extended as a disk.

Since these are all closed conditions, it follows from Lemma 4.2 that there is a unique maximal n-pruning disk $D(\alpha_k, y)$. This pruning disk is denoted $D^{(n)}(x)$.

4.3 The Bestvina-Handel algorithm

Algorithms for computing train tracks for surface homeomorphisms have been given by Bestvina-Handel [5], Franks-Misiurewicz [10], and Los [16]. In this section it is explained how the Bestvina-Handel algorithm can be recast in such a way that the only isotopies which it involves are prunings and thickenings: in particular, this means that no new dynamics can be created at any point during the algorithm. The essential observation is that pruning is a more general operation than *folding* (the driving operation of Bestvina-Handel) — every folding can be interpreted as a pruning, but not vice-versa: in particular, it often requires several foldings to destroy the dynamics which can be removed with a single pruning.

Rather than give a formal description of the algorithm, (thereby duplicating the work of Bestvina and Handel), a list is given below of the operations involved in the Bestvina-Handel algorithm, and how they should be reinterpreted in terms of pruning. Before this, however, an example is given. It is hoped that the algorithm (repeatedly finding and pruning non-trivial pruning disks) will be clear before the details of the translation are given.

Example 6 This is an example in which a single 4-pruning (and two trivial 1-prunings) produce a pseudo-Anosov map: applying the Bestvina-Handel algorithm to the same example requires many more moves (eight successive foldings and tightenings in one computer implementation [12]).

Consider the horseshoe map $F\colon (S^2, \mathbb{G}, A) \to (S^2, \mathbb{G}, A)$, where A is the period 8 orbit with code $\overline{10011010}$. The Thurston representative of the isotopy class of F relative to A will be computed. The initial thick graph map, the invariant set A, and the induced graph endomorphism $f\colon G \to G$ are shown in Figure 19. The points of A are labelled 1 to 8.

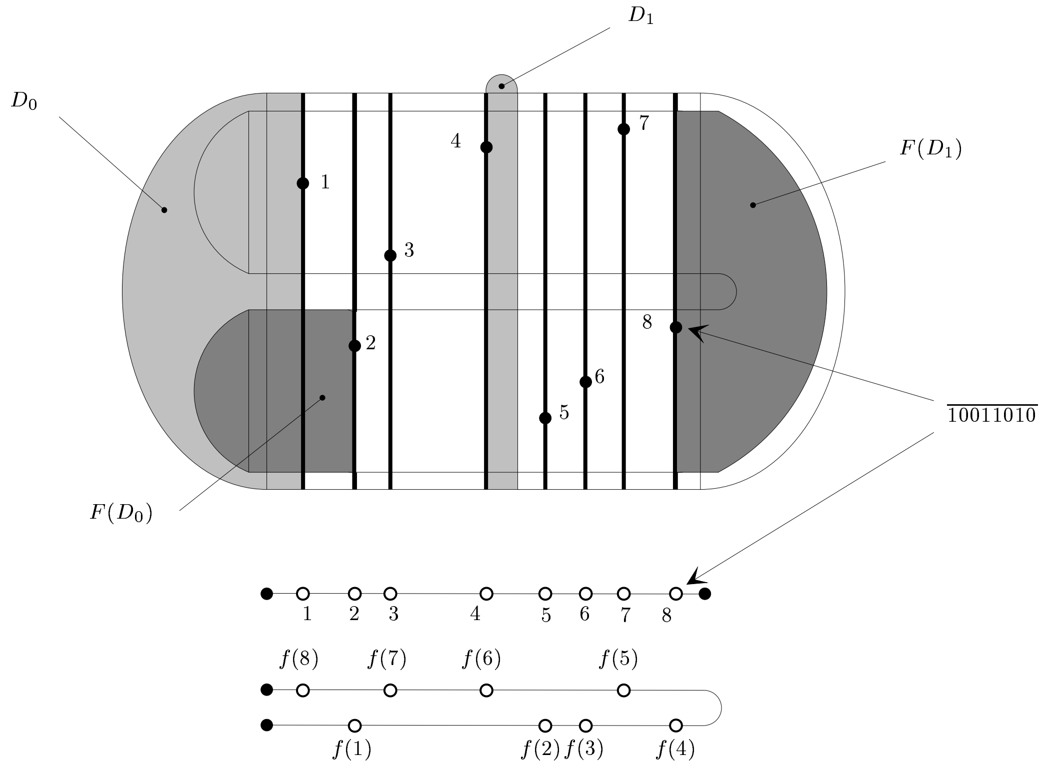

Figure 19: A period 8 orbit in the horseshoe, and two 1-pruning disks

The first, and essentially trivial, step, is to prune away the 1-pruning disks D_0 and D_1 depicted in Figure 19. On the level of the graph endomorphism, this corresponds, in the language of Bestvina-Handel, to tightening the graph endomorphism with respect to A. The result of these prunings can be seen in Figure 20.

The remainder of the algorithm will be demonstrated first by using kneading theory on the graph endomorphism: this is the way that one would carry out the algorithm in practice. At the end, pictures will be given of the relevant pruning disks and their images in S^2.

In the graph endomorphism $f\colon G \to G$ of Figure 20, there is a unique prunable turning point, namely p which is 4-turning to the vertex 8. The algorithm proceeds by finding the largest connected 4-kneading set containing p which is disjoint from A: this corresponds to the maximal 4-pruning disk $D^{(4)}(p)$ described in Section 4.2.

