SYMBOLIC DYNAMICS AND TOPOLOGICAL MODELS IN DIMENSIONS 1 AND 2

ANDRÉ DE CARVALHO AND TOBY HALL

1. Introduction

Symbolic dynamics has a long and distinguished history, going back to Hadamard's work on the geodesic flow on negatively curved surfaces. Because of its success in describing the dynamics of systems with Markov partitions, it is most closely associated with the study of such systems. However, in the 1970's, beginning with the work of Metropolis, Stein and Stein and culminating with the *kneading theory* of Milnor and Thurston, symbolic dynamics was used in a quite different way in the study of 1-dimensional discrete dynamics: the partition was fixed (given by the critical points) and the maps were allowed to vary in families, as long as the critical points remained the same. This is fundamentally different from the use of symbolic dynamics in the presence of Markov partitions. Not only does it not require the maps being studied to have such partitions, but also, by fixing the partition, it permits the description of all maps in a family in terms of the same symbols, thus allowing the comparison of different maps.

One of the conclusions of kneading theory is that every family of *unimodal maps* (i.e. continuous piecewise monotone self-maps of the unit interval with exactly two monotone pieces) presents essentially the same dynamical behaviour as it passes from trivial to chaotic dynamics. This is sometimes called *topological universality* for 1-dimensional dynamics. In dimension 2 the situation is more complicated. In this article we present some topological aspects of kneading theory together with an analogous 2-dimensional theory called *pruning*. The ideas are presented in such a way as to emphasize the parallels between the two theories. The concept of pruning was introduced by Cvitanović to describe families of 2-dimensional diffeomorphisms (or homeomorphisms) such as the Hénon family. Unlike their 1-dimensional counterparts, such families do not have critical points and usually do not have any obvious substitutes (much of the hard work in some papers about the Hénon family (e.g. [BC91, WY01]) consists in finding parameter values for which there are such substitutes.) This absence of critical points is one of the main reasons for the additional complexity of the theory in dimension 2. The intention of this article is to give an outline of both theories, omitting many of the technical details, in order to bring out those analogies which do exist. There are many good detailed accounts of kneading theory in the literature (e.g. [MT88, dMvS93]): the approach based on abstract models taken in this article, though folkloric, does not seem to have appeared in print before. For more detail on pruning, see [Cvi91, CGP88, dCH, dCH01, dCH02].

Table 1 indicates the correspondence between the relevant parallel concepts in dimensions 1 and 2.

2. Kneading Theory

2.1. **Unimodal maps.** In this section the basic concepts of kneading theory for unimodal maps are presented. We omit those details which are not essential to the present discussion.

Date: May 2002.

Dimension 1	Dimension 2
$f_\mu(x) = \mu x(1-x), \mu > 4$	Horseshoe map
Kneading	Pruning
Kneading sequence	Pruning front
Unimodal order	Ordering along stable and unstable manifolds
0-entropy equivalence for endomorphisms	0-entropy equivalence for homeomorphisms
Piecewise affine maps with constant slope	Generalized pseudo-Anosov homeomorphisms
Permutation of a periodic orbit	Braid type of a periodic orbit

TABLE 1. Parallel concepts in the 1- and 2-dimensional theories

More complete accounts of kneading theory can be found in [MT88, Dev89, dMvS93, KH95].

Definitions 1. Let $I = [a, b]$. A *unimodal map* on I is a continuous map $f\colon I \to \mathbb{R}$ with $f(a) = f(b) = a$, for which there exists $c \in (a, b)$ such that f is increasing on $[a, c]$ and decreasing on $[c, b]$ (not necessarily strictly). The point c is the *critical* or *turning point* of f, and $f(c)$ is its *critical value.*

Given a unimodal map $f\colon I \to \mathbb{R}$, we are interested from a dynamical point of view in the set

$$\Lambda^+ = \Lambda^+_f = \{x \in I;\ f^n(x) \in I \text{ for all } n \in \mathbb{N}\}$$

of points whose orbits remain within I. Clearly f restricts to a continuous map $f\colon \Lambda^+ \circlearrowleft$.

Example 1. The best-known example of a family of unimodal maps is the *logistic family* $\{f_\mu\}_{\mu>0}$, defined on $I = [0, 1]$ by

$$f_\mu(x) = \mu x(1-x).$$

Here $c = 1/2$ for all μ, and $f_\mu(I) \subseteq I$ if and only if $\mu \leq 4$.

Symbolic dynamics is introduced by defining the *itinerary* of a point $x \in \Lambda^+$ to be the sequence

$$i(x) = s_0 s_1 s_2 \ldots \in \{0, C, 1\}^{\mathbb{N}}$$

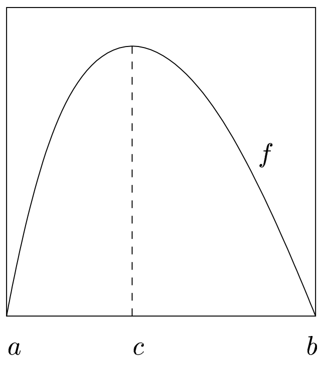

FIGURE 1. A unimodal map.

given by

$$s_j = \begin{cases} 0 & \text{if } f^j(x) < c, \\ C & \text{if } f^j(x) = c, \\ 1 & \text{if } f^j(x) > c. \end{cases}$$

This defines a map $i\colon \Lambda^+ \to \{0, C, 1\}^{\mathbb{N}}$ with the property that $i \circ f = \sigma \circ i$, where $\sigma\colon \{0, C, 1\}^{\mathbb{N}} \circlearrowleft$ is the *shift map* given by $\sigma(s_0 s_1 s_2 \dots) = s_1 s_2 s_3 \dots$. The symbol C can only occur in itineraries if $f(I) \subseteq I$.

Example 2. Consider a logistic map f_μ with $\mu > 4$, so that $I = [0, 1]$ is not invariant under f_μ: points of the interval $A = (1/2 - \epsilon, 1/2 + \epsilon)$, where $\epsilon = \sqrt{\frac{\mu-4}{4\mu}}$, have images outside of I. The preimage $f_\mu^{-1}(A)$ consists of two smaller intervals A_0 and A_1 to the left and right of A: points in these intervals are mapped out of I by f_μ^2. More generally, the set of points of I which are mapped out of I by f_μ^n but not by f_μ^{n-1} consists of 2^n open intervals with disjoint closures. This construction of 'excluded' intervals is similar to that of the middle-thirds Cantor set, and indeed it can be shown that $\Lambda^+_{f_\mu}$ is a Cantor set. (For μ large enough that $|f'_\mu(x)| > 1$ for $x \notin A$, it is straightforward to prove this statement. The proof in the general case is more subtle.) In this case the itinerary map $i\colon \Lambda^+_{f_\mu} \to \{0, 1\}^{\mathbb{N}}$ is a homeomorphism which conjugates f_μ and the shift map.

The itinerary of the critical value of a unimodal map $f\colon I \to \mathbb{R}$ with $f(I) \subseteq I$ plays a particularly important role in kneading theory.

Definition 2. The *kneading sequence* $\kappa(f)$ of a unimodal map $f\colon I \to \mathbb{R}$ with $f(I) \subseteq I$ is the itinerary $\kappa(f) = i(f(c))$ of its critical value.

