Split, knead, fold: a story of Markovian dynamics in one and two dimensions

Ethan Farber

A dissertation submitted to the Faculty of the Department of Mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

March 2023

Split, knead, fold: a story of Markovian dynamics in one and two dimensions

Ethan Farber

Advisor: Kathryn Lindsey, Ph.D.

We use interval maps to construct pseudo-Anosovs and relate important invariants of each regime. This work builds on techniques of André de Carvalho, Toby Hall, Bill Thurston, and others. We introduce a new perspective on the pseudo-Anosovs created in this way, showing how they constitute the vertices of a tree whose edges encode relations between them. We also characterize the pseudo-Anosovs arising from interval maps, and use this result to reprove a universal lower bound on their stretch factors originally due to Boissy-Lanneau.

Contents

List of Figures

Acknowledgments

I imagine myself as a tree, whose roots draw in life from those around me, and whose fruit will sustain them in turn.

My parents, **Sheri** and **Nuri**, form the bedrock in which I took root. Self-confidence, financial stability, and a quiet resilience are a few of the gifts they have given me. My sister, **Stephanie**, was there to model, comfort, and advise. The rest of my foundation is supplied by my broader family: my grandparents **Arona** and **Bob** have taught me not only to savor moments of joy, but to make them myself. My aunts, uncles, and cousins reinforce this lesson inexhaustibly.

My advisor, **Kathryn Lindsey**, was precisely the figure I needed to develop into a professional mathematician. Confidence, focus, imagination, thoughtfulness, and generosity are merely a few of the skills I hope to have learned from her tutelage.

My cohorts were the primary drivers of my growth in graduate school. **Jacob Caudell**'s joie de vivre was a constant exhortation to push my limits. **Dalton Fung** introduced me to so many games and never tired of my ceaseless questions. **Marius Huber** is the sweetest man you will ever meet, and one of the best friends I could ever ask for. **Matt Kownacki** was somehow half older sibling, half partner in crime, and half meme-lord extraordinaire. **Gage Martin** taught me countless lessons in solidarity, self-care, friendship, and general badassery.

The other graduate students in the Boston College math department provided a supportive ecosystem. **Ross Goluboff** and **Clayton McDonald** welcomed me to the program as storytellers. **Stella Ajluni-Gastineau** has taught me more about life than any other individual, except for **Braeden Reinoso**. In the chaotic maelstroms these two created, **Martha McAlister-Raeburn-Gastineau** was a solid rock and a welcome companion.

Friends in other departments attached me to a broader community: **Michael Bailey**, **Sam Levinson**, **Vera Sharunova**, and **Bryn Spielvogel**.

My roommates, **Keith** and **Shannon**, were my link to non-academic life. The pandemic, dreadful in so many ways, helped us establish a living space that was comforting. Baked treats, video games, hikes in the woods, and putting up wallpaper were welcome respites from the storms I weathered at work.

And finally, life is less full without animal companions: my dog **Storm**, her new sister **Skye**, the snakes **Cornelius** and **Val**, the cats **Phoenix** and **Jack**, and King **Harrold** himself in all his glory.

Chapter 1

Introduction

1.1 Background and context

In the study of discrete dynamical systems, we are interested in a transformation f sending a set X into itself. The set X is often a topological or metric space, and in these contexts f might be a continuous map, a diffeomorphism, or an isometry. Perhaps the first question a dynamicist will ask of a given system $f : X \to X$ is: "Which points $x \in X$ return to their starting position?" Stated symbolically,

Question 1.1.1 (Fundamental Question of Dynamics)**.** What is the set

$$\mathrm{Per}(f) = \{x \in X : f^k(x) = x \text{ for some } k \geq 1\}?$$

We call an element x of this set a *periodic point* of f, and we call its "return time" the *period*: that is, if $k \geq 1$ is the smallest number so that $f^k(x) = x$, then the period of x is k.

Happily, understanding the periodic points of a system f is often quite informative of the whole picture. This is especially true for the systems of study in this thesis, which admit a measurable conjugacy to a mixing subshift of finite type.

Definition 1.1.2. Let M be an $n \times n$ matrix with entries in $\{0, 1\}$. The *subshift of* *finite type* (or *SFT*) determined by M is the set of sequences

$$\Sigma_M^+ = \left\{(x_0, x_1, \ldots) : x_i \in \{0, \ldots, n-1\} \text{ and } M_{x_{i+1}, x_i} = 1 \text{ for all } i \geq 0\right\}.$$

The SFT is *mixing* if M^k has all positive entries for k large enough (cf. Definition 3.3.3).

One often thinks of the matrix M as describing a directed graph on n vertices, labeled $0, 1, \ldots, n-1$. There is a directed edge from vertex j to vertex i exactly when $M_{i,j} = 1$. An element of Σ_M^+ then corresponds to an infinitely long directed path in this graph. Dynamicists refer to this graph G_M as the *adjacency graph* of M. From this perspective, the SFT is mixing exactly when there is a $k \geq 1$ such that there is a directed path of length k from any node to any other node.

Every SFT is equipped with a natural topology generated by the *cylinder sets*

$$\mathrm{Cyl}(a_0, \ldots, a_k) = \{x \in \Sigma_M^+ : x_i = a_i \text{ for } 0 \leq i \leq k\}.$$

We denote by $\mathcal{C}$ the associated Borel σ-algebra. The SFT also possesses a natural homeomorphism $\sigma : \Sigma_M^+ \to \Sigma_M^+$, called the *shift map*. The shift map deletes the first entry of a sequence and shifts the remaining entries to the left by one:

$$\sigma((x_0, x_1, x_2, \ldots)) = (x_1, x_2, \ldots).$$

Let us return to Question 1.1.1. In the case $f = \sigma$ and $X = \Sigma_M^+$, we completely understand the set of periodic points. They are given by sequences (i.e., elements of Σ_M^+) whose entries are periodic. In terms of the adjacency graph G_M, directed loops of minimal length k correspond to periodic points of σ of period k.

The matrix M dictates the dynamics of $\sigma : \Sigma_M^+ \to \Sigma_M^+$ even more strongly. The following proposition synthesizes several well-known statements in the field. See, for example, [Par64] and [Par66]

Proposition 1.1.3. *Let Σ_M^+ be a mixing SFT.*

1. *The set $Per(\sigma) \subseteq \Sigma_M^+$ is dense.*

2. *For each $k \geq 1$ the number of periodic points for σ whose period divides k is equal to the trace of M^k.*

3. *The eigenvalue λ of M of largest modulus is real and greater than 1.*

4. *The Perron-Frobenius Theorem (cf. Theorem 2.1.3) implies that the eigenvalue λ is simple and admits a left eigenvector whose entries are positive real numbers. Let $w = (w_0, \ldots, w_{n-1})$ be the unique left eigenvector for λ such that*

$$\sum_{i=0}^{n-1} w_i = 1.$$

   Then we obtain a probability measure μ defined on $\mathcal{C}$ as follows: for each cylinder set $\mathrm{Cyl}(a_0, \ldots, a_k)$, declare

$$\mu\left(\mathrm{Cyl}(a_0, \ldots, a_k)\right) = \prod_{i=0}^{k} w_{a_i}.$$

5. *The measure μ is an ergodic (in fact, mixing) invariant measure for σ.*

6. *The measure μ has metric entropy $\log \lambda$, and is the unique measure of maximal entropy for σ. Therefore, the topological entropy of σ is*

$$\mathrm{h}(\sigma) = \log \lambda.$$

We give a name to systems modeled on a mixing SFT.

Definition 1.1.4. Let X be a topological space and $\mathcal{B}$ the associated Borel σ-algebra. A measure-preserving system $(f, X, \mathcal{B}, \nu)$ is a *hyperbolic Markov system* if it admits a measurable semi-conjugacy to some mixing SFT $(\sigma, \Sigma_M^+, \mathcal{C}, \mu)$:

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \big\downarrow{\scriptstyle \mathrm{It}_f} & & \big\downarrow{\scriptstyle \mathrm{It}_f} \\ \Sigma_M^+ & \xrightarrow{\sigma} & \Sigma_M^+ \end{array}$$

The *weak Markov partition* of f is the collection of topological closures

$$X_i = \overline{\mathrm{It}_f^{-1}(\mathrm{Cyl}(i))} \quad \text{for } 0 \le i \le n-1,$$

where n is the dimension of the matrix M. We call $\mathrm{It}_f : X \to \Sigma_M^+$ the *itinerary map*, since it records the sequence of Markov partition elements visited by a point's orbit.

Remark 1.1.5. In general we can only expect that the domain and range of It_f are full-measure sets in X and Σ_M^+, respectively. Ambiguity arises when points lie in the boundary of multiple X_i. We will point this out in the examples below.

In this thesis we study the connection between two types of hyperbolic Markov systems:

1. a piecewise-linear map $f : [0,1] \to [0,1]$ whose postcritical orbits comprise a set of cardinality p, such that the slope of f is always $\pm\lambda$, for some $\lambda > 1$; and

2. a pseudo-Anosov ψ defined on a sphere with $p+1$ punctures, with stretch factor $\lambda > 1$.

The map f is a *PCF λ-expander*, and λ is the *stretch factor* of f. The double use of λ is intentional: both systems have topological entropy $\log \lambda$, and we will be concerned with producing from a PCF λ-expander f a pseudo-Anosov ψ_f with stretch factor λ, via a procedure called *thickening*.

It is not always possible to thicken a given PCF λ-expander to a pseudo-Anosov. When it is, however, we will see that both systems are measurably semi-conjugate to the same SFT. For concreteness, we pause here to provide such a pair of systems. The following examples are adapted from [Thu14]. Both are measurably semi-conjugate to the SFT given by

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Thus the stretch factor λ in these examples is the largest eigenvalue of this matrix: namely, the golden mean

$$\lambda = \frac{1+\sqrt{5}}{2}.$$

First we introduce the interval map.

Example 1.1.6. Define $f : I \to I$ to be the continuous function

$$f(x) = \begin{cases} \lambda^{-2} + \lambda x & \text{if } 0 \leq x < \lambda^{-2} \\ \lambda(1-x) & \text{if } \lambda^{-2} \leq x \leq 1. \end{cases}$$

The graph of f is shown in Figure 1.1. The number λ has the special property that $1 = \lambda^{-1} + \lambda^{-2}$. Therefore, the orbit of the single critical point $k = \lambda^{-2}$ is

$$\begin{aligned} f^1(k) &= \lambda^{-2} + \lambda \cdot \lambda^{-2} = 1 \\ f^2(k) &= \lambda \cdot (1-1) = 0 \\ f^3(k) &= \lambda^{-2} + \lambda \cdot 0 = k. \end{aligned}$$

So $f^3(k) = k$, meaning that k is periodic of period 3. Furthermore, the points in the orbit of k cut I into two pieces: the subintervals $I_0 = [0, k]$ and $I_1 = [k, 1]$. In Figure 1.1 the dashed lines show this "cutting" of the interval, and help us see how

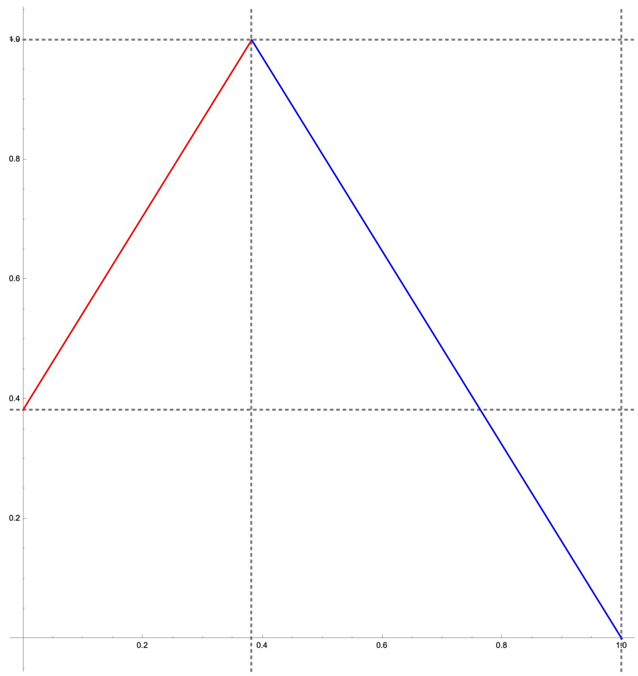


Figure 1.1: The graph of f in Example 1.1.6. The dashed lines demarcate the subintervals of the Markov partition for f.

f acts on each piece:

$$\begin{aligned} f(I_0) &= I_1 \\ f(I_1) &= I_0 \cup I_1. \end{aligned} \tag{1.1}$$

These subintervals will be the elements of the weak Markov partition for f. Indeed, if $x \neq k$ then we define the *address* of x to be

$$A(x) = \begin{cases} 0 & \text{if } x \in I_0 \\ 1 & \text{if } x \in I_1. \end{cases}$$

We avoid the case $x = k$ because its address is ambiguous, although there are solutions to this problem. To define the measurable semi-conjugacy It_f, we compile the addresses of each forward image of x into a single sequence:

$$\mathrm{It}_f(x) = (A(x), A(f(x)), A(f^2(x)), \ldots).$$

Here we must take care to avoid the set K of all preimages of k, which is a countable but dense subset of I. Thus It_f is only defined up to a set of Lebesgue measure 0 (cf. Remark 1.1.5). Using the relations in (1.1) we may quickly verify that the image of It_f is a subset of the SFT defined by

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

We can say even more. Since $\lambda > 1$ one can argue that It_f is injective. Also, the image of It_f is exactly

$$\mathrm{It}_f\left(I \setminus K\right) = \Sigma_M^+ \setminus \mathcal{K},$$

where

$$\mathcal{K} = \bigcup_{n=0}^{\infty} \sigma^{-n}\left((1,1,0)^\infty\right).$$

The sequence $(1,1,0)^\infty$ is the only sensible choice for $\mathrm{It}_f(k)$, since it is actually an element of Σ_M^+. Making this choice allows us to extend It_f to all of I, obtaining a measurable *conjugacy*, everywhere defined, from f to σ. In particular, we can answer Question 1.1.1 for f: its periodic points are in one-to-one correspondence with those of σ, since the latter are simply the codes of the former.

Now we pass to the pseudo-Anosov ψ_f obtained by thickening the PCF interval map f from the previous example. We will not discuss the thickening procedure here: see Chapter 3 for more.

Example 1.1.7. We will need the other eigenvalue of M besides λ. This is the number

$$\alpha = \frac{1-\sqrt{5}}{2} = -\lambda^{-1}.$$

Now define the two-dimensional transformation $f_G : I \times \mathbb{R} \to I \times \mathbb{R}$ as follows:

$$f_G(x,y) = \begin{cases} (\lambda^{-2} + \lambda x, \alpha^{-2} + \alpha y) & \text{if } 0 \leq x < \lambda^{-2} \\ (\lambda(1-x), \alpha(1-y)) & \text{if } \lambda^{-2} \leq x \leq 1. \end{cases}$$

We restrict our attention to a natural invariant subset of f_G, called the *limit set* Λ_f. This is the set of accumulation points of orbits of f_G, and turns out to be a finite union of Euclidean rectangles (cf. Figure 1.2).

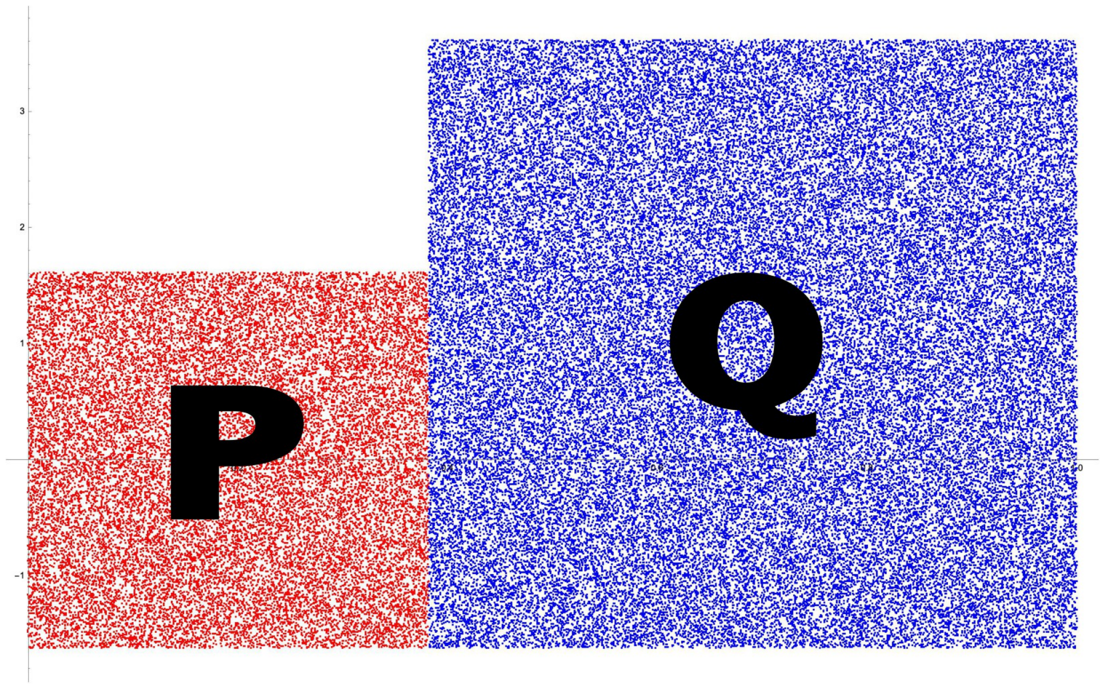
Note that f_G acts on the first coordinate as the interval map f from Example 1.1.6: that is, if $\pi : I \times \mathbb{R} \to I$ maps $(x,y) \mapsto x$, then $\pi \circ f_G = f \circ \pi$. We can use π to obtain a semi-conjugacy $\mathrm{It}_f \circ \pi : \Lambda_f \to \Sigma_M^+$ from f_G to σ. We also obtain a weak Markov partition for f_G by pulling back the weak Markov partition for f by π. Explicitly, we set

$$P = \pi^{-1}(I_0) \quad \text{and} \quad Q = \pi^{-1}(I_1).$$

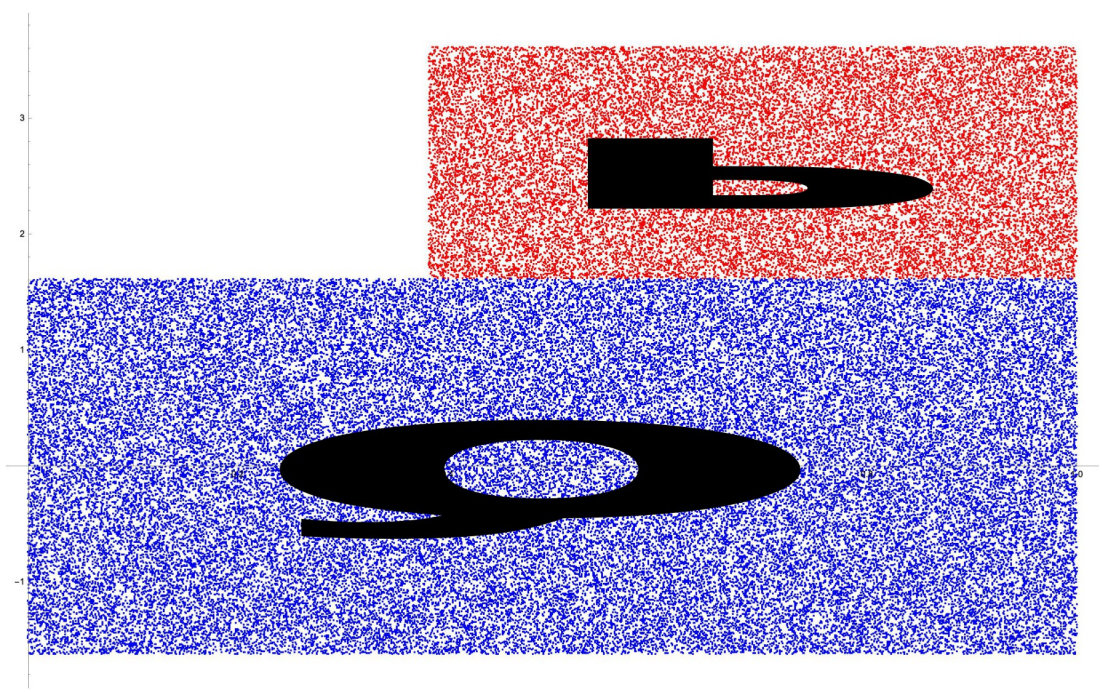
See Figure 1.2a.

To obtain the desired pseudo-Anosov ψ_f from f_G, we identify segments of the boundary of Λ_f so that the f_G becomes a homeomorphism in the quotient. The resulting topological space is the sphere S^2, and we obtain a semiconjugacy $\mathrm{It}_{\psi_f} : S^2 \to \Sigma_M^+$ from $\mathrm{It}_f \circ \pi$. In order to do so, however, we must restrict to points whose forward orbit does not intersect a horizontal boundary component of Λ_f, since the quotient map from Λ_f to S^2 makes coding these points ambiguous.

We return once again to Question 1.1.1. It turns out that there is a near-correspondence between the periodic points of $\sigma : \Sigma_M^+ \to \Sigma_M^+$ and those of ψ_f. Indeed, every periodic point of σ corresponds to a periodic point of ψ_f, and there is a distin-



(a) The limit set Λ_f of f_G, with its weak Markov partitions labeled.



(b) The weak Markov partition of Λ_f after being pushed forward by f_G. In addition to being scaled vertically and horizontally, the box P is reflected vertically, while the box Q is reflected horizontally.

Figure 1.2: The action of f_G on its limit set Λ_f. The points in these figures constitute the first 100 000 elements of the forward orbit of a random point, which is equidistributed in Λ_f.

guished periodic orbit of σ that corresponds to a fixed point of ψ_f. We refer to this point as ∞, and understanding the dynamics of ψ_f near ∞ for more general f will be crucial to our methods in Chapter 4.

At this point, the reader will likely be confused. Why, exactly, did we define f_G the way that we did? What choices are generalizable, and when? The author posed these same questions upon first seeing this example in [Thu14]. Chapter 3 answers them.

The map f_G is a shortcut of sorts, bypassing a final hyperbolic Markov system that unites f and ψ_f: a *train track map*. Briefly, a train track map is a nice type of transformation of a graph smoothly embedded in a surface S, called a *train track*. In our case, we can obtain a train track τ from a PCF interval map $f : I \to I$ by thinking of each element of $\mathrm{PC}(f)$ as a vertex in the graph I, and then adding a loop edge at each vertex. The idea is then to embed τ into a sphere in such a way that

1. $S^2 \setminus \tau$ consists of topological once-punctured discs, and

2. after thickening τ, f extends to a homeomorphism ψ_f.

If we are able to do this, then something exciting happens. The weak Markov partition for f thickens into a weak Markov partition for ψ_f, and the periodic orbits of f produce periodic orbits for ψ_f, almost always of the same length. It is perhaps unsurprising, then, that other invariants of f are reborn as invariants of ψ_f, but this is indeed what happens.

In Chapter 3 we give several equivalent characterizations for a subset of interval maps that we can thicken into pseudo-Anosovs. Along the way we introduce two invariants, $\Phi(f)$ and $D_f(t)$. In Chapter 4 we reinterpret these one-dimensional invariants using invariants of the two-dimensional map ψ_f. In Chapter 5 we reverse course, investigating when a pseudo-Anosov is the thickening of an interval map.

1.2 Summary of results

We include here a brief summary of the main results in this thesis. For more information on notation and definitions, consult the introductory section of the relevant chapter.

Chapter 3: Constructing pseudo-Anosovs from expanding interval maps

We investigate when a special kind of interval map, called a λ-*zig-zag map*, can be thickened into a pseudo-Anosov homeomorphism. We provide several equivalent characterizations of this property by introducing two invariants of a zig-zag map f: the fraction $\Phi(f)$ and the polynomial $D_f(t)$.

Theorem A. *Let $f : I \to I$ be a λ-zig-zag map with $\lambda > 2$. Then f is of pseudo-Anosov type if and only if the following conditions are satisfied:*

1. *f is PCP and the digit polynomial D_f has λ^{-1} as a root, and*
2. *the limit set Λ_f of f_G is rectangular (cf. Definition 3.2.5).*

In this case, the invariant generalized train track τ_L of F_L is finite, and recovers the action of f_G on Λ_f in the following way. Let S' be the compact topological disc obtained by performing the gluings indicated by the non-loop infinitesimal edges of τ_L. Let $\tilde{f} : S' \to S'$ be the map induced by F_L. Then there is a homeomorphism $i : S' \to \Lambda_f$ such that the following diagram commutes:

$$\begin{tikzcd} S' \arrow[r, "\tilde{f}"] \arrow[d, "i"] & S' \arrow[d, "i"] \\ \Lambda_f \arrow[r, "f_G"] & \Lambda_f \end{tikzcd}$$

Moreover, i sends the stable and unstable foliations of S' to those of Λ_f. Therefore, after identifying segments of boundary in each set so as to obtain pseudo-Anosovs

$\psi_1 : S \to S$ and $\psi_2 : \overline{\Lambda}_f \to \overline{\Lambda}_f$, these systems are conjugate via a homeomorphism that sends the (un)stable foliation of ψ_1 to the (un)stable foliation of ψ_2.

Theorem B. *Fix $m \geq 2$ and let $\Phi : \mathrm{PA}(m) \to \mathbb{Q} \cap (0, 1)$ be the map defined by*

$$\Phi(f) = \frac{n-k}{n-1} \quad \text{if } \rho(f) = \rho_m(n, k).$$

Then Φ is a bijection. Moreover, for each $p \geq 4$ the image $\Phi(\mathrm{PA}(m, p))$ consists of the set of reduced rationals in $(0, 1)$ of denominator $p - 2$.

Theorem C. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 2$ with $\Phi(f) = \frac{a}{b} \in \mathbb{Q} \cap (0, 1)$ in lowest terms. Define $L : [0, b] \to \mathbb{R}$ by $L(t) = \frac{a}{b} \cdot t$. Then*

$$D_f(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^{b+1-i},$$

where the c_i satisfy

$$c_i = \begin{cases} m & \text{if } L(t) \in \mathbb{N} \text{ some } t \in [i-1, i] \\ m-2 & \text{otherwise} \end{cases} \tag{1.2}$$

In particular, $c_i = c_{b-i}$, so D_f is reciprocal: that is,

$$D_f(t) = t^{b+1} D_f(t^{-1}).$$

Theorem D. *For each $g \geq 1$ let $f_g \in \mathrm{PA}(2)$ be the map such that $\Phi(f_g) = \frac{1}{2g}$. Let λ_g be the stretch factor of f_g. Then the following are true for each $g \geq 1$.*

1. *The number λ_g is a Salem number of degree $2g$.*

2. *The pseudo-Anosov ψ_g obtained from f_g is defined on $S_{0,2g+2}$.*

3. *The translation surface* (X_g, ω_g) *obtained as the hyperelliptic double cover of* $S_{0,2g+2}$ *is of genus* g*, and hence algebraically primitive.*

Chapter 4: A Farey tree structure on a family of pseudo-Anosovs

We deepen our understanding of the invariants $\Phi(f)$ and $D_f(t)$. These are *a priori* one-dimensional invariants defined in terms of f, but we show that they can be reinterpreted using invariants of the associated pseudo-Anosov ψ_f. We also show how $\Phi(f)$ and $D_f(t)$ transport the structure of a number theoretic tree on $\mathbb{Q} \cap (0,1)$ to the family $\mathrm{PA}(m)$.

Theorem E. *Fix* $f \in \mathrm{PA}(m)$*. Then* $\Phi(f)$ *measures the local clockwise rotation of* ψ_f *at the fixed singularity* ∞*: that is,*

$$\Phi(f) = 1 - \mathrm{rot}_\infty(\psi_f).$$

Theorem F. *Given* $f, g \in \mathrm{PA}(m)$*, we have*

$$\Phi(f) < \Phi(g) \iff \lambda(\psi_f) < \lambda(\psi_g).$$

Theorem G. *Let* $f \in \mathrm{PA}(m)$ *and set* $n = |\,\mathrm{PC}(f)| = 1 + \deg(D_f)$*. Set* $\lambda = \lambda(f) = \lambda(\psi_f)$*.*

1. *The digit polynomial determines the Artin-Mazur zeta function of* f*:*

$$\zeta_f(t) = \frac{1}{\mathcal{R}(D_f(t))} = \frac{1}{\mathcal{R}(\det(tI - W_f))}.$$

2. *The digit polynomial is equal to the strong Markov polynomial of* f*:*

$$D_f(t) = \chi_M(f; t).$$

3. *The digit polynomial determines the homology, symplectic, and puncture polynomials of ψ_f, defined by Birman-Brinkmann-Kawamuro ([BBK12]):*

$$D_f(t) = h(\psi_f;t) = \begin{cases} s(\psi_f;t) & \text{if } n \text{ is odd} \\ s(\psi_f;t)(t+1) & \text{if } n \text{ is even.} \end{cases}$$

4. *Let $\mathbb{S}$ be the surface obtained from the orientation double cover of ψ_f by filling in the lifts of the punctured 1-prong singularities of ψ_f. Denote by $\chi_+(t)$ (resp., $\chi_-(t)$) the characteristic polynomial of the lift ψ_+ (resp., ψ_-) acting on $H_1(\mathbb{S};\mathbb{Z})$. Then*

$$D_f(t) = \chi_+(t) \quad \text{and} \quad D_f(-t) = \chi_-(t).$$

5. *Let β_f be any n-braid representative of ψ_f obtained by ripping open $\infty \in S_{0,n+1}$ to a boundary circle. Let $\mathbb{B}(\beta_f, z)$ denote the reduced Burau matrix for β_f, and set $\chi(\beta_f;t) = \det(tI - \mathbb{B}(\beta_f,-1))$. Then*

$$\chi(\beta_f;t) = \begin{cases} D_f(t) & \text{if } \chi(\beta_f;t) \text{ has } \lambda \text{ as a root} \\ D_f(-t) & \text{if } \chi(\beta_f;t) \text{ has } -\lambda \text{ as a root.} \end{cases}$$

Chapter 5: Pseudo-Anosovs from interval maps

Having thoroughly examined when an interval map can be the train track map of a pseudo-Anosov, we consider the reverse direction: beginning with a pseudo-Anosov with the proper singularity data, when does it admit an interval-like train track? The answer is simply, “always.” This chapter contains work joint with Braeden Reinoso and Luya Wang, and other work joint with Karl Winsor.

Theorem H. *If ψ is a pseudo-Anosov with singularity data $(1^p, p-2)$ for some $p \geq 3$,* *then* ψ *is interval-like.*

Chapter 2

Definitions

2.1 Perron-Frobenius theory

We introduce the basics of Perron-Frobenius theory. The interested reader can refer to [Gan59] for details.