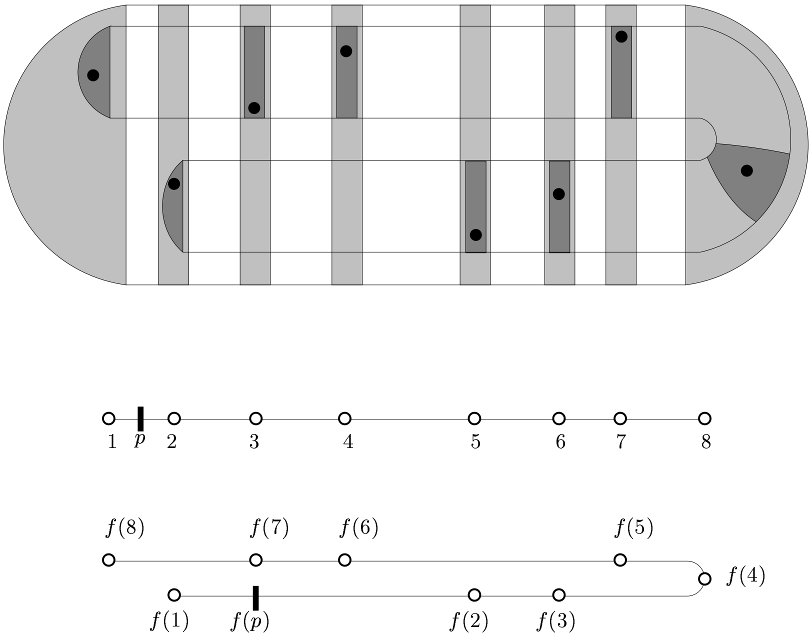

Figure 20: The thick graph and graph maps after the 1-prunings

Observe that there is a period 3 point α in e_1, with the property that $\beta = f(\alpha) \in e_3$ and $\gamma = f(\beta) \in e_7$. The periodic point α lies to the right of p in e_1.

Let K be the connected 1-kneading set for f^4 which contains p and has α as its right endpoint. This interval, together with its first 4 iterates, is shown in Figure 21.

Because G is a tree and K is connected, the kneading condition is satisfied if and only if $f^k(\overline{K}) \cap K = \emptyset$ for $1 \leq j \leq 3$, which is clearly the case (notice that K is as large as it could be while satisfying the kneading conditions, since if $\alpha \in K$ then $f^3(\overline{K}) \cap K \neq \emptyset$). Thus K is a 4-kneading set. Drawing $f^4(G)$ taking into account the thick graph structure is a tedious exercise which shows that K is an innermost 4-kneading set, and hence is F-compatible by Lemma 3.2. Hence, by Theorem 3.11, there is a 4-pruning disk D for $F\colon (S^2, \mathbb{G}, A) \to (S^2, \mathbb{G}, A)$ with $p(\mathrm{Int}\,(D \cap \mathbb{G})) = K$: and by Theorem 3.12, pruning away D is equivalent to pulling away K.

The first step in pulling away K is to identify points which lie in the same $\sim_3$ equivalence class: in this case, as in Example 5, these are exactly points of $f^3(\overline{K})$ which have the same image under f. The identifications are illustrated in Figure 22, in which lines have been drawn joining pairs of points which have the same image. This yields a new graph endomorphism $f_1\colon G_1 \to G_1$, where $G_1 = G/{\sim_3}$, with a 3-kneading set $K_1 = K/{\sim_3}$. For the sake of notational clarity, subscripts will be dropped after this and every stage of the procedure. This new graph endomorphism is depicted in Figure 23, together with the identifications arising from $\sim_2$. Carrying out this second identification gives the graph endomorphism of Figure 24, for which K is a 2-kneading set: the identifications arising from $\sim_1$ are shown on this figure.