Notice that $\kappa(f)$ describes the tail of the itinerary of any preimage of c: if $\sigma^n(i(x)) = C \dots$ for any $n \in \mathbb{N}$, then $\sigma^{n+1}(i(x)) = \kappa(f)$.

In Section 2.3 it will be seen that the case where $\kappa(f)$ is *periodic* (i.e. $\kappa(f) = \overline{w}$, the infinite repetition of some word $w \in \{0, C, 1\}^n$) must be treated separately from the case where it is not periodic.

2.2. **The unimodal order.** The unimodal order $\prec$ is a linear order on $\{0, C, 1\}^{\mathbb{N}}$ which reflects the usual ordering of points in I. A finite word $w \in \{0, 1\}^n$ is *even* or *odd* according as the sum of its entries is even or odd. Let $s = s_0 s_1 \dots, t = t_0 t_1 \dots \in \{0, C, 1\}^{\mathbb{N}}$ and assume that $s_j = t_j \neq C$ for all $j < n$ and that $s_n \neq t_n$ (note that if s and t are itineraries of points under a given unimodal map with $s_j = t_j = C$, then $s_n = t_n$ for all $n > j$). Set $0 < C < 1$ and define

$$s \prec t \text{ if } \quad \begin{cases} s_0 s_1 \dots s_{n-1} & \text{is even and } s_n < t_n \text{ or} \\ s_0 s_1 \dots s_{n-1} & \text{is odd and } s_n > t_n. \end{cases}$$

The following straightforward lemma states that the unimodal order does indeed reflect the usual order on I:

Lemma 2.1. *Let $f\colon I \to \mathbb{R}$ be a unimodal map, and $x, y \in \Lambda^+_f$. If $i(x) \prec i(y)$ then $x < y$.*

Of course it follows immediately that if $x < y$ then $i(x) \preceq i(y)$: the possibility of equality in this partial converse cannot be excluded. Thus if we define

$$I(x) = \{y \in \Lambda^+_f;\ i(y) = i(x)\},$$

for $x \in \Lambda^+_f$, then $I(x)$ is an interval (possibly degenerate) containing x on which all iterates of f are monotone, and it is maximal with respect to these properties.

The next step is to describe the set of all itineraries of a given unimodal map $f\colon I \circlearrowleft$ (for the remainder of this section, we restrict to the case $f(I) \subseteq I$). Notice that if $s = s_0 s_1 \ldots$ is the itinerary of some point $x \in I$ then it must satisfy the following conditions:

1) If $s_n = C$ then $\sigma^{n+1}(s) = \kappa(f)$
2) $\sigma^n(s) \preceq \kappa(f)$ for all $n \geq 1$.

Condition 2) is a consequence of the observation that, for a unimodal map, the image of the critical point is greater than or equal to the image of any other point in the interval.

If the inequality in 2) is strict, these two conditions are also sufficient for the existence of a point $x \in I$ with itinerary s:

Theorem 2.2. *Let $f\colon I \circlearrowleft$ be a unimodal map, and suppose that $\kappa(f)$ is not periodic. Suppose $s = s_0 s_1 \ldots \in \{0, C, 1\}^{\mathbb{N}}$ satisfies*

1) If $s_n = C$ then $\sigma^{n+1}(s) = \kappa(f)$, and
2) $\sigma^n(s) \prec \kappa(f)$ for all $n \geq 1$.

Then there exists $x \in I$ such that $i(x) = s$.

Remarks.

a) Inequality in 2) is necessary: if $f(x) = f_4(x) = 4x(1-x)$, the sequences $s = 01000\ldots$ and $s = 11000\ldots$ both satisfy $\sigma(s) = \kappa(f_4) = 1000\ldots$, but neither is the itinerary of a point under f_4.
b) With some small provisos, the statement also holds for periodic kneading sequences (see [Dev89]).

For a proof, see for example [Dev89].

2.3. **Abstract synthetic models.** In this section we develop an abstract symbolic discussion and use it to construct synthetic models for unimodal maps.

Definitions 3. Let Σ^+ denote the 1-sided 2-shift space $\{0,1\}^{\mathbb{N}}$. An element $\kappa \in \Sigma^+$ is a *kneading sequence* if it satisfies

$$\sigma^n(\kappa) \preceq \kappa \text{ for all } n \geq 0.$$

If κ is a kneading sequence, define

$$\Sigma_\kappa^+ = \{s \in \Sigma^+;\ \sigma^n(s) \preceq \kappa \text{ for all } n \geq 1\}.$$

Σ_κ^+ is clearly σ-invariant. We also denote by σ the restriction of the shift map to Σ_κ^+.

We now introduce an equivalence relation $\sim_\kappa$ on Σ_κ^+ by setting

$$s_0 \ldots s_{n-1} 0 \kappa \sim s_0 \ldots s_{n-1} 1 \kappa$$

for all $s_0 \ldots s_{n-1}$ for which these sequences lie in Σ_κ^+, and extending to the largest equivalence relation (i.e. with smallest equivalence classes) $\sim_\kappa$ with closed equivalence classes. Let I_κ denote the quotient space $\Sigma_\kappa^+/{\sim_\kappa}$. If $\kappa = \overline{s_0 \ldots s_n}$ is periodic, define $\kappa(C) = \overline{s_0 \ldots s_{n-1} C}$ and

$$\tilde{\kappa} = \max\{\gamma \in \Sigma^+;\ \gamma \text{ is a periodic even kneading sequence}, \gamma \preceq \kappa\}$$

(where the maximum is taken with respect to the unimodal order).

Theorem 2.3. *I_κ is either a point or an interval. If it is an interval, then $\sigma\colon \Sigma_\kappa^+ \circlearrowleft$ projects to a unimodal map $\phi_\kappa\colon I_\kappa \circlearrowleft$. Moreover, the kneading sequence $\kappa(\phi_\kappa)$ of ϕ_κ is given by*

$$\kappa(\phi_\kappa) = \begin{cases} \kappa & \text{if } \kappa \text{ is not periodic} \\ \kappa(C) & \text{if } \kappa \text{ is periodic even} \\ \tilde{\kappa}(C) & \text{if } \kappa \text{ is periodic odd.} \end{cases}$$

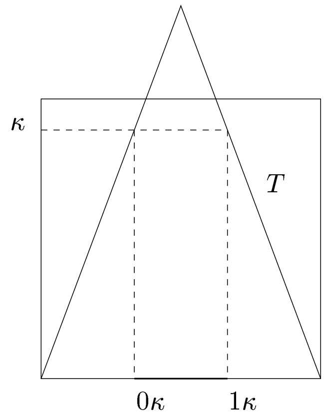

FIGURE 2. The first collapsed interval $A_0 = (0\kappa, 1\kappa)$.

Sketch Proof. Using the itinerary map, identify $\sigma\colon \Sigma^+ \circlearrowleft$ with the restriction of the tent map $T\colon I \to \mathbb{R}$ defined by

$$T(x) = \begin{cases} 3x & 0 \le x \le 1/2 \\ 3 - 3x & 1/2 \le x \le 1 \end{cases}$$

to its invariant Cantor set Λ_T^+ (which is precisely the standard middle-thirds Cantor set). From now on we will not distinguish between σ and T, or between points of Σ^+ and points of $\Lambda_T^+ \subseteq I$.