Definition 2.1.1. Let M be an $n \times n$ matrix with non-negative integer entries. We say that M is *irreducible* if $M + M^2 + \cdots + M^r$ has positive entries for some $r > 0$, and *primitive* if M^r has positive entries for some $r > 0$. A directed graph is *strongly connected* if there is a directed path from any vertex to any other vertex.

If G_M is the directed graph with vertex set $\{1, \ldots, m\}$ and an edge from i to j if and only if the (i, j)-entry of M is positive, then M is irreducible if and only if G_M is strongly connected.

Definition 2.1.2. Let M be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The *spectral radius* of M is the quantity

$$\rho(M) = \max_i |\lambda_i|.$$

One of the fundamental theorems in the study of Markovian dynamics is the Perron-Frobenius theorem. There are many variations of the theorem: we state the most pertinent one for our purposes.

Theorem 2.1.3 (Perron-Frobenius)**.** *Let M be a primitive matrix with spectral radius $\lambda = \rho(M)$. Then the following are true:*

1. *The number λ is greater than 1 and is an eigenvalue of M, called the dominant or Perron-Frobenius eigenvalue.*

2. *If $\alpha \neq \lambda$ is another eigenvalue of M, then $|\alpha| < \lambda$.*

3. *The number λ is a simple eigenvalue.*

4. *There exist left- and right-eigenvectors of M for λ whose entries are all positive. No other eigenvalue of M has this property.*

5. *The number λ is bounded from below (resp. above) by the minimum (resp., maximum) row sum of M. By taking the transpose of M, the same is true for the minimum and maximum column sums of M.*

In spirit, the Perron-Frobenius theorem describes how dynamical properties of a hyperbolic systems is determined by the dominant eigenvalue (and eigenvectors) of the Markov matrix. Indeed, the theorem governs the dynamics of both PCF interval maps and pseudo-Anosov homeomorphisms. In both cases, if M is an associated Markov matrix and λ the dominant eigenvalue, then

1. the topological entropy of the system is $\log \lambda$, and

2. reasonable choices for positive left- and right-eigenvectors for λ determine the ergodic (in fact, mixing) absolutely continuous invariant measure for the system.

See Sections 2.3 and 2.6 for more details.

The dominant eigenvalue λ of a primitive matrix M has several interesting properties. It is a root of the charateristic polynomial, a monic polynomial with integer coefficients. Therefore, λ is an algebraic integer. As we have already seen, $\lambda > 1$, and since it is the spectral radius of M it must be greater in absolute value than any of its Galois conjugates. Numbers with the above properties have a name.

Definition 2.1.4. A *Perron number* is a real algebraic integer $\lambda > 1$ such that if $\alpha \neq \lambda$ is a Galois conjugate of λ, then $|\alpha| < \lambda$.

Thus the dominant eigenvalue of a primitive matrix is always a Perron number. Conversely, Lind proved in [Lin92] that every Perron number is the dominant eigenvalue of some primitive matrix.

Remark 2.1.5. There are analogous narratives in each of the two dynamical regimes studied in this thesis. For PCF interval maps, it is known that the topological entropy is always $\log \lambda$, where λ is a *weak Perron number*. The converse statement, analogous to Lind's theorem, is a theorem of W. Thurston in [Thu14]. For pseudo-Anosovs, the topological entropy is always $\log \lambda$ where λ is a *bi-Perron unit*. The converse statement is sometimes called Fried's Question, and is still open.

The following theorem will be useful for proving Theorem H, but we state it now.

Proposition 2.1.6. *Let n be a positive integer, and fix $B > 0$. There are only finitely many $n \times n$ primitive matrices with and spectral radius at most B.*

Proof. Let $M = [m_{i,j}]$ be an $n \times n$ primitive matrix and spectral radius $\rho(M) \leq B$. Set I to be the $n \times n$ identity matrix. Since G_M is strongly connected, the entries of $A = (M + I)^{n-1} = [a_{i,j}]$ are positive. The spectral radius of $B = A^2 = [b_{i,j}]$ is $(\rho(M) + 1)^{2n-2}$, so we have

$$\sum_{i,j=1}^{m} m_{i,j} \leq \sum_{i,j=1}^{m} a_{i,j} \leq \min_{j} \sum_{i=1}^{m} b_{i,j} \leq (\rho(M) + 1)^{2n-2} \leq (B + 1)^{2n-2}.$$

Since each $m_{i,j}$ is a non-negative integer, there are only finitely many possibilities for M. $\square$

2.2 Interval dynamics: first principles

Let $f : I \to I$ be a continuous interval map. By a *critical point* we will mean a point $x \in I$ such that f does not restrict to a homeomorphism on any neighborhood of x. We say f is *m-modal* if it has exactly m critical points. Throughout this thesis, every interval map f will be m-modal for some m, and we will denote its critical points by $k_1, \ldots, k_m$.

Definition 2.2.1. The *postcritical set* of $f : I \to I$ is the set

$$\mathrm{PC}(f) = \{f^n(k) : n \geq 1 \text{ and } k \text{ a critical point of f}\}.$$

The elements of $\mathrm{PC}(f)$ are the *postcritical points* of f. The map f is *postcritically finite* (or *PCF* for short) if $\mathrm{PC}(f)$ is a finite set.

For the purpose of studying a pseudo-Anosov ψ_f created from f, we will usually be interested with studying $\mathrm{PC}(f)$. It is, however, sometimes more convenient to use a larger set.

Definition 2.2.2. The *weak postcritical set* of f is the set

$$\mathrm{WPC}(f) = \{f^n(k) : n \geq 0 \text{ and } k \text{ a critical point of } f.\}$$

Thus $\mathrm{WPC}(f)$ contains $\mathrm{PC}(f)$, as well as all critical points of f that are not postcritical. Since we always assume that f is m-modal, $\mathrm{WPC}(f)$ is finite if and only if $\mathrm{PC}(f)$ is finite. In either case, one obtains a dynamical partition of I. We begin with the partition obtained from $\mathrm{WPC}(f)$, since this is slightly easier from a technical perspective.

Definition 2.2.3. Suppose that f is PCF, and put the elements of $\mathrm{WPC}(f)$ in increasing order: $w_0 < w_1 < \cdots < w_r$. The *weak Markov partition* of f is the collection of subintervals

$$I_j = [w_j, w_{j+1}] \quad \text{for } 0 \le j \le r-1.$$

See Definition 1.1.4 and Example 3.4.10. The *weak Markov matrix* of f is the $r \times r$ matrix W_f whose (i,j)-entry is

$$(W_f)_{i,j} = \begin{cases} 1 & \text{if } f(I_j) \supseteq I_i \\ 0 & \text{otherwise.} \end{cases}$$

The *weak Markov polynomial* of f is the characteristic polynomial of W_f:

$$\chi_W(f;t) = \det(tI - W_f).$$

The Perron-Frobenius theorem states that the spectral radius λ of W_f is in fact an eigenvalue of W_f. Moreover, using the natural measurable semiconjugacy from f to the SFT $\sigma : \Sigma^+_{W_f} \to \Sigma^+_{W_f}$ one can show that the topological entropy of f satisfies $\mathrm{h}(f) = \log \lambda$. Thus $e^{\mathrm{h}(f)}$ is the largest root of $\chi_W(f;t)$.

As noted above, the weak Markov partition is finer than necessary for most applications in this thesis. Instead, we cut I only at the points of $\mathrm{PC}(f)$.

Definition 2.2.4. Suppose that f is PCF, and put the elements of $\mathrm{PC}(f)$ in increasing order: $p_0 < p_1 < \cdots < p_n$. The *strong Markov partition* of f is the collection of subintervals

$$I_j = [p_j, p_{j+1}] \quad \text{for } 0 \le j \le n-1.$$

The *strong Markov matrix* of f is the $n \times n$ matrix M_f whose (i,j)-entry is

$$(M_f)_{i,j} = \text{the number of times } f(I_j) \text{ crosses } f(I_i).$$

The *strong Markov polynomial* of f is $\chi_M(f;t) = \det(tI - M_f)$.

An exercise in linear algebra demonstrates the following relationship.

Proposition 2.2.5. *Set $r = |\operatorname{WPC}(f)|$ and $n = |\operatorname{PC}(f)|$. Then*

$$\chi_W(f;t) = t^{r-n}\chi_M(f;t).$$

In particular, $\mathrm{e}^{\mathrm{h}(f)}$ is a root of $\chi_M(f;t)$.

The matrices W_f and M_f share most properties that are important for this thesis. As we have seen, they each have the same set of non-zero eigenvalues, counting multiplicity. The powers of W_f and M_f have something else in common, as well.

Definition 2.2.6. An $n \times n$ matrix M is *positive* if all entries of M are positive. The matrix M is *primitive* if M^k is positive for some $k \geq 1$.

Proposition 2.2.7. *For any $k \geq 1$, the matrix W_f^k is positive if and only if M_f^k is positive. In particular, W_f is primitive if and only if M_k is primitive.*

2.3 Interval dynamics: uniform expanders

It turns out that the essential dynamics of an m-modal map $f : I \to I$ can be modeled by a continuous piecewise-linear map F. More explicitly, we have the following theorem, which is a combination of Theorem 5 in [Par66] and Theorem 7.4 in [MT88].

Theorem 2.3.1. *Let $f : I \to I$ be an m-modal map with topological entropy $\mathrm{h}(f) > 1$. Then there exists a non-decreasing surjective continuous map $C : I \to I$ and a piecewise-linear map $F : I \to I$ such that the following diagram commutes:*

$$\begin{tikzcd} I \arrow[r, "f"] \arrow[d, "C"] & I \arrow[d, "C"] \\ I \arrow[r, "F"] & I \end{tikzcd}$$

Moreover, the slope of F is everywhere $\pm e^{\mathrm{h}(f)}$, and $\mathrm{h}(F) = \mathrm{h}(f)$.

Finally, if f has a dense orbit then C is a homeomorphism.

Briefly, we say that F is the *linear model* of f. Because of Theorem 2.3.1, we will focus entirely on the case when $f : I \to I$ is piecewise-linear with slope equal to $\pm\lambda$ everywhere, for some $\lambda > 1$. Such maps have a name.

Definition 2.3.2. Fix $\lambda > 1$. A *uniform λ-expander*, or *uniform expander* or *λ-expander*, is a piecewise-linear m-modal map $f : I \to I$ such that every linear branch of f has slope λ or $-\lambda$.

Misiurewicz and Szlenk proved that the topological entropy of an m-modal map f can be computed using special covers for f (cf. Theorem 1 in [MS80]). In particular, one can use their techniques to show the following theorem.

Proposition 2.3.3. *The topological entropy of a uniform λ-expander is* $\log \lambda$*.*

In the case that f is PCF with primitive weak Markov matrix W_f, then we may construct the linear model for f combinatorially. As we have seen, $\lambda = e^{\mathrm{h}(f)}$ is the dominant eigenvalue of W_f. Theorem 2.1.3 implies the existence of a positive left-eigenvector $v = (v_1, \ldots, v_r)$ for W_f with eigenvalue λ. We may assume $\sum_i v_i = 1$.

We now scale each element I_i of the weak Markov partition for f to have length (i.e., Lebesgue measure) v_i. Because v is an eigenvector for λ, it follows that the resulting interval map $F : I \to I$ is a uniform λ-expander. Thus, in a sense, we have "linearized" f using the spectral properties of W_f. This is extremely analogous to the construction of the invariant foliations for a pseudo-Anosov from its action on an invariant train track (cf. Section 2.6). This is the kernel of the main relationship studied in this thesis.

2.4 Kneading theory

Kneading theory provides a more general framework for coding the orbits of an m-modal map, even if the map is not PCF. We present here one variation of the theory suitable for our techniques in Chapter 4.

In this section, and whenever discussing the kneading theory of an m-modal map f, we use the notation

$$I_0 = [0, k_1), \;\; I_m = (k_m, 1], \text{ and } I_j = (k_j, k_{j+1}) \text{ for } j = 1, \dots, m-1.$$

These are the intervals of monotoniciy for f. Put $\mathcal{A} = \mathcal{A}_m = \{0, k_1, 1, k_2, 2, \dots, k_m, m\}$, an alphabet on $2m+1$ letters.

Definition 2.4.1. For any $x \in I$, the *address* of x is

$$A(x) = \begin{cases} j & \text{if } x \in I_j \\ k_j & \text{if } x = k_j. \end{cases}$$

The *itinerary* of x is the sequence

$$\mathrm{It}_f(x) = (A(x), A(f(x)), A(f^2(x)), \dots).$$

Given a critical point k_j, its image $f(k_j)$ is a *critical value.* The *kneading sequences* of f are the itineraries of these critical values:

$$\mathcal{K}_j(f) = \mathrm{It}_f(f(k_j)) \;\; \text{for } j = 1, \dots, m.$$

The *kneading data* of f is the vector of kneading sequences

$$\mathcal{K}_f = (\mathcal{K}_1(f), \dots, \mathcal{K}_m(f)).$$

Definition 2.4.2. The *sequence space* of f is the set

$$\Sigma_f = \{\mathrm{It}_f(x) : x \in I\}.$$

The *shift* on Σ_f is the unique map $\sigma : \Sigma_f \to \Sigma_f$ such that the following diagram commutes:

$$\begin{tikzcd} I \arrow[r, "f"] \arrow[d, "\mathrm{It}_f"'] & I \arrow[d, "\mathrm{It}_f"] \\ \Sigma_f \arrow[r, "\sigma"] & \Sigma_f \end{tikzcd}$$

Remark 2.4.3. The space Σ_f is a subshift of finite type if and only if f is PCF.

Definition 2.4.4. The *sign function* of f is the function $E : \mathcal{A} \to \{-1, 0, +1\}$ such that

$$E(j) = \begin{cases} +1 & \text{if } f \text{ is increasing on } I_j \\ -1 & \text{if } f \text{ is decreasing on } I_j \\ 0 & \text{if } j = k_i \text{ is a critical point.} \end{cases}$$

Note that since f is continuous, E is determined by $E(0)$. In the following discussion we work with a fixed f and the sign function E it defines.

Definition 2.4.5. The *cumulative sign vector* of a (possibly infinite) word A in the alphabet $\mathcal{A}$ is the sequence $(s_i(A))_i$ such that

$$s_i(A) = \begin{cases} +1 & \text{if } i = 0 \\ E(A_{i-1}) \cdot s_{i-1}(A) & \text{if } i \geq 1. \end{cases}$$

We now define a partial order on Σ_f.

Definition 2.4.6. Let f be an m-modal map with sign vector E. We define the *twisted lexicographic order* $\leq_E$ on Σ_f as follows. First, we declare

$$0 <_E k_1 <_E 1 <_E k_2 <_E 2 <_E \cdots <_E k_m <_E m.$$

For $A, B \in \Sigma_f$ we now define $A \leq_E B$ if either $A = B$ or else, for the minimal index l such that $A_l \neq B_l$, either

$$\begin{cases} A_l <_E B_l & \text{if } s_l(A) = +1, \text{ or} \\ A_l >_E B_l & \text{if } s_l(A) = -1. \end{cases}$$

While we technically defined the ordering $\leq_E$ on the shift space for a single map f, observe that this definition only relies on the input data of the sign function E. Therefore, it makes sense to compare sequences from the shift spaces of two m-modal maps f and g with the same sign function E. With this in mind, we now define a partial order on the space of kneading data of maps with the same sign function.

Definition 2.4.7. Let f, g be m-modal maps with the same sign function E. We say that $\mathcal{K}_f \ll \mathcal{K}_g$ if for all $j = 1, \ldots, m$ we have

$$\begin{cases} \mathcal{K}_j(f) \leq_E \mathcal{K}_j(g) & \text{if } E(j) = -1, \text{ and} \\ \mathcal{K}_j(f) \geq_E \mathcal{K}_j(g) & \text{if } E(j) = +1. \end{cases} \tag{2.1}$$

The following proposition shows that this ordering on kneading data has implications for the topological entropies of maps with the same sign function.

Proposition 2.4.8 (Corollary 4.5 in [MT00])**.** *If $\mathcal{K}_f \ll \mathcal{K}_g$, then* $\mathrm{h}(f) \leq \mathrm{h}(g)$*.*

2.5 Pseudo-Anosovs

Let $S = S_{g,p}$ be the closed, connected, oriented surface of genus g with p marked points, which we will often treat as punctures. Assume that $6g - 6 + 2p > 0$.

Definition 2.5.1. A *pseudo-Anosov* is a homeomorphism $\psi : S \to S$ for which there exist (1) two transverse singular measured foliations $(\mathcal{F}^s, \mu^s)$ and $(\mathcal{F}^u, \mu^u)$ of S and (2) a real number $\lambda > 1$, such that:

- $\psi \cdot (\mathcal{F}^s, \mu^s) = (\mathcal{F}^s, \lambda^{-1}\mu^s)$, and
- $\psi \cdot (\mathcal{F}^u, \mu^u) = (\mathcal{F}^u, \lambda\mu^u)$.

That is, ψ preserves both $\mathcal{F}^s$ and $\mathcal{F}^u$, while scaling the former's transverse measure by λ^{-1} and the latter's by λ.

There is a lot to unpack in this definition. First, let us treat the singular measured foliations $\mathcal{F}^s$ and $\mathcal{F}^u$. These are the *stable* and *unstable foliations* of ψ, respectively. This naming convention comes from the action of ψ on the space $\mathcal{PMF}$ of projective measured foliations of S: the classes of $\mathcal{F}^s$ and $\mathcal{F}^u$ are the two fixed points of ψ, with the former being the unique sink and the latter the unique source. See [Thu88] for more.

The foliations are "singular" because of the existence of a finite number of points where $\mathcal{F}^s$ and $\mathcal{F}^u$ simultaneously fail to be true foliations.

Definition 2.5.2. Let $k \geq 1$ be an integer. A *k-prong singularity of ψ* is a point $s \in S$ such that $k = k(s)$ leaves of $\mathcal{F}^u$ (equivalently, $\mathcal{F}^s$) land at s and either

1. $k \geq 3$, or
2. $k = 1$ or 2, and s is a marked point of S.

We say s is a *regular point* if it is not a singularity.

Remark 2.5.3. The restriction on 1-prong singularities is technical: as we shall see, pseudo-Anosovs are unique up to isotopy rel marked points, and a one-prong singularity at an unmarked point of S would vanish under isotopy.

We only allow 2-pronged singularities to occur at marked points of S, since otherwise these points are not topologically distinct from regular points.

The Euler-Poincaré formula gives a constraint on singularities of ψ in terms of the topology of S.

Proposition 2.5.4 (Euler-Poincaré)**.** *We have*

$$2\chi(S) = \sum_s (2 - k(s)),$$

where the sum is over all singularities of ψ*.*

We will occasionally need to refer to the collection of numbers $k(s)$, where s ranges over the singularity set of ψ. We introduce notation for this purpose.

Definition 2.5.5. The *singularity data* of ψ is the unordered tuple $(k(s))_s$. For brevity, we will use the notation k^n to denote n distinct k-prong singularities.

We will often reference the following situation in Chapters 3, 4, and 5.

Example 2.5.6. Suppose that $\psi : S^2 \to S^2$ is a pseudo-Anosov having $p \geq 3$ distinct 1-prong singularities, as well as one additional singularity labelled ∞. The Euler-Poincaré formula implies that $k(\infty) = p - 2$, so the singularity data of ψ is $(1^p, p - 2)$.

We return again to Definition 2.5.1. The number $\lambda > 1$ is a fundamental invariant of ψ, called its *stretch factor* or *dilatation*. Suppose ψ is orientation-preserving. Away from its singularities, one may think of ψ as acting in local coordinates by an affine map having derivative

$$D\psi = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Here the first coordinate is in the direction of $\mathcal{F}^u$, and the second in the direction of $\mathcal{F}^s$. The number $\lambda > 1$, called the *stretch factor* or *dilatation* of ψ, is a fundamental invariant. As it turns out, λ is an algebraic integer of degree at most $6g - 6 + 2p$, and the topological entropy of ψ is $\mathrm{h}(\psi) = \log \lambda$.

More is true: λ is the spectral radius of a primitive integer matrix M associated to ψ, so the Perron-Frobenius theorem (Theorem 2.1.3) implies that λ is a Perron number. In fact, Fried proved in [Fri85] that λ is a *bi-Perron unit*.

Definition 2.5.7. A *bi-Perron unit* is a real algebraic integer $\lambda > 1$ whose Galois conjugates are contained in the annulus

$$A(\lambda) = \{z : \lambda^{-1} \leq |z| \leq \lambda\},$$

and such that at most one conjugate exists on each boundary component.

It is an open question whether every bi-Perron unit is the stretch factor of a pseudo-Anosov. Researchers sometimes refer to this question as "Fried's Question," seemingly confusing it with a weaker problem posed by Fried.

Question 2.5.8 (Problem 2 in [Fri85])**.** Does every bi-Perron unit have a power that is the stretch factor of some pseudo-Anosov?

2.6 Train tracks

Any pseudo-Anosov ψ may be reconstructed, along with its invariant foliations, by investigating its action on particular graphs embedded in S, called *train tracks*.

Definition 2.6.1. A *train track* is a graph τ smoothly embedded in S satisfying the following properties:

1. For each vertex v of τ, each smooth path $p : [-1, 1] \to \tau$ with $p(0) = v$ has the same tangent line at v. This tangency condition creates cusps between adjacent incoming edges with the same unit tangent vector at v.

2. Each connected component of $S \setminus \tau$ is either

   (a) an un-punctured disc with at least three cusps on its boundary, or

   (b) a once-punctured disc with at least one cusp on its boundary.

If the image $\psi(\tau)$ is homotopic to τ in a restricted sense, then we say τ is an *invariant train track* for ψ, and by concatenating ψ with the homotopy we obtain a graph map $f : \tau \to \tau$. We call f a *train track map*, and it can be used to reconstruct ψ along with the foliations $\mathcal{F}^s$ and $\mathcal{F}^u$. Indeed, the edges of τ form a 1-dimensional Markov partition for f. The subset of *expanding edges* of τ, meaning those edges e such that $f^n(e) = \tau$ for large n, produce a 2-dimensional Markov partition for ψ consisting of rectangles. See [BH95] and Section 3.3.2 for more.

Remark 2.6.2. If τ is an invariant train track for ψ, then the complementary components of $S \setminus \tau$ determine the singularity data of ψ. Explicitly, ψ has exactly one k-prong singularity for each component with k cusps. Components of type 2a correspond to unpunctured singularities, while components of type 2b correspond to punctured singularities.

The dynamics of $f : \tau \to \tau$ thus determine the dynamics of $\psi : S \to S$, up to a null set. We are interested in producing pseudo-Anosovs ψ from suitable graph maps f, and relating invariants from the two different regimes.

Chapter 3

Pseudo-Anosovs from expanding interval maps

3.1 Introduction

In this chapter we investigate when an element of a class of PCF interval maps can be used to create a pseudo-Anosov homeomorphism. The impetus for this investigation is an example of W. Thurston in [Thu14], in which he packages an initial interval map f with a "Galois conjugate" action in a second coordinate to obtain a piecewise affine map (cf. Example 1.1.7). This affine map has a natural compact invariant set, which in this case is a finite union of Euclidean rectangles. Moreover, appropriately identifying segments of the boundaries of these rectangles produces a pseudo-Anosov ψ_f with the same stretch factor as f, and hence the same topological entropy.

To understand this procedure, we turn to the machinery of *generalized pseudo-Anosovs* developed by André de Carvalho in [dC05]. Briefly, this more flexible framework allows one to "thicken" a ergodic graph map into a hyperbolic surface homeomorphism, where the surface need not be finite-type. When the surface is finite-type, the resulting homeomorphism is a pseudo-Anosov with the same topological entropy as the initial graph map. With these methods, de Carvalho and Toby Hall were able to combinatorially describe the set of tent maps that thicken to a pseudo-Anosov.

Our first result connects the machinery of de Carvalho-Hall to the example of Thurston. An m-modal map f is *of pseudo-Anosov type* if it can be thickened to a pseudo-Anosov via de Carvalho's technology. We say f is *PCP* (*postcritically periodic*) if it acts on $\mathrm{PC}(f)$ by a (finite) permutation. The *Galois lift* is the map f_G produced from f by a generalization of Thurston's method. The *limit set* Λ_f is a natural compact invariant set associated to f_G.

Theorem A. *Let $f : I \to I$ be a λ-zig-zag map with $\lambda > 2$. Then f is of pseudo-Anosov type if and only if the following conditions are satisfied:*

1. *f is PCP and the digit polynomial D_f has λ^{-1} as a root, and*
2. *the limit set Λ_f of f_G is rectangular (cf. Definition 3.2.5).*

In this case, the invariant generalized train track τ_L of F_L is finite, and recovers the action of f_G on Λ_f in the following way. Let S' be the compact topological disc obtained by performing the gluings indicated by the non-loop infinitesimal edges of τ_L. Let $\tilde{f} : S' \to S'$ be the map induced by F_L. Then there is a homeomorphism $i : S' \to \Lambda_f$ such that the following diagram commutes:

$$\begin{array}{ccc} S' & \xrightarrow{\tilde{f}} & S' \\ \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} \\ \Lambda_f & \xrightarrow{f_G} & \Lambda_f \end{array}$$

Moreover, i sends the stable and unstable foliations of S' to those of Λ_f. Therefore, after identifying segments of boundary in each set so as to obtain pseudo-Anosovs $\psi_1 : S \to S$ and $\psi_2 : \overline{\Lambda}_f \to \overline{\Lambda}_f$, these systems are conjugate via a homeomorphism that sends the (un)stable foliation of ψ_1 to the (un)stable foliation of ψ_2.

Essentially, when Λ_f is rectangular the Galois lift provides an explicit choice of coordinates for a generalized pseudo-Anosov produced from f. If, moreover, a particular polynomial D_f associated to f vanishes at the inverse of the stretch factor of f, then this generalized pseudo-Anosov is simply a pseudo-Anosov. We will return to D_f in Theorem C.

At this point we leave the Galois lift behind. The machinery of generalized pseudo-Anosovs is more amenable to study, presenting a series of combinatorial problems. If we restrict to the case of zig-zag maps, which are simple due to having only two critical values, then these combinatorial problems become tractable. We consider the following sets:

$$\mathrm{PA}(m,p) = \{m\text{-modal zig-zags of pseudo-Anosov type with } |\,\mathrm{PC}(f)| = p\}$$

$$\mathrm{PA}(m) = \bigcup_{m \geq 4} \mathrm{PA}(m,p)$$

As we have seen, in order for f to be of pseudo-Anosov type it must act on $\mathrm{PC}(f)$ by a permutation. We are able to derive explicit formulas for this permutation action, denoted $\rho_m(n,k)$. Note that this permutation action depends on the modality m of f. In particular, we show that $f \in \mathrm{PA}(m)$ if and only if f acts by on the orbit of $x = 1$ by some $\rho_m(n,k)$ where $n-1$ and $n-k$ are coprime (cf. for example Proposition 3.7.6). This realization leads us to the next main result of this chapter.

Theorem B. *Fix $m \geq 2$ and let $\Phi : \mathrm{PA}(m) \to \mathbb{Q} \cap (0,1)$ be the map defined by*

$$\Phi(f) = \frac{n-k}{n-1} \quad \text{if } \rho(f) = \rho_m(n,k).$$

Then Φ is a bijection. Moreover, for each $p \geq 4$ the image $\Phi(\mathrm{PA}(m,p))$ consists of the set of reduced rationals in $(0,1)$ of denominator $p-2$.

Theorem B can be seen as a generalization of the results of Toby Hall in [Hal94] for tent maps of pseudo-Anosov type, i.e. the case when $m = 1$. We stress that this bijection is more rightly a parameterization using the dynamics of f. As we will see in Chapter 4, $\Phi(f)$ also describes the rotation of the pseudo-Anosov ψ_f around a distinguished point (cf. Theorem E). We will also see that $\Phi(f)$ increases monotonically in the stretch factor $\lambda(f)$ (cf. Theorem F).

With a clearer understanding of the dynamics of $f \in \mathrm{PA}(m)$, we return to the *digit polynomial*, D_f. This is a polynomial with integer coefficients, and has the stretch factor $\lambda(f)$ as a root. Additionally, D_f is determined by the orbit of $x = 1$ under f. Thus, Theorem B allows us to relate the coefficients of D_f to the fraction $\Phi(f)$.

Theorem C. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 2$ with $\Phi(f) = \frac{a}{b} \in \mathbb{Q} \cap (0,1)$ in lowest terms. Define $L : [0,b] \to \mathbb{R}$ by $L(t) = \frac{a}{b} \cdot t$. Then*

$$D_f(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^{b+1-i},$$

where the c_i satisfy

$$c_i = \begin{cases} m & \text{if } L(t) \in \mathbb{N} \text{ some } t \in [i-1, i] \\ m-2 & \text{otherwise} \end{cases} \tag{3.1}$$

In particular, $c_i = c_{b-i}$, so D_f is reciprocal: that is,

$$D_f(t) = t^{b+1} D_f(t^{-1}).$$

Theorem C generalizes Lemmas 2.5 and 2.6 of [Hal94], which again treat the case $m = 1$.

Finally, in Section 3.9 we consider an infinite family of zig-zags producing an algebraically primitive translation surface for each positive genus.

Theorem D. *For each $g \geq 1$ let $f_g \in \mathrm{PA}(2)$ be the map such that $\Phi(f_g) = \frac{1}{2g}$. Let λ_g be the stretch factor of f_g. Then the following are true for each $g \geq 1$.*

1. *The number λ_g is a Salem number of degree $2g$.*
2. *The pseudo-Anosov ψ_g obtained from f_g is defined on $S_{0,2g+2}$.*
3. *The translation surface (X_g, ω_g) obtained as the hyperelliptic double cover of $S_{0,2g+2}$ is of genus g, and hence algebraically primitive.*

To summarize: in Section 3.3 we review the construction of generalized pseudo-Anosovs, focusing particularly on the concept of a thick interval map. This is followed in Section 3.4 by a discussion on the different ways of thickening an interval map to a thick interval map. Here we prove that, when attempting to construct pseudo-Anosovs, the only thickening of a zig-zag we need consider is the exterior left-veering map F_L (cf. Proposition 3.4.19).