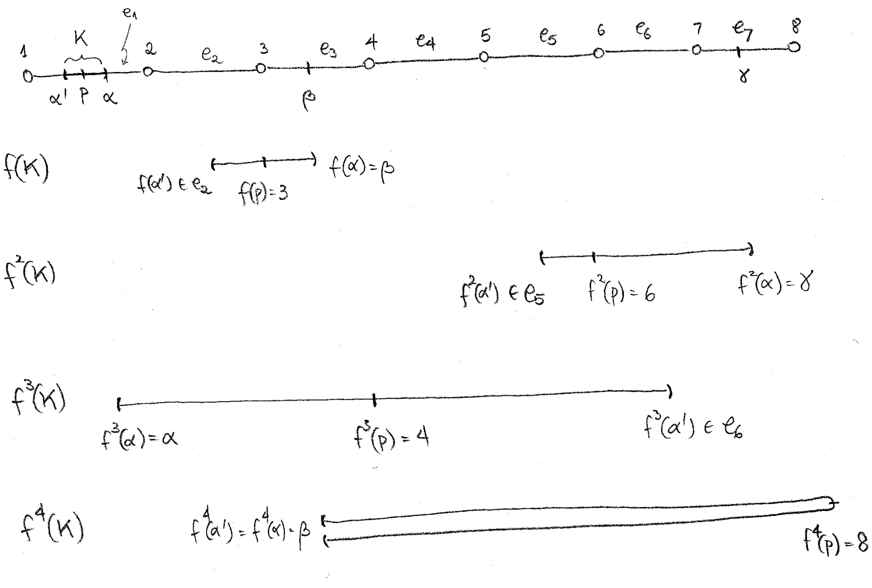

Figure 21: A maximal 4-kneading set and its iterates

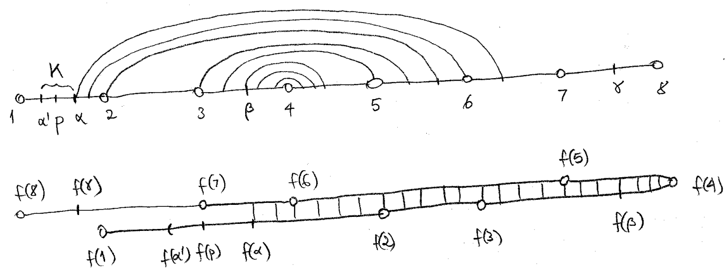

Figure 22: Identifications under $\sim_3$

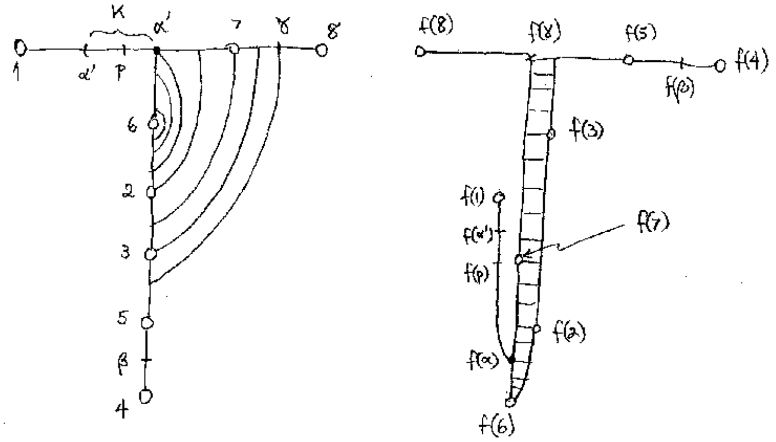

Figure 23: Identifications under $\sim_2$

Carrying out the final identifications yields the graph endomorphism of Figure 25: observe that K is now a 1-kneading set, which can be pulled away as in the figure. The resulting graph endomorphism $f\colon G \to G$ is now *efficient* in the terminology of Bestvina and Handel: that is, there are no prunable turning points. An invariant train track for the isotopy class can now be calculated as described in [5]: it, and its image, are shown in Figure 26.

The remaining figure in this example illustrates the pruning which corresponds to this kneading. In order to simplify the diagram, the gaps in the invariant Cantor set Λ have been collapsed (another way of viewing this is that the initial thick graph map $F\colon \mathbb{G} \to \mathbb{G}$ has been chosen in such a way that $F(S)$ very nearly fills S, and the two 1-junctions J_1 and J_2 are so narrow as to be indiscernible). Figure 27 shows the 4-pruning disk D, together with its images $F(D)$, $F^2(D)$, $F^3(D)$, and $F^4(D)$, which project to K and its images (compare with Figure 21). The period 8 orbit A is shown with small squares, and the period 3 orbit whose points lie in $p^{-1}(\alpha)$, $p^{-1}(\beta)$, and $p^{-1}(\gamma)$ with larger circles.

Notice that $F^4(C)$ is contained in a single decomposition element. The thick graph map obtained by pruning away D has trivial dynamics in $\bigcup_{i=-\infty}^{\infty} F^i(D)$, and is equal to the horseshoe elsewhere. The induced graph endomorphism is the efficient graph endomorphism which was obtained by pulling away K. The pseudo-Anosov map in the isotopy class of F relative to A can be obtained by collapsing the closures of the wandering domains of F_D to points. Symbolic dynamics, as a subshift of the full 2-shift, can be introduced as in Examples 1 and 2: note that this is different from introducing symbolic dynamics using the train track, where 10 symbols are required. This makes it possible to compare this pseudo-Anosov representative with other maps in the pruning family of the horseshoe.