By definition, Σ_κ^+ is obtained from Σ^+ by removing $A_0 = T^{-1}((\kappa, \infty))$ and all of its preimages. Notice that $A_0 = (0\kappa, 1\kappa)$ (Figure 2). We can thus think of the equivalence relation $\sim_\kappa$ as an equivalence relation on the interval I which collapses to points the closures of the connected components of A_0 and its preimages. It is therefore a monotone decomposition of the interval and, since it has closed equivalence classes (by definition), the quotient space I_κ is either a point or an interval. Since the equivalence is σ-invariant, σ induces a continuous map $\phi_\kappa\colon I_\kappa \circlearrowleft$. That ϕ_κ is unimodal follows from the monotonicity of the canonical projection $\pi\colon I \to I_\kappa$ (i.e. preimages of points are connected), and the unimodality of the tent map T. Moreover, denoting by $[s]$ the $\sim_\kappa$ equivalence class of $s \in \Sigma_\kappa^+$, the critical point of ϕ_κ is $[0\kappa] = [1\kappa]$, and its critical value is $[\kappa]$.

In order to determine the kneading sequence of ϕ_κ, it is necessary to consider the way in which Σ_κ^+ depends on κ:

Lemma 2.4. *Σ_κ^+ has isolated points if and only if κ is periodic odd.*

Sketch Proof. Because κ is a kneading sequence, connected components of preimages of A_0 under T are either contained in or disjoint from A_0. Since these components are open intervals, it follows that Σ_κ^+ has isolated points if and only if A_0 shares an endpoint with some component, not contained in A_0, of one of its preimages. This can happen if and only if κ is periodic (with period n, say) and has the property that nearby points exchange sides under the n-th iterate of T. This is equivalent to saying κ is periodic odd. ∎

Assume first that Σ_κ^+ has no isolated points. Then it is a Cantor set (since it is certainly compact and totally disconnected), and $\sim_\kappa$ is the equivalence which identifies two points if and only if they are the endpoints of a connected component of $[0,1] \setminus \Sigma_\kappa^+$, i.e. points of the form

$$s_0 \dots s_{n-1} 0\kappa \sim_\kappa s_0 \dots s_{n-1} 1\kappa.$$

Thus equivalence classes consist either of a single point (if the point is not a preimage of κ), or of exactly two points of the above form. In this situation, it is straightforward to describe the itinerary map i associated with $\phi_\kappa\colon I_\kappa \circlearrowleft$. If s is not a preimage of κ under σ, then $i([s]) = s$. Otherwise, there are two possibilities: if κ is not periodic, then

$$i([s_0\dots s_n\kappa]) = s_0\dots s_{n-1}C\kappa;$$

while if $\kappa = \overline{t_0\dots t_k}$, then

$$i([s_0\dots s_n\kappa]) = i([s_0\dots s_n\overline{t_0\dots t_k}]) = s_0\dots s_{n-1}\overline{Ct_0\dots t_{k-1}}$$

(since the critical point $[0\kappa] = [1\kappa]$ of ϕ_κ is periodic).

In particular, the itinerary of the critical value $[\kappa]$ is either κ (if κ is not periodic), or $\overline{t_0\dots t_{k-1}C}$ if $\kappa = \overline{t_0\dots t_k}$ is periodic even. This deals with the first two possibilities for $\kappa(\phi_\kappa)$ in the statement of the theorem.

The final case is where κ is periodic odd: it is illustrated by Examples 3–5, and the reader is encouraged to refer to them while reading the following argument. Suppose then that $\kappa = \overline{s_1\dots s_n}$ is periodic odd. Then $\kappa \succ (s_1\dots\hat{s}_n)^k\overline{s_1\dots s_n}$, for any $k > 0$ (where we write $\hat{0} = 1$ and $\hat{1} = 0$). Therefore

$$\begin{array}{rcl}
\kappa & = & s_1\dots s_n\overline{s_1\dots s_n} \\
 & \sim_\kappa & s_1\dots\hat{s}_n\overline{s_1\dots s_n} \\
 & \vdots & \\
 & \sim_\kappa & (s_1\dots\hat{s}_n)^k\overline{s_1\dots s_n} \\
 & \vdots & \\
 & \sim_\kappa & \overline{s_1\dots\hat{s}_n}
\end{array}$$

where the final equivalence is due to the equivalence classes being closed and the fact that

$$(s_1\dots\hat{s}_n)^k\overline{s_1\dots s_n} \to \overline{s_1\dots\hat{s}_n}, \text{ as } k\to\infty.$$

The sequence $\kappa_1 = \overline{s_1\dots\hat{s}_n}$ is clearly periodic even and it is also a kneading sequence. It may be the case, however, that $s_1\dots\hat{s}_n = (s_1\dots s_{j_1})^{n/j_1}$ (see examples below). If this is the case and $\kappa_1 = \overline{s_1\dots s_{j_1}}$ is odd, the same argument shows that $\kappa_1 \sim \overline{s_1\dots\hat{s}_{j_1}}$. Continuing this process produces finitely many periodic odd kneading sequences $\kappa_1,\dots,\kappa_{m-1}$, with (minimal) periods satisfying $j_\ell \,|\, j_{\ell-1}$ for $\ell = 1,\dots,m-1$, until we arrive at a periodic even kneading sequence $\kappa_m = \tilde{\kappa}$ (which is not a repetition of a shorter word) and

$$\kappa = \kappa_0 \sim_\kappa \kappa_1 \sim_\kappa \dots \sim_\kappa \kappa_{m-1} \sim_\kappa \kappa_m = \tilde{\kappa}.$$

That $\tilde{\kappa}$ is the largest periodic even kneading sequence smaller than κ follows from the observation that there exists no kneading sequence strictly between two kneading sequences of the forms $\overline{t_1\dots t_n}$ and $\overline{t_1\dots\hat{t}_n}$.

This argument shows that the $\sim_\kappa$ equivalence class of κ consists of the points

$$\kappa, \kappa_1, \dots, \kappa_{m-1}, \tilde{\kappa}$$

together with some transitional orbits (see examples below). The points in this equivalence class are precisely those $s \in \Sigma_\kappa^+$ for which $s \succeq \tilde{\kappa}$. It follows that every $\sim_\kappa$ equivalence class contains a point of $\Sigma_{\tilde{\kappa}}^+$: $\Sigma_\kappa^+ \setminus \Sigma_{\tilde{\kappa}}^+$ consists of a countable collection of points contained in the complementary components of $\Sigma_{\tilde{\kappa}}^+$ in I; such points are exactly the preimages under T of those $s \in \Sigma_\kappa^+$ for which $s \succeq \tilde{\kappa}$; if A is an open interval in $I \setminus \Sigma_{\tilde{\kappa}}^+$, then all points in the closure of $A \cap \Sigma_\kappa^+$ are $\sim_\kappa$-equivalent and this closure contains both endpoints of A. Therefore, $\sim_\kappa$ 'collapses' the difference between Σ_κ^+ and $\Sigma_{\tilde{\kappa}}^+$. From this it follows that I_κ and $I_{\tilde{\kappa}}$ are homeomorphic, that ϕ_κ and $\phi_{\tilde{\kappa}}$ are topologically conjugate and, therefore, that the kneading sequences of both are equal to $\tilde{\kappa}(C)$. ∎

The examples that follow clarify the last possibility in the statement of Theorem 2.3.