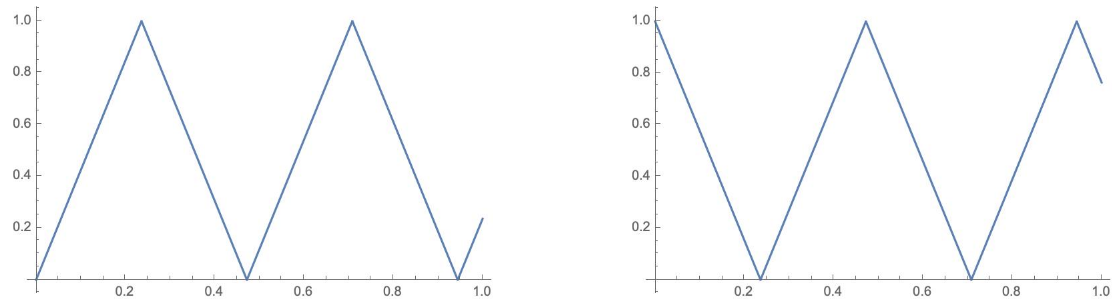
In Section 3.5 we prove Theorem A, reconciling Thurston's construction with that of de Carvalho. Beginning in Section 3.7, we turn our attention to classifying zig-zag maps of pseudo-Anosov type, observing several necessary conditions on their postcritical orbit structure. Section 3.6 in effect demonstrates that these conditions are also sufficient, proving Theorem B.

Section 3.8 investigates the digit polynomial $D_f(t)$, establishing Theorem C.

Section 3.9 turns briefly to considerations of flat geometry, providing a terse introduction to part of the theory. As an application, we prove Theorem D.

3.2 Preliminaries

Definition 3.2.1. We say that a PCF map $f : I \to I$ is *postcritically periodic* (or *PCP*) if each critical value $f(c)$ is periodic. In other words, f acts on $\mathrm{PC}(f)$ by a permutation.



(a) A positive zig-zag map.

(b) A negative zig-zag map.

Figure 3.1: The positive and negative zig-zag maps of slope λ equal to the Perron root of $x^2 - 4x - 1$.

As we will see (cf. Theorem 3.4.7), an interval map $f : I \to I$ being PCP is necessary for producing a pseudo-Anosov by thickening f.

Definition 3.2.2. We call a uniform λ-expander $f : I \to I$ a *zig-zag map* (or *zig-zag* or λ-*zig-zag*) if the only critical points of f are $c_i = i \cdot \lambda^{-1}$ for $i = 1, \ldots, \lfloor \lambda \rfloor$. In other words, $f(0) \in \partial I$, and $c \in \operatorname{int}(I)$ is a critical point of f if and only if $f(c) \in \partial I$. Note that for a fixed λ there are two distinct λ-zig-zags: the *positive* zig-zag satisfies $f(0) = 0$, while the *negative* zig-zag satisfies $f(0) = 1$.

Observe that a zig-zag map f is PCP if and only if $x = 1$ is periodic.

Definition 3.2.3. Let $f : I \to I$ be a PCF λ-expander such that $\mathrm{WPC}(f) \subseteq \mathbb{Q}(\lambda)$. Denote by $w_0 < w_1 < \cdots < w_r$ the elements of $\mathrm{WPC}(f)$, and set $I_j = [w_j, w_{j+1}]$ for $j = 0, \ldots, r-1$. Let $f_j(x) = a_j(\lambda) \pm \lambda x$ denote the linear map defining f on I_j, where $a_j(\lambda) \in \mathbb{Q}(\lambda)$. Set $\widetilde{f}_j(x) = a_j(\lambda^{-1}) \pm \lambda^{-1} x$.

The *Galois lift* of f is the piecewise-affine transformation $f_G : I \times \mathbb{R} \to I \times \mathbb{R}$ such that

$$f_G(x, y) = \begin{cases} \Big(f_j(x), \widetilde{f}_j(y)\Big) & \text{if } x \in I_j \setminus \{w_{j+1}\} \\ \Big(f_{r-1}(1), \widetilde{f}_{r-1}(1)\Big) & \text{if } x = 1. \end{cases} \tag{3.2}$$

The *limit set* of f_G is the smallest closed set Λ_f containing all accumulation points of orbits under f_G.

Remark 3.2.4. The Galois lift f_G has two key properties:

1. If $\pi : I \times \mathbb{R} \to I$ is the natural projection, then f is a factor of f_G:

$$\pi \circ f_G = f \circ \pi.$$

2. The Galois lift sends horizontal (resp., vertical) lines to horizontal (resp., vertical) lines, scaling by a factor of $\lambda > 1$ (resp., $\lambda^{-1} < 1$).

Note that f_G is discontinuous at w_j for $1 \leq j \leq r-1$, but is still continuous from the right at these points. The slightly tricky notation involved in (3.2) is merely to establish a choice of which branch of f to use at a critical point when defining f_G.

In this chapter, we investigate when Λ_f satisfies the following property.

Definition 3.2.5. Let $f : I \to I$ be a PCF uniform λ-expander with weak Markov partition $\{I_j\}$ and let $\pi : \Lambda_f \to I$ be projection onto the first coordinate. We say that Λ_f is *rectangular* if it has connected interior and each $R_j = \pi^{-1}(I_j)$ is a Euclidean rectangle.

Despite the terminology "Galois lift," we do not explicitly require λ and λ^{-1} to be Galois conjugate. There is, however, a natural integral polynomial $D_f(t)$ associated to a PCP λ-zig-zag f, called the *digit polynomial*, and under the condition that $f_G : \Lambda_f \to \Lambda_f$ defines a pseudo-Anosov this polynomial has both λ and λ^{-1} as roots (cf. Theorem A). Indeed, in all observed examples λ and λ^{-1} are Galois conjugate.

Definition 3.2.6. Let f be a PCP λ-zig-zag, and let $n \in \mathbb{N}$ be minimal such that $f^n(1) \in \partial I$. For each $0 \leq k < n$ let $f_k : \mathbb{C}^2 \to \mathbb{C}$ be the map of the form $f_k(x, z) = c_k \pm zx$ such that the restriction $f_k|_\lambda : x \mapsto f_k(x, \lambda)$ coincides with f on a neighborhood of $f^k(1)$ in I. Then the *digit polynomial* of f is the degree n polynomial $D_f : \mathbb{C} \to \mathbb{C}$ defined by

$$D_f(t) = \epsilon\big[f_{n-1}|_{z=t} \circ \cdots \circ f_0|_{z=t}(1) - f^n(1)\big]$$

where $\epsilon = \pm 1$ is a normalization factor to make D_f monic. By definition, $D_f(\lambda) = 0$.

Remark 3.2.7. The term "digit" polynomial is chosen to reflect the relation between the coefficients of $D_f(t)$ and the digits of the f-expansion of $x = 1$ (cf. Section 3.8) and Theorem C). Since $D_f(t) \in \mathbb{Z}[t]$, the minimal polynomial of λ divides the digit polynomial. The definition of $D_f(t)$ resembles that of the *Parry polynomial* of the λ-expander f (cf. [Tho16]).

3.3 Review of generalized pseudo-Anosovs

The purpose of this section is to review the theory of generalized pseudo-Anosovs, following the work of de Carvalho in [dC05] and de Carvalho and Hall in [dCH04]. A reader who is already familiar with the theory may skip this section.

3.3.1 Thick intervals

In this subsection we introduce thick interval maps and the procedure of *thickening* an interval map $f : I \to I$ to a thick interval map F.

A *thick interval* is a closed topological 2-disc $\mathbb{I} \subseteq S^2$ consisting of *decomposition elements* which come in two types: a *leaf*, homeomorphic to the interval $I = [0, 1]$, and a *junction*, homeomorphic to the closed 2-disc. The intersection of the boundary of a junction with $\mathbb{I}$ may consist of one or two connected components. We allow only finitely many junctions in a thick interval.

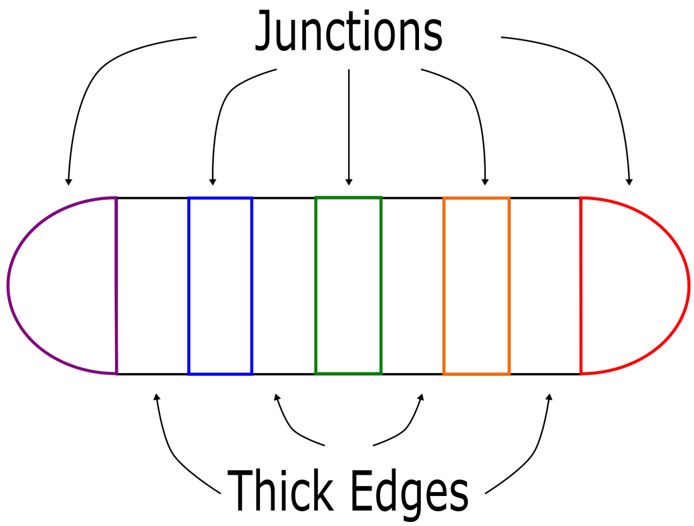


Figure 3.2: A typical thick interval consists of alternating thick edges and two-sided junctions, bookended by a pair of one-sided junctions.

We denote by $\mathbb{V}$ the union of the junctions of $\mathbb{I}$, and we refer to the connected components of $\mathbb{I} \setminus \mathbb{V}$ as *strips*. Each strip is homeomorphic to $(0,1) \times [0,1]$, and is a union of leaves. We put coordinates $h_s : \overline{s} \to [0,1] \times [0,1]$ on the closure of each strip s such that the leaves of s are precisely the sets

$$h_s^{-1}(\{x\} \times [0,1]), \quad x \in (0,1).$$

Following [dCH04] we denote by $\mathbb{E}$ the union of the closures of the strips. See Figure 3.2.

The notation $\mathbb{V}$, $\mathbb{E}$ is purposefully suggestive. As we shall see, a thick interval map is meant as a dynamical blow-up of an unrestricted interval map $f : I \to I$. The junctions correspond to elements of the weak postcritical set $\mathrm{WPC}(f)$ (cf. Section 3), whereas the strips correspond to the subintervals between these points.

We refer to [dCH04] for the following definition. We write $F : (X, A) \to (X, A)$ to represent a map $F : X \to X$ of topological spaces such that $F(A) \subseteq A$ for some $A \subseteq X$.

Definition 3.3.1. A *thick interval map* is an orientation-preserving homeomorphism

$F : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ such that

1. $F(\mathbb{I})$ is contained in the interior of $\mathbb{I}$,

2. if γ is a leaf of $\mathbb{I}$ then $F(\gamma)$ is contained in a decomposition element, and the diameter of $F^n(\gamma)$ with respect to the coordinates h_s tends to 0 as $n \to \infty$,

3. if J is a junction of $\mathbb{I}$ then $F(J)$ is contained in a junction,

4. F is linear with respect to the coordinates h_s: in each connected component of $s_i \cap F^{-1}(s_j)$, where s_i and s_j are strips, F contracts vertical coordinates uniformly by a factor $\mu_j < 1$ and expands horizontal coordinates uniformly by a factor $\lambda_j > 1$,

5. if J and J' are junctions such that $F(J) \subseteq J'$ then $F(\partial J \setminus \partial\mathbb{I}) \subseteq (\partial J' \setminus \partial\mathbb{I})$,

6. if J is a junction with $F^n(J) \subseteq J$ for some $n \geq 1$ then J has an attracting periodic point of period n in its interior whose basin contains the interior of J.

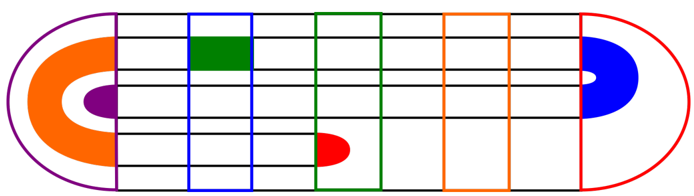


Figure 3.3: A thick interval map. The images of the junctions have been shaded darker for clarity. The first junction is mapped into itself, the second into the fifth, the third into the second, the fourth into the first, and the fifth into the third.

Definition 3.3.2. We associate to a thick interval map F a *transition matrix* $M = (m_{i,j})$ such that if $s_1, \ldots, s_n$ are the strips of $\mathbb{I}$ then $m_{i,j}$ is the number of times $F(s_j)$ crosses s_i. Note that since strips are separated by junctions and since junctions are mapped by F to other junctions, the $m_{i,j}$ are integers; there are no partial crossings.

Definition 3.3.3. A non-negative matrix M is *primitive* if there exists some $m \in \mathbb{N}$ such that M^m is positive, i.e. $(M^m)_{i,j}$ is positive for each i, j.

Definition 3.3.4. Let $F : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ be a thick interval map. Collapse each decomposition element of $\mathbb{I}$ to a point, obtaining an interval $\tilde{I}$ and an induced map $\tilde{f} : \tilde{I} \to \tilde{I}$. If $x \in \tilde{I}$ corresponds to a leaf that F maps into a junction, then $\tilde{f}$ will be constant in a neighborhood of x. Further collapsing these intervals of constancy produces either a single point or a new interval I with an induced map $f : I \to I$. We say that f is a *thinning* of F. Similarly, we say that F is a *thickening* of f, and *thickening* f refers to the process of associating to f a thick interval map F that thins to f.

Proposition 3.3.5. *[Theorem 2 in [dC05]] If F has primitive transition matrix M then thinning F always produces a nontrivial interval I. Moreover, the induced map $f : I \to I$ also has transition matrix M.*

Remark 3.3.6. Observe that while there is a unique thinning of a thick interval map, there are multiple thickenings of an interval map with at least one critical point. In Section 3 we investigate the proper thickening to choose for a zig-zag map if one hopes to produce a pseudo-Anosov.

3.3.2 Train tracks and generalized pseudo-Anosovs

In this subsection we describe how to associate to a given thick interval map F a branched 1-manifold τ invariant under F, up to isotopy. This τ is called a *generalized invariant train track* (cf. Definition 3.3.7), and provides the blueprints for constructing a *generalized pseudo-Anosov* $\psi : S^2 \to S^2$ (cf. Definition 3.3.10). In particular, τ dictates the structure of the singular invariant foliations of ψ.

To a thick interval $\mathbb{I}$ we associate the data of a finite set A of points, called *punctures*. Each puncture is contained in a junction, and each junction contains at most one puncture. For a strip s we define the arc γ_s to be the path $\gamma_s(t) = h_s^{-1}(t, 1/2)$.

Let RE denote the set of such paths. The endpoints of each arc γ_s are on the boundary components of s and are called *switches*. We denote by L the set of switches.

We again take the following definition from [dCH04].

Definition 3.3.7. Given a thick interval $\mathbb{I} \subseteq S^2$ with a set of punctures A, a *generalized train track* $\tau \subseteq \mathbb{I} \setminus A$ is a graph with vertex set L and countably many edges, each of which intersects $\partial\mathbb{V}$ only at L, such that

1. The edges of τ which intersect the interior of $\mathbb{E}$ are precisely the elements of RE, and
2. no two edges e_1, e_2 contained in a given junction J are *parallel*: that is, e_1 and e_2 may only bound a disc if it contains a point of A or another edge of τ.

Two generalized train tracks τ and τ' are *equivalent*, denoted $\tau \sim \tau'$, if they are isotopic by an isotopy supported on $\mathbb{V} \setminus A$.

Condition 1 says that the edges of τ contained in the strips of $\mathbb{I}$ are simple to describe: they are elements γ_s of RE, which are called *real edges*. The more complicated edges are those contained in the junctions, which are called *infinitesimal edges*. The collection of infinitesimal edges will be denoted by IE.

The infinitesimal edges will provide extra information not already given by the (finite) incidence matrix M for some thick interval map F. Indeed, we associate to F a specific generalized train track as follows. Let τ_0 denote the (disconnected) generalized train track given by the real edges $\gamma_s \in$ RE. We apply F to τ_0 and then perform the following series of pseudo-isotopies:

1. On each strip s we define the pseudo-isotopy $\eta_s : \overline{s} \times [0, 1] \to \overline{s}$ by

$$\eta_s(x, y, t) = (x, (1 - t)y + t/2)$$

2. Within each junction we define another pseudo-isotopy η_{e_1,e_2} for each pair of infinitesimal edges that are parallel, which homotopes e_1 and e_2 together.

The effect of the first set of pseudo-isotopies is to collapse all components of $F(\tau_0)$ contained within $\mathbb{E}$ to the real edges γ_s, while the second set homotopes parallel infinitesimal edges and is only supported on a disc containing the relevant junction. Composing these pseudo-isotopies produces a new generalized train track, denoted $\tau_1' := F_*(\tau_0)$. One may check that τ_1' is isotopic, relative to A, to a generalized train track τ_1 containing τ_0. Continuing in this way, we obtain an increasing sequence $\tau_0 \subseteq \tau_1 \subseteq \cdots$ of train tracks, and the union $\tau = \cup_{n\geq 0}\tau_n$ is *F-invariant*, i.e., $F_*(\tau)$ is isotopic to τ. See Figure 3.4.

Definition 3.3.8. Let $\eta : S^2 \times I \to S^2$ denote the composition of the pseudo-isotopies in steps 1 and 2 above. The generalized train track $\tau = \cup_{n\geq 0}\tau_n$ is called the *invariant generalized train track* for F. The *train track map* associated to F is the map $f : \tau \to \tau$ defined by $f(x) = \eta(F(x), 1)$.

The choice to denote the generalized train track by f is a slight but intentional abuse of notation. If F is the thickening of a PCF interval map $f : I \to I$ then we think of $f : \tau \to \tau$ as an extension of the transformation to a graph with a richer structure.

Let $\mathbb{I}$ be a thick interval and $F : (S^2, \mathbb{I}, A) \to (S^2, \mathbb{I}, A)$ be a thick interval map with A a finite invariant set of F. Let τ be the associated invariant generalized train track. τ has at most countably many edges. Label the finitely many real edges $e_1, \ldots, e_n$ and then label the possibly infinitely many infinitesimal edges e_k for $k \geq n+1$. We form an *extended transition matrix* $N = (n_{i,j})$ by setting

$$n_{i,j} = \text{the number of times } f(e_j) \text{ crosses } e_i.$$

We may write this as the block matrix

$$N = \begin{pmatrix} M & 0 \\ B & \Pi \end{pmatrix}$$

where M is the incidence matrix of F, B records the transitions from real to infinitesimal edges, and Π the transitions from infinitesimal edges to other infinitesimal edges.

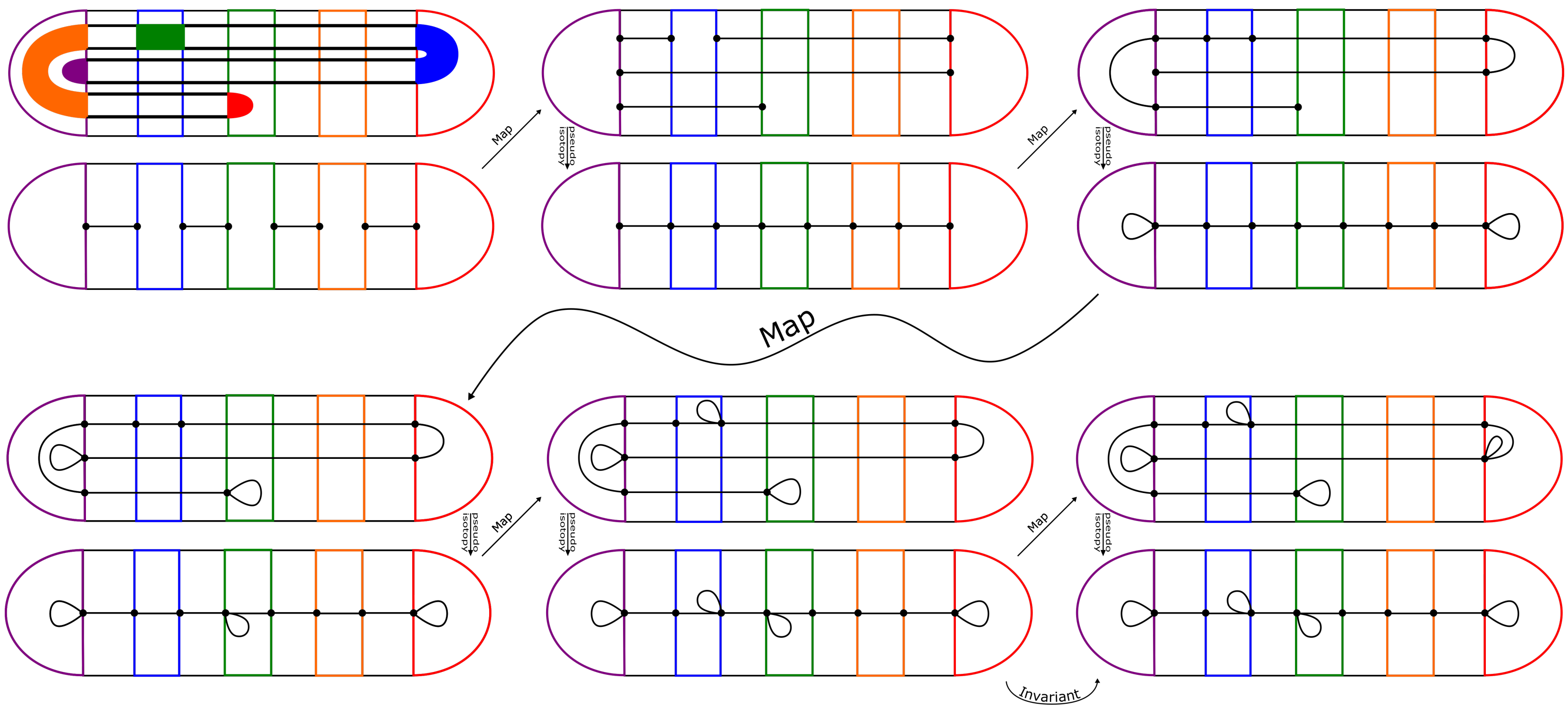


Figure 3.4: The process for generating an invariant train track for a given thick interval map. One alternately applies the map to τ_n and the pseudo-isotopies to obtain τ_{n+1}. In this case, the process terminates with τ_4, which is invariant. Since the invariant generalized train track is finite, the resulting surface homeomorphism will be a pseudo-Anosov (cf. Figure 3.6). In general, however, this process may continue indefinitely, producing a *generalized* pseudo-Anosov with infinitely many singularities. See Figures 3.7 and 3.8 for such an example.

We assume from now on that M is primitive. In this case, the Perron-Frobenius theorem states that the spectral radius $\rho(M) > 0$ of M is in fact an eigenvalue of M, called the *dominant eigenvalue* of M. The Perron-Frobenius theorem further states that the dominant eigenvalue is simple, and that the associated one-dimensional eigenspace is spanned by a positive eigenvector $u = (u_i)$, while no other eigenspace contains a positive eigenvector. We normalize this eigenvector to have unit L^1-norm: that is, $\sum_i u_i = 1$.

If $\lambda = \rho(M)$ is the Perron eigenvalue of M, let $x' = (x_1, \ldots, x_n)$ denote the canonical positive left λ-eigenvector associated to M such that $\sum_i x_i = 1$, and let $y' = (y_1, \ldots, y_n)$ denote the positive right λ-eigenvector of M such that $\sum_i x_i y_i = 1$. One shows that these can be extended to left- and right- λ-eigenvectors of N, i.e. that there exist possibly infinite vectors x, y such that $xN = \lambda x$ and $Ny = \lambda y$. Moreover, it is not hard to see that $x = (x_1, \ldots, x_n, 0, 0, \ldots)$.

For $i = 1, \ldots, n$, construct rectangles R_i of dimensions $x_i \times y_i$. These are the building blocks of the surface on which the generalized pseudo-Anosov will act. The infinitesimal edges incident to one endpoint of a real edge e_i, along with their weights, encode how to identify segments of the corresponding vertical boundary of R_i. While this process is visually intuitive, a precise explanation is nonetheless elusive in the literature, so we describe it here for completeness.

In what follows, we fix a junction J between two adjacent real edges e_L and e_R. Denote by v_L the endpoint of e_L on ∂J, and similarly define v_R.

Definition 3.3.9. Let e be an infinitesimal edge contained in J, considered as a smooth parameterized arc $e : [0, 1] \to J$. Observe that $e(\{0, 1\}) \subseteq \{v_L, v_R\}$. We define an *end* of e to be an arc of the form

$$\alpha = e\left(\left[0, \frac{1}{2}\right]\right) \text{ or } \alpha = e\left(\left[\frac{1}{2}, 1\right]\right).$$

If α is an end of e, then for $0 < \epsilon < \frac{1}{2}$ we define the *ϵ-subend* of α to be

$$\alpha_\epsilon = \begin{cases} e([0,\epsilon]) & \text{if } \alpha = e\left(\left[0,\frac{1}{2}\right]\right) \\ e([1-\epsilon,1]) & \text{if } \alpha = e\left(\left[\frac{1}{2},1\right]\right). \end{cases}$$

If e, f are two infinitesimal edges of J, not necessarily distinct, with ends α, β incident to v_L, we set

$$\alpha \leq_L \beta \text{ if } \alpha_\delta \text{ is below } \beta_\delta \text{ for all } \delta > 0 \text{ sufficiently small.}$$

If instead α, β are incident to v_R, we set

$$\alpha \leq_R \beta \text{ if } \alpha_\delta \text{ is below } \beta_\delta \text{ for all } \delta > 0 \text{ sufficiently small.}$$

It is not difficult to see that $\leq_L$ is a total order on the set of ends incident to v_L. That is, for any two arcs α, β incident to v_L, we either have $\alpha \leq_L \beta$ or $\beta \leq_L \alpha$. Moreover, if

- $e_{j_1}, e_{j_2}, \ldots$ are the infinitesimal edges with one end incident to e_L,
- $e_{k_1}, e_{k_2}, \ldots$ are the infinitesimal edges with two ends incident to e_L, and
- the rectangle R_{e_L} has height $y(L)$,

then the fact that y is a right-λ-eigenvector for N implies that we have

$$y(L) = \sum_i y_{j_i} + 2 \cdot \sum_l y_{k_l} \tag{3.3}$$

Equation (3.3) is often referred to as a *switch condition.*

For an end α, denote by $e(\alpha)$ the infinitesimal edge of which α is an end, and for an infinitesimal edge e, let $y(e)$ denote the entry of y corresponding to e. Then equation (3.3) and the fact that $\leq_L$ is a total order together imply that there is a unique way to partition the right vertical boundary of R_{e_L} into segments d_α of length $y(e(\alpha))$ such that for two ends α, β incident to e_L,

$$d_\alpha \text{ is below } d_\beta \iff \alpha \leq_L \beta$$

The same argument shows how to partition the left vertical boundary of R_{e_R}. It remains to describe how to identify these boundary segments. If e is an infinitesimal edge incident to both e_L and e_R, and α, β are the corresponding ends of e, then $y(e(\alpha)) = y(e(\beta)) = y(e)$ and we identify d_α on $\partial_V R_{e_L}$ with d_β on $\partial_V R_{e_R}$ by an orientation-preserving isometry. If instead both ends α, β of e are incident to e_L then we identify the segments d_α, d_β on $\partial_V R_{e_L}$ by an orientation-reversing isometry, and similarly if both ends α, β are incident to e_R. See Figure 3.5 for an example with finitely many infinitesimal edges, and Figure 3.8 for an example with infinitely many.

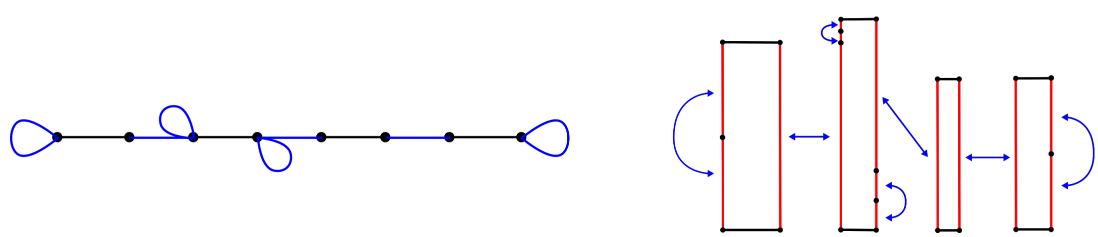


Figure 3.5: The rectangle decomposition of a surface from the invariant train track. The black edges are real edges, while the blue edges are infinitesimal edges. The arrows in the right-hand picture point to pairs of segments that are to be identified isometrically.

The train track map f induces an endomorphism $\tilde{\Psi} : \mathcal{R} \to \mathcal{R}$ which stretches the foliation of $\mathcal{R}$ by horizontal lines by a factor of λ, and it scales the vertical foliation by a factor of λ^{-1}. This map $\tilde{\Psi}$ is a homeomorphism except on the boundary of $\mathcal{R}$. This boundary is a topological circle and contains a periodic orbit of $\tilde{\Psi}$. After identifying adjacent segments of this circle that eventually map to the same segment, we obtain a homeomorphism Ψ in the quotient. This new quotient surface is homeomorphic to S^2, and the periodic orbit on $\partial\mathcal{R}$ becomes a single point, called the *point at infinity.* The induced map $\psi : S^2 \to S^2$ inherits stable and unstable foliations, and is a *generalized* *pseudo-Anosov.*

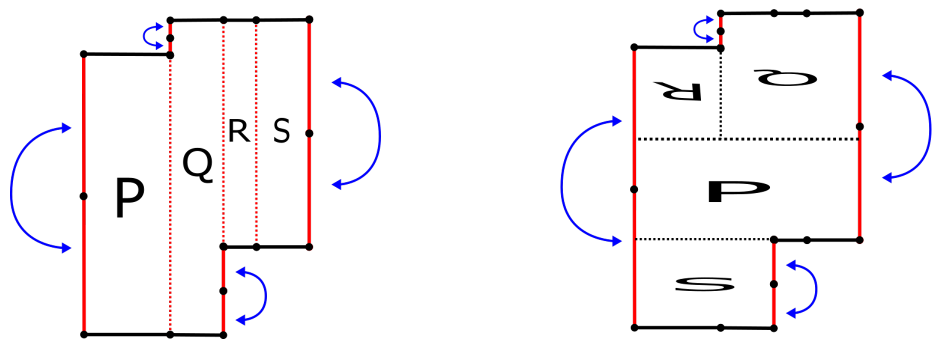


Figure 3.6: The transformation $\tilde{\Psi} : \mathcal{R} \to \mathcal{R}$. Identifying points on the horizontal boundary that are eventually mapped to the same point produces a homeomorphism of S^2. Since there are only finitely many singularities, this is a pseudo-Anosov.

Definition 3.3.10. A *generalized pseudo-Anosov* is a homeomorphism ψ of a compact surface S such that the following hold:

1. there exists a number $\lambda > 1$ and two transverse singular measured foliations $(\mathcal{F}_u, \mu_u)$, $(\mathcal{F}_s, \mu_s)$ of S such that

$$\psi \cdot (\mathcal{F}_u, \mu_u) = (\mathcal{F}_u, \lambda\mu_u), \quad \psi \cdot (\mathcal{F}_s, \mu_s) = (\mathcal{F}_s, \lambda^{-1}\mu_s).$$

2. the singularities of $\mathcal{F}_u$ and $\mathcal{F}_s$, while potentially infinite in number, accumulate on only finitely many points of S.

Remark 3.3.11. Observe that in the case that the foliations have only finitely many singularities, ψ is a pseudo-Anosov.