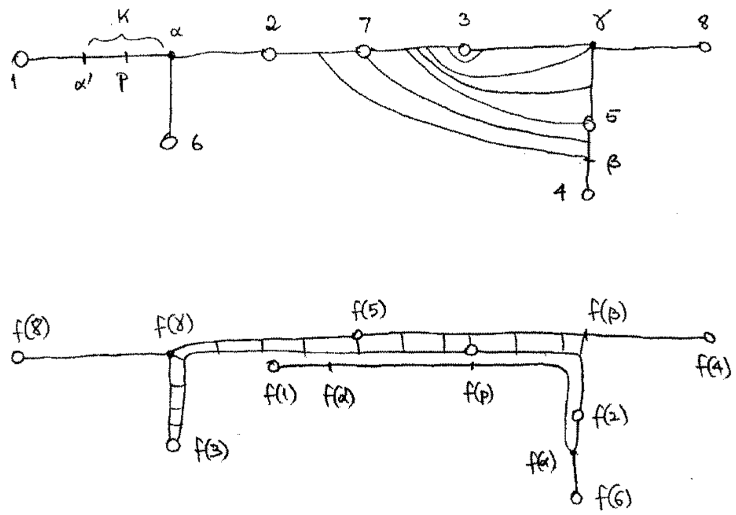

Figure 24: Identifications under $\sim_1$

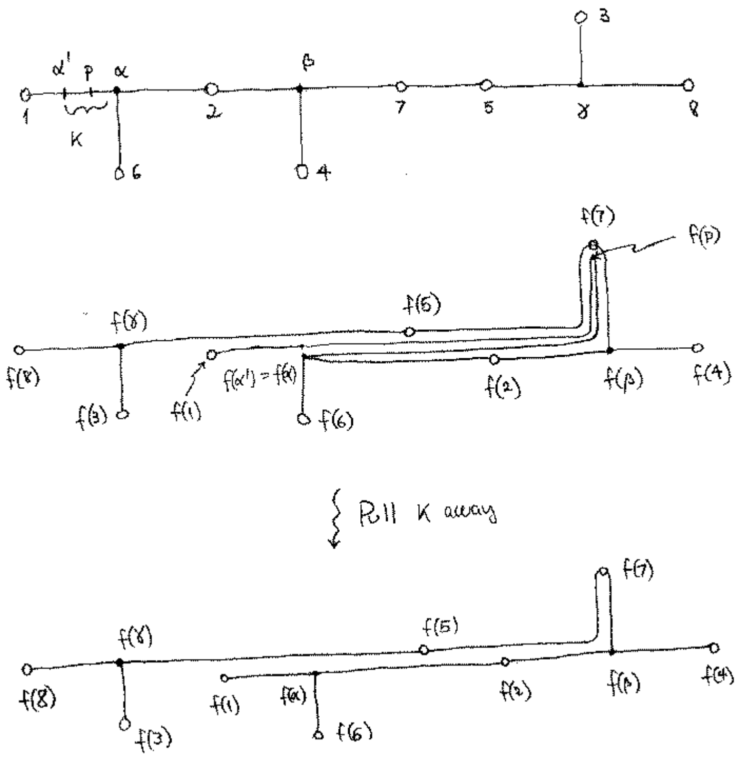

Figure 25: Pulling away the resulting 1-kneading set

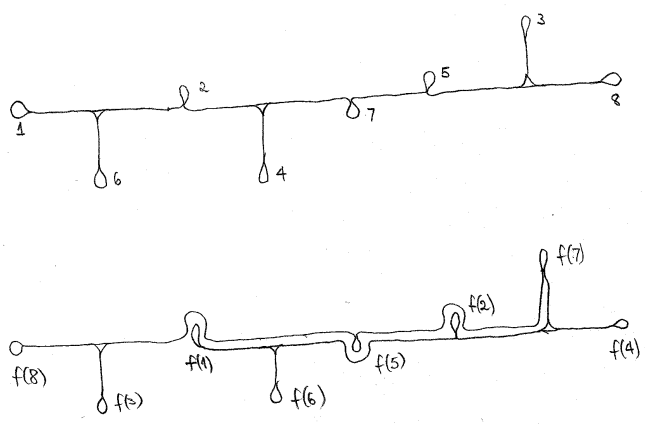

Figure 26: An invariant train track

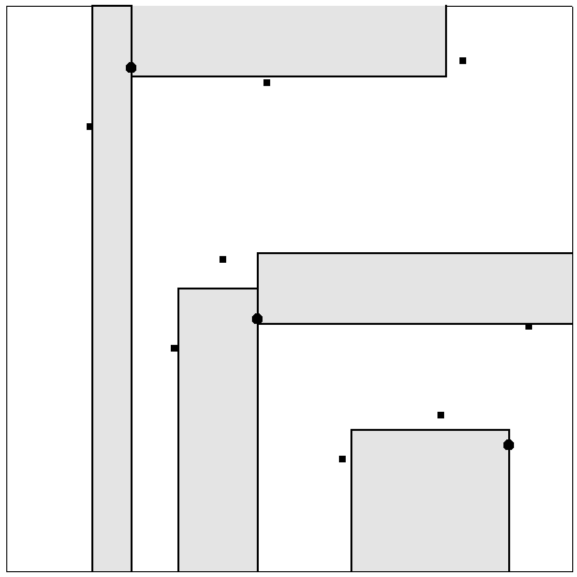

Figure 27: The 4-pruning disk and its images

Example 7 The same algorithm can be carried out for other invariant sets A in the horseshoe. Figure 28 shows a 2-pruning disk D for the horseshoe map $F\colon (S^2, \mathbb{G}, A) \to (S^2, \mathbb{G}, A)$, where A consists of the homoclinic orbit $\overline{0}1011\overline{0}$ and the fixed point $\overline{0}$, together with the images $F(D)$ and $F^2(D)$. Once this 2-pruning has been carried out, there are no more non-trivial pruning disks. In fact, this is precisely the 2-pruning disk of Example 2. The remainder of this section deals only with pruning up to finite invariant sets.

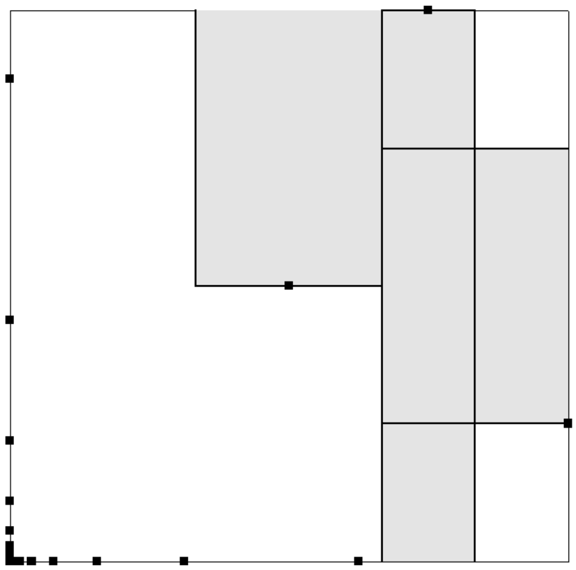

Figure 28: Pruning up to a homoclinic orbit

Remark: For the sake of simplicity and brevity, all of the examples in this paper have been of homeomorphisms of S^2. However, the theory applies equally to orientable surfaces of any genus.