Example 3. Consider first some simple examples:

i) If $\kappa = \overline{0}$ then $\Sigma^+_\kappa = \{\overline{0}\}$.
ii) If $\kappa = \overline{1}$ then $\Sigma^+_\kappa = \{\overline{0}, 0^k\overline{1}; k \geq 0\} = \Sigma^+_{\overline{0}} \cup \{0^k\overline{1}; k \geq 0\}$.
iii) If $\kappa = \overline{10}$ then $\Sigma^+_\kappa = \{\overline{0}, 0^k\overline{1}, 0^k1^\ell\overline{10}; k, l \geq 0\} = \Sigma^+_{\overline{1}} \cup \{0^k1^\ell\overline{10}; k, l \geq 0\}$.

Although these are trivial examples from the point of view of the theorem — the quotient spaces I_κ are single points — they are representative of the general case: as κ increases from $\overline{0}$ to $\overline{1}$ to $\overline{10}$, the orbits which existed in a previous stage (e.g. $\overline{0}$) cease to be isolated, a new isolated periodic orbit (e.g. $\overline{1}$) is born, and transitional orbits (e.g. $0^k\overline{1}$) are created.

Example 4. For a more complicated example, in which I_κ is a nondegenerate interval, note that $\Sigma^+_{\overline{101}}$ is a Cantor set (since $\overline{101}$ is periodic even). It is easy to check that the sequences $s \in \Sigma^+_{\overline{100}}$ satisfying $\overline{101} \prec s \prec \overline{100}$ are of the form $(101)^k\overline{100}$, for $k \geq 1$. Notice that such points form a discrete set in $\Sigma^+_{\overline{100}}$ accumulating on $\overline{101}$ as $k \to \infty$. From this it follows that $\Sigma^+_{\overline{100}} \setminus \Sigma^+_{\overline{101}}$ is discrete.

Consider the equivalence relation $\sim = \sim_{\overline{100}}$. By definition, it satisfies

$$\overline{100} = 100\,\overline{100} \sim 101\,\overline{100} = 101\,100\,\overline{100} \sim 101\,101\,\overline{100} \sim \cdots \sim (101)^k\,\overline{100}.$$

Since the points of the form $(101)^k\overline{100}$ accumulate on $\overline{101}$ as $k \to \infty$ and the equivalence classes are closed, $\overline{101}$ is also equivalent to all of the points of this form. Since $\overline{101}$ is alone in its $\sim_{\overline{101}}$ equivalence class, and $\Sigma^+_{\overline{100}} \supset \Sigma^+_{\overline{101}}$ it follows that

$$\left\{(101)^k\overline{100}, k \geq 0\right\} \cup \left\{\overline{101}\right\}$$

is a complete equivalence class and that $\phi_{\overline{100}}$ is topologically conjugate to $\phi_{\overline{101}}$: their kneading sequence is $\overline{10C}$.

Example 5. The final example is one in which two steps taking closures are involved. Let $\kappa = \overline{100101}$. Then the points $s \in \Sigma^+_\kappa$ with $s \succeq \overline{101}$ are of the form

$$(101)^k\overline{100} \text{ and } (101)^k(100)^\ell\overline{100101} \text{ for } k, \ell \geq 0.$$

Here points of the form $(101)^k(100)^\ell\overline{100101}$ are isolated for all $k, \ell \geq 0$, but those of the form $(101)^k\overline{100}$ are accumulation points of $(101)^k(100)^\ell\overline{100101}$ as $\ell \to \infty$. Reasoning as in the previous example, we see that

$$\overline{100101} \sim (100)^\ell\overline{100101} \sim (101)^k(100)^\ell\overline{100101} \sim \overline{100} \sim \overline{101},$$

so that $\phi_{\overline{100101}}$ is also conjugate to $\phi_{\overline{101}}$.

Remark. While the symbolic approach taken above is rather messy, the results are beautifully explained by renormalization theory [CE80].

We can now compare these abstractly constructed maps to unimodal maps in general. Let $f\colon I \circlearrowleft$ be a unimodal map and assume first that its turning point is not periodic. Then the kneading sequence $\kappa = \kappa(f)$ is an element of Σ^+ and we can go through the construction above to obtain a unimodal map $\phi_\kappa\colon I_\kappa \circlearrowleft$. If the turning point of f is periodic, then $\kappa = \overline{s_0 \dots s_{n-1}C}$. The discussion above illustrates that changing the last symbol of κ to 0 or 1 produces the same quotient map: $\phi_{\kappa(0)} = \phi_{\kappa(1)}$ (where $\kappa(0), \kappa(1)$ denote the kneading sequences obtained changing the last symbol of κ to 0, 1 respectively). Thus the theorem below is quite general.

Theorem 2.5. *Let $f\colon I \circlearrowleft$ be a unimodal map with turning point c. If c is not periodic then let $\kappa = \kappa(f)$, and if c is periodic then let κ be the periodic sequence obtained by changing the symbol C in $\kappa(f)$ to either 0 or 1. Then the composition $I \to \Sigma^+_\kappa \to I_\kappa$ of the itinerary map with the quotient map is a semi-conjugacy from $f\colon I \circlearrowleft$ to $\phi_\kappa\colon I_\kappa \circlearrowleft$.*

Thus every unimodal map $f\colon I \circlearrowleft$ can be modelled by a quotient of a subshift of the 1-sided 2-shift. The reader who has followed the arguments presented so far should have no difficulty providing a proof of this theorem. Observe that the fibres of the semi-conjugacy $I \to \Sigma^+_\kappa \to I_\kappa$ in the statement are the equivalence classes of the largest equivalence relation on the interval whose equivalence classes are closed and either contain or are disjoint from intervals $I(x)$ of points with the same itinerary as x.

3. Pruning

We now start to consider 2-dimensional analogues of the theory presented in Section 2. There are, however, some fundamental differences between the two cases and it is not possible to define exact analogues of several concepts. In particular, the very definition of a '2-dimensional unimodal map' is lacking. The Hénon family $\{H_{a,b}\}_{a,b\in\mathbb{R}}$ defined by

$$H_{a,b}(x,y) = (f_\mu(x) - by, x)$$

is a family of diffeomorphisms of $\mathbb{R}^2$ which is a 2-dimensional counterpart of the logistic family; however, since the Hénon maps are invertible, they do not have any defining topological trait which entitles them to be called 'unimodal'. Nonetheless, it is possible to develop a number of analogies despite such basic difficulties. We will concentrate on orientation-preserving Hénon maps (i.e. we assume that $b > 0$).