In what follows we will investigate closely the construction of generalized pseudo-Anosovs from thick interval maps, focusing particularly on the circumstances under which this process produces a pseudo-Anosov. This occurs if and only if the generalized invariant train track is finite, i.e. has only finitely many edges. In this case τ is simply a train track on a punctured sphere.

Example 3.3.12. It will be instructive to see an example where the invariant generalized train track has infinitely many edges, so that the resulting homeomorphism ψ is not a pseudo-Anosov. Let $\lambda = 1 + \sqrt{2}$ and $f : I \to I$ the uniform λ-expander defined by

$$f(x) = \begin{cases} \lambda x & \text{if } 0 \leq x < \lambda^{-1} \\ 2 - \lambda x & \text{if } \lambda^{-1} \leq x < 2\lambda^{-1} \\ \lambda x - 2 & \text{if } 2\lambda^{-1} \leq x \leq 1. \end{cases}$$

We see that f has two critical points, $k_1 = \lambda^{-1}$ and $k_2 = 2\lambda^{-1}$. The first of these is 2-periodic, with $f(k_1) = 1$ and $f(1) = k_1$. On the other hand, $f(k_2) = 0$ is a fixed point. Thus the weak postcritical set of f is $\mathrm{WPC}(f) = \{0, k_1, k_2, 1\}$. We thicken this set to junctions and the intermediate subintervals to thick edges.

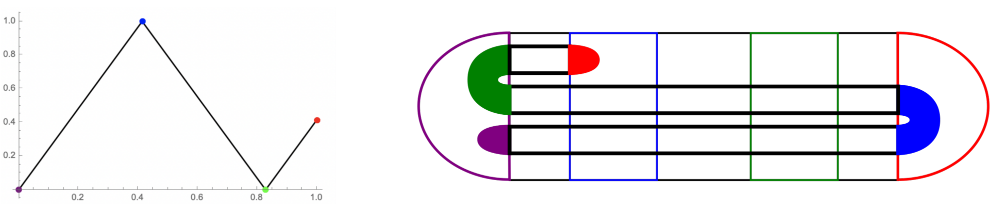


Figure 3.7: A uniform λ-expander for $\lambda = 1 + \sqrt{2}$ and a thickening of it. Note that in the first junction the tightening pseudo-isotopies will produce non-parallel loops of the generalized invariant train track.

Figure 3.7 shows an example of a thickening F of f. In this case the generalized invariant train track for F has infinitely many edges, corresponding to infinitely many singularities for the resulting sphere homeomorphism. See Figure 3.8. It is not hard to show that any orientation-*preserving* thickening of f will fail to produce a pseudo-Anosov. Interestingly enough, there does exist an orientation-reversing map that accomplishes this, and indeed if we define the Galois lift f_G of f by replacing all instances of λ with its conjugate $-\lambda^{-1}$, then the limit set Λ_f is rectangular.

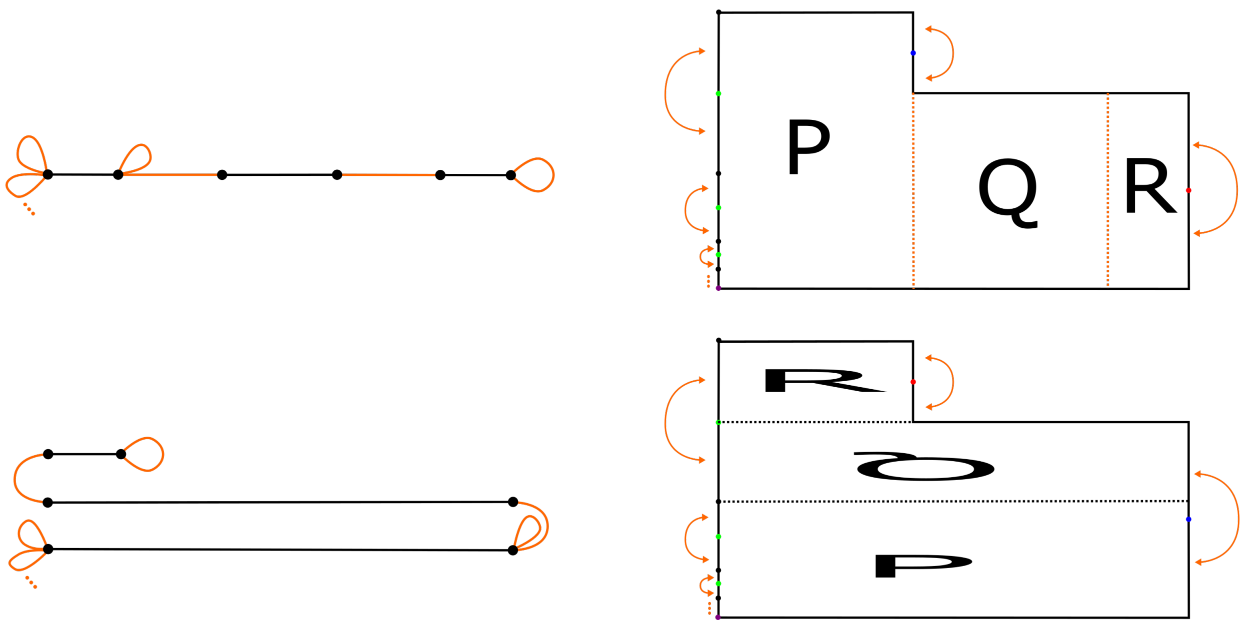


Figure 3.8: The action of F on its generalized invariant train track, and the induced map on the corresponding surface.

3.4 First constructions

We turn our attention to the task of constructing a pseudo-Anosov on S^2 from a zig-zag map (cf. Definition 3.2.2). Our goal in this section is to show that for a zig-zag map there is a unique pair of thickenings to consider when attempting to construct a pseudo-Anosov: that is, no other possible thickenings can give a pseudo-Anosov (cf. Theorem 3.4.19). Moreover, these thickenings produce conjugate generalized pseudo-Anosovs, so it suffices to only consider one of them.

Before we proceed to the proof of Theorem 3.4.19, we must consider how to thicken zig-zag maps, and whether they can be thickened in the first place. For this we will rely on Proposition 3.3.5.

3.4.1 Some ergodic theory for zig-zag maps

In this subsection we show that every zig-zag map is weak-mixing. It follows that the transition matrix for a postcritically finite λ-zig-zag with $\lambda > 2$ is primitive.

In [Hal94] Hall classifies the λ-zig-zags of pseudo-Anosov type for $1 < \lambda \leq 2$. A unimodal zig-zag map is called a *tent map*. Much of our focus will be on PCP λ-zig-zags for $\lambda > 2$. It is therefore important for the generalized pseudo-Anosov construction that we ensure that the strong Markov matrix of such a map is primitive. Theorem 3.4.1 essentially accomplishes this goal, and its proof uses a result by Wilkinson [Wil75].

Theorem 3.4.1. *If $f : I \to I$ is a λ-zig-zag for some $\lambda > 2$, then f is weak-mixing with respect to Lebesgue measure.*

Proof. Let $\lfloor \lambda \rfloor = m \geq 2$ be the number of critical points of f and set $P_0 = [0, k_1]$, $P_m = [k_m, 1]$, and $P_i = [k_i, k_{i+1}]$ for $i = 1, \ldots, m-1$. Following [Wil75], set

$$\Delta(j_1, \ldots, j_n) = P_{j_1} \cap f^{-1} P_{j_2} \cap \cdots \cap f^{-(n-1)} P_{j_n}$$

We say $\Delta(j_1, \ldots, j_n)$ is *full of rank n* if $\mu(f^n(\Delta(j_1, \ldots, j_n))) = 1$, where μ is Lebesgue measure; otherwise $\Delta(j_1, \ldots, j_n)$ is said to be *non-full*.

For the n-tuple $(j_1, \ldots, j_n)$ let $\mathfrak{l}(j_1, \ldots, j_n)$ be the number of non-full intervals of positive measure of the form $\Delta(j_1, \ldots, j_n, i)$, for $1 \leq i \leq m+1$. Define

$$\mathfrak{l}_n = \sup \mathfrak{l}(j_1, \ldots, j_n),$$

where the supremum is taken over all n-tuples $(j_1, \ldots, j_n)$ such that $\Delta(j_1, \ldots, j_n)$ has positive Lebesgue measure. Wilkinson shows in [Wil75] that f is weak-mixing as long as

$$\mathfrak{l} = \sup_n \mathfrak{l}_n < \lambda.$$

However, note that all P_i are full for $i = 0, \ldots, m-1$. The full subsets of $\Delta(j_1, \ldots, j_n)$ of rank $n+1$ and positive measure are of the form $\Delta(j_1, \ldots, j_n, j)$ for $0 \leq j \leq J$, where J depends on the ordered n-tuple $(j_1, \ldots, j_n)$. If $m \geq 2$ the only one of these that can be non-full is $\Delta(j_1, \ldots, j_n, J)$, hence $\mathfrak{l} \leq 1 < \lambda$. □

Definition 3.4.2. A subset $E \subseteq \mathbb{N}$ is said to have *density 0* if

$$\lim_{n\to\infty} \frac{|E \cap \{1, \ldots, n\}|}{n} = 0.$$

The union of two sets of density 0 also has density 0. It is well-known that a dynamical system $(X, \mathcal{B}, \mu, T)$ is weak-mixing if and only if for every $A, B \in \mathcal{B}$ there exists a set $E = E(A, B) \subseteq \mathbb{N}$ of density 0 such that

$$\lim_{E \not\ni n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

Recall that a non-negative matrix M is *primitive* if there is some $k \in \mathbb{N}$ such that every entry of M^k is positive.

Corollary 3.4.3. *If $f : I \to I$ is a PCP λ-zig-zag map for some $\lambda > 2$, then the transition matrix of f is primitive.*

Proof. Let I_j, I_k be two subintervals in the weak Markov partition of f. Since f is weak-mixing and both I_j and I_k have positive measure, there exists a subset $E_{j,k} \subseteq \mathbb{N}$ of density 0 such that $\mu(I_j \cap f^{-i}(I_k))$ is positive for large $i \notin E_{j,k}$. The set $E = \cup_{j,k} E_{j,k}$ also has density 0, so there exists some large i so that $\mu(I_j \cap f^{-i}(I_k)) > 0$ for all j, k. In particular, W_f^i is positive. □

3.4.2 Passing between intervals and thick intervals

Recall that from a thick interval map $F : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ we obtain an interval map $f : I \to I$ by first collapsing all decomposition elements of $\mathbb{I}$ to points, and then further collapsing any subintervals on which the induced map is constant. We call this composition of quotient maps the *thinning projection* and denote it by $\pi : \mathbb{I} \to I$.

Note that by definition, $\pi \circ F|_{\mathbb{I}} = f \circ \pi$.

Lemma 3.4.4. *Let $F : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ be a thickening of $f : I \to I$. Then for each $x \in \mathrm{PC}(f)$ there exists a junction $J_x \subseteq \mathbb{I}$ such that $\pi(J_x) = x$. In particular, since thick intervals have only finitely many junctions, f cannot be thickened unless it is PCF.*

Proof. Let $\pi : \mathbb{I} \to I$ denote the thinning projection. Let k be a critical point of f, and let $K = \pi^{-1}(k)$ be the set of decomposition elements projecting to k. Similarly, let $K' = \pi^{-1}(f(k))$.

We may assume without loss of generality that k is a local maximum of f. Therefore, there exist points x_1, x_2 satisfying $x_1 < k < x_2$ such that $f(x_1) = f(x_2) < f(k)$. Let $X_i = \pi^{-1}(x_i)$ for $i = 1, 2$. Then each $F(X_i)$ lies to the left of K', whereas $F(K) \subseteq K'$. Since F is linear with respect to the coordinates h_s (cf. Definition 3.3.1), $F(\mathbb{I})$ must pass through a junction J after $F(X_1)$ and before $F(X_2)$. Picking x_i arbitrarily close to k shows that in fact $F(K) \subseteq J$. But now $J \cap K' \neq \emptyset$, and since K' is the collection of all decomposition elements that project to $f(k)$ it follows that $J \subseteq K'$.

F maps junctions into junctions, so $F^n(K')$ must be contained in a junction for each $n \geq 1$. But $\pi(F^n(K')) = f^n(f(k))$, so by the same argument as before, $\pi^{-1}(f^n(f(k)))$ contains a junction for each n. Repeating this procedure for all critical points shows that $\pi^{-1}(x)$ contains a junction for each $x \in \mathrm{PC}(f)$. $\square$

Example 3.4.5. We revisit Example 3.3.12, demonstrated in Figure 3.7. $\mathrm{PC}(f)$ consists of the first, second, and fourth heavily drawn points, and the thick interval map F has a junction for each of these. Note that F also has a junction for the third point $k_2 = 2\lambda^{-1}$, despite the fact that k_2 is not a postcritical point. This junction is not strictly necessary, since no junction maps into it, but it does allow us to identify the weak Markov matrices of F and f: each is given by

$$W_f = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}.$$

One checks that $\chi_W(f;t) = t(t^2 - 2t - 1)$ has dominant root $\lambda = 1 + \sqrt{2}$, the growth rate of f. Eliminating the extraneous junction over k_2 has the effect of combining the second and third subintervals, producing the strong Markov matrix

$$M_f = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

with characteristic polynomial $\chi_M(f;t) = t^2 - 2t - 1$. See Remark 3.4.9 below.

Definition 3.4.6. We say that a map $f : I \to I$ is *of pseudo-Anosov type* if the following two conditions hold:

1. the strong Markov matrix of f is primitive.

2. there is a thickening $F : \mathbb{I} \to \mathbb{I}$ of f whose invariant generalized train track is finite, i.e., is a train track in the classical sense.

Note that a map of pseudo-Anosov type is necessarily PCF. As the next theorem shows, such a map must in fact be PCP.

Theorem 3.4.7. *Let $f : I \to I$ be of pseudo-Anosov type, and let F be a thickening of f with a finite invariant generalized train track τ. Then τ has exactly one loop in each junction corresponding to an element of* $\mathrm{PC}(f)$*. Moreover, these loops determine the one-pronged singularities of the pseudo-Anosov ψ, and f is in fact PCP.*

Proof. Each loop of τ determines a one-pronged singularity of ψ. Multiple loops in a single junction J would imply that ψ has singularities connected by a leaf of its stable (vertical) foliation. It is well known that such a thing, called a *saddle connection*, cannot occur in an invariant foliation of a pseudo-Anosov. Thus each junction V, corresponding to a point $v \in \mathrm{WPC}(f)$, contains either at most one loop of τ. We claim that V contains a loop of τ if and only if v is in the subset $\mathrm{PC}(f)$.

To see this, observe that V contains a loop if v is a critical value of f, since in this case $F(\tau)$ makes a turn through V, and after pseudo-isotopy this turn pinches to a loop. By the invariance of τ, it now follows that V contains a loop if v is in the forward orbit of a critical point, which is to say that $v \in \mathrm{PC}(f)$. The only other possibility remaining for v is that it is an element of $\mathrm{WPC}(f) \setminus \mathrm{PC}(f)$, i.e. a critical point that is not in the forward orbit of any critical point. But then no junction of $\mathbb{I}$ maps into V, so no loop of τ maps into V. Furthermore, $F(\tau)$ does not make a turn through V, since that would imply that v is a critical value, which is it not. These are the only ways a loop of τ will appear in V during the process of constructing τ, so in fact V does not contain a loop of τ if $v \notin \mathrm{PC}(f)$.

To finish the proof, note that a pseudo-Anosov permutes its 1-prong singularities. Hence f acts on the elements of $\mathrm{PC}(f)$ by a permutation: in other words, f is PCP. $\square$

Definition 3.4.8. Let $f : I \to I$ be of pseudo-Anosov type, and let τ be the invariant train track associated to a thickening F of f. We say an infinitesimal edge of τ is a *connecting edge* if it is not a loop, i.e. if it joins distinct adjacent real edges.

Remark 3.4.9. Theorem 3.4.7 demonstrates that the important points of f to consider are the elements of $\mathrm{PC}(f)$, rather than those of $\mathrm{WPC}(f)$, since these are the points that correspond to the singularities of any generalized pseudo-Anosov obtained from a thickening of f.

Example 3.4.10. Here is an example of a map of pseudo-Anosov type (cf. Figure 3.4). Let $f : I \to I$ be the positive zig-zag map of growth rate $\lambda = (3 + \sqrt{5})/2$, the dominant root of $x^2 - 3x + 1$. The orbit of $x = 1$ is periodic of period 3 and includes

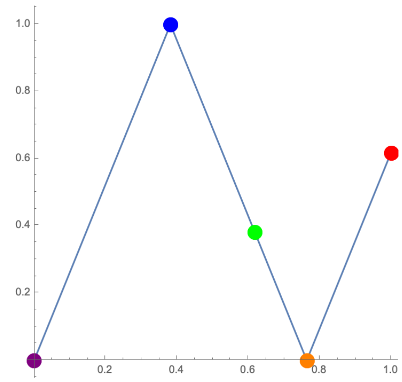


Figure 3.9: The positive zig-zag map for $\lambda = (3+\sqrt{5})/2$. Compare with Figures 3.4 and 3.10.

the first critical point $k_1 = \lambda^{-1}$. The other postcritical orbit is the forward orbit of $x = k_2$, which maps to the fixed point at $x = 0$. See Figure 3.9. Thus

$$\mathrm{PC}(f) = \{0, k_1, v, 1\} \quad \text{and} \quad \mathrm{WPC}(f) = \{0, k_1, v, k_2, 1\}.$$

Thus we may verify that f is PCP, acting on $\mathrm{PC}(f)$ by the permutation $(1)(2,4,3)$. The strong and weak Markov matrices for f are

$$M_f = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad W_f = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

The strong and weak Markov polynomials of f are $\chi_M(f;t) = (t+1)(t^2-3t+1)$ and $\chi_W(f;t) = t(t+1)(t^2-3t+1)$. Both matrices are primitive: we have

$$M_f^2 = \begin{pmatrix} 3 & 2 & 2 \\ 3 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad \text{and} \quad W_f^2 = \begin{pmatrix} 3 & 2 & 1 & 1 \\ 3 & 2 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 \end{pmatrix}.$$

3.4.3 The exterior left-veering thickening

In this subsection we investigate the possible thickenings of a PCP zig-zag, and show that only two of these thickenings have a chance of producing a pseudo-Anosov. These are the *exterior left-* and *exterior right-veering* thickenings F_L and F_R (cf. Definition 3.4.16). These thickenings "swirl" out from the center, turning either always left or always right, respectively. Since these maps are topologically conjugate, we therefore restrict our analysis to F_L in future sections.

Let f be a zig-zag map of growth rate $\lambda > 2$ and critical points $k_i = i \cdot \lambda^{-1}$ for $i = 1, \ldots, \lfloor \lambda \rfloor = m$. We assume that f is PCP, which is equivalent to the orbit of $x = 1$ being periodic. Let $0 \leq v_1 < v_2 < \ldots < v_n = 1$ be the orbit of 1. These, along with $x = 0$ if it is not already among the v_i, are precisely the points to which the junctions of our thick interval $\mathbb{I}$ will project. To specify a particular thick interval map F projecting to f, however, it remains to determine how F folds $\mathbb{I}$, i.e. how the image $F(\mathbb{I})$ turns within $\mathbb{I}$.

If $F : \mathbb{I} \to \mathbb{I}$ is a thickening of $f : I \to I$ with thinning projection π, then for brevity we will denote by $\mathbf{0}$ the junction satisfying $\pi(\mathbf{0}) = 0$. Similarly, $\mathbf{1}$ will denote the junction satisfying $\pi(\mathbf{1}) = 1$.

The image $F(\mathbb{I}) \subseteq \mathbb{I}$ is a collection of thick intervals stacked vertically and stretching between $\mathbf{0}$ and $\mathbf{1}$, with at most a single exception. Orienting $\mathbb{I}$ from left to right, we number these thick subintervals $\mathbb{I}_j$ as we travel along $F(\mathbb{I})$, beginning at $\mathbf{0}$ if f is a positive zig-zag and at $\mathbf{1}$ if f is a negative zig-zag.

Definition 3.4.11. The height of $\mathbb{I}_j$ is defined to be $h(\mathbb{I}_j) = i$ if $\mathbb{I}_j$ is the i-th thick subinterval from the bottom. The *type* of the thick interval map F is then the element σ of the permutation group $\mathfrak{S}_{m+1}$ such that

$$\sigma(j) = i \iff h(\mathbb{I}_j) = i.$$

For an example, see Figure 3.10 below.

Example 3.4.12. Let f be the positive zig-zag map for $\lambda = (3+\sqrt{5})/2$ from Example 3.4.10. We have $\lfloor \lambda \rfloor = 2$, and f is postcritically periodic of length 3 with $\mathrm{PC}(f)$ given by

$$0 < k_1 = v_1 < v_2 < k_2 < v_3 = 1.$$

More specifically, the orbit of $x = 1$ is $1 \mapsto v_2 \mapsto v_1 \mapsto 1$. The six possible thickenings of f are pictured in Figure 3.10. These are paired according to the topological conjugacy class of the resulting generalized pseudo-Anosov. Note that the conjugacy classes in this example are the elements of the orbit space of ι_3 acting on $\mathfrak{S}_3$, where ι_3 is the order-reversing permutation

$$\iota_3 = (1,3)(2).$$

See Definition 3.4.14 below.

Example 3.4.13. Not every permutation $\sigma \in \mathfrak{S}_{m+1}$ gives a valid thick interval map. For example, let f be a PCP λ-zig-zag map such that $\lfloor \lambda \rfloor = 3$. Then regardless of the orbit of $x = 1$, there can be no thickening of f of type

$$\sigma = (1)(2,3)(4),$$

since the image of this thick interval map would necessarily intersect itself.

Definition 3.4.14. The *orientation-reversing permutation* $\iota_n \in \mathfrak{S}_n$ is the permutation such that

$$\iota_n(i) = n + 1 - i \quad \text{for all } 1 \leq i \leq n.$$

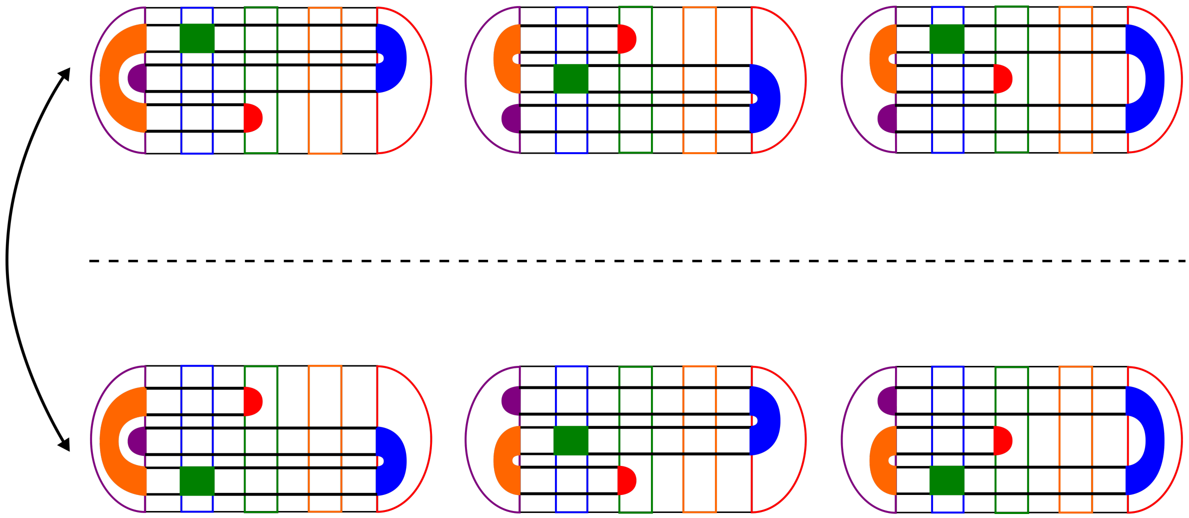


Figure 3.10: The six possible thickenings of the positive zig-zag for $\lambda = (3+\sqrt{5})/2$. The permutation type of each is an element of $\mathfrak{S}_3$, determined by the heights of the horizontal layers of $F(\mathbb{I})$. For example, the top left thick interval map has permutation type $\sigma_1 = (1,2,3)$, and the top right has permutation type $\sigma_2 = (1)(2,3)$. Each column is an orbit of the action of i_3.

We will often suppress notation and write $\iota_n = \iota$ when n is understood from context. Note that $\iota^2 = \mathrm{id}$, so if we let ι_n act on $\mathfrak{S}_n$ on the left then the orbit space is parameterized by pairs of elements of $\mathfrak{S}_n$. In general ι_n is not central in $\mathfrak{S}_n$, so the orbit space is not a group. The following proposition says that the elements of the orbit space correspond to topological conjugacy classes of generalized pseudo-Anosovs.

Proposition 3.4.15. *Let f be a PCP zig zap map and F_1, F_2 thickenings of f of type σ_1, σ_2 respectively. If $\sigma_1 = \iota\sigma_2$ then F_1 is topologically conjugate to F_2 via the orientation-reversing homeomorphism $i : \mathbb{I} \to \mathbb{I}$ that reflects through the horizontal midline of $\mathbb{I}$. Consequently, if ψ_i is the generalized pseudo-Anosov induced by F_i, then ψ_1 is topologically conjugate to ϕ_2.*

Proof. One immediately checks that if $\sigma_1 = \iota\sigma_2$ then $F_1 = i \circ F_2 \circ i^{-1}$. Following through the details of the construction given in Section 2, we see that the invariant generalized train tracks τ_i of F_i satisfy $\tau_1 \sim i_*(\tau_2)$, and so the resulting generalized pseudo-Anosovs are conjugate by the homeomorphism induced by i. □

Definition 3.4.16. The *positive exterior left-veering permutation* is the element σ_L^+ of $\mathfrak{S}_n$ defined as follows: if n is even, then

$$\sigma_L^+(k) = \begin{cases} \frac{n+k}{2} & \text{if } k \text{ is even} \\ \frac{n+1-k}{2} & \text{if } k \text{ is odd.} \end{cases}$$

If instead n is odd, then we define

$$\sigma_L^+(k) = \begin{cases} \frac{n+1+k}{2} & \text{if } k \text{ is even} \\ \frac{n-k}{2}+1 & \text{if } k \text{ is odd.} \end{cases}$$

The *negative exterior left-veering permutation* $\sigma_L^- \in \mathfrak{S}_n$ is defined by

$$\sigma_L^-(k) = n + 1 - \sigma_L^+(k).$$

The *positive (resp. negative) exterior right-veering permutation* is $\sigma_R^+ = \iota \circ \sigma_L^+$ (resp., $\sigma_R^- = \iota \circ \sigma_L^-$). If F is a thick interval map of type $\sigma_L^\pm$ we also say F is *exterior left-veering*, and write $F = F_L^\pm$. Similarly we denote by $F_R^\pm$ the *exterior right-veering* thick interval of permutation type $\sigma_R^\pm$.

Example 3.4.17. For $\mathfrak{S}_3$, σ_L^+ and σ_R^+ are shown in Figure 3.10 as the top and bottom of the leftmost column, respectively. For $\mathfrak{S}_4$ we have

$$\sigma_L^+ = (1,2,3)(4) \quad \text{and} \quad \sigma_R^+ = (1,3,4)(2).$$

Remark 3.4.18. Given a PCP zig-zag f, only one pair of thickenings is defined: either F_L^+ and F_R^+ if f is positive, or F_L^- and F_R^- if f is negative. In this case we will drop the superscripts from the notation and simply refer to F_L and F_R.

Proposition 3.4.19. *Let $f : I \to I$ be a PCP zig-zag map of pseudo-Anosov type with growth rate $\lambda > 2$, and let $F : \mathbb{I} \to \mathbb{I}$ be a thickening of f that induces a pseudo-*

Anosov. Then $F = F_L$ or $F = F_R$.

Proof. Since F induces a pseudo-Anosov, Theorem 3.4.7 implies that each of $\mathbf{0}$ and $\mathbf{1}$ contains a single loop of the invariant generalized train track τ. For each of these junctions, the turns of F that pass through them must be concentric, since otherwise a second loop would appear. Orienting $F(\mathbb{I})$ from $F(\mathbf{0})$ to $F(\mathbf{1})$, it follows inductively that each turn has to be in the same direction as the previous one. Thus if the first turn is to the left, all turns are to the left and $F = F_L$. Similarly, if the first turn is to the right then $F = F_R$. □

It is always possible to construct a thick interval map F having permutation type σ_L or σ_R projecting to a given PCP zig-zag map. Indeed, such a thick interval map swirls out from the center, turning to the left if it is has type σ_L and to the right if it has type σ_R. In particular, it follows from Propositions 3.4.15 and 3.4.19 that associated to a given PCP zig-zag map $f : I \to I$ is a canonical thick interval map to consider when investigating whether f is of pseudo-Anosov type: namely, F_L.

3.5 Reconciling two constructions

In this section we prove Theorem A, establishing the connection between generalized pseudo-Anosovs and the Galois lift f_G of Thurston (cf. Definition 3.2.3.) We recall the statement of the theorem.

Theorem A. *Let $f : I \to I$ be a PCP λ-zig-zag map with $\lambda > 2$. Then f is of pseudo-Anosov type if and only if the following conditions are satisfied:*

1. *the digit polynomial D_f of f has λ^{-1} as a root, and*
2. *the limit set Λ_f of f_G is rectangular.*

In this case, the invariant generalized train track τ_L of F_L is finite, and recovers the action of f_G on Λ_f in the following way. Let S' be the closed topological disc *obtained by performing the gluings indicated by the non-loop infinitesimal edges of* τ_L*. Let* $\tilde{f} : S' \to S'$ *be the map induced by* F_L*. Then there is an isometry* $i : S' \to \Lambda_f$ *such that the following diagram commutes:*

$$\begin{array}{ccc} S' & \xrightarrow{\tilde{f}} & S' \\ \big\downarrow{\scriptstyle i} & & \big\downarrow{\scriptstyle i} \\ \Lambda_f & \xrightarrow{f_G} & \Lambda_f \end{array}$$

Moreover, i *sends the stable and unstable foliations of* S' *to those of* Λ_f*. Therefore, after identifying segments of boundary in each set so as to obtain pseudo-Anosovs* $\psi_1 : S \to S$ *and* $\psi_2 : \overline{\Lambda}_f \to \overline{\Lambda}_f$*, these systems are conjugate via a homeomorphism that sends the (un)stable foliation of* ψ_1 *to the (un)stable foliation of* ψ_2*.*

3.5.1 The reverse direction

Throughout this subsection we assume that $f : I \to I$ is a PCP λ-zig-zag whose Galois lift f_G has rectangular limit set Λ_f and whose digit polynomial D_f satisfies $D_f(\lambda^{-1}) = 0$. We will use I_i to denote a subinterval in the weak Markov partition of f, and we write $R_j = \pi^{-1}(I_j) \subseteq \Lambda_f$ to denote the corresponding Euclidean rectangle. Note that for each j, f_G is defined by a single affine map $\widetilde{f_i}$ on $\mathrm{int}(R_j)$, and is continuous from the right at each point of $\mathrm{WPC}(f)$.