The remainder of this section describes the steps which are needed to convert the Bestvina-Handel algorithm to a pruning algorithm in the general case. It is assumed that S is a closed orientable surface, that $B \subseteq S$ is a finite non-empty set of punctures, and that α is an isotopy class of orientation-preserving homeomorphisms of S relative to B (the extension to the case where S is unpunctured is described in Section 5 of [5]). As in [5], it is possible to find a thick graph $(S, \mathbb{G}, B)$, whose induced graph G can also be considered as a subset of S, and an initial thick graph map $F\colon (S, \mathbb{G}, B) \to (S, \mathbb{G}, B)$ which is in the isotopy class α.

a) There are two ways to deal with the puncture set B. In the first, described in detail in [5], choose one of the orbits of B under α, and designate a *peripheral* *subgraph* P of G which consists of a circle surrounding each of the punctures which does not belong to the chosen orbit: the remainder of $\mathbb{G}$ is chosen so that each component of $S \setminus \mathbb{G}$ contains exactly one puncture (which is the unique periodic point in that component). The initial thick graph map $F(S,\mathbb{G}) \to (S,\mathbb{G})$ is chosen so that $f\colon G \to G$ restricts to a homeomorphism of P, and this condition is required to be maintained throughout the algorithm.

The second approach is again to choose one of the orbits of B, and to let A be the set of punctures not in this orbit. Choose $(S,\mathbb{G},A)$ so that each component of $S \setminus \mathbb{G}$ contains exactly one point of $B \setminus A$, and such that the points of A are contained in distinct decomposition elements of $\mathbb{G}$. The initial thick graph map is then taken to be $F\colon (S,\mathbb{G},A) \to (S,\mathbb{G},A)$.

The advantage of the second approach is that two of the more complicated operations of the Bestvina-Handel algorithm, namely *absorbing into* P and *making irreducible*, can be avoided: these operations ensure that the entries of the transition matrix of $f\colon G \to G$ which are decreased by folding correspond to *main* edges (i.e. those edges e such that $f^n(e)$ is not contained in P for any n). The disadvantage is that it is necessary to keep track, throughout the algorithm, of how the points of A lie relative to the image $F(\mathbb{G})$.

It is also possible to adopt a combination of these approaches, surrounding some punctures with peripheral loops and choosing others to constitute the invariant set A. Here it is assumed that the second approach has been adopted, since the first is discussed in [5].

b) The Bestvina-Handel operation *pulling tight* can be accomplished by 1-prunings.

c) The Bestvina-Handel operation *collapsing an invariant forest disjoint from* P is interpreted as follows: if $f\colon G \to G$ has an invariant forest, each component T of which contains at most one point of A, then collapse this invariant forest. In terms of the thick graph map, redeclare the decomposition elements of $\mathbb{G}$ so that each set $p^{-1}(T)$ is a single junction: thicken these junctions if necessary to ensure that $\operatorname{diam}(F^n(J)) \to 0$ as $n \to \infty$ for each junction J.

d) Suppose that $F\colon (S,\mathbb{G},A) \to (S,\mathbb{G},A)$ is tight, and that there are no invariant forests whose components contain no more than one point of A. If the transition matrix for $f\colon G \to G$ is reducible, then there exists a corresponding reduction of the isotopy class α.

e) It is possible to replace the Bestvina-Handel operation *performing a valence 2 isotopy* with a new operation called *cleaning a valence 2 vertex*, which can be implemented by a (perhaps infinite) sequence of prunings. This has the advantage that it enables the whole algorithm to be carried out with prunings and thickenings, and in particular in such a way that dynamics is monotonically destroyed. It also avoids the need to calculate eigenvalues and eigenvectors of transition matrices. Cleaning is, however, a complicated procedure: it is described in Section 5. In practice, if the aim is to calculate a train track, it is more straightforward to perform valence 2 isotopies.

f) The Bestvina-Handel operation *Folding to decrease* λ should be replaced by finding (as described in section 4.2) and pruning a non-trivial pruning disk.

Rewriting the Bestvina-Handel algorithm in this way makes it clear that all Thurston representatives can be obtained by pruning. More precisely

Theorem 4.3 *Let* $F\colon (S,\mathbb{G}) \to (S,\mathbb{G})$ *be a thick graph map. Then the pruning family* $\mathcal{P}(F)$ *of* F *contains maps semi-conjugate to the Thurston representatives of each isotopy class of* F *relative to a finite* F*-invariant set.*

The semi-conjugacy mentioned in the statement of the theorem above consists in collapsing the wandering domains of the pruned map. This is quite a mild operation – in particular, the fibres of the semi-conjugacy carry no entropy. Maps in $\mathcal{P}(F)$ have invariant *laminations* instead of *foliations* as is the case with pseudo-Anosov maps.

5 Cleaning a valence two vertex

The main modification necessary to the algorithms of Bestvina-Handel and Franks-Misiurewicz, if they are to be carried out entirely in the language of pruning, arises from the need to avoid performing valence two isotopies (*dragging*, in the terminology of Franks-Misiurewicz). It is necessary to avoid the proliferation of valence two vertices in order that the number of edges in the underlying graph remains bounded: this ensures that the set of possible spectral radii of transition matrices is discrete, and hence that if the spectral radius continually decreases the algorithm is guaranteed to terminate.