3.1. **The horseshoe.** The horsehoe map was introduced by Smale [Sma67]. In the context of this article, it should be thought of as a 2-dimensional analogue of the logistic map f_μ with $\mu > 4$. The horseshoe is a homeomorphism $F\colon S^2 \circlearrowleft$ of the sphere, all of whose interesting dynamics is contained in the stadium-shaped region $\mathbb{I}$ which is the union of a central square S and two half-disks as shown in Figure 3. It maps $\mathbb{I}$ into itself in a horseshoe shape, mapping vertical and horizontal line segments in $S \cap F^{-1}(S)$ into vertical and horizontal line segments respectively. In each component V_0 and V_1 of $S \cap F^{-1}(S)$ (Figure 4) the horseshoe is linear, contracting vertical segments by the factor $1/3$ and expanding horizontal segments by the factor 3. There is an attracting fixed point in the half-disk on the left denoted by x (shown as $\blacksquare$) and the point at infinity in S^2 is a repeller whose basin contains all points outside $\mathbb{I}$.

Notice that collapsing vertical segments and the two half-disks of $\mathbb{I}$ to points we obtain an interval and the horseshoe projects to a 'flat top' tent map $f\colon I \circlearrowleft$. This is also shown in Figure 3: the image $f(I)$ is shown slightly separated so it is possible to see what f does to the interval.

We now describe the dynamics of F inside the square S: see for example [Dev89] for a more detailed description. In Figure 5, $F^n(S) \cup F^{-n}(S)$ are shown for $n = 1, 2$. The intersection $F^n(S) \cap F^{-n}(S)$ consists precisely of the set of points which remain inside S for n iterates of both F and F^{-1}. In general, $F^n(S) \cap F^{-n}(S)$ consists of 2^{2n} squares of side length 3^{-n}. The set of points all of whose iterates stay in the square is

$$\Lambda = \bigcap_{n=0}^{\infty} F^n(S) \cap F^{-n}(S).$$

Λ is a Cantor set and, as in Section 2, we define an itinerary map $i\colon \Lambda \to \{0,1\}^{\mathbb{Z}}$ by setting

$$i(x) = \begin{cases} 0 & \text{if } F(x) \in V_0 \\ 1 & \text{if } F(x) \in V_1 \end{cases}$$

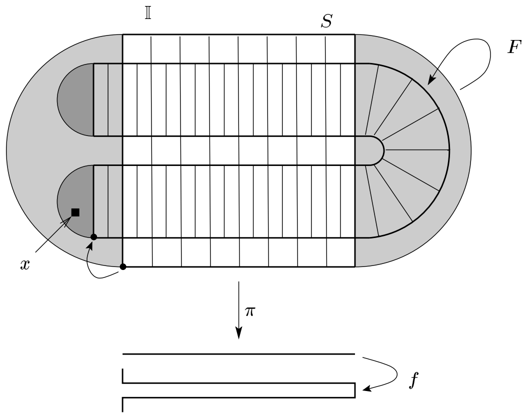

FIGURE 3. The horseshoe map.

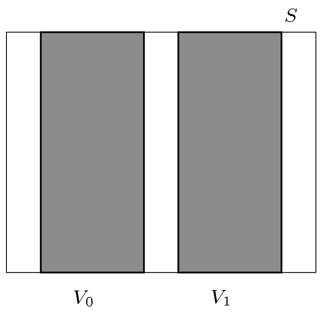

FIGURE 4. The vertical strips of $S \cap F^{-1}(S)$.

As in Example 2, the itinerary map conjugates the restriction of F to Λ with the shift map $\sigma\colon \{0,1\}^{\mathbb{Z}} \circlearrowleft$, defined by

$$\sigma(\ldots s_{-2}s_{-1} \cdot s_0 s_1 \ldots) = \ldots s_{-1}s_0 \cdot s_1 s_2 \ldots$$

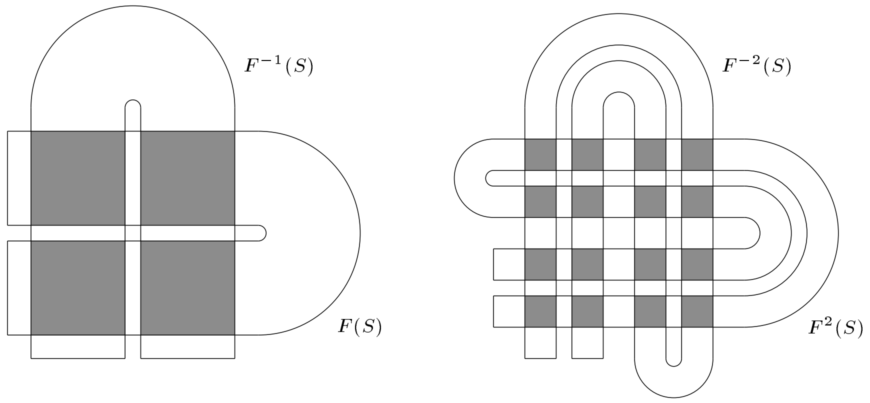

FIGURE 5. First steps in the construction of the horseshoe-invariant Cantor set

3.2. **Models abstract and concrete.** The models in Section 2.3 were constructed starting from symbolic dynamics. It was then shown that they model actual unimodal maps. In dimension 2, it is necessary to build models which are homeomorphisms of the sphere. Once this is done, it is possible to describe them using symbolic dynamics just as in dimension 1. Because there are more technical problems in dimension 2, symbolic dynamics will be relegated to a secondary role. We first state the pruning theorem [dC99], which describes how it is possible to destroy the dynamics of a surface homeomorphism in a controlled way by isotopy. Our models will consist of all such *prunings* of the horseshoe and will form a family depending on infinitely many parameters.

Let $F\colon S \circlearrowleft$ be an orientation-preserving homeomorphism of an orientable surface S (which in this paper will always be taken to be $\mathbb{R}^2$ or S^2). The pruning theorem 3.1 gives conditions on a topological disk $D \subseteq S$ (the *pruning conditions*) which ensure that it is possible to perform an isotopy of F which destroys all of the dynamics in the (orbit of the) interior of D, while leaving the dynamics unchanged elsewhere. If this theorem is to be applied repeatedly then it is essential that it should only require continuity, since it is impossible in general for such pruning isotopies to preserve differentiability. Despite this, we start with an intuitive motivation for one of the pruning conditions in the context of diffeomorphisms.

Suppose for the time being, then, that F is a diffeomorphism. Roughly speaking, the first pruning condition for a disk D states that ∂D is the union of a segment C of stable manifold and a segment E of unstable manifold (perhaps of some periodic point of F). If this is the case, then a pruning isotopy can be thought of as uncrossing the stable and unstable manifolds within D (Figure 6), since if any such crossings remain then there are local horseshoes, and hence non-trivial dynamics, inside D. The second pruning condition, which should be thought of as a 2-dimensional analogue of the requirement that a sequence be a kneading sequence, ensures that such uncrossing is possible. To understand this condition on an intuitive level, consider the disk D depicted in Figure 7, together with an iterate $F^n(D)$. Uncrossing stable and unstable manifolds within D means uncrossing them within all of its images and preimages, including $F^n(D)$. However, the configuration of unstable manifolds in Figure 7 means that it is not possible to remove the circled crossings.