Lemma 3.5.1. *Let* μ *be the finite measure on* Λ_f *inherited from Lebesgue measure on* $\mathbb{R}^2$*. Then* $f_G : (\Lambda_f, \mu) \to (\Lambda_f, \mu)$ *is measure-preserving.*

Proof. For each i the affine map $\tilde{f}_i$ defining f_G on the rectangle R_i is measure-preserving, having Jacobian

$$D\tilde{f}_i = \pm \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Indeed, these maps are invertible with measure-preserving inverse, so $\mu(\tilde{f}_i(R_i)) = \mu(R_i)$. Therefore we have

$$\mu(\Lambda_f) = \sum_{i=0}^{k} \mu(R_i) = \sum_{i=0}^{k} \mu(\tilde{f}_i(R_i)) \geq \mu(f_G(\Lambda_f)).$$

If the final inequality is strict then $A = \Lambda_f \setminus f_G(\Lambda_f)$ has positive measure. But since Λ_f and $f_G(\Lambda_f)$ are each finite unions of rectangles, so is A, hence if $\mu(A) > 0$ then A has non-empty interior U. No point $x \in U \subseteq \Lambda_f$ can be a limit point of an orbit of f_G, contradicting the definition of Λ_f. Hence in fact $\mu(f_G(\Lambda_f)) = \mu(\Lambda_f)$, and in particular

$$\sum_{i=0}^{k} \mu(\tilde{f}_i(R_i)) = \mu(f_G(\Lambda_f)).$$

In other words, $\mu(\tilde{f}_i(R_i) \cap \tilde{f}_j(R_j)) = 0$ for all indices $i \neq j$, and so for any measurable $B \subseteq \Lambda_f$ we have

$$\mu(f_G^{-1}(B)) = \sum_{i=0}^{k} \mu(\tilde{f}_i^{-1}(B)) = \sum_{i=0}^{k} \mu(B \cap R_i) = \mu(B).$$

Thus f_G preserves μ. □

Let M_g be the weak Markov matrix for an ergodic PCF uniform λ-expander g. The map g has a unique invariant measure that is absolutely continuous, i.e. that is in the equivalence class of Lebesgue measure. Denote this measure by ν_0. If M_g is primitive, then results of Parry in [Par64] on subshifts of finite type imply that the Perron eigenvectors of M_g determine ν_0. Namely, let $u = (u_1, \ldots, u_n)$ denote the left λ-eigenvector of M with $\sum_i u_i = 1$, and let $v = (v_1, \ldots, v_n)$ denote the right λ-eigenvector of M_g such that $u \cdot v = 1$. Then the density of ν_0 is (up to a null set) a step function whose value is v_i on the subinterval

$$\left[\sum_{j=1}^{i-1} u_i,\ \sum_{j=1}^{i} u_i\right],$$

which has Lebesgue measure u_i. Note that in general if we use the right λ-eigenvector $v' = cv$ for $c > 0$ then we obtain the invariant measure $\nu_0' = c{\cdot}\nu_0$, which is a probability measure exactly when $c = 1$.

Lemma 3.5.2. *Let $\pi : \Lambda_f \to I$ denote projection onto the first coordinate, and define $\nu = \pi_*\mu$. Then $\nu = \nu_0(I) \cdot \nu_0$.*

Recall from Definition 3.2.3 that if $w_j \in \mathrm{WPC}(f)$, then f_G is defined by one affine map on $[w_{j-1}, w_j)$ and another on $[w_j w_{j+1})$. Moreover, these two affine maps differ if and only if w_j is a critical point of f.

Proof. Since μ is f_G-invariant and $\pi \circ f_G = f \circ \pi$, ν is invariant with respect to f. Furthermore, because Λ_f is a connected finite union of rectangles of positive measure, we see that ν is equivalent to Lebesgue measure, and in fact has invariant density a step function whose values are given by the heights of the rectangles of Λ_f. The result now follows by uniqueness of ν_0. Note in general that $\nu(I) = \mu(\Lambda_f) \neq 1$. □

Lemma 3.5.3. *Let $\alpha = \alpha_0$ be a periodic point of f of least period p. Set $\alpha_i = f^i(\alpha_0)$ for $0 \leq i \leq p-1$, and let $g_i : \mathbb{R} \to \mathbb{R}$ be a choice of affine map defining f at $x = \alpha_i$ so that $f(\alpha_i) = g_i(\alpha_i) = c_i \pm \lambda\alpha_i = \alpha_{i+1}$, where the indices are understood modulo p. Define $\widetilde{g}_i : \mathbb{R} \to \mathbb{R}$ by $\widetilde{g}_i(y) = c_i \pm \lambda^{-1} y$.*

Then there is a unique periodic point of f_G projecting to α_0, and it is given by (α_0, y_0) where y_0 is the unique solution to the equation $\widetilde{g}_{p-1} \circ \cdots \circ \widetilde{g}_0(y) = y$. In the case $\alpha_0 = 1$ this periodic point is $(1, 1)$.

Proof. Observe first that $\tilde{g}_{p-1} \circ \cdots \circ \tilde{g}_0 : \mathbb{R} \to \mathbb{R}$ is a contraction of factor $\lambda^{-p} < 1$. This map has a unique fixed point y_0. By definition,

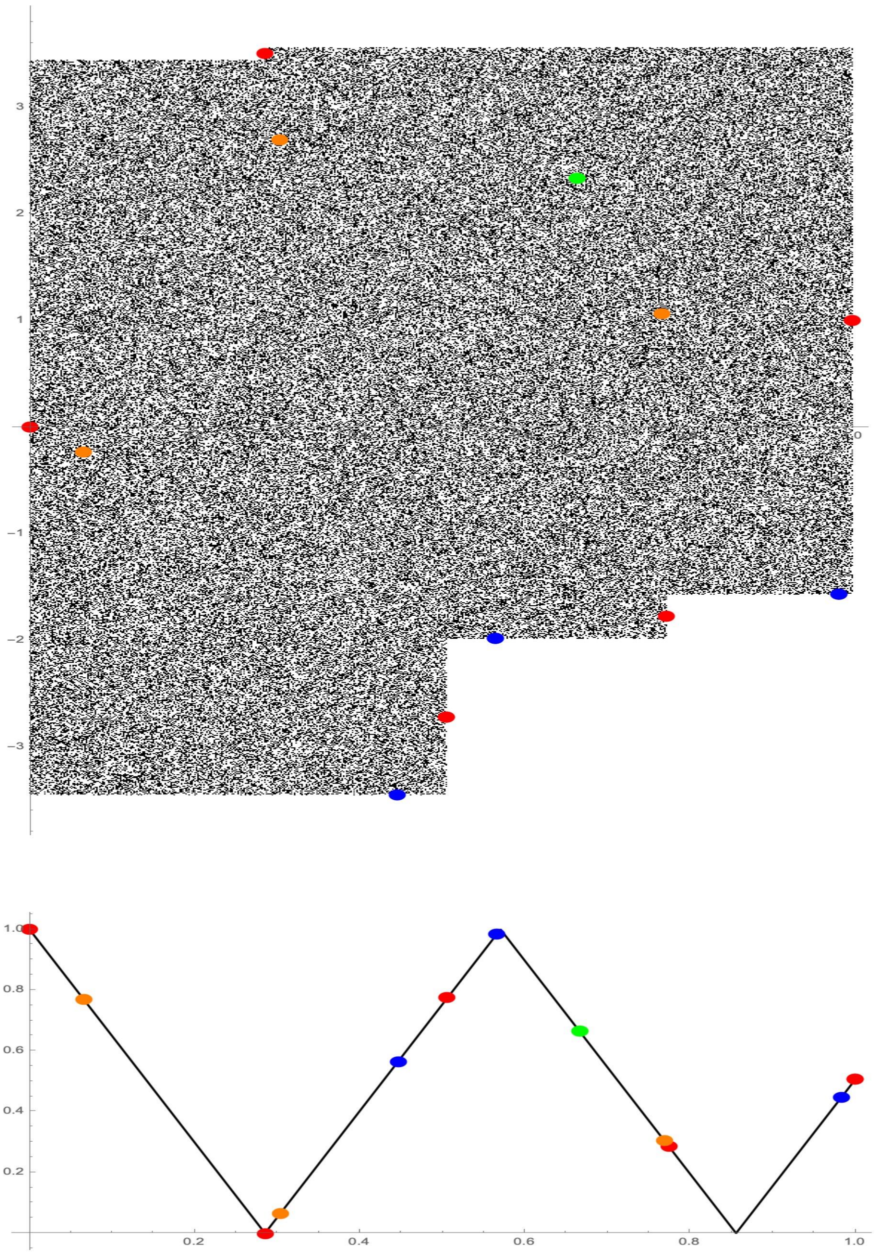


Figure 3.11: (Below) The graph of the negative λ-zig-zag f, for λ the Perron root of $D_f(t) = t^4 - 3t^3 - t^2 - 3t + 1$. (Above) The limit set Λ_f of the Galois lift of f. In each picture, heavily marked points of the same color belong to the same periodic orbit, and the points of one color in Λ_f are the unique periodic lifts of the points of I of the same color. This limit set was drawn by plotting the orbit under f_G of a single point with transcendental coordinates, to ensure that it is not eventually periodic. The same method was used to draw all limit sets in the paper.

$$f_G(\alpha_0, y_0) = (g_0(\alpha_0), \widetilde{g}_0(y_0)).$$

Indeed, inductively defining $y_i = \widetilde{g}_{i-1}(y_{i-1})$ we have $f_G^i(\alpha_0, y_0) = (\alpha_i, y_i)$ for $0 \le i \le p-1$. Since α_0 and y_0 are periodic of period p, it now follows that $f_G^p(\alpha_0, y_0) = (\alpha_0, y_0)$.

Suppose that (α_0, z) is a periodic point of f_G of least period p. Then

$$f_G^i(\alpha_0, z) = (\alpha_i, \widetilde{g}_i \circ \cdots \circ \widetilde{g}_0(z)) \quad \text{for } 0 \le i \le p-1.$$

In particular, z is a fixed point of $\widetilde{g}_{p-1} \circ \cdots \circ \widetilde{g}_0$, hence is equal to y_0.

The fact that $(1,1)$ is the periodic point of f_G projecting to $1 \in I$ follows immediately from the assumption that $D_f(\lambda^{-1}) = 0$. □

Definition 3.5.4. Let $R = [a,b] \times [c,d]$ be a Euclidean rectangle. The *horizontal and vertical boundaries of* R are, respectively,

$$\partial_H R = [a,b] \times \{c,d\} \quad \text{and} \quad \partial_V R = \{a,b\} \times [c,d].$$

Two rectangles $R_1 = [a_1,b_1] \times [c_1,d_1]$, $R_2 = [a_2,b_2] \times [c_2,d_2]$ are *lower- (upper-)aligned* if $c_1 = c_2$ ($d_1 = d_2$).

Definition 3.5.5. Let $f : I \to I$ be a PCP zig-zag whose Galois lift f_G has rectangular limit set Λ_f. Let $R_0, \ldots, R_k$ be the rectangles defined by the strong Markov partition of f which subdivide Λ_f. The *vertical boundary of* Λ_f is the set

$$\partial_V \Lambda_f = \left(\partial \Lambda_f \cap \bigsqcup_{i=0}^{k} \partial_V R_i \right) \setminus X,$$

where X is the set of isolated points of $\partial \Lambda_f \cap \bigsqcup_i \partial_V R_i$. The *horizontal boundary of* Λ_f is the set

$$\partial_H \Lambda_f = \bigsqcup_{i=0}^{k} \partial_H R_i$$

A *vertical (horizontal) component* of $\partial \Lambda_f$ is a connected component of $\partial_V \Lambda_f$ ($\partial_H \Lambda_f$).

Lemma 3.5.6. *Vertical components of $\partial \Lambda_f$ project to postcritical points of f.*

Proof. If $x = a \in \mathrm{WPC}(f) \setminus \mathrm{PC}(f)$ then a is a critical point of f that is not in the forward image of any $b \in \mathrm{WPC}(f)$. Therefore if R is a rectangle projecting to an element of the weak Markov partition of f and $f_G(R)$ intersects the line $x = a$, the in fact $f_G(R)$ must cross the line. In other words, if R_i and R_{i+1} are the rectangles bordering the line $x = a$, then for any rectangle R the image $f_G(R)$ crosses R_i if and only if it also crosses R_{i+1}. Consequently R_i and R_{i+1} are both upper- and lower-aligned, and hence there is no vertical component of $\partial\Lambda_f$ projecting to $a \in I$. $\square$

Lemma 3.5.7. *If $a \in \mathrm{PC}(f)$ then the unique periodic point of f_G projecting to a is contained in $\partial_V\Lambda_f$.*

Proof. If $a = 1$ then the conclusion holds. Since the point $x = 1$ is periodic under f, its forward orbit contains a critical point $a \in \mathrm{PC}(f)$ such that $f(a) \in \{0, 1\}$. We consider the case when $f(a) = 1$, since this is more complicated.

We know from Lemma 3.5.3 that $(1, 1)$ is the periodic point of f_G projecting to $x = 1$. Moreover, if f_i and f_{i+1} are the linear branches of f on left and right of $a \in I$, respectively, then in fact $f_i(x) = c_i + \lambda x$ and $f_{i+1}(x) = 2 - c_i - \lambda x$ for some integer c_i. Therefore the affine maps defining f_G on either side of the line $x = a$ are

$$\tilde{f}_i(x, y) = (c_i + \lambda x, c_i + \lambda^{-1} y), \quad \tilde{f}_{i+1}(x, y) = (2 - c_i - \lambda x, 2 - c_i - \lambda^{-1} y).$$

If $(a, y) \in \Lambda_f$ then the images $\tilde{f}_i(a, y) = (1, c_i + \lambda^{-1} y)$ and $\tilde{f}_{i+1}(1, 2 - c_i - \lambda^{-1} y)$ are on the line $x = 1$ and are symmetric about the point $(1, 1)$. In particular, the maps $\tilde{f}_i$ and $\tilde{f}_{i+1}$ agree precisely at the periodic point $\tilde{a}$ of f_G projecting to a. It follows that if $\tilde{a} \in \mathrm{int}(R_i \cup R_{i+1})$ and $U \subseteq \mathrm{int}(R_i \cup R_{i+1})$ is an open rectangle symmetric about $\tilde{a}$ then

$$f_G(U \cap \mathrm{int}(R_i)) = f_G(U \cap \mathrm{int}(R_{i+1})).$$

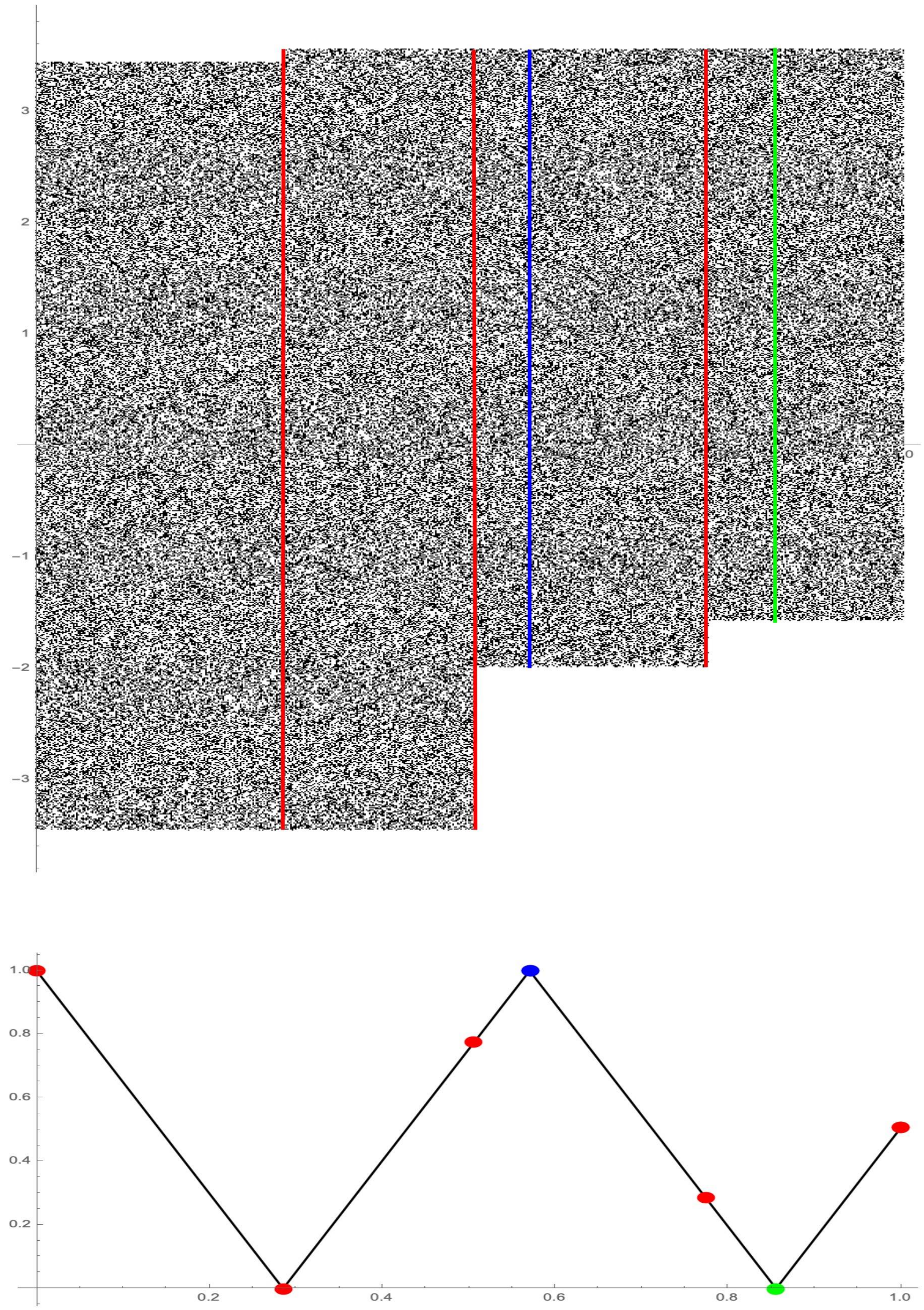


Figure 3.12: The weak postcritical set of f pulls back to a rectangular decomposition of Λ_f. Note that the vertical components of $\partial\Lambda_f$ project to the points of $\mathrm{PC}(f)$, in red. Observe that each rectangle is either upper- or lower-aligned with its neighbors (cf. Lemma 3.5.8). Moreover, each vertical edge in this case contains at its metric center the periodic lift of the postcritical point to which it projects (cf. Figure 3.11 and Lemma 3.5.9).

This contradicts the fact that f_G is measure-preserving, so in fact $\tilde{a} \notin \operatorname{int}(R_i \cup R_{i+1})$. In other words, $\tilde{a} \in \partial_V \Lambda_f$.

We now proceed inductively, going through the periodic orbit of $(1, 1)$ in reverse order. Suppose that $\tilde{a} \in I \times \mathbb{R}$ is a point in the orbit of $(1, 1)$ that is contained in $\partial_V \Lambda_f$, and let $\tilde{b}$ be the periodic point such that $f_G(\tilde{b}) = \tilde{a}$. If $\pi(\tilde{b}) = b$ is a critical point of f, then we repeat the above argument. Otherwise f_G is defined by a single affine map in a neighborhood of $\tilde{b}$. In particular, if R_j and R_{j+1} are the rectangles bordering the line $x = b$, if $\tilde{b} \in \operatorname{int}(R_j \cup R_{j+1})$ then f_G maps an open rectangular neighborhood of $\tilde{b}$ to an open rectangular neighborhood of $f_G(\tilde{b}) = \tilde{a}$ in $\mathbb{R}^2$. This neighborhood cannot be a subset of Λ_f, since $\tilde{a} \in \partial_V \Lambda_f$ by assumption, but this contradictions the invariance of Λ_f.

Therefore $\tilde{b} \notin \operatorname{int}(R_j \cup R_{j+1})$, hence $\tilde{b} \in \partial_V \Lambda_f$. $\square$

Lemma 3.5.8. *There is exactly one vertical component of $\partial \Lambda_f$ projecting to a point $a \in \mathrm{PC}(f)$. Therefore, all adjacent rectangles R_i, R_{i+1} of Λ_f are either upper- or lower-aligned.*

Proof. Suppose first that $a \notin \{0, 1\}$. There is a unique point $b \in \mathrm{PC}(f)$ satisfying $f(b) = a$, and b is not a critical point. Hence, if R_i and R_{i+1} are the two rectangles of Λ_f intersecting $\pi^{-1}(b)$, then f_G acts on R_i and R_{i+1} by the same affine map $\widetilde{f}$. In particular, f_G acts by a homeomorphism on a neighborhood of $\pi^{-1}(b)$. We claim that each connected component of $\pi^{-1}(b) \cap \partial_V \Lambda_f$ maps homeomorphically onto a connected component of $\pi^{-1}(a) \cap \partial_V \Lambda_f$. Indeed, if B is a component of $\pi^{-1}(b) \cap \partial_V \Lambda_f$ that maps into the interior of $\pi^{-1}(a) \cap \partial_V \Lambda_f$, then there must be some other rectangle R whose vertical boundary intersects the line $x = c$, for some $c \in \mathrm{WPC}(f)$ satisfying $f(c) = a$. This point c cannot be in $\mathrm{PC}(f)$, by the uniqueness of b, so c is a critical point of f. But we assumed $a \notin \{0, 1\}$ so this is impossible.

Thus we conclude that if $a \notin \{0, 1\}$ then f_G maps each connected component of $\partial_V \Lambda_f$ over b to a connected component of $\partial_V \Lambda_f$ over A. Since f_G acts by a homeomorphism in a neighborhood of $\pi^{-1}(b)$, we see that the number of connected components over b is equal to the number of connected components over a.

If instead $a \in \{0,1\}$, then certainly $\partial_V \Lambda_f \cap \pi^{-1}(a)$ consists of a single connected component. To finish the proof, therefore, we can travel forward through each postcritical orbit. There is a single component over a, thus by our argument this component maps homeomorphically to the unique component over $f(a)$, then to the unique component over $f^2(a)$, etc.

That all adjacent rectangles are either upper- or lower-aligned now follows immediately: if two such rectangles are neither upper- nor lower-aligned, then there would exist at least two components of $\partial_V \Lambda_f$ along their intersection. □

Lemma 3.5.9. *Let $a \in \mathrm{PC}(f)$ and $\tilde{a} \in \Lambda_f$ the unique periodic point projecting to it. Then $\tilde{a}$ is at the metric center of the vertical component of $\partial \Lambda_f$ containing it.*

Proof. We begin by proving the statement for $a = 1$. By Lemma 3.5.3 we know $\tilde{a} = (1,1)$ in this case. If $f(c) = 1$ then c is a critical point that either is or is not periodic.

If c is not periodic then $c \notin \mathrm{PC}(f)$ and the two rectangles of Λ_f intersecting the line $x = c$ are both upper- and lower-aligned, by Lemma 3.5.6. Denote these two rectangles by R_i, R_{i+1} and let $\tilde{f}_i$, $\tilde{f}_{i+1}$ denote the affine maps by which f_G acts on these rectangles, respectively. Then $\tilde{f}_i(R_i \cap R_{i+1})$ and $\tilde{f}_{i+1}(R_i \cap R_{i+1})$ are vertical line segments of equal length lying on the line $x = 1$. Furthermore, these line segments are contained in Λ_f and symmetric about $\tilde{a}$.

If c' is another non-postcritical point satisfying $f(c') = 1$ and R_j, R_{j+1} the rectangles intersecting the line $x = c'$, then the images $\tilde{f}_j(R_j \cap R_{j+1})$ and $\tilde{f}_{j+1}(R_j \cap R_{j+1})$ do not intersect $\tilde{f}_i(R_i \cap R_{i+1})$ and $\tilde{f}_{i+1}(R_i \cap R_{i+1})$ except perhaps at their endpoints: otherwise two rectangles, say R_i and R_j, would have

$$f_G(\operatorname{int} R_i) \cap f_G(\operatorname{int} R_j) \neq \emptyset,$$

contradicting the fact that f_G is measure-preserving.

If c is the unique periodic critical point of f, then there is a single vertical component of $\partial\Lambda_f$ projecting to c, and this component contains the periodic lift $\tilde{c}$ of c. Denote by R_l, R_{l+1} the rectangles of Λ_f intersecting at $x = c$. As before, $\tilde{f}_l(R_l \cap R_{l+1})$ and $\tilde{f}_{l+1}(R_l \cap R_{l+1})$ are vertical line segments symmetric about $\tilde{a}$. Furthermore, f_G maps the single vertical component V to a line segment containing $\tilde{a}$. The union $V \cup (R_l \cap R_{l+1})$ is a connected line segment, hence $\tilde{f}_l(V \cup (R_l \cap R_{l+1}))$ and $\tilde{f}_{l+1}(V \cup (R_l \cap R_{l+1}))$ are connected line segments such that

$$\tilde{f}_l\left(V \cup (R_l \cap R_{l+1})\right) \cap \tilde{f}_{l+1}\left(V \cup (R_l \cap R_{l+1})\right) = f_G(V).$$

It now follows that $f_G(V)$ is symmetric about $\tilde{a}$, and consequently so is the entire component of $\partial_V\Lambda_f$ containing $\tilde{a}$.

We now prove the claim for all points in the forward orbit of $1 \in \mathrm{PC}(f)$, proceeding inductively. As we argued in the proof of Lemma 3.5.8, f_G maps the component of $\partial_V\Lambda_f$ projecting to $a \in \mathrm{PC}(f)$ homeomorphically onto the component projecting to $f(a)$. In particular, if the unique periodic lift $\tilde{a}$ of a lies at the metric center of the vertical component containing it, then so does $f_G(\tilde{a})$, since f_G contracts the vertical direction uniformly by λ^{-1}.

If f is a negative zig-zag, then it has a single postcritical orbit and so the proof is complete. If, however, f is a positive zig-zag then we must still prove the claim for the fixed point $0 \in \mathrm{PC}(f)$. The exact argument used for $a = 1$ applies here, after observing that the unique periodic lift of $a = 0$ is the point $\tilde{a} = (0, 0)$. □

Lemma 3.5.10. *The action of f_G on Λ_f can be recovered via the de Carvalho-Hall construction in the sense of Theorem A. In particular, f is of pseudo-Anosov type.*

Proof. We reverse engineer the invariant train track τ for the exterior left-veering thick interval map F projecting to f. Construct τ to be the train track with a

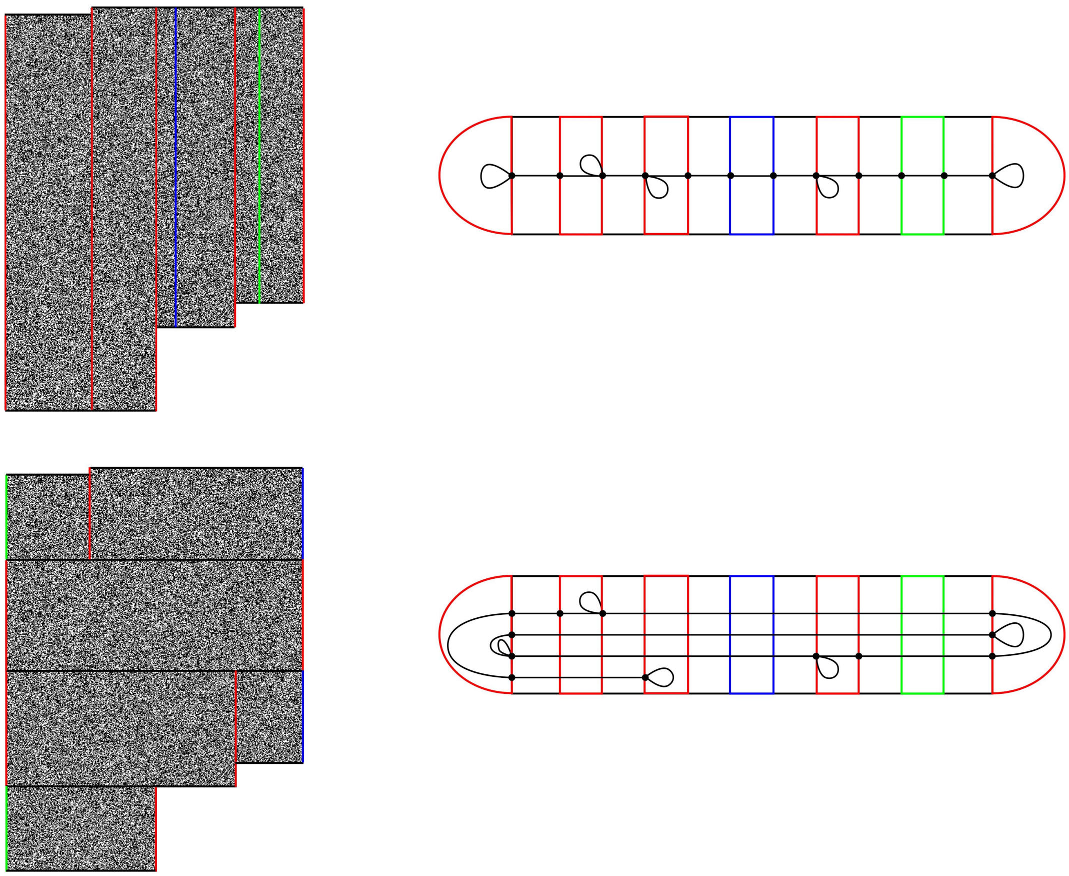


Figure 3.13: Reverse-engineering the invariant train track τ for the exterior left-veering thickening of a zig-zag f with rectangular limit set Λ_f. On the top left is the rectangular decomposition of Λ_f projecting to $\mathrm{WPC}(f)$, and on the top right is a train track with junctions corresponding to the vertical boundaries of the rectangles. From left to right, each junction corresponds to a vertical boundary of the same color. Moreover, a junction contains a loop if and only if it corresponds to a line projecting to an element of $\mathrm{PC}(f)$. In this case, the loop is above (resp. below) the spine of τ if the rectangles adjoining the line are lower- (resp. upper-)aligned, and is on the left (resp. right) of the junction if the rectangle intersecting $\partial_V \Lambda_f$ is on the left (resp. right) of the line.

real edge e_j for each rectangle R_j, a connecting infinitesimal edge between the real edges e_j, e_{j+1} for adjacent rectangles R_j, R_{j+1}, and an infinitesimal loop for each e_j corresponding to a p_i. The loop is on the left (resp. right) of e_j if p_i is on the left (resp. right) vertical boundary of R_j, and if the loop is on the same side of e_j as a connecting infinitesimal edge then the loop is above (resp. below) the connecting edge if the vertical boundary component of L that contains p_i is above (resp. below) the adjacent rectangle.