Valence two isotopies, while they remove valence two vertices and never increase the spectral radius, may create dynamics (periodic orbits for example) which didn't exist before. This is incompatible with pruning, in which dynamics is only ever destroyed. Thus an alternative to performing valence two isotopies, called *cleaning a valence two vertex*, is presented here. The technicalities involved in giving a formal description of this operation, and in showing that it terminates, are very intricate and degenerate into a plethora of cases: they are also quite tedious, and do not introduce any new techniques which make the effort worthwhile.

In practice, however, it is relatively easy to clean valence two vertices: an informal description of the operation follows. Moreover, the use of pruning rather than folding in the way described by Bestvina-Handel (which requires frequent subdivision of edges and subsequent erasure of the resulting valence two vertices) makes valence two vertices much rarer. It is also worth noting that cleaning does not require the calculation of eigenvalues and eigenvectors of transition matrices, in which respect it can be easier to carry out by hand than a valence two isotopy.

Suppose, then, that a valence two vertex v (corresponding to a 2-junction J in $\mathbb{G}$) is created during the course of the algorithm. If there are no turning points to v, then it is possible to split the junctions corresponding to all vertices of valence other than two in its backward orbit (this requires type b) splittings). This ensures that all vertices which eventually map to v have valence two. Since there are no turning points to v, there are no turning points to any of the vertices which eventually map to it: thus all of these vertices can be erased (and the corresponding 2-junctions decomposed into leaves).

If there are turning points to v, then the aim is to prune them away and proceed as above. There can only be finitely many turning points to v so v can be erased after finitely many prunings, decreasing the total number of valence two vertices, provided that no new turning points to v, or valence two vertices which are not preimages of v, are created during pruning.

It is always possible to carry out the prunings without introducing new turning points or valence two vertices, with the exception of one important case: a prototypical example of this case will now be described. It is possible to have much more complicated situations (for example, with vertices of valence greater than three), but they are treated in exactly the same way.

Suppose, then, that under some iterate of $f\colon G \to G$ the situation depicted in Figure 29 arises. Here, and in subsequent figures, the upper part shows segments of G, and the lower part the images of these segments under the endomorphism in question.

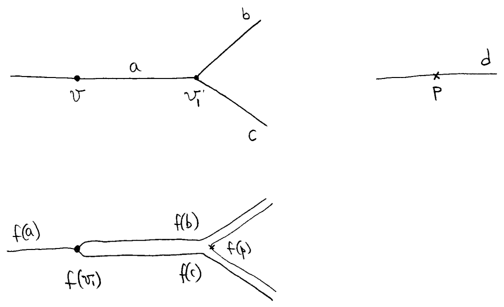

Figure 29: Prototypical situation leading to an infinite sequence of prunings

The point $p \in d$ is 2-turning to v: assume that K is a 2-kneading set about p which is as large as this situation allows: namely K is an open interval with $f^2(K) = a$. Pulling K away yields the graph endomorphism shown in Figure 30. Thus in attempting to pull away the 2-turning point p to v, this same point has become a 3-turning point to v. This leads to an infinite sequence of kneadings (with a corresponding infinite sequence of prunings), which at each step turns an n-turning point to v into an $(n + 1)$-turning point to v.

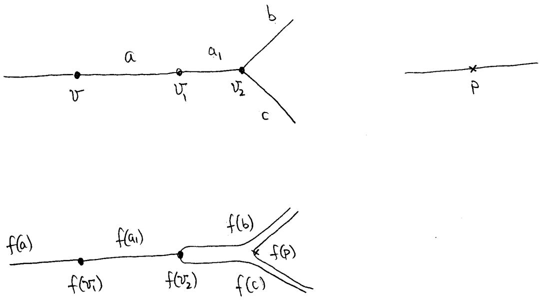

Figure 30: The result of kneading the graph endomorphism of Figure 29

Notice that, by the Markov assumption, $f(b)$ strictly contains b and $f(c)$ strictly contains c. It follows that f has fixed points in b and c: let $q_1 \in b$ and $q_2 \in c$ be those which are closest to v_1. The sequence of kneadings converges to give a graph endomorphism in which q_1 and q_2 are identified to the same point q. The limiting graph endomorphism is shown in Figure 31.

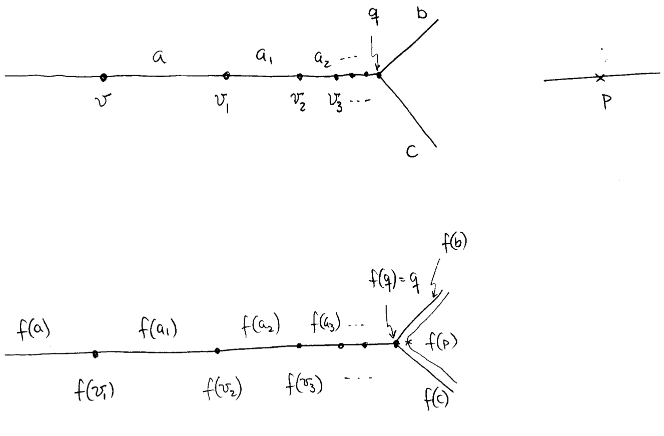

Figure 31: The limit of an infinite sequence of kneadings

In the limit, v has an infinite backward orbit consisting of valence two vertices, but there is no longer any turning point to v. This orbit of valence two vertices can therefore be erased to obtain the graph endomorphism shown in Figure 32.