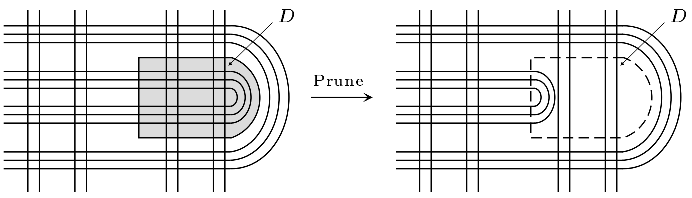

FIGURE 6. Uncrossing stable and unstable manifolds inside a pruning disk

After a little thought it can be seen that the problem in this figure arises because there is an integer $n > 0$ such that $F^n(C)$ intersects both $\mathrm{Int}(D)$ and $S \setminus D$, and by symmetry the same problem arises if $F^{-n}(E)$ intersects both $\mathrm{Int}(D)$ and $S \setminus D$ for some $n > 0$. The purpose of the second pruning condition is to prohibit such intersections. In fact, it goes further, and prohibits *any* intersection of $F^n(C)$ or $F^{-n}(E)$ with $\mathrm{Int}(D)$ for $n > 0$. Thus, in addition to bad intersections such as that of Figure 7, it also prohibits intersections where $F^n(C) \subseteq D$ or $F^{-n}(E) \subseteq D$. The reason for this is not that such intersections make it impossible to prune the dynamics in D, but rather that if they occur then D can be replaced by an equivalent disk D' for which there are no such intersections.

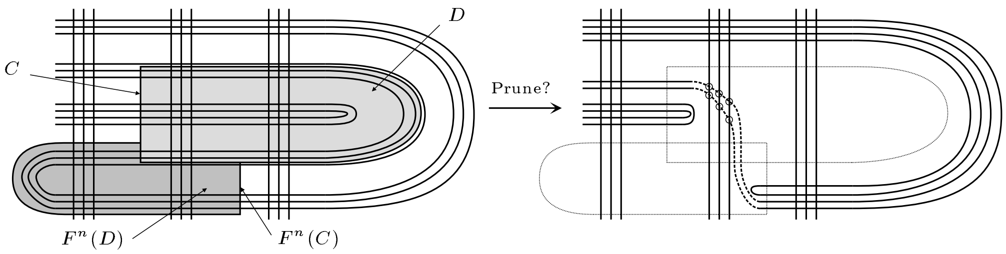

FIGURE 7. A disk which is not a pruning disk

A sketch of the construction of such a disk D' is given in Figure 8. Here $F^n(C) \subseteq \mathrm{Int}(D)$: in particular, the circled intersections to the left of C are mapped by F^n to the circled intersections to the right of $F^n(C)$, which are contained in D. Hence any isotopy which removes all of the intersections in D also removes those immediately to the left of D. The disk D can therefore be enlarged along the unstable segment of its boundary to a new disk D', without changing the set of intersections to be removed. Such an enlargement can be continued until such a time as $F^n(C') \subseteq C'$, when C' is disjoint from $\mathrm{Int}(D)$) and contains a fixed point p of F^n.

Hence all intersections of $F^n(C)$ with $\mathrm{Int}(D)$ (and analogously all those of $F^{-n}(E)$ with $\mathrm{Int}(D)$) can be prohibited by the conditions of the pruning theorem.

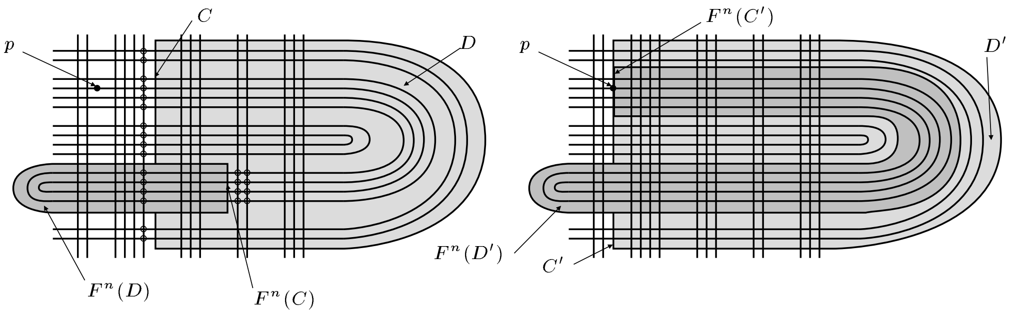

FIGURE 8. Equivalent pruning disks

Definition 4. A *pruning disk* for F is a closed topological disk D whose boundary can be written as the union of two arcs intersecting only at their endpoints, $\partial D = C \cup E$, satisfying the following *pruning conditions*:

i)
$$\begin{aligned} &\mathrm{diam}(F^n(C)) \to 0 \qquad \text{as } n \to \infty \\ &\mathrm{diam}(F^{-n}(E)) \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

ii)
$$\begin{aligned} F^n(C) \cap \mathrm{Int}(D) &= \emptyset \qquad \text{for all } n > 0 \\ F^{-n}(E) \cap \mathrm{Int}(D) &= \emptyset \qquad \text{for all } n > 0. \end{aligned}$$

Remark. When dealing with non-differentiable maps in general, condition ii) above is more technical. The interested reader can see [dC99] for more details.

Example 6. In Figure 9, two pruning disks for Smale's horseshoe map $F\colon S^2 \to S^2$ are depicted. In a), the boundary of the disk is made up of stable and unstable segments of the fixed point with itinerary $\overline{0}$, while in b) the other fixed point, with itinerary $\overline{1}$, is used. In both cases, pruning condition i) is trivially satisfied (where C is the segment of stable manifold in ∂D, and E is the segment of unstable manifold). Straightforward arguments suffice to show that pruning condition ii) is also satisfied: in a), it can be seen by a simple calculation that the forward iterates $F^n(C)$ of C (depicted with bolder lines) are all disjoint from D (since they are attracted to the fixed point, only finitely many need be considered) — a similar argument applies to the backward iterates $F^{-n}(E)$ of E; in b), all iterates $F^n(C)$ are contained in the horizontal segment through the fixed point, and all iterates $F^n(C)$ (respectively $F^{-n}(E)$) are contained in the vertical (respectively horizontal) segment through the fixed point, and hence cannot intersect $\mathrm{Int}(D)$. This argument generalizes readily to any configuration where D is bounded by segments C and E of the stable and unstable manifolds of a fixed point p, and where both the segment of stable manifold between p and C and the segment of unstable manifold between p and E are disjoint from the interior of D.