The map induced by f_G on τ leaves τ invariant under the pseudo-isotopies defined in section 2, hence τ is the invariant train track for a thick interval map projecting to f. By Proposition 3.4.19 this thick interval map must be either F_L or F_R. One easily checks that the Galois conjugate coordinates imply that it must be F_L.

To complete the proof, observe that by Lemma 3.5.2 we may choose a right λ-eigenvector of the transition matrix M of f such that the rectangles obtained from τ have the exact dimensions of those in the rectangle decomposition of L. Therefore we may define an isometry $i : S' \to \Lambda_f$ taking the horizontal (resp. vertical) foliation of S' to that of Λ_f. By construction, the map induced on Λ_f is f_G. □

3.5.2 The forward direction

Lemma 3.5.10 completes the proof of the reverse direction of Theorem A. Now it remains to prove the forward direction.

We consider the case of $f : I \to I$ a positive λ-zig-zag of pseudo-Anosov type, the case for f negative being essentially identical. Let F be the exterior left-veering thickening of f. Let S' be as in the statement of the theorem. F induces a map $\tilde{f} : S' \to S'$. Let $p \in S'$ be the unique fixed point of $\tilde{f}$ projecting to $x = 0 \in I$, and let $q \in S'$ be the unique periodic point projecting to $x = 1 \in I$.

We wish to define a map $i : S' \to I \times \mathbb{R}$ such that $i \circ \tilde{f} = f_G \circ i$ and such that i sends the horizontal and vertical foliations of S' to those of $I \times \mathbb{R}$ in an orientation-preserving way. Fix $s, t \in \mathbb{R}$ distinct and define $i(p) = (0, s)$ and $i(q) = (1, t)$. Observe that these two choices determine i, and moreover the number $d = t - s$ controls the area of $i(S')$: if A_d is the Lebesgue measure of $i(S')$ for any choice of i satisfying $t - s = d$, then $A_d = |d| \cdot A_1$.

By construction, there is a piecewise affine $G : I \times \mathbb{R} \to I \times \mathbb{R}$ satisfying $i \circ \tilde{f} = G \circ i$. In particular, if $f_0, \ldots, f_k$ denote the linear maps defining the original zig-zag f then G is of the form

$$G(x, y) = (f_i(x), g_i(y)) \text{ if } f(x) = f_i(x)$$

where $g_i(y) = a_i(s, t) + (-1)^i \cdot \lambda^{-1} y$. This follows from the fact that $\tilde{f}$ is a piecewise affine orientation-preserving map projecting to f. The remainder of the proof is an analysis of the constants $a_i(s, t)$.

Note that $i(p) = (0, s)$ is a fixed point for G, and hence $y = s$ is fixed by $g_0(y) = a_0(s, t) + \lambda^{-1} y$. We thus have $a_0(s, t) = (1 - \lambda^{-1})s$. To compute the remaining a_i we require the following lemma.

Lemma 3.5.11. *Let $k_i = i \cdot \lambda^{-1}$ be a critical point of f and set g_{i-1}, g_i to be the defining y-coordinate maps of G on either side of the line $x = k_i$. Then $g_{i-1}(y_i) = g_i(y_i)$ for a unique number y_i, and moreover this number satisfies $G(k_i, y_i) = s$ or $G(k_i, y_i) = t$ depending on whether i is even or odd, respectively.*

Proof. Since g_{i-1} and g_i are affine with slopes of opposite sign they agree at a unique point y_i. To understand this number y_i, consider the corresponding situation for $\tilde{f} : S' \to S'$.

If k_i is not a postcritical point of f, then the line $x = k_i$ partitions a rectangle of S'. Depending on the parity of i, the affine maps defining $\tilde{f}'$ on either side of the vertical line send the line to either $x = 0$ or $x = 1$, with opposite orientations. The fact that the unique periodic lifts of $x = 0$ and $x = 1$ are points of cone angle π lying in the center of the corresponding vertical sides of S' implies that these two affine maps would send the same point of $k_i \times \mathbb{R}$ to the singularity if their domains were extended to include this point.

If k_i is a postcritical point, then the unique periodic lift of k_i lies on the vertical boundary of exactly one of the corresponding rectangles of S', and moreover it lies in the center of this boundary component. Since the image of this singularity is the sole singularity of $\tilde{f}$ on the corresponding vertical leaf of S', the two affine maps defining $\tilde{f}$ on either side of $x = k_i$ must map each point of this line to points on $x = 0$ or $x = 1$ that are equal distances from the singularity. Because all vertical distances are scaled by the same factor, it now follows that both maps agree on the singularity projecting to k_i.

In either case, the same is true of the maps g_{i-1} and g_i after mapping into $I \times \mathbb{R}$. □

We return to computing the $a_i(s,t)$, thereby completing the proof of Theorem A.

Proof of Theorem A. By the lemma, $g_1(y) = 2t - g_0(y)$, and in particular

$$a_1(s,t) = 2t - a_0(s,t) = 2t - (1 - \lambda^{-1})s.$$

Similarly, we have $g_2(y) = 2s - g_1(y)$, hence $a_2(s,t) = 2s - a_1(s,t)$, and in general we have

$$a_i(s,t) = \begin{cases} (i + 1 - \lambda^{-1})s - it & \text{if } i \text{ is even} \\ (\lambda^{-1} - i)s + (i+1)t & \text{if } i \text{ is odd.} \end{cases}$$

Setting $s = 0$ and $t = 1$ (and hence $d = 1$), we specialize to the case

$$a_i(s,t) = \begin{cases} -i & \text{if } i \text{ is even} \\ i + 1 & \text{if } i \text{ is odd.} \end{cases}$$

In other words, each affine piece of $G : I \times \mathbb{R} \to I \times \mathbb{R}$ is of the form

$$G_i(x, y) = \begin{cases} (\lambda x - i, \lambda^{-1} y - i) & \text{if } i \text{ is even} \\ (i + 1 - \lambda x, i + 1 - \lambda^{-1} y) & \text{if } i \text{ is odd.} \end{cases}$$

Thus G is precisely the Galois lift f_G of Thurston. Condition (2) of the theorem is immediately verified, so it remains to argue that the digit polynomial D_f has λ^{-1} as a root. We can define a "vertical" digit polynomial using the orbit of $y = 1$ under the g_i. Observe that this new polynomial is precisely D_f, and is necessarily satisfied by λ^{-1}. This completes the proof of Theorem A. $\square$

Remark 3.5.12. One quickly verifies that

$$-\frac{k}{1 - \lambda^{-1}} a_0(s, t) + \sum_{i=1}^{k} a_i(s, t) = 0.$$

and hence the map $(s, t) \mapsto (a_0(s, t), \ldots, a_k(s, t))$ is a linear map of $\mathbb{R}^2 \setminus \Delta$ into $\mathbb{R}^{k+1}$ whose image is a plane minus the line corresponding to $d = 0$. This missing line is the image of the diagonal $\Delta \subseteq \mathbb{R}^2$, and we can foliate the image plane with lines parallel to this one, each such line corresponding to a different value of $d = t - s$. The area of the limit set $i_d(S')$ of G_d scales linearly with $|d|$, i.e. if A_d is the area of $i_d(S')$ then $A_d = |d| \cdot A_1$. In this way we can interpret the prohibited case $d = 0$ as a degenerate limit set of area zero. Moreover, the two half planes defined by $d > 0$ and $d < 0$ correspond to the underlying train track map τ being left- or right-veering, respectively.

3.6 The postcritical orbit of a zig-zag of pseudo-Anosov type

In light of Theorem A, it is natural to ask when an exterior left-veering thickening F_L has a finite generalized invariant train track. This section obtains necessary conditions on the structure of the thick interval map F_L associated to a PCP zig-zag of pseudo-Anosov type. In particular, Proposition 3.6.10 will be instrumental to our proof of Theorem B in Section 6.

Recall that if $F_L : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ is a thickening of a PCF interval map f, then we denote by $\mathbf{0}$ the junction projecting to $0 \in I$, and similarly we denote by $\mathbf{1}$ the vertex projecting to $1 \in I$.

Definition 3.6.1. Let $f : I \to I$ be a λ-zig-zag map and $k_i = i \cdot \lambda^{-1}$ the critical points of f, for $i = 1, \ldots, \lfloor \lambda \rfloor$. Let $F_L : (S^2, \mathbb{I}) \to (S^2, \mathbb{I})$ be the exterior left-veering thickening of f. For each i we denote by K_i the junction projecting to k_i.

Definition 3.6.2. Let τ be the generalized invariant train track for the left-veering thickening F_L of the PCP zig-zag map f. By a *connecting infinitesimal edge* we will mean an infinitesimal edge of τ connecting two distinct real edges. The remaining infinitesimal edges of τ, namely those that connect a real edge to itself, are called *loops*.

We define the *spine* τ' of τ to be the union of all real edges and connecting infinitesimal edges of τ. We orient τ' from $\mathbf{0}$ to $\mathbf{1}$. A loop $\gamma \subseteq \tau$ contained in an intermediate vertex of $\mathbb{I}$ is called *interior* if γ is to the left of τ' and *exterior* otherwise. See Figure 3.14.

Remark 3.6.3. The interior loops γ of τ are precisely the loops whose image $F(\gamma)$ points toward the horizontal midline of $\mathbb{I}$. One can see this by noting that F is exterior left-veering and preserves orientation. It is for this reason that we refer to such loops

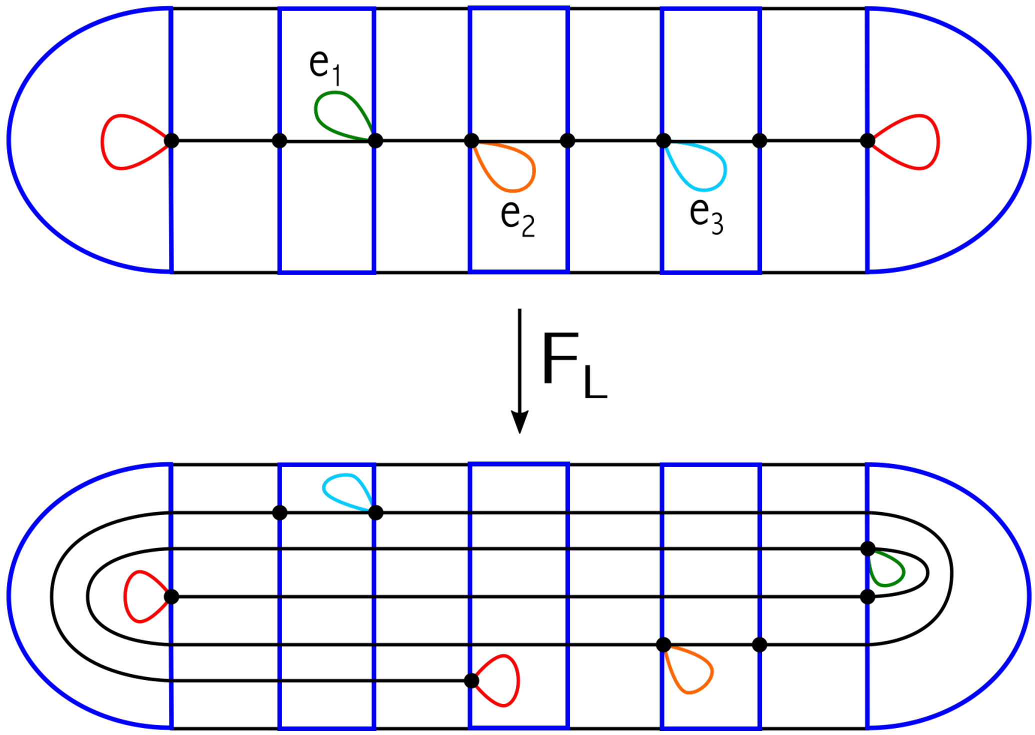


Figure 3.14: A (finite) generalized train track τ and an exterior left-veering thick interval map preserving it. Here the spine of τ is the union of the black edges. Of the intermediate loops, e_1 is interior while e_2 and e_3 are exterior. Observe that under the action of F_L, the image of an interior loop points in toward the horizontal midline of $\mathbb{I}$, hence will be enclosed by arcs after pseudo-isotopy unless it maps into $\mathbf{0}$ or $\mathbf{1}$.

as "interior." In the arguments of this section we will see that interior loops are rare, and the existence of more than one for τ is an obstruction to the finiteness of τ.

Recall (cf. Definition 3.2.2) that a zig-zag f is *positive* if $f(0) = 0$, and *negative* if $f(0) = 1$.

Proposition 3.6.4. *Let $f : I \to I$ be a PCP zig-zag map of growth rate $\lambda > 2$. If f is of pseudo-Anosov type, then $k_1 = \lambda^{-1}$ is in the forward orbit of $x = 1$. Moreover, the loop in the junction C_1 must be interior.*

Proof. Suppose first that f is positive, so that $f(0) = 0$. Since f is PCP, some $k_i = i \cdot \lambda^{-1}$ is in the forward orbit of $x = 1$. By Theorem 3.4.7, the corresponding junction C_i contains an edge ϵ and a loop γ of the invariant train track τ. If γ is exterior then after pseudo-isotopy the images $F(\epsilon)$ and $F(\gamma)$ will be non-parallel loops in **1**, contradicting Theorem 3.4.7. See Figure 3.15 below. Thus γ must be interior.

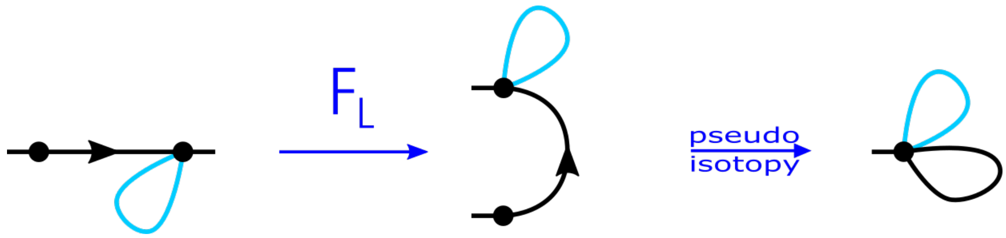


Figure 3.15: An exterior loop in C_i produces two non-parallel loops in **1**

If $i \neq 1$ then F will map the connecting edge ϵ' of τ that is within C_1 into **1** further interior than the images of ϵ and γ, and thus after pseudo-isotopy we will obtain two non-parallel loops enclosed by a third, again contradicting Lemma 3.4.7. See Figure 3.16. On the other hand, if $i = 1$ then the images of ϵ and γ will be the furthest interior in **1** and all other edges that map into **1** will pinch to parallel loops, which then combine under pseudo-isotopy.

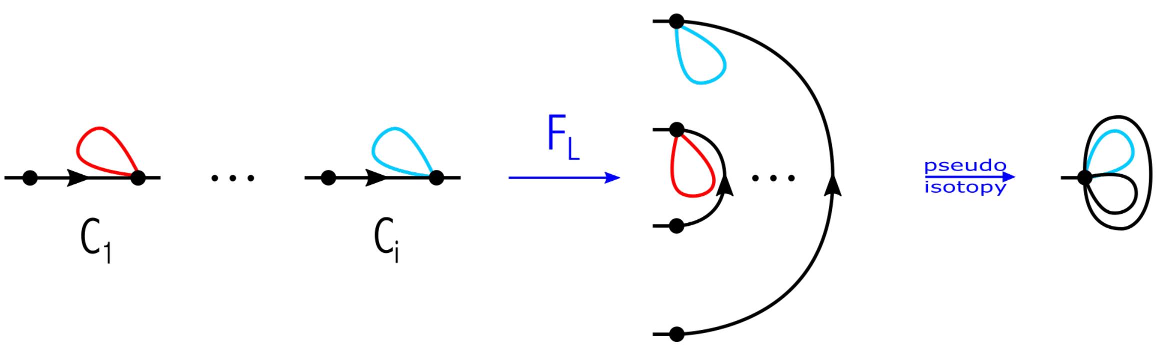


Figure 3.16: The image of all edges in C_1 is further interior than that of any other C_i mapping into **1**, so any loop in C_i generates a loop in **1** that persists after pseudo-isotopy.

Now suppose that f is negative, so that $f(0) = 1$. The same argument applies, except that the role of **1** is taken by **0**, which maps into **1**. □

Proposition 3.6.5. *Let $f : I \to I$ be a PCP zig-zag map of growth rate $\lambda > 2$. If f is of pseudo-Anosov type, then the invariant train track $\tau = \tau_L$ has exactly one interior loop, namely the loop contained in the junction C_1.*

Proof. Proposition 3.6.4 tells us that C_1 contains an interior loop, so it remains to show that τ has no other interior loops. Furthermore, our argument in the proof of 3.6.4 demonstrates that no other critical junction C_i can contain an interior loop, so if τ contains another interior loop γ then $F(\gamma)$ cannot lie in $\mathbf{0}$ or in $\mathbf{1}$. In other words, $F(\gamma)$ lies in some intermediate vertex V. However, since γ is interior the image $F(\gamma)$ faces toward the center of $\mathbb{I}$. See Figure 3.17. Thus after pseudo-isotopy $F(\gamma)$ will be enclosed by multiple connecting edges, contradicting the structure of τ described in Theorem 3.4.7. $\square$

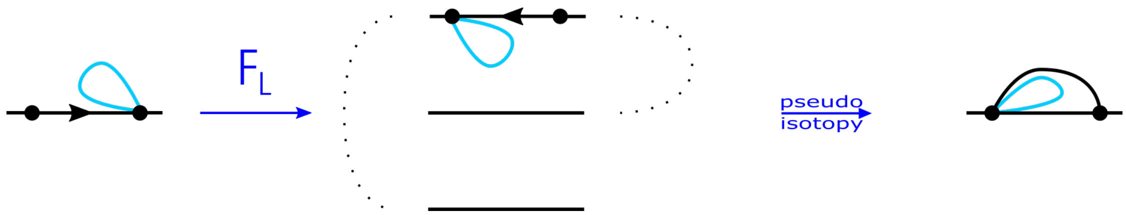


Figure 3.17: The image of an interior loop points inward, and because all horizontal layers of $F(\tau)$ span the full length of $\mathbb{I}$ except for the last, the image of an interior loop will be trapped by other edges after pseudo-isotopy unless it maps into $\mathbf{0}$ or $\mathbf{1}$, where it can be absorbed into the parallel loop formed by the turn.

Corollary 3.6.6. *Let $f : I \to I$ be a PCP zig-zag map of growth rate $\lambda > 2$, and suppose f is of pseudo-Anosov type.*

1. *If f is positive and $\lfloor \lambda \rfloor = m$ is odd, then the orbit of $x = 1$ under f is*

$$1 \mapsto \lambda^{-1} \mapsto 1.$$

2. *If f is negative and $\lfloor \lambda \rfloor = m$ is even, then the orbit of $x = 1$ under f is*

$$1 \mapsto \lambda^{-1} \mapsto 0 \mapsto 1.$$

In either case, λ has minimal polynomial $p(x) = x^2 - (m+1)x + 1$.

Proof. First suppose f is positive and m is odd. The image $F_L(\gamma)$ of the loop $\gamma \subseteq \tau$ contained in $\mathbf{1}$ is above the other strands of $F_L(\tau)$, and so after pseudo-isotopy it will be an interior loop γ'. This interior loop must necessarily be contained in C_1 and must map into $\mathbf{1}$ by Proposition 3.6.5. It follows that $x = 1$ has period 2. The linear branch of f that applies to $x = 1$ is $x \mapsto (m+1) - \lambda x$, hence we have $\lambda^{-1} = m+1-\lambda$.

The case when f is negative and m is even is similar. The image $F_L(\gamma)$ is again an interior loop after pseudo-isotopy, and so must be contained in C_1. Since the linear branch of f that applies to $x = 1$ is $x \mapsto (m+1) - \lambda x$, it follows that $\lambda^{-1} = m + 1 - \lambda$. □

Remark 3.6.7. The examples described in Corollary 3.6.6 are the simplest examples of uniform expanders of pseudo-Anosov type. The corresponding pseudo-Anosovs all live on the four-punctured sphere $\Sigma_{0,4}$ and lift via a branched double cover to hyperbolic automorphisms of the torus. To avoid these simple cases, we make the following definition (see Remark 3.6.9 below).

Definition 3.6.8. For $m \geq 2$ and $p \geq 3$, define $\mathrm{PA}(m,p)$ to be the set of zig-zags f of pseudo-Anosov type such that

1. f has m critical points, and
2. $\#\,\mathrm{PC}(f) = p$.

We also define the set

$$\mathrm{PA}(m) = \bigcup_{p \geq 4} \mathrm{PA}(m,p).$$

Remark 3.6.9. An element $f \in \mathrm{PA}(m,p)$ defines a pseudo-Anosov on the $(p+1)$-punctured sphere, where the point at infinity is fixed. Ripping open this point to a boundary component, we may also think about this pseudo-Anosov as a braid on the p-punctured disc, up to multiplication by a full twist Δ^2 around the boundary. The simple examples described in Corollary 3.6.6 are precisely the maps $f \in \mathrm{PA}(m,3)$, and the definition of $\mathrm{PA}(m)$ allows us to avoid these examples in the future.

We now turn our attention to the kneading theory of zig-zags of pseudo-Anosov type. Refer to Section 2.4 for notation.

Proposition 3.6.10. *Suppose $f \in \mathrm{PA}(m)$ with slope λ. Then $\mathrm{PC}(f)$ always contains the following 3 types of points:*

(E) The extremal points $x = 0$ and $x = 1$

(C) The critical point $k_1 = \lambda^{-1}$

(R) Exactly one point in the interior of I_{m-1}

If $\mathrm{PC}(f)$ contains other points, then they are of the following 2 types:

(P_{m-2}) Points in the interior of I_{m-2}

(P_m) Points in the interior of I_m

Proof. The existence of type E points is obvious, and the existence of type C follows from Proposition 3.6.4. The single point of type R is precisely the postcritical point that maps to k_1. Since the loop of τ in C_1 must be interior, it follows that the image of the loop of τ that maps into C_1 must be interior, hence lies over a postcritical point at which f is orientation-reversing. Since there is only one interior loop of τ, this is the only postcritical point at which f reverses orientation.

To finish the proof, suppose f has another critical point $x = p$ other than type E, C, and R. Then the loop of τ over this critical point must have an image that is exterior so that $F(\tau)$ does not trap the loop after tightening. Since F is exterior veering, it follows that the image of this loop must come after the third-to-last turn of F, hence p is in the interior of $I_{m-2} \cup I_{m-1} \cup I_m$. Moreover, since p is not of type R, f preserves orientation at p, hence $p \notin I_{m-1}$. $\square$

Example 3.6.11. Consider the thick interval map in Figure 3.14. This is the exterior left-veering thickening of the positive λ zig-zag f, where λ is the Perron root of $D_f(t) = t^4 - 4t^3 - 2t^2 - 4t + 1$. The graph of f is shown in Figure 3.18.

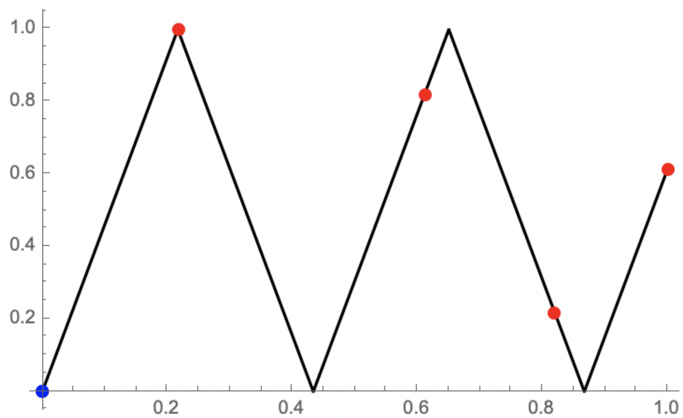


Figure 3.18: The positive zig-zag for λ the Perron root of $t^4 - 4t^3 - 2t^2 - 4t + 1$. The map f is of pseudo-Anosov type, as demonstrated by Figure 3.14, and the orbit of $x = 1$ is shown in red. In particular, $\mathrm{PC}(f)$ has points of type E, C, R, and P_{m-2}.

Corollary 3.6.12. *Suppose $f \in \mathrm{PA}(m)$ has slope λ. If $\mathrm{PC}(f)$ contains no points of type P_{m-2} or P_m then λ is a quadratic integer.*

Proof. If f is positive and has no postcritical points of type P_i then the orbit of $x = 1$ is

$$1 \mapsto \lambda - m \mapsto m + m\lambda - \lambda^2 \mapsto 2 - m\lambda - m\lambda^2 + \lambda^3 = 1,$$

hence λ satisfies the relation

$$0 = \lambda^3 - m\lambda^2 - m\lambda + 1 = (\lambda + 1)\left(\lambda^2 - (m+1)\lambda + 1\right).$$

Since $\lambda > 1$ and $x^2 - (m+1)x + 1$ is irreducible over $\mathbb{Q}$, we see that λ is a quadratic integer. Similarly, if f is negative and has no postcritical points of type P_i then from the orbit of $x = 1$ we obtain the relation

$$0 = \lambda(\lambda+1)\left(\lambda^2 - (m+1)\lambda + 1\right)$$

Hence λ is again a quadratic integer. $\square$

3.7 Classifying zig-zags of pseudo-Anosov type

In this section we consider the set $\mathrm{PA}(m)$ of zig-zags of pseudo-Anosov type with modality $m \geq 2$. The case $m = 1$ was considered by Hall in Theorem 2.1 of [Hal94], in which he gave an explicit bijection between $\mathrm{PA}(1)$ and $\mathbb{Q} \cap (0, 1/2)$. We give an explicit bijection between $\mathrm{PA}(m)$ and $\mathbb{Q} \cap (0, 1)$ for each $m \geq 2$.

The proof of Theorem B naturally breaks into 3 cases: (1) $m \geq 4$ even, (2) $m \geq 3$ odd, and (3) $m = 2$. In preparation for the proof, we investigate each of these cases separately. We treat case (1) first, since it is essential to understanding cases (2) and (3).

Throughout this section, f is a PCP λ-zig-zag map with modality $m = \lfloor \lambda \rfloor$. Moreover, we always assume that $\mathrm{PC}(f) \geq 4$, so that it is possible for f to belong to $\mathrm{PA}(m)$ (cf. Definition 3.6.8).

3.7.1 The case $m \geq 4$ even

Definition 3.7.1. Given $n \geq 3$ and $2 \leq k \leq n-1$, define $\rho_e(n,k) \in \mathfrak{S}_n$ to be the permutation such that

$$\rho_e(n,k)(i) = \begin{cases} n & \text{if } i = 1 \\ i + (n-k) & \text{if } 2 \leq i \leq k-1 \\ i - (k-1) & \text{if } k \leq i \leq n. \end{cases}$$

Example 3.7.2. Here are some examples of $\rho_e(n,k)$ for $n = 7$.

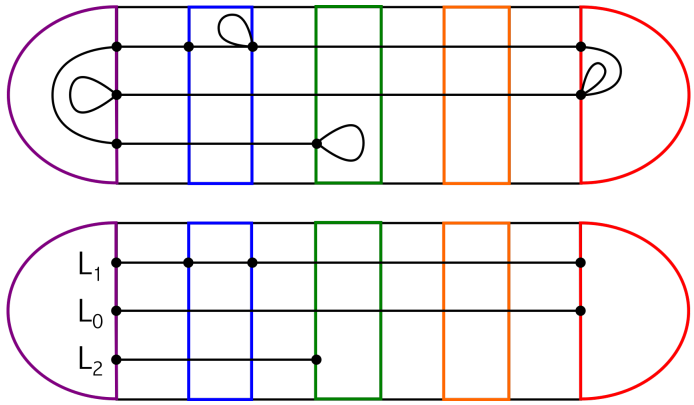


Figure 3.19: The layers of $F(\tau)$ for the exterior left-veering thickening of a bimodal zig-zag.

$$\rho_e(7,2) = (7,6,5,4,3,2,1)$$

$$\rho_e(7,3) = (7,5,3,1)(6,4,2)$$

$$\rho_e(7,6) = (7,2,3,4,5,6,1).$$

Note that $\rho_e(7,2)$ and $\rho_e(7,6)$ are full 7-cycles, whereas $\rho_e(7,3)$ is not. As we will see, $\rho_e(n,k)$ is an n-cycle if and only if $n-k$ and $n-1$ are coprime (cf. Proposition 3.7.7).

Definition 3.7.3. Let F be a thickening of a PCF map f, and let τ denote the invariant generalized train track of F. We define a *layer* of the image $F(\tau)$ to be a connected component of the complement of $\mathbf{0} \cup \mathbf{1}$ in $F(\tau')$. See Figure 3.19.

If f is an m-modal zig-zag and F is any thickening of f, then $F(\tau)$ has $m+1$ layers. Indeed, if L is a layer of $F(\tau)$, then $F^{-1}(L)$ projects to one of the $m+1$ different intervals of monotonicity for f.

Definition 3.7.4. Let f be an m-modal zig-zag and F the exterior left-veering thickening of f. For each $0 \le j \le m$, the *j-th layer* of $F(\tau)$ is the layer L_j such that

$\pi(F^{-1}(L_j))$ is the j-th interval of monotonicity for f.

Definition 3.7.5. Let f be a PCP interval map, and let $x_1 < x_2 < \ldots < x_n = 1$ denote the elements of the forward orbit of $x = 1$. The *permutation type* of f is the permutation $\rho(f) \in \mathfrak{S}_n$ such that $f(x_i) = x_{\rho(f)(i)}$.

Proposition 3.7.6. *Let $f : I \to I$ be a PCP λ-zig-zag map with $\lfloor\lambda\rfloor = m \geq 4$ even. Then $f \in \mathrm{PA}(m)$ if and only if the following hold:*

1. *the permutation type of f is have $\rho(f) = \rho_e(n, k)$ for some k, n.*
2. *the set $\{x_i\}$ satisfies the conclusion of Proposition 3.6.10, with x_1 being the sole point of type (C) and x_k being the sole point of type (R).*

Proof. We begin first by assuming that f is of pseudo-Anosov type. Let τ be the invariant train track for the exterior left-veering thick interval map F. By Proposition 3.6.4 the junction C_1 over $k_1 = \lambda^{-1}$ contains the sole interior loop γ of τ, and this loop must map into $\mathbf{1}$. Moreover, since $\lfloor\lambda\rfloor \geq 4$, the layers L_0, L_1, and L_2 stretch fully between $\mathbf{0}$ and $\mathbf{1}$, with L_1 above L_3 and below L_2. In particular, there can be no intermediate loops of τ before γ, since otherwise these would be trapped by L_2 or L_3. Thus if ρ is the permutation describing the action of f on the orbit of $x = 1$, we must have $\rho(1) = n$.