Notice in particular that $Df\colon \mathrm{Lk}(q) \to \mathrm{Lk}(q)$ is the identity map. This is the main mechanism whereby fixed prong singularities of pseudo-Anosov maps are created.

There is also another, more complicated, situation which might lead to an infinite sequence of prunings. Here again, the resulting sequence of graph endomorphisms converges to one in which it is possible to erase all of the infinitely many valence two vertices which are created. However, by contrast with the previous case, this situation

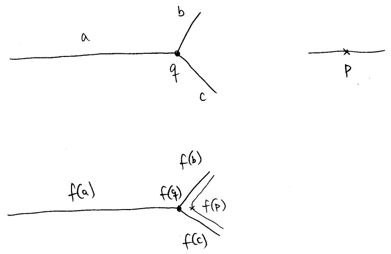

Figure 32: The graph endomorphism after erasing the orbit of valence two vertices

is inessential in that the infinite sequence can be avoided altogether by a clever choice of pruning disk. In the particular example described below (which again, could be made much more complicated but shows the only essential point of the process) it is shown that, instead of choosing a 2-pruning disk to be as 'long' as possible, it is more sensible to choose a 3-pruning disk which must be 'shorter', but is, however, 'thicker'. (The interpretation of these terms will become clear from the example).

Suppose then that $f\colon G \to G$ is a graph endomorphism (perhaps an iterate of another graph endomorphism) with the behaviour indicated in Figure 33. In this

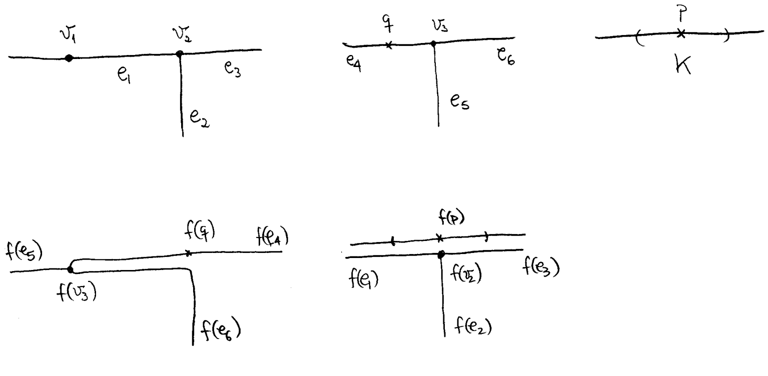

Figure 33: A situation leading to an inessential infinite sequence of prunings

figure, p is 2-turning to (v_1, e_1), and q is 3-turning to (v_1, e_1). Suppose that it is decided to prune a maximal 2-pruning disk about p (or to pull away a maximal interval about p). Then q becomes a valence three vertex, since a maximal kneading set K about p satisfies $f^2(K) = e_1$, with the left-hand endpoint of K, as shown in Figure 33, mapping to q (under a single iterate). Thus pulling K away yields the endomorphism shown in Figure 34. Here e is the new edge caused by the identification of the segment of e_4 between q and v_3 with a segment of e_5, as dictated by the equivalence relation $\sim_1$ on $f(\overline{K})$.

Now both p and q are no longer turning to (v_1, e_1). However, a new point $q_1 \in e_1$ has been created which is 4-turning to (v_1, e_1): this leads to an infinite sequence of

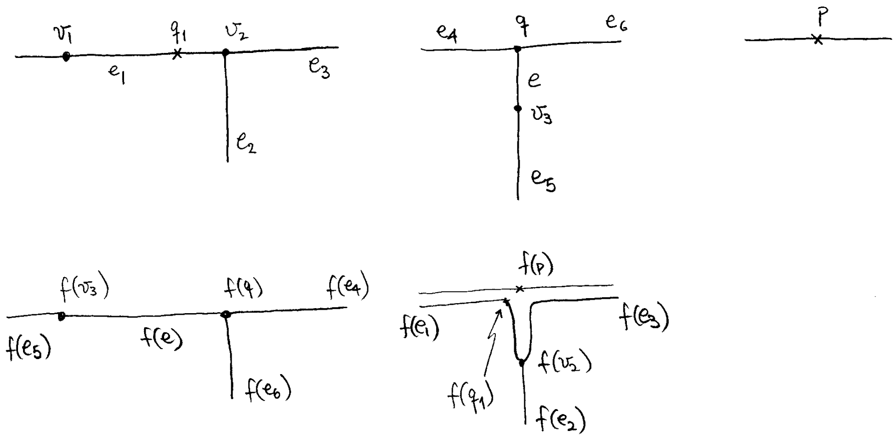

Figure 34: The result of kneading a maximal 2-kneading set

prunings. There is a way to avoid this, however, by pulling both p and q away at the same time, instead of starting with p, which made it impossible to pull away q afterwards. The following argument establishes that there exists a 3-kneading set which contains q and all preimages of p.

Notice that there is a period 2 orbit, one of whose two points x_1, x_2 lie in e_1 and e_4 respectively, x_2 lying between q and v_3. Let $K' \subseteq e_4$ be the symmetric interval about q with right hand endpoint x_2 (Figure 35). It is not hard to see that K' is a 3-kneading set for f. By Theorem 3.11, there exists a 3-pruning disk which projects to a kneading set containing K'. After pruning it away (Figure 36) there are no longer any turning points to (v_1, e_1).