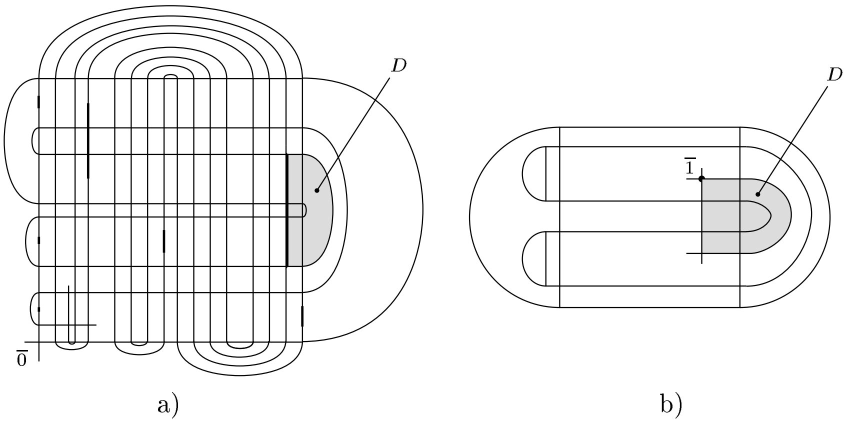

FIGURE 9. Pruning disks for the horseshoe

Theorem 3.1. *Let $F\colon S \circlearrowleft$ be a uniformly continuous orientation-preserving homeomorphism of an orientable surface, and let D be a pruning disk for F. Then there is an isotopy, supported on $U = \bigcup_{n\in\mathbb{Z}} F^n(Int(D))$, from F to a homeomorphism $F_D\colon S \circlearrowleft$ such that the non-wandering set of F_D is disjoint from U.*

The fact that the isotopy is supported in U means that the dynamics outside of U is unchanged by it. On the other hand, at the end of the isotopy there are no non-wandering points left inside U.[1] For example, in each of the two examples of Figure 9, an isotopy can be performed on the horseshoe which destroys precisely those orbits which enter the interior of the pruning disk.

Remark. A suitably generalized version of the pruning theorem holds for finite unions of pruning disks, which we refer to as *pruning fronts* (see [dC99] for details). In the language of Cvitanović [Cvi91], it would be more appropriate to call them *primary pruned regions.*

Let F be the horseshoe and D be a pruning disk for F. Denote by Δ the image of $\Lambda \cap D$ under the itinerary map. The set of nonwandering points of F_D in S is given symbolically by

$$\Sigma_D = \{s \in \{0,1\}^{\mathbb{Z}};\ \sigma^n(s) \notin \Delta \text{ for all } n \in \mathbb{Z}\}.$$

[1]The two papers [dC99] and [dCH01] have slightly different definitions of pruning fronts and of the construction of the pruning isotopy. This statement is true for the formulation of [dC99]. For that of [dCH01], all of the dynamics is destroyed with the possible exception of a single periodic orbit.

Clearly, Σ_D is σ-invariant and the map $\sigma\colon \Sigma_D \circlearrowleft$ is a direct analogue of the maps constructed in Section 2.3 using kneading sequences. Proving that an appropriately defined quotient of such a subshift is a surface homeomorphism is much harder than in dimension 1. For this reason, we construct the 2-dimensional models directly from the pruned maps F_D: this is an equivalent construction.

Definition 5. Let $\mathcal{D}$ denote the set of finite sequences $(D_1, \ldots, D_n)$ of disks with the property that D_1 is a pruning disk for F, and D_j is a pruning disk for $F_{D_1 \ldots D_{j-1}}$ for each j with $2 \le j \le n$. The *pruning family* associated to the horseshoe map F is

$$\mathcal{P} = \overline{\{F_{D_1 \ldots D_n};\, (D_1, \ldots, D_n) \in \mathcal{D}\}},$$

where the closure is taken in the uniform topology.

We now define an equivalence relation for surface homeomorphisms, the 0-*entropy equivalence relation*: it plays a role analogous to that of $\sim_\kappa$ in Section 2.3. The quotient of an interval under a closed equivalence relation with connected classes is either a point or an interval, but in dimension 2 the situation is more complex, requiring more careful definitions and more elaborate proofs.

3.2.1. *The 0-entropy equivalence relation.* We start by recalling Bowen's definition of topological entropy [Bow71]. If (X, d) is a metric space and $F\colon X \circlearrowleft$ is uniformly continuous, we say that a subset S of X is (n, ϵ)-*separated* if it is possible to distinguish between the orbits of any two points $x, y \in S$ up to $n-1$ iterates with precision ϵ: $d(F^j(x), F^j(y)) > \epsilon$ for some $0 \le j < n$. If X is compact, the topological entropy $h(F)$ of F is defined by

$$h(F) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \ln N(n, \epsilon),$$

where $N(n, \epsilon)$ denotes the maximum cardinality of an (n, ϵ)-separated subset of X. For general X and $K \subset X$ compact, counting only those orbits which start in K, we obtain the entropy of F *in* K, denoted $h_F(K)$. Thus if we denote by $N(n, \epsilon, K)$ the maximum cardinality of an (n, ϵ)-separated subset of K, then

$$h_F(K) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \ln N(n, \epsilon, K).$$

The topological entropy of F in X can then be defined as

$$h(F) = \sup\{h_F(K); K \subseteq X, K \text{ compact}\}.$$

Definition 6. Let $F\colon X \circlearrowleft$ be a homeomorphism. Then two points $x, y \in X$ are 0-*entropy equivalent* if there exists a continuum (i.e. a compact connected set) $K \subseteq X$ containing both x and y with

$$h_F(K) = 0 = h_{F^{-1}}(K).$$

Remarks.

a) That this indeed defines an equivalence relation follows from two facts: 1) $h_F(K \cup K') = \max\{h_F(K), h_F(K')\}$ and 2) the union of two continua containing a point in common is also a continuum.
b) If K is a proper subset of X, then it is not necessarily the case that $h_F(K) = h_{F^{-1}}(K)$.
c) In general, we can consider the family of equivalence relations $\sim_\alpha$, indexed by positive reals α, declaring two points to be $\sim_\alpha$-equivalent if there is a continuum containing both and carrying entropy smaller than α.
d) If $f\colon X \circlearrowleft$ is not invertible, we can define 0-entropy equivalence by omitting the condition on f^{-1}. It can be shown that the fibres of the semi-conjugacy given by Theorem 2.5 are exactly the 0-entropy equivalence classes under the unimodal map $f\colon I \circlearrowleft$.

The equivalence classes of the α equivalence relations mentioned in the previous item can also be understood for unimodal maps in terms of renormalization.

The 0-entropy equivalence relation is most interesting for two-dimensional systems.

Example 7. In this example we describe the equivalence classes for the horseshoe map F. Denote by $\mathcal{H}^u$ and $\mathcal{H}^s$ the closures of the unstable and stable manifolds of the fixed point with itinerary $\overline{0}$ (or indeed of any other periodic point, since their closures coincide) and let $\mathcal{H} = \mathcal{H}^s \cup \mathcal{H}^u$. Equivalence classes are of four kinds:

a) Closures of connected components of $S^2 \setminus \mathcal{H}$.
b) Closures of connected components of $\mathcal{H}^u \setminus \mathcal{H}^s$ (not already contained in sets in a)).
c) Closures of connected components of $\mathcal{H}^s \setminus \mathcal{H}^u$ (not already contained in sets in a)).
d) Single points which are in none of the sets in a), b) or c).

To see that these sets do not carry entropy, notice that all points in any connected component of $S^2 \setminus \mathcal{H}$ (before taking the closure) converge to the attracting fixed point x. It is not hard to see that nothing significant happens on taking the closure, so that the sets in a) indeed carry no entropy. The same holds for sets of types b) and c). To see that any larger continuum must contain entropy, notice that if C is a connected set that contains two distinct sets among the ones described above, then it must intersect a Cantor set's worth of invariant manifolds, either stable or unstable (or both). It follows that either its ω- or its α-limit set contains the whole nonwandering set of the horseshoe, so that either $h_F(C) = \ln 2$ or $h_{F^{-1}}(C) = \ln 2$.