Now set $\rho^{-1}(1) = k$. By Proposition 3.6.5, all loops in τ corresponding to $i \neq k$ must have exterior image, so their images must lie in L_{m-2} or L_m. In particular, if $k = 2$ then the remaining $n - k$ loops must be sent in order to the remaining $k - 2$ junctions containing a loop. In particular, we have

$$\rho = (1, n, n-1, n-2, \ldots, 3, 2) = \rho_e(n, 2).$$

If $k > 2$, then the loops labelled 2 through $k-1$ must map to the last $k-3$ loops not in **1**, in order. This is again because these loops must all have exterior images, and so in particular cannot be covered from below by loops on L_m. Now the loops labelled $k+1$ through n must map to the remaining loops labelled 2 through $n-k+1$ in order, and thus we see that $\rho = \rho_e(n,k)$.

Suppose instead that f satisfies properties (1) and (2) above. Let F be the exterior left-veering thickening of f. We construct the invariant generalized train track τ of F as described in Section 2. We begin with τ_0, the (disconnected) union of real edges, and after the tightening pseudo-isotopies, $\tau_1 = F_*(\tau_0)$ contains a loop γ projecting to $x_n = 1$. Consider the successive images $F_*^j(\gamma)$, which by conditions (1) and (2) are exterior until some minimal index $j = a$. Moreover, each of these initial forward images is guaranteed to be uncovered from below by layers of $F(\tau_{j-1})$, and so are the only infinitesimal edges within the corresponding junctions aside from a single connecting edge between the two adjacent real edges.

$F_*^a(\gamma)$ is an interior loop, and by assumption must project to $x_1 = \lambda^{-1}$, since x_k is the sole postcritical point of type (R) and $\rho_e(n,k)(k) = 1$. In particular, $F_*^{a+1}(\gamma)$ is a loop over x_n and since $x_1 = \lambda^{-1}$ this loop is interior to every turn of $F_*(\tau_a)$. Hence after tightening we recover a single loop over x_n and the resulting generalized train track $\tau_{a+1} = \tau$ is invariant. It follows that τ is finite and that f is of pseudo-Anosov type. $\square$

In order to completely characterize positive PCP zig-zags with modality $m \geq 4$ even, it remains to determine when $\rho_e(n,k)$ is an n-cycle.

Proposition 3.7.7. *The permutation $\rho_e(n,k) \in \mathfrak{S}_n$ is an n-cycle if and only if* $\gcd(n-k, n-1) = 1$.

Proof. Set $\rho = \rho_e(n,k)$, and define $\rho' \in \mathfrak{S}_{n-1}$ to be the permutation on $\{2,\ldots,n\}$ defined by

$$\rho'(i) = \begin{cases} \rho(i) & \text{if } 2 \leq i \neq k \\ n & \text{if } i = k. \end{cases}$$

Thus ρ' is the element of $\mathfrak{S}_{n-1}$ obtained by deleting the symbol 1 from the cycle decomposition of ρ. Observe that, since 1 is in the orbit of n, the permutation ρ is an n-cycle if and only if ρ' is an $(n-1)$-cycle. Shifting all labels down by 1, we have

$$\rho'(i) = \begin{cases} i + (n-k) & \text{if } 1 \leq i \leq k-1 \\ i - (k-1) & \text{if } k \leq i \leq n-1. \end{cases}$$

Interpreting this modulo $n-1$, we see that ρ' acts via addition by $n-k$. Thus this action is transitive if and only if $n-k$ and $n-1$ are coprime. □

3.7.2 The case $m \geq 3$ odd

We next consider the case when f is a PCP zig-zag of modality $m \geq 3$ odd. In this case we observe that the orbit of $x = 1$ ends with $k_1 \mapsto 0 \mapsto 1$.

Definition 3.7.8. Given $n \geq 3$ and $2 \leq k \leq n-1$, define $\rho_o(n,k) \in \mathfrak{S}_{n+1}$ to be the permutation such that

$$\rho_o(n,k)(i) = \begin{cases} n & \text{if } i = 0 \\ 0 & \text{if } i = 1 \\ i + (n-k) & \text{if } 2 \leq i \leq k-1 \\ i - (k-1) & \text{if } k \leq i \leq n. \end{cases}$$

Remark 3.7.9. Note that if we define $\rho' \in \mathfrak{S}_n$ to be the permutation obtained from $\rho = \rho_o(n,k)$ by deleting 0 from the orbit of n, then ρ' is an n-cycle if and only if ρ is an $(n+1)$-cycle. Since $\rho' = \rho_e(n,k)$, we see that this is again the case if and only if

$\gcd(n-k, n-1) = 1$. We have therefore proven the following statement.

Proposition 3.7.10. *The permutation $\rho_o(n,k) \in \mathfrak{S}_{n+1}$ is an $(n+1)$-cycle if and only if $\gcd(n-k, n-1) = 1$.*

A nearly identical argument to that of Proposition 3.7.6 proves the following analogue.

Proposition 3.7.11. *Let $f : I \to I$ be a PCP λ-zig-zag map with $\lfloor \lambda \rfloor = m \geq 3$ odd. Then $f \in \mathrm{PA}(m)$ if and only if the following hold:*

1. *under the ordering $0 = x_0 < x_1 < \ldots < x_n = 1$ of the orbit of $x = 1$, we have $\rho(f) = \rho_o(n,k)$ for some $k.\rho(f) = \rho_o(n,k)$ for some k, n.*
2. *the set $\{x_i\}$ satisfies the conclusion of Proposition 3.6.10, with x_1 being the sole point of type (C) and x_k being the sole point of type (R).*

3.7.3 The case $m = 2$

It remains to consider the case of unrestricted PCP zig-zags of modality $m = 2$.

Definition 3.7.12. Let $n \geq 3$. For $2 \leq k \leq n-1$ set the permutation

$$\kappa(n,k) = (1, 2, \ldots, k-1) \in \mathfrak{S}_n$$

and define

$$\rho_2(n,k) = [\kappa(n,k)]^{-1} \circ \rho_e(n,k) \circ \kappa(n,k).$$

Lemma 3.7.13. *Fix $n \geq 3$ and $2 \leq k \leq n-1$ and set $\rho_2 = \rho_2(n,k)$. Then the following conditions hold:*

1. *$\rho_2(k) = k - 1$.*

2. $\rho_2(k-1) = n$.

3. If $i \leq j < k$ *then* $\rho_2(i) \leq \rho_2(j)$.

4. If $k < i \leq j$ *then* $\rho_2(i) \leq \rho_2(j)$.

5. If $i < k < j$ *then* $\rho_2(j) < \rho_2(i)$.

Moreover, $\rho_2(n,k)$ *is the only element of* $\mathfrak{S}_n$ *to satisfy these conditions. Finally,* $\rho_2(n,k)$ *is an* n*-cycle if and only if* $\gcd(n-k, n-1) = 1$.

Proof. Set $\rho = \rho_e(n,k)$ and $\kappa = \kappa(n,k)$ so that $\rho_2 = \kappa^{-1}\rho\kappa$. One readily checks conditions (1) and (2) from this equation. Now suppose $i \leq j < k$. Then we have $\rho_2(i) = \kappa^{-1}(i+1+(n-k))$ and $\rho_2(j) = \kappa^{-1}(j+1+(n-k))$. Since $i \leq j$ and κ preserves this ordering except at k, we see that $\rho_2(i) \leq \rho_2(j)$ unless $i+1+(n-k) = 1$, which is impossible. Similarly, if $k < i \leq j$, then $\rho_2(i) = \kappa^{-1}(i-(k-1))$, and $\rho_2(j) = \kappa^{-1}(j-(k-1))$. Again we see that $\rho_2(i) \leq \rho_2(j)$ unless $i-(k-1) = 1$, i.e. that $i = k$. This is also impossible.

To prove condition (5), it is enough by (4) to argue that if $i < k$ then $\rho_2(n) < \rho_2(i)$. $\rho_2(i) = \kappa^{-1}(i+1+(n-k))$, whereas $\rho_2(n) = \kappa^{-1}\rho(n) = \kappa^{-1}(n-k+1)$. Again by the fact that κ preserves ordering except at k, it follows that $\rho_2(n) < \rho_2(i)$.

Suppose that $\rho \in \mathfrak{S}_n$ satisfies conditions (1)-(5) above. We wish to show that $\rho = \rho_2(n,k)$. Condition (1) implies that this second set has image including n, and condition (2) determines the image of the last remaining index. The monotonicity conditions (3) and (4) impose a strict ordering on the two subsets $\{1,\ldots,k-1\}$ and $\{k+1,\ldots,n\}$ of remaining indices, and condition (5) implies that the image of the second subset must be completely below that of the first. It follows that $\rho(i)$ is determined for each i, and hence $\rho = \rho_2(n,k)$.

Finally, since ρ_2 is conjugate to ρ the two share the same cycle type, and hence by Proposition 3.7.7, ρ_2 is an n-cycle if and only if $\gcd(n-k, n-1) = 1$. □

Proposition 3.7.14. *Let $f : I \to I$ be a PCP λ-zig-zag map with $\lfloor \lambda \rfloor = 2$. Then $f \in \mathrm{PA}(2)$ if and only if the following hold:*

1. *the permutation type of f is have $\rho(f) = \rho_2(n,k)$ for some k, n.*

2. *the set $\{x_i\}$ satisfies the conclusion of Proposition 3.6.10, with x_{k-1} being the sole point of type (C) and x_k being the sole point of type (R).*

Proof. Suppose that f is of pseudo-Anosov type. By Proposition 3.6.10, f has exactly one postcritical point in the interior of I_1, as well as another at $x = \lambda^{-1}$. Let $\rho \in \mathfrak{S}_n$ describe the action of f on the periodic orbit of $x = 1$. Let F be the exterior left-veering thick interval map projecting to f and let τ be the invariant generalized train track for F. Since f is of pseudo-Anosov type, τ is finite and has structure described by Lemma 3.4.7 and Proposition 3.6.5. In particular, τ has a single interior loop, and hence there is a single loop γ of τ whose image is this interior loop. Let x_k be the postcritical point of f to which this loops projects. Since the single interior loop of τ necessarily maps to the loop projecting to $x_n = 1$, we have $\rho^2(k) = n$.

Since γ maps to the unique interior loop of τ, γ is the unique loop such that $F(\gamma) \in L_2$. Moreover, since $\rho^2(k) = n$ we see that $F(F_*(\gamma))$ is after L_1 and before L_2. In particular, there are no loops of τ after $F_*(\gamma)$ and before γ. Therefore $F_*(\gamma)$ projects to x_{k-1}, and so $\rho(k) = k-1$. Since $\rho^2(k) = n$ it now follows that $\rho(k-1) = n$. It is now readily checked that ρ satisfies the five conditions of Lemma 3.7.13, and hence $\rho = \rho_2(n,k)$.

Suppose on the other hand that conditions (1) and (2) above hold. Let F be the exterior left-veering thick interval map projecting to f, and let τ be the invariant generalized train track of F. Note that τ has a loop γ projecting to $x = 1$. Let F^a be the first iterate that sends γ to an interior loop. All forward images of γ before this must be exterior loops, by the definition of ρ_2 and the fact that F is exterior left-veering. Note that $a \geq 2$ and that $F^{a-1}(\gamma)$ is a loop projecting to x_k, since its image is an interior loop.

We claim that $F^{a+1}(\gamma) = \gamma$. Indeed, $x = \lambda^{-1}$ must be the $(k-1)$-st postcritical point of f from left to right, since $\rho_2(n,k)(k-1) = n$. Moreover, since f has a single postcritical point in I_1, this point must be the k-th postcritical point, immediately proceeding $x = \lambda^{-1}$, and since $\rho_2(n,k)(k-1) = n$ we see that this point maps to $x = \lambda^{-1}$. Thus the single interior loop of τ is mapped into the fat vertex over $x = 1$, and hence after isotopy becomes simply γ. In particular, τ is finite, and hence f is of pseudo-Anosov type. $\square$

Remark 3.7.15. Propositions 3.7.6, 3.7.11, and 3.7.14 imply that if $f \in \mathrm{PA}(m)$ then

$$\rho(f) = \begin{cases} \rho_e(n,k) & \text{if } m \geq 4 \text{ even} \\ \rho_o(n,k) & \text{if } m \geq 3 \text{ odd} \\ \rho_2(n,k) & \text{if } m = 2. \end{cases}$$

Therefore, when m is unspecified we will write $\rho(f) = \rho_m(n,k)$.

3.7.4 The proof of Theorem B

We are now ready to prove Theorem B, which we restate here for convenience.

Theorem B. *Fix $m \geq 2$ and let $\Phi : \mathrm{PA}(m) \to \mathbb{Q} \cap (0,1)$ be the map defined by*

$$\Phi(f) = \frac{n-k}{n-1} \quad \text{if } \rho(f) = \rho_m(n,k).$$

Then Φ is a bijection. Moreover, for each $p \geq 4$ the image $\Phi(\mathrm{PA}(m,p))$ consists of the set of reduced rationals in $(0,1)$ of denominator $p-2$.

Proof. To fix ideas, let $m = 2$. Suppose f and g are two bimodal unrestricted PCP zig-zags of pseudo-Anosov type, and that $\rho(f) = \rho(g) = \rho_2(n,k)$ for some n, k satisfying $(n-k, n-1) = 1$. We claim that $f = g$. Indeed, $\rho_2(n,k)$ and bimodality determine the strong Markov matrix M_f of f, by Proposition 3.7.14. In particular, f and g are uniform expanders with the same topological entropy, and hence the same slope. It follows that $f = g$, and so the map is injective.

To prove surjectivity, it is enough to show that for every $\rho_2(n, k)$ with $(n-k, n-1)$ there is a bimodal unrestricted PCP zig-zag f of pseudo-Anosov type that acts on the orbit of $x = 1$ as $\rho_2(n, k)$. We do this by constructing the exterior left-veering thick interval map and then projecting onto the horizontal coordinate to obtain a zig-zag with the necessary combinatorics. Indeed, fix some $\rho_2(n, k)$ satisfying $(n-k, n-1) = 1$ and let τ be a train track consisting of real edges $e_1, \ldots, e_n$ and infinitesimal edges $f_1, \ldots, f_{n-1}$, with f_j joining e_j to e_{j+1}. Further adorn τ with infinitesimal loops as follows:

1. Loops γ_0 and γ_n on the left of e_1 and the right of e_n, respectively.

2. An upward-pointing loop γ_{k-1} attached to the left of e_k.

3. Downward-pointing loops γ_j attached to the right of e_j for $j = 1, \ldots, n-1$, except for $j = k-1$.

Define F to be the exterior left-veering train track map permuting the γ_k according to $\rho_2(n, k)$, while also fixing γ_0. It is not hard to see by the construction of F and the structure of $\rho_2(n, k)$ that F preserves τ.

We claim that the transition matrix M of the e_j is irreducible, i.e. for every pair of edges e_{j_1}, e_{j_2} some iterate of F maps e_{j_1} across e_{j_2}. Since each γ_j for $j \geq 1$ maps to γ_n, every e_j eventually maps across e_n, and so it is enough to prove that each e_j is eventually covered by some forward image of e_n. Since γ_n is on the right of e_n and γ_n maps to γ_j for each $j \geq 1$, it follows that e_n must map across the edge to the left of γ_j for each positive $j \neq k$: that is, e_n eventually maps across e_j for each $j \neq k-1$.

It remains to prove that e_n eventually maps across e_{k-1}. But some e_j does this, and since e_n already maps to every edge besides e_{k-1} we will be done unless e_{k-1} is the only edge that maps across e_{k-1}. This is impossible: the image of e_{k-1} is entirely before the first turn of τ, since γ_{k-1} maps to γ_n, and so the image of some other e_j covers e_{k-1} between the first and second turns. Thus M is irreducible.

Let λ be the spectral radius of M. By the Perron-Frobenius theorem, λ is a simple eigenvalue for M, and M has positive left- and right-eigenvectors for λ. Let $u = (u_1, \dots, u_n)$ be the unique left eigenvector such that $\sum_i u_i = 1$. Then assigning each e_j the length u_j (and declaring each infinitesimal edge to be length 0) we obtain a uniform λ-expander f by projecting the action of F onto the horizontal coordinate. Since u is positive none of the e_j are collapsed, and so f is a PCP zig-zag having the necessary combinatorics: $x = 1$ is periodic and f acts on this orbit by $\rho_2(n, k)$. Thus the map is surjective for $m = 2$.

Finally, consider the image $\Phi(\mathrm{PA}(2, p))$. For a map $f \in \mathrm{PA}(2, p)$, the fraction $\Phi(f)$ is reduced and has denominator $p - 2$, since $\rho(f) = \rho_2(p - 1, k)$ for some k such that $k - 1$ is coprime to $p - 2$. The fact that $\Phi(\mathrm{PA}(2, p))$ contains all such fractions follows because Φ is a bijection onto $\mathbb{Q} \cap (0, 1)$.

A similar argument works for $m \geq 3$ odd and $m \geq 4$ even. In this case, the placement of infinitesimal loops as in steps 1-3 above is chosen accordingly to ensure that τ is invariant under the exterior left-veering thickening.

$\square$

3.8 The digit polynomial D_f

In this section we prove Theorem C, restated below for convenience. This theorem generalizes Lemma 2.5 in [Hal94], which treats the unimodal case. In [Hal94], the result is phrased in terms of the kneading sequence of the unique critical point.

Theorem C. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 2$ with $\Phi(f) = \frac{a}{b} \in \mathbb{Q} \cap (0, 1)$ in lowest terms. Define $L : [0, b] \to \mathbb{R}$ by $L(t) = \frac{a}{b} \cdot t$. Then*

$$D_f(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^{b+1-i},$$

where the c_i satisfy

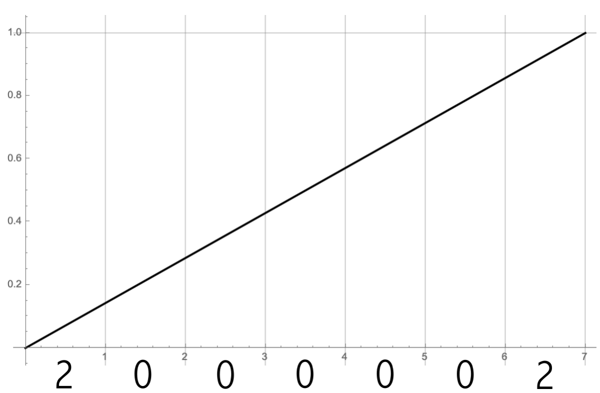
$$c_i = \begin{cases} m & \text{if } L(t) \in \mathbb{N} \text{ some } t \in [i-1, i] \\ m-2 & \text{otherwise.} \end{cases} \tag{3.4}$$

In particular, $c_i = c_{b-i}$, so D_f is reciprocal: that is,

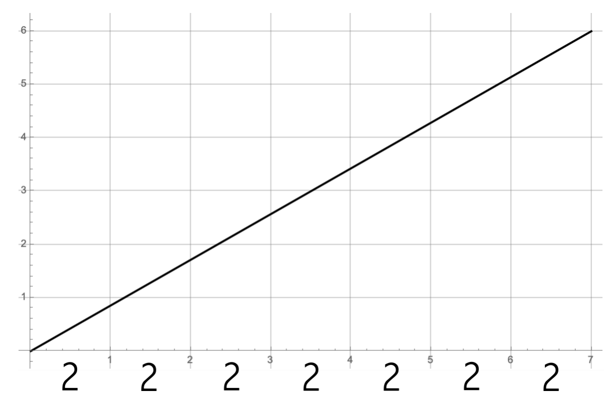
$$D_f(t) = t^{b+1} D_f(t^{-1}).$$

Example 3.8.1. Let $f, g \in \mathrm{PA}(2)$ be the zig-zags such that $\Phi(f) = \frac{1}{7}$ and $\Phi(g) = \frac{6}{7}$. According to Theorem C and Figure 3.20, we have

$$D_f(t) = t^8 - 2t^7 - 0t^6 - 0t^5 - 0t^4 - 0t^3 - 0t^2 - 2t + 1$$
$$D_g(t) = t^8 - 2t^7 - 2t^6 - 2t^5 - 2t^4 - 2t^3 - 2t^2 - 2t + 1.$$



(a) D_f for the $f \in \mathrm{PA}(2)$ with $\Phi(f) = \frac{1}{7}$.



(b) D_g for the $g \in \mathrm{PA}(2)$ with $\Phi(g) = \frac{6}{7}$.

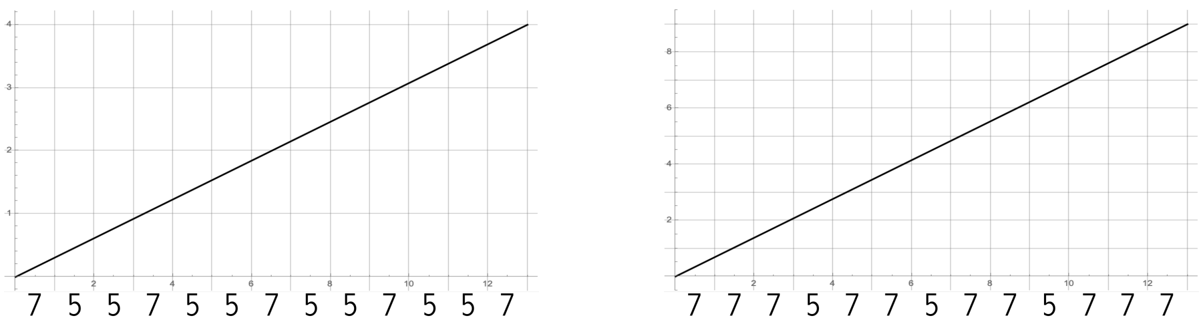
Figure 3.20: Computing the digit polynomials of bimodal maps using Theorem C.

Example 3.8.2. Let $p, q \in \mathrm{PA}(7)$ be the zig-zags such that $\Phi(p) = \frac{4}{13}$ and $\Phi(q) = \frac{9}{13}$.

According to Theorem C and Figure 3.21, we have

$$D_p(t) = t^{14} - 7t^{13} - 5t^{12} - 5t^{11} - 7t^{10} - 5t^9 - 5t^8 - 7t^7 - 5t^6 - 5t^5 - 7t^4 - 5t^3 - 5t^2 - 7t + 1$$

$$D_q(t) = t^{14} - 7t^{13} - 7t^{12} - 7t^{11} - 5t^{10} - 7t^9 - 7t^8 - 5t^7 - 7t^6 - 7t^5 - 5t^4 - 7t^3 - 7t^2 - 7t + 1.$$



(a) D_p for the $p \in \mathrm{PA}(7)$ with $\Phi(p) = \frac{4}{13}$.

(b) D_q for the $q \in \mathrm{PA}(7)$ with $\Phi(q) = \frac{9}{13}$.

Figure 3.21: Computing the digit polynomials of 7-modal maps using Theorem C.

The true impact of Theorem C will not be felt until Chapter 4, where we prove that $D_f(t)$ coincides with several polynomial invariants related to both f and ψ_f. See Theorem G for more information.

We first prove Theorem C for the case that f has modality $m \geq 4$ even. This case being completed, we will then use it to deduce the cases for $m \geq 3$ odd and $m = 2$. We proceed via a sequence of lemmas.

Lemma 3.8.3. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 4$ even. Let n be minimal such that $f^n(1) = 1$, and let $f_i(x)$ be the defining linear branch of f at $f^{i-1}(1)$. Then there exist constants c_i for $i = 1, \dots, n-2$ such that*

$$f_i(x) = \begin{cases} \lambda x - c_i & i = 1, \dots, n-2 \\ m - \lambda x & i = n-1 \\ 2 - \lambda x & i = n \end{cases}$$

Proof. The sole point of type R (cf. Proposition 3.6.10) in the orbit of $x = 1$ must map to the point of type C, by Proposition 3.6.5, and this latter point is precisely $x = \lambda^{-1}$, and hence maps to $x = 1$ by $f_n(x) = 2 - \lambda x$. At the point of type R we have $f_{n-1}(x) = m - \lambda x$.

All other points in the orbit of $x = 1$ are of type P_1 or P_2, and the corresponding f_i are $f_i(x) = \lambda x - (m-2)$ and $f_i(x) = \lambda x - m$, respectively. □

Lemma 3.8.4. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 4$ even. Let $D_f(t)$ be the digit polynomial of f. Then*

$$D_f(t) = t^n + 1 - \sum_{i=1}^{n-1} c_i t^{n-i}$$

where, as in Lemma 3.8.3, the c_i are defined by $f_i(x) = \lambda x - c_i$ for $1 \leq i \leq n-2$ and $f_{n-1}(x) = c_{n-1} - \lambda x$. In particular, $c_1 = c_{n-1} = m$.

Proof. For $i = 0, \ldots, n$ define $g_i(t) \in \mathbb{Z}[t]$ by $g_i(\lambda) = f^i(\lambda)$. Thus for example $g_0(\lambda) = 1$ and $g_1(\lambda) = \lambda - m$. Here it is important that we are treating λ as a formal variable, rather than an algebraic integer satisfying a polynomial relation.

Observe that $g_{i+1}(\lambda) = f_{i+1}(g_i(\lambda))$ by definition. In particular, by Lemma 3.8.3 for $i = 0, \ldots, n-3$ we have $g_{i+1}(\lambda) = \lambda g_i(\lambda) - c_i$, and so inductively we see that for

$$g_i(\lambda) = \lambda^i - c_1\lambda^{i-1} - \cdots - c_{i-1}\lambda - c_i \quad \text{for } 0 \leq i \leq n-2$$

We therefore have the equalities

$$g_{n-1}(\lambda) = c_{n-1} - \lambda^{n-1} + c_1\lambda^{n-2} + \cdots + c_{n-3}\lambda^2 + c_{n-2}\lambda$$
$$g_n(\lambda) = 2 - c_{n-1}\lambda - c_{n-2}\lambda^2 - \cdots - c_1\lambda^{n-1} + \lambda^n.$$

By the definition of $D_f(t)$ and the fact that $g_n(\lambda) = f^n(1) = 1$, we have $D_f(\lambda) =$

$g_n(\lambda) - 1$. $\square$

Lemma 3.8.5. *Suppose $f \in \mathrm{PA}(m)$ for $m \geq 4$ even. Let n be minimal such that $f^n(1) = 1$ and suppose that f acts on the orbit of $x = 1$ by the permutation $\rho(f) = \rho_e(n,k)$ with $\gcd(n-k, k-1) = 1$. Let $D_f(t) = t^n + 1 - \sum_{i=1}^{n-1} c_i t^{n-i}$. Then for $i = 1, \dots, n-1$ we have*

$$c_i = \begin{cases} m & \text{if } \rho(f)^{i-1}(n) \geq k \\ m-2 & \text{if } 2 \leq \rho(f)^{i-1}(n) \leq k-1. \end{cases}$$

Proof. The permutation $\rho(f)$ is defined such that postcritical points corresponding to symbols i satisfying $k+1 \leq i \leq n$ are of type P_2, whereas those corresponding to i satisfying $2 \leq i \leq k-1$ are of type P_1. Finally, the point corresponding to the symbol k is of type R, specifically $f^{n-2}(1)$. The result follows by Lemmas 3.8.3 and 3.8.4. $\square$

Lemma 3.8.6. *Theorem C holds for $f \in \mathrm{PA}(m)$ with $m \geq 4$ even.*

Proof. We have already shown that $c_0 = c_n = 1$ and that $c_i \in \{m-2, m\}$ for $1 \leq i \leq n-1$. By Lemma 3.8.5 it is enough to understand when $\rho^i(n) \geq k$. Let $\rho' \in \mathfrak{S}_{n-1}$ be the permutation obtained by deleting the symbol 1 from the cycle decomposition of ρ and decreasing all remaining labels by 1, as in the proof of Proposition 3.7.7. As observed previously, ρ' acts on $\{1, \dots, n-1\}$ as addition by $n-k$ modulo $n-1$. Therefore, if we set $q = \frac{n-k}{n-1}$, the values $L_q(t) = qt$ for $t = 1, \dots, n-1$ satisfy

$$L_q(t) - \lfloor L_q(t) \rfloor = \frac{(\rho')^t(n-1)}{n-1}.$$

To see why this is true, observe first that the left-hand side is the fractional part of $L_q(t)$. This quantity is a rational number with denominator $n-1$, and the numerator increases by $n-k$ modulo $n-1$. Since $(\rho')^t(n-1)$ also changes in this way, it remains to note that

$$L_q(1) = \frac{n-k}{n-1} = \frac{(\rho')(n-1)}{n-1}.$$

Now we observe that

$$\begin{aligned} c_i = m &\iff \rho^{i-1}(n) \geq k \\ &\iff (\rho')^{i-1}(n-1) \geq k-1 \\ &\iff L_q(i-1) - \lfloor L_q(i-1) \rfloor \geq \frac{k-1}{n-1} \\ &\iff L_q(t) \in \mathbb{N} \text{ for some } t \in [i-1, i]. \end{aligned}$$

The last equivalence holds because $L_q(t) = qt$ is a line of slope $\frac{n-k}{k-1}$. The proof is complete. □

It remains to prove Theorem C for the case when the modality of f is $m = 2$ or $m \geq 3$ odd. Recall that $\rho_2(n,k) = \kappa^{-1}(n,k) \circ \rho_e(n,k) \circ \kappa(n,k)$, where $\kappa(n,k) = (1, 2, \dots, k-1) \in \mathfrak{S}_n$.

Lemma 3.8.7. *For any $l \geq 0$, $\rho_2^l(n) \geq k$ if and only if $\rho_e^l(n) \geq k$, and in this case $\rho_2^l(n) = \rho_e^l(n)$.*

Proof. Suppose $\rho_e^l(n) \geq k$. Then $\rho_2^l(n) = \kappa^{-1}\rho_e^l\kappa(n) = \rho_e^l(n)$, since $\kappa(j) = j$ for all $j \geq k$. Similarly, if $\rho_2^l(n) \geq k$ it follows that $\rho_e^l(n) \geq k$ as well. □

We now complete the proof of Theorem C.

Proof of Theorem C. One quickly verifies that Lemmas 3.8.3 through 3.8.5 hold for f an m-modal zig-zag of pseudo-Anosov type for $m = 2$ and $m \geq 3$ odd, after replacing all instances of $\rho_e(n,k)$ with either $\rho_2(n,k)$ or $\rho_o(n,k)$. Lemmas 3.8.5 and 3.8.7 now imply the Theorem for $m = 2$.

For $m \geq 3$ odd, recall that if we delete the symbol 0 from the cycle type of $\rho_o(n,k)$ we obtain the permutation $\rho_e(n,k)$. Deleting this symbol corresponds to ignoring the linear map $f_{n+1}(x) = 1 - \lambda x$. This linear branch is not used to compute D_f, since the fact that $f \in \mathrm{PA}(m)$ for $m \geq 3$ odd implies that $f^n(1) = 0$, terminating the process of constructing D_f before the $(n+1)$th step (cf. Definition 3.2.6).