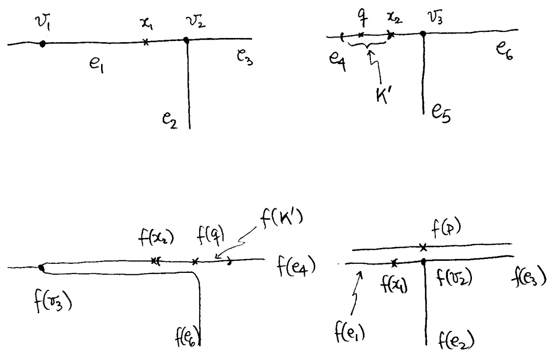

Figure 35: A better choice of kneading set

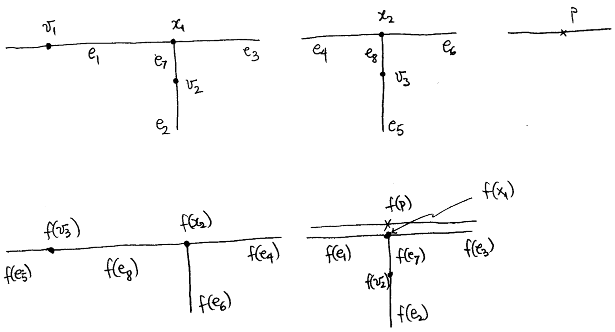

Figure 36: No remaining turning points to v

The common feature in the two examples described in this section is that pulling away an n-turning point to (v, e) creates a new $(n+1)$-turning point to (v, e). It has been explained that both situations give rise to infinite sequences of prunings: however, these sequences converge, and in the second example can in fact be dealt with by a finite sequence of prunings.

References

[1] Alves, J. and Sousa Ramos, J. "Kneading theory for tree maps." Instituto Superior Técnico Preprint 33/2000 (2000).

[2] Baillif, M. "Dynamical zeta functions for tree maps." *Nonlinearity* **12** (1999), 1511–1529.

[3] Barge, M. and Martin, J. "The construction of global attractors." *Proc. Amer. Math. Soc.* **110** (1990), 523–525.

[4] Benedicks, M. and Carleson, L. "The dynamics of the Hénon map." *Ann. of Math. (2)* **133** (1991), 73–169.

[5] Bestvina, M. and Handel, M. "Train-tracks for surface homeomorphisms." *Topology* **34** (1995), 109–140.

[6] Boyland, P. "Braid types and a topological method of proving positive entropy." *Unpublished* (1984).

[7] de Carvalho, A. "Pruning fronts and the formation of horseshoes." *Ergodic Theory Dynam. Systems* **19** (1999), 851–894.

[8] Cvitanović, P., Gunaratne, G., and Procaccia, I. "Topological and metric properties of Hénon-type strange attractors." *Phys. Rev. A (3)* **38** (1988), 1503–1520.

[9] Devaney, R. "An introduction to chaotic dynamical systems." Second edition. Addison-Wesley Studies in Nonlinearity. Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1989.

[10] Franks, J. and Misiurewicz, M. "Cycles for disk homeomorphisms and thick trees." Nielsen theory and dynamical systems (South Hadley, MA, 1992), 69–139, *Contemp. Math.* **152**, Amer. Math. Soc., Providence, RI, 1993.

[11] Hall, T. "The creation of horseshoes." *Nonlinearity* **7** (1994), 861–924.

[12] Hall, T. "Trains, an implementation of the Bestvina-Handel algorithm for finding train-tracks of surface automorphisms". Software available from `http://www.liv.ac.uk/maths/PURE`.

[13] Holmes, P. and Whitley, D. "Bifurcations of one- and two-dimensional maps." *Philos. Trans. Roy. Soc. London Ser. A* **311** (1984), 43–102

[14] Holmes, P. and Williams, R. "Knotted periodic orbits in suspensions of Smale's horseshoe: torus knots and bifurcation sequences." *Arch. Rational Mech. Anal.* **90** (1985), 115–194.

[15] Ishii, Y. "Towards a kneading theory for Lozi mappings. I. A solution of the pruning front conjecture and the first tangency problem." *Nonlinearity* **10** (1997), 731–747.

[16] Los, J. "Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm." *Proc. London Math. Soc.* **66** (1993), 400–430.

[17] Milnor, J. and Thurston, W. "On iterated maps of the interval." Dynamical systems (College Park, MD, 1986–87), 465–563, Lecture Notes in Math., 1342, Springer, Berlin-New York, 1988.

[18] Shub, M., Fathi, A. and Langevin, R. "Global stability of dynamical systems." Translated from the French by Joseph Christy. Springer-Verlag, New York-Berlin, 1987.

[19] Smale, S. "Differentiable dynamical systems." *Bull. Amer. Math. Soc.* **73** (1967), 747–817.

[20] van Strien, S. "On the bifurcations creating horseshoes." Dynamical systems and turbulence, Warwick 1980 (Coventry, 1979/1980), pp. 316–351, Lecture Notes in Math., 898, Springer, Berlin-New York, 1981.

[21] Thurston, W. "On the geometry and dynamics of diffeomorphisms of surfaces." *Bull. Amer. Math. Soc. (N.S.)* **19** (1988), 417–431.

[22] Williams, R. "Classification of one dimensional attractors." 1970 Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968) pp. 341–361 Amer. Math. Soc., Providence, R.I.