If F is a $C^{1+\epsilon}$ diffeomorphism of a surface S, then its 0-entropy equivalence classes form an upper semi-continuous decomposition of S into continua [dCP]. The main idea of the proof is that, by results of Katok and Pesin theory [Kat80, KH95], any continuum C carrying positive entropy must intersect a Cantor set of segments of stable or unstable manifolds of a periodic orbit which contains transverse homoclinic intersections. In particular, carrying positive entropy is stable under small perturbations of C, and so if a sequence (C_n) of continua carrying 0 entropy converges in the Hausdorff topology, then the limiting continuum also carries 0 entropy. This is equivalent to saying that the equivalence classes form an upper semi-continuous decomposition of S.

By results of Moore [Moo62] and Roberts-Steenrod [RS38], it follows immediately that $S/\sim$ is a *cactoidal surface* (intuitively, a surface with nodes). Since the equivalence is dynamically defined, F induces a homeomorphism $F/\sim\colon S/\sim\ \circlearrowleft$, which has the same topological entropy as F (by a result of Bowen [Bow71], using the fact that the equivalence classes carry 0 entropy). Moreover, any non-trivial continuum in $S/\sim$ must carry entropy of either $F/\sim$ or its inverse.

Intuitively, we can think of $F/\sim$ as a 'tight' version of F, in which all of the wandering domains have been collapsed to points. As an example, the quotient of the plane by the 0-entropy equivalence relation of the horseshoe (Example 7) is depicted in Figure 10. The quotient space is a sphere (since everything outside the homoclinic tangle is collapsed to a point), obtained by identifying the solid boundary in the figure along the dotted semi-circular arcs from the mid-point at the top to the corner point on the lower left. The stable and unstable manifolds of the horseshoe project to a transverse pair foliations with singularities, represented by solid and dashed lines, respectively. These foliations carry holonomy-invariant measures on transverse arcs, whose product gives a measure on the sphere. The quotient map preserves both foliations, dividing one of the measures by 2 and multiplying the other by 2, so that the product measure is invariant. This map is a *generalized pseudo-Anosov map*: it has all the defining characteristics of a pseudo-Anosov map, except that it has infinitely many singularities (and an accumulation point of singularities).

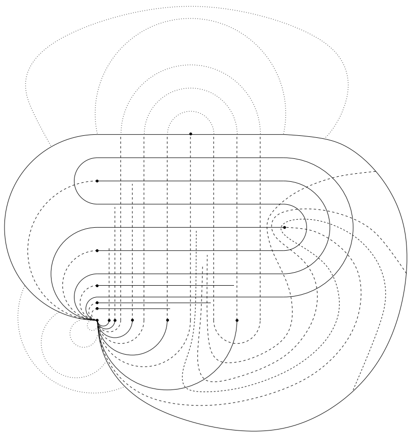

FIGURE 10. The quotient of the plane under the 0-entropy equivalence relation for the horseshoe

We finish with a conjectural analogue of Theorem 2.5. This is a precise statement of a conjecture due to Cvitanović [Cvi91].

Pruning Front Conjecture. *Let F be the horseshoe and $H_{a,b}$ denote the Hénon family. Then, for each choice of parameters a, b there exists a map $F_{a,b} \in \mathcal{P}$ such that $F_{a,b}/\sim$ and $H_{a,b}/\sim$ are topologically conjugate, where $\sim$ is the 0-entropy equivalence relation.*

In other words, up to taking quotients by the 0-entropy equivalence relation, the Hénon family is contained in the pruning family of the horseshoe.

A similar conjecture has been proved by Ishii [Ish97] for the Lozi family

$$L_{a,b}(x, y) = (1 - a|x| - by, x),$$

which is more tractable than the Hénon family: it is piecewise affine and its non-differentiability locus $x = 0$ divides the plane naturally into two parts which can be used to introduce symbolic dynamics.

REFERENCES

[BC91] M. Benedicks and L. Carleson. The dynamics of the Hénon map. *Ann. of Math. (2)*, 133(1):73–169, 1991.

[Bow71] R. Bowen. Entropy for group endomorphisms and homogeneous spaces. *Trans. Amer. Math. Soc.*, 153:401–414, 1971.

[CE80] P. Collet and J.-P. Eckmann. *Iterated maps on the interval as dynamical systems.* Birkhäuser Boston, Mass., 1980.

[CGP88] P. Cvitanović, G. Gunaratne, and I. Procaccia. Topological and metric properties of Hénon-type strange attractors. *Phys. Rev. A (3)*, 38(3):1503–1520, 1988.

[Cvi91] P. Cvitanović. Periodic orbits as the skeleton of classical and quantum chaos. *Phys. D*, 51(1-3):138–151, 1991. Nonlinear science: the next decade (Los Alamos, NM, 1990).

[dC99] A. de Carvalho. Pruning fronts and the formation of horseshoes. *Ergodic Theory Dynam. Systems*, 19(4):851–894, 1999.

[dCH] A. de Carvalho and T. Hall. The forcing relation for horseshoe braid types. To appear in Experiment. Math.

[dCH01] A. de Carvalho and T. Hall. Pruning theory and Thurston's classification of surface homeomorphisms. *J. Eur. Math. Soc. (JEMS)*, 3(4):287–333, 2001.

[dCH02] A. de Carvalho and T. Hall. How to prune a horseshoe. *Nonlinearity*, 15(3):R19–R68, May 2002.

[dCP] A. de Carvalho and M. Paternain. Monotone quotients of surface diffeomorphisms. In preparation.

[Dev89] R. Devaney. *An introduction to chaotic dynamical systems.* Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, second edition, 1989.

[dMvS93] W. de Melo and S. van Strien. *One-dimensional dynamics.* Springer-Verlag, Berlin, 1993.

[Ish97] Y. Ishii. Towards a kneading theory for Lozi mappings. I. A solution of the pruning front conjecture and the first tangency problem. *Nonlinearity*, 10(3):731–747, 1997.

[Kat80] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, (51):137–173, 1980.

[KH95] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems.* Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[Moo62] R. Moore. *Foundations of point set theory.* American Mathematical Society, Providence, R.I., 1962.

[MT88] J. Milnor and W. Thurston. On iterated maps of the interval. In *Dynamical systems (College Park, MD, 1986–87)*, pages 465–563. Springer, Berlin, 1988.

[RS38] J. Roberts and N. Steenrod. Monotone transformations of two-dimensional manifolds. *Ann. of Math.*, 39:851–862, 1938.

[Sma67] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747–817, 1967.

[WY01] Q. Wang and L-S. Young. Strange attractors with one direction of instability. *Comm. Math. Phys.*, 218(1):1–97, 2001.

Institute for Mathematical Sciences, State University of New York, Stony Brook, NY 11794-3660, USA

E-mail address: an re@math.sunysb.edu

Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, United Kingdom

E-mail address: T.Hall@liverpool.ac.uk