The arguments of Lemmas 3.8.3 through 3.8.6 now prove the Theorem in this case. □

3.9 A family of pseudo-Anosovs with Salem dilatation

In this section we provide an application of the theory developed over the course of this paper. Recall that a *Salem number*, introduced in [Sal83], is a real algebraic integer $\lambda > 1$ such that all Galois conjugates of λ are contained within the closed unit disc, with at least one of these conjugates lying on the unit circle. It is not hard to show that λ^{-1} must be among the Galois conjugates of λ in this case, and that all other conjugates lie on the unit circle. In particular, a Salem number is a Perron number of even degree $d = 2g$. If $p(x) \in \mathbb{Z}[x]$ is the minimal polynomial of a Salem number, then $p(x)$ is *reciprocal*, i.e.

$$p_*(x) := x^{\deg(p)} p(x^{-1}) = p(x).$$

It is well-known that if $f(x) \in \mathbb{Z}[x]$ is a reciprocal polynomial of degree $d = 2g$ then $f(x) = x^g q(x + x^{-1})$ for some integral polynomial q. We call $q(x)$ the *companion polynomial* to $f(x)$. If $\deg(f) = 2g + 1$ then $f(x) = (x+1)f_1(x)$ for $f_1(x)$ reciprocal of even degree, and therefore $f(x) = (x+1)x^g q(x + x^{-1})$ for some $q(x) \in \mathbb{Z}[x]$. In this case we again call q the companion polynomial of f.

Note that there is a bijection between roots of q and pairs of roots of f: if λ, λ^{-1} are roots of f then $\lambda + \lambda^{-1}$ is a root of q. Moreover, if $|\lambda| = 1$, then $\lambda^{-1} = \overline{\lambda}$ and so the root $\lambda + \lambda^{-1}$ of q is a real number contained in the interval $[-2, 2]$. We refer to this interval, with or without its endpoints, as the *critical interval*.

In the case of a Salem number λ of degree $2g$, the companion polynomial is irreducible of degree g with dominant root $\lambda + \lambda^{-1} > 2$ and the remaining $g-1$ roots in the critical interval. In particular, the companion polynomial has all real roots, so $\lambda + \lambda^{-1}$ is a totally real algebraic integer of degree g.

Recall that a *translation surface* is a pair (X, ω) of a Riemann surface X equipped with a non-zero abelian differential ω, i.e. a holomorphic one-form. If X is of genus g, then ω has $2g - 2$ zeros, counting multiplicity. Let Σ be the collection of these zeros. Then away from Σ, X has a Euclidean structure: in other words, $X \setminus \Sigma$ admits an atlas of charts to $\mathbb{C}$ whose transition functions are translations. In the neighborhood of a zero p of order k, X has the structure of $2(k+1)$ metric half-discs glued together, so that the total angle about p is $2\pi(k + 1)$.

Fixing $g > 1$, let μ be a positive integer partition of $2g - 2$. We think of μ as describing the multiplicities of the zeros of an abelian differential ω on X. The *stratum* $\mathcal{H}(\mu)$ is the collection of genus g translations surfaces with zero orders specified by μ.

Given some $(X, \omega) \in \mathcal{H}(\mu)$ and $A \in \mathrm{SL}(2, \mathbb{R})$, define $A \cdot (X, \omega)$ to be the translation surface obtained by post-composing the charts of (X, ω) into $\mathbb{R}^2 \cong \mathbb{C}$ by A.

The *Veech group* of a translation surface (X, ω), denoted $\mathrm{SL}(X, \omega)$, is the stabilizer of (X, ω) under the action by $\mathrm{SL}(2, \mathbb{R})$. The *trace field* of (X, ω) is the field K obtained by adjoining to $\mathbb{Q}$ the traces of all elements of $\mathrm{SL}(X, \omega)$. By a result of Möller in [MÏ3], the degree of the trace field satisfies

$$[K : \mathbb{Q}] \leq g(X).$$

If the degree of this extension is maximal, i.e. is equal to the genus of X, then we say that (X, ω) is *algebraically primitive*. One remarks that such surfaces cannot arise as covers of translation surfaces of lower genus: if $\pi : (X, \omega) \to (Y, \eta)$ is a translation cover, then the trace fields of X and Y coincide, and the result of Möller mentioned above now shows that if X is algebraically primitive, then $g(X) = g(Y)$.

A translation surface is called *Veech* if its Veech group is a lattice in $\mathrm{SL}(2, \mathbb{R})$, i.e. if the quotient

$$\mathrm{SL}(2, \mathbb{R}) / \mathrm{SL}(X, \omega)$$

has finite volume. The $\mathrm{GL}(2, \mathbb{R})$-orbit of a Veech surface is called a *Teichmüller curve*. By results of Möller [Mö08] and Apisa [Api18], there are only finitely many algebraically primitive Teichmüller curves in any genus $g \geq 3$. Therefore, it is interesting to find algebraically primitive surfaces with non-trivial Veech group of arbitrarily high genus.

We now restate our last main result.

Theorem D. *For each $g \geq 1$ define $f_g : I \to I$ to be the bimodal PCP zig-zag map of pseudo-Anosov type corresponding to $r_g = \frac{1}{2g} \in \mathbb{Q} \cap (0, 1)$. Let λ_g be the growth rate of f_g. Then the following are true for each $g \geq 1$:*

1. *The stretch factor λ_g is a Salem number of degree $2g$.*
2. *The pseudo-Anosov ψ_g obtained from f_g is defined on $S_{0,2g+2}$.*
3. *The translation surface (X_g, ω_g) obtained as the hyperelliptic double cover of $S_{0,2g+2}$ is of genus g, and hence algebraically primitive.*

Remark 3.9.1. One might reasonably object that there is no such thing as a Salem number of degree 2. Indeed, Salem numbers are normally defined to have at least one Galois conjugate on the unit circle, in which case all Salem numbers must be of degree at least four. Here we choose to view a quadratic unit $\lambda > 1$ as a "degenerate" Salem number of degree 2.

$q_2(w) = w^2 - 3w + 1$
$q_3(w) = w^3 - 3w^2 + 3$
$q_4(w) = w^4 - 3w^3 - w^2 + 6w - 1$
$q_5(w) = w^5 - 3w^4 - 2w^3 + 9w^2 - w - 3$

Table 3.1: The first few companion polynomials q_g

3.9.1 λ_g is Salem of degree $2g$

As before, the proof of Theorem D will proceed in a sequence of lemmas. The bulk of our efforts will be focused on establishing statement (1) of Theorem D.

Lemma 3.9.2. *Let $D_g(t)$ be the digit polynomial of f_g. Then for all $g \geq 1$ we have*

$$D_g(t) = t^{2g+1} - 2t^{2g} - 2t + 1 = (t+1)d_g(t),$$

where

$$d_g(t) = t^{2g} + 1 + 3\sum_{i=1}^{2g-1}(-1)^i t^i.$$

Proof. The fact that $D_g(t) = t^{2g+1} - 2t^g - 2t + 1$ follows from Theorem C. One readily checks that $(t+1)d_g(t) = D_g(t)$. □

We wish to show that $d_g(t)$ is the minimal polynomial of a Salem number. To do this, we must prove that $d_g(t)$ has $2g-2$ roots on the unit circle, and also that the polynomial is irreducible. As we have seen, to accomplish the first task it will be enough to show that the companion polynomial $q_g(t)$, defined by $d_g(t) = t^g q_g(t + t^{-1})$, has $g-1$ roots in the critical interval. This is the content of the next two lemmas.

Lemma 3.9.3. *The companion polynomials $q_g(w)$ satisfy the following properties:*

1. *For all $g \geq 4$ w have the recurrence relation*

$$q_{g+2}(w) = wq_{g+1}(w) - q_g(w). \tag{3.5}$$

2. *$q_g(2) = -1$ for all n.*

3. *Each q_g has a real root $\alpha_g > 2$.*

4. *For all $g \geq 2$ we have $(-1)^g q_g(-2) > 0$. In particular, $q_g(-2)$ and $q_{g+1}(-2)$ have opposite signs.*

Proof. By definition, q_g satisfies $q_g(t + t^{-1}) = t^{-g} d_g(t)$, so relation (3.5) is equivalent to the recurrence

$$d_{g+2}(t) = t^2 \left[d_{g+1}(t) - d_g(t)\right] + d_{g+1}(t).$$

This formula is a straightforward consequence of the pattern of the coefficients of d_g, proving statement (1). Statement (2) follows inductively after noting that it holds for q_2 and q_3. Now note q_2 and q_3 are monic, so by equation (3.5) q_g is monic for all g. In particular, $\lim_{w\to\infty} q_g(w) = \infty$ for all g, so the Intermediate Value Theorem and statement (2) together imply statement (3).

Finally, to prove statement 4 observe that

$$d_g(-1) = 1 + 3(2g - 1) + 1 = 6g - 1 > 0.$$

for all $g \geq 2$. Since $d_g(-1) = (-1)^g q_g(-2)$, the result follows. ☐

Lemma 3.9.4. *For each $g \geq 2$ the polynomial q_g has $g-1$ roots in $(-2, 2)$. Moreover, if these roots are denoted $-2 < a_1 < \cdots < a_{g-1} < 2$, then the g roots of q_{g+1} in $(-2, 2)$, denoted $b_1, \ldots, b_g$, satisfy the ordering*

$$-2 < b_1 < a_1 < b_2 < a_2 < \cdots < b_{g-1} < a_{g-1} < b_g < 2.$$

Proof. We proceed inductively. The quadratic polynomial q_2 has a single root $a_1 = \frac{3-\sqrt{5}}{2}$ in $(-2, 2)$. By Lemma 3.9.4(3), this is in fact the only root of q_2 in the critical interval. Since $q_3(-1) = -1$, $q_3(0) = 3$, $q_3(1) = 1$, and $q_3(2) = -1$, we see that q_3 has two roots $b_1, b_2 \in (-2, 2)$ satisfying $-1 < b_2 < 0$ and $1 < b_2 < 2$. Since $0 < a_1 < 1$, the claim is satisfied in this case.

Suppose now that the claim holds for all $n \leq g+1$. We may assume without loss of generality that $q_{g+1}(-2) < 0$: the other case is essentially identical. Lemma 3.9.3(4) implies that both q_g and q_{g+2} are positive at $w = -2$. Let $b_1 < \ldots < b_g$ be the roots of q_{g+1} in the critical interval. Then we have

$$q_{g+2}(b_1) = b_1 q_{g+1}(b_1) - q_g(b_1) = -q_g(b_1).$$

By assumption, the smallest root a_1 of q_g in $(-2, 2)$ is greater than b_1. Since $q_g(-2) > 0$, it follows that $q_g(b_1) > 0$ and thus $q_{g+2}(b_1) < 0$. Therefore q_{g+2} has a root $c_1 \in (-2, b_1)$. Next we observe that

$$q_{g+2}(b_2) = b_2 q_{g+1}(b_2) - q_g(b_2) = -q_g(b_2).$$

Since $a_1 < b_2 < a_2$, we see that $q_g(b_2) < 0$, so $q_{g+2}(b_2) > 0$, implying that q_{g+2} has a root $c_2 \in (b_1, b_2)$. Continuing in this fashion we find roots $c_i \in (b_{i-1}, b_i)$ for $2 \leq i \leq g$. Finally, note that since $q_g(2) = -1$ we have $q_g(w) < 0$ for all $w \in (a_{g-1}, -2]$. Therefore, since $a_{g-1} < b_g < 2$ we have

$$q_{g+2}(b_g) = -q_g(b_g) > 0.$$

Since $q_{g+2}(2) = -1$, it follows that q_{g+2} has a root $c_{g+1} \in (b_g, 2)$. Lemma 3.9.3(4) implies that q_{g+2} has another root $\alpha_{g+2} > 2$, so there cannot be any other roots of q_{g+2}. The proof is complete. □

Lemma 3.9.4 implies that d_g has $g-1$ pairs of roots on the unit circle in addition to a pair of positive real roots λ_g and λ_g^{-1}. This is already enough to conclude that λ_g is a Salem number. Indeed, λ_g must be Galois conjugate to λ_g^{-1}, since otherwise

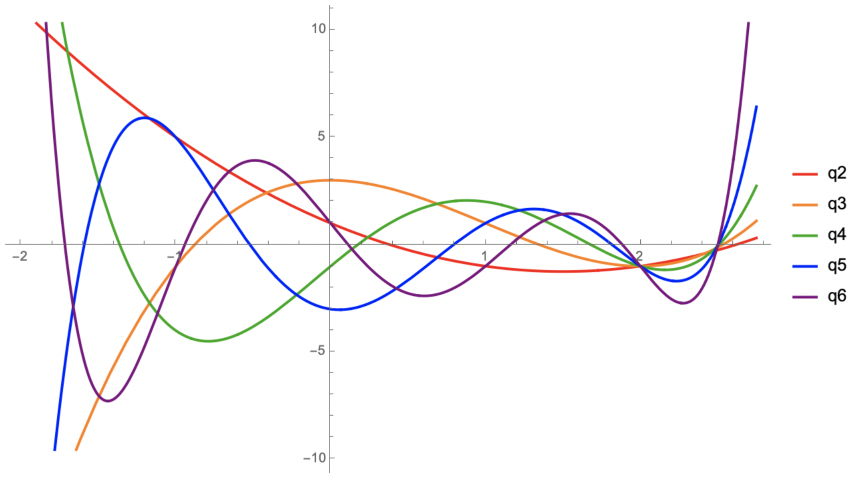


Figure 3.22: The interlacing property of the roots of the companion polynomials q_g.

the conjugates of λ_g^{-1} would be contained in the closed unit disc, implying that they are all roots of unity, by Kronecker's theorem. This last is impossible, since $\lambda_g^{-1} < 1$. Hence λ_g and λ_g^{-1} are Galois conjugate and λ_g is a Salem number. It remains to determine whether d_g is irreducible.

We follow a similar proof given by Shin in [Shi16]. Interestingly, Shin's proof shows that the "dual" Perron roots ψ_g corresponding to the fraction $\tilde{r}_g = 1 - r_g = \frac{2g-1}{2g}$ are also Salem of degree $2g$.

Lemma 3.9.5. *Each $d_g(t)$ is irreducible over $\mathbb{Z}[t]$. Consequently, λ_g is a Salem number of degree $2g$ for all $g \geq 1$.*

Proof. Since λ_g is necessarily Galois conjugate to λ_g^{-1}, any factor of $d_g(t)$ other than the minimal polynomial of λ_g must have all roots on the unit circle, and therefore must be cyclotomic, by Kronecker's theorem. Suppose therefore that $e^{2\pi i/m}$ is a root of $d_g(t)$. Then we have

$$D_g(e^{2\pi i/m}) = e^{(2g+1)\cdot 2\pi i/m} - 2e^{2g\cdot 2\pi i/m} - 2e^{2\pi i/m} + 1 = 0.$$

We take real and imaginary parts to obtain the system of equations

$$\begin{cases} \cos\left(\frac{(2g+1)2\pi}{m}\right) - 2\cos\left(\frac{2g\cdot 2\pi}{m}\right) - 2\cos\left(\frac{2\pi}{m}\right) + 1 = 0 \\ \sin\left(\frac{(2g+1)2\pi}{m}\right) - 2\sin\left(\frac{2g\cdot 2\pi}{m}\right) - 2\sin\left(\frac{2\pi}{m}\right) = 0. \end{cases} \tag{3.6}$$

For the first equation, we use the formula $\cos(2x) = 2\cos^2(x) - 1$ for the first cosine term and the formula $\cos(a) + \cos(b) = 2\cos(\frac{a+b}{2})\cos(\frac{a-b}{2})$ for the latter two terms to obtain

$$\cos\left(\frac{(2g+1)\pi}{m}\right)\left[\cos\left(\frac{(2g+1)\pi}{m}\right) - 2\cos\left(\frac{(2g-1)\pi}{m}\right)\right] = 0. \tag{3.7}$$

Similarly for the second equation in (3.6) we use the formula $\sin(2x) = 2\sin(x)\cos(x)$ and the formula $\sin(a) + \sin(b) = 2\sin(\frac{a+b}{2})\cos(\frac{a-b}{2})$ to find

$$\sin\left(\frac{(2g+1)\pi}{m}\right)\left[\cos\left(\frac{(2g+1)\pi}{m}\right) - 2\cos\left(\frac{(2g-1)\pi}{m}\right)\right] = 0. \tag{3.8}$$

Since $\sin(x)$ and $\cos(x)$ do not have any common roots, it must be the case that

$$\cos\left(\frac{(2g+1)\pi}{m}\right) - 2\cos\left(\frac{(2g-1)\pi}{m}\right) = 0.$$

Setting $\varphi = \frac{(2g-1)\pi}{m}$ we rewrite this as

$$\cos\left(\varphi + \frac{2\pi}{m}\right) - 2\cos(\varphi) = 0. \tag{3.9}$$

It follows that $-1/2 \leq \cos(\varphi) \leq 1/2$. In other words, since $\cos(\pi/3) = 1/2$ and $\cos(2\pi/3) = -1/2$ we must have

$$-\frac{2\pi}{3} \leq \varphi \leq -\frac{\pi}{3} \text{ or } \frac{\pi}{3} \leq \varphi \leq \frac{2\pi}{3}.$$

Moreover, by equation (3.9), φ and $\varphi + \frac{2\pi}{m}$ are angles on the same side of the y-axis, since their cosines have the same sign. We claim that φ must be in either the second or fourth quadrant. Suppose for contradiction that φ is in the first quadrant, so that $\frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$. We may assume $m \geq 3$, since $x = \pm 1$ are clearly not roots of d_g, and thus $\frac{2\pi}{m} < \pi$. In particular, $\varphi + \frac{2\pi}{m} < \frac{\pi}{2} + \pi = \frac{3\pi}{2}$, hence cannot be in the fourth quadrant.

Since both φ and $\varphi + \frac{2\pi}{m}$ are on the same side of the y-axis and we assumed φ is in the first quadrant, it follows that $\frac{\pi}{3} \leq \varphi + \frac{2\pi}{m} \leq \frac{\pi}{2}$ as well. Since $\cos(x)$ is decreasing on this interval we have

$$\cos(\varphi) > \cos\left(\varphi + \frac{2\pi}{m}\right) > 0 \implies 2\cos(\varphi) > \cos\left(\varphi + \frac{2\pi}{m}\right),$$

contradicting equation (3.9). A similar argument shows that φ cannot be in the third quadrant. Thus we revise the restrictions on φ to be

$$-\frac{\pi}{2} < \varphi \leq -\frac{\pi}{3} \text{ or } \frac{\pi}{2} < \varphi \leq \frac{2\pi}{3}.$$

Appealing to the formula $\cos(\theta) = \sin(\theta + \pi/2)$, we may equivalently consider the equation $\sin(\psi + 2\pi/m) - 2\sin(\psi) = 0$ with

$$0 < \psi \leq \frac{\pi}{6} \text{ or } \pi < \psi \leq \frac{7\pi}{6}.$$

Suppose ψ is in the first quadrant and write

$$\psi = \varphi + \frac{\pi}{2} = \frac{(2g-1)\pi}{m} + \frac{\pi}{2} \equiv \frac{j\pi}{2m} \pmod{2\pi}.$$

for some positive integer $j \leq 2m - 1$ such that $0 < \frac{j\pi}{2m} \leq \frac{\pi}{6}$. Using the subadditivity of $\sin(x)$ on $[0, \pi]$ now gives

$$\begin{aligned}\sin\left(\psi + \frac{2\pi}{m}\right) - 2\sin(\psi) &\leq \sin(\psi) + \sin\left(\frac{2\pi}{m}\right) - 2\sin(\psi)\\ &= \sin\left(\frac{2\pi}{m}\right) - \sin\left(\frac{j\pi}{2m}\right).\end{aligned}$$

This expression cannot be zero unless $j = 4$ because of the restriction on j. In this case,

$$\begin{aligned}\sin\left(\psi + \frac{2\pi}{m}\right) - 2\sin(\psi) &= \sin\left(\frac{4\pi}{m}\right) - 2\sin\left(\frac{2\pi}{m}\right)\\ &= 2\sin\left(\frac{2\pi}{m}\right)\cos\left(\frac{2\pi}{m}\right) - 2\sin\left(\frac{2\pi}{m}\right)\\ &= 2\sin\left(\frac{2\pi}{m}\right)\left[\cos\left(\frac{2\pi}{m}\right) - 1\right].\end{aligned}$$

This quantity can only be zero if $m = 1$, which we quickly rule out after observing that $d_g(1) = 6g - 1$. The same argument works if we assume ψ is in the third quadrant. Therefore $d_g(t)$ has no cyclotomic factor, and hence is irreducible. $\square$

3.9.2 Completing the proof of Theorem D

Proof of Theorem D. Lemma 3.9.5 proves statement (1). Since the λ_g-uniform expander f_g corresponds to the fraction $q_g = \frac{1}{2g}$, the point $x = 1$ is periodic of length $2g+1$. Each of these points lifts to a one-pronged singularity of the pseudo-Anosov ψ_g on a punctured sphere, as does the fixed postcritical point $x = 0$. There are no other one-pronged singularities, so taking the double cover of the surface by branching at each of these $2g + 2$ points produces a surface on which the lift of each point is a flat point, i.e. has cone angle 2π. The only other cone point downstairs is the fixed point at infinity, with cone angle $2g \cdot \pi$. This point lifts to two points of angle $g \cdot 2\pi$.

The derivative of the lifted pseudo-Anosov is

$$D\tilde{\psi}_g = \begin{pmatrix} \lambda_g & 0 \\ 0 & \lambda_g^{-1} \end{pmatrix},$$

an element of the Veech group $\mathrm{SL}(X_g, \omega_g)$. The trace of this matrix is $\lambda_g + \lambda_g^{-1}$, contained in the trace field K_g by definition. But now

$$g \geq [K_g : \mathbb{Q}] \geq [\mathbb{Q}(\lambda_g + \lambda_g^{-1}) : \mathbb{Q}] = g,$$

so in fact we have $[K_g : \mathbb{Q}] = g$, implying that (X_g, ω_g) is algebraically primitive. $\square$

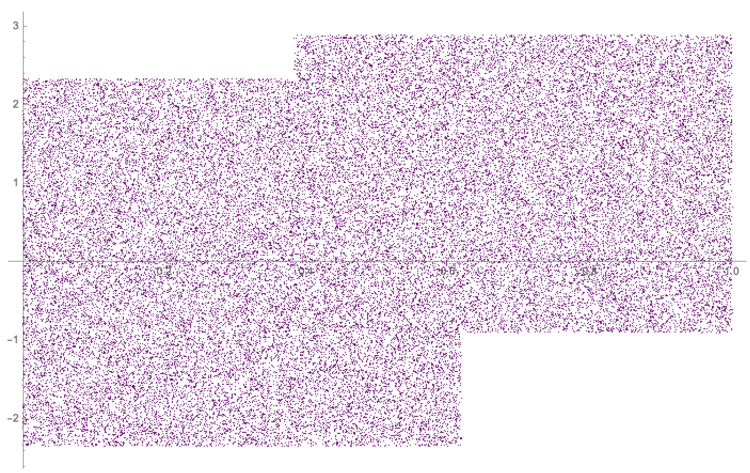


Figure 3.23: The limit set of the Galois lift of f_1. This glues to a sphere with 4 marked points, which then lifts to a torus. The lift of the pseudo-Anosov ψ_1 is a linear Anosov diffeomorphism of the torus with stretch factor $\lambda_1 = \frac{3+\sqrt{5}}{2}$, hence is conjugate to Arnold's cat map.

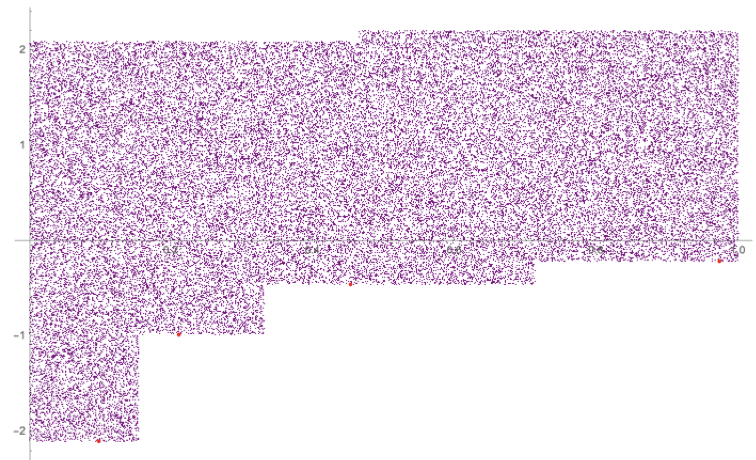


Figure 3.24: The limit set of the Galois lift of f_2. This glues to a sphere with 6 cone points of angle π, one at the center of each vertical edge. Taking the double cover of this sphere branched at the 6 cone points produces a genus 2 surface.

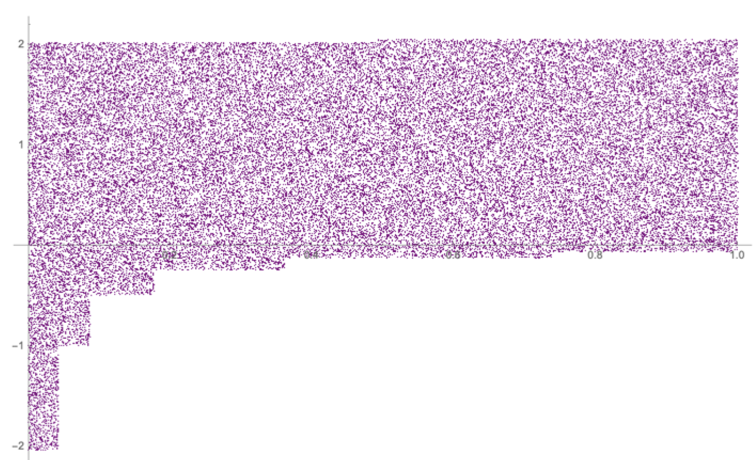


Figure 3.25: The limit set of the Galois lift of f_3. This glues to a sphere with 8 marked points, which then lifts to a genus 3 surface.

Chapter 4

A Farey tree structure on a family of pseudo-Anosovs

4.1 Introduction

The goal of this chapter is to further explore the structure of the families $\mathrm{PA}(m)$ defined in Chapter 3. Recall that if $f \in \mathrm{PA}(m)$ then f is an m-modal zig-zag map of pseudo-Anosov type. The action of f on the periodic orbit of $x = 1$ is particularly restricted, and allowed us to construct a dynamical parameterization $\Phi : \mathrm{PA}(m) \to \mathbb{Q} \cap (0, 1)$. The pseudo-Anosov ψ_f generated by f has stretch factor $\lambda(\psi_f) = \lambda(f)$, and the singularities of ψ_f correspond to dynamically significant points for f:

- There is a conjugacy between the action of ψ_f on its 1-prong singularities and the action of f on $\mathrm{PC}(f)$.
- The only other singularity of ψ_f corresponds to a particular periodic orbit of f, of length $|\,\mathrm{PC}(f)| - 2$. We will always denote this point by ∞, and it is necessarily a fixed point of ψ_f.

Our first main result presents a geometric interpretation of this parameterization. Given a pseudo-Anosov ψ fixing a b-prong singularity q, the *rotation number* of ψ at q is $\text{rot}_q(\psi) = \frac{a}{b}$ if ψ rotates the prongs at q by a spaces counterclockwise.

Theorem E. *Fix $f \in \text{PA}(m)$. Then $\Phi(f)$ measures the local clockwise rotation of ψ_f at the fixed singularity ∞: that is,*

$$\Phi(f) = 1 - \text{rot}_\infty(\psi_f).$$

André de Carvalho suggested to us the possibility of Theorem E in the context of his work with Toby Hall on generating pseudo-Anosovs from *unimodal maps*, i.e. PCF λ-expanders with a single critical point. Indeed, we would be remiss to not mention that our techniques in 3 and in the present chapter take great inspiration from those of de Carvalho and Hall in [dCH04]. We prove Theorem E in Section 4.3. While we are not currently aware of a proof of the analogous statement in the case of unimodal maps, we suspect that the statement holds and is amenable to a similar proof strategy.

Let us further contextualize Theorem E. Thickening an interval map $f : I \to I$ to a pseudo-Anosov $\psi_f : S \to S$ involves embedding I into S as an invariant train track τ for ψ_f. To do so, we add a loop of length 1 to I at each point in the forward orbit of a critical point of f, representing the fact that some iterate of f makes a turn at that point. We add a puncture to S within each loop to make it non-trivial in homology. In the language of train tracks, these loops are *infinitesimal edges*. The remaining edges of τ (i.e., the subintervals of I) are *expanding edges*, and produce Markov partitions for both f and ψ_f. In particular, f and ψ_f are both Bernoulli processes with the same transition matrix, and so the dynamics of f determine the dynamics of ψ_f almost entirely.

Thus, one may hope to transport the techniques of one-dimensional dynamics to study the pseudo-Anosovs that are thickenings of interval maps. More specifically, one desires to interpret invariants of f using invariants of ψ_f. Theorem E is a first step in this direction. Indeed, while $\Phi(f)$ is initially defined using the permutation action of f on the orbit of $x = 1$, we see that it has a new interpretation via the dynamics of ψ_f.

Additionally, it is often fruitful to study dynamical systems as members of a larger family. Interval maps come in natural families determined by variations in their *kneading data*. The maps in $\mathrm{PA}(m)$ constitute one such family (cf. Section **??**). If the interval maps in a given family thicken to pseudo-Anosovs, it is natural then to study the pseudo-Anosovs as a collection. This perspective provides an alternative to the standard ways of creating families of pseudo-Anosovs, such as

1. composing Dehn twists along a given set of curves according to a pattern, or
2. taking branched covers of pseudo-Anosovs.

Indeed, in case (1) the pseudo-Anosovs all act on the same surface S, while in case (2) the pseudo-Anosovs all have the same stretch factor. Varying the kneading data of the interval map generating a pseudo-Anosov, however, will generally change both the stretch factor of the pseudo-Anosov and the number of punctures of the surface on which it acts.

In this context, Theorem E proposes analyzing the pseudo-Anosovs generated by $\mathrm{PA}(m)$ using $\mathbb{Q} \cap (0,1)$. As we will see, this is the vertex set of a special graph $\mathcal{F}$ whose edges encode relations between pairs of maps in $\mathrm{PA}(m)$. We refer to $\mathcal{F}$ as the *Farey tree*, and in order to understand its structure we will make use of a second invariant of a map $f \in \mathrm{PA}(m)$: the *digit polynomial* $D_f(t)$. The digit polynomial has integral coefficients and satisfies $D_f(\lambda(f)) = 0$. Its coefficients encode the orbit of $x = 1$ under f, which is enough to determine the entire kneading data of f due to the simplicity of zig-zag maps.

By studying how D_f depends on the location of $\Phi(f)$ in $\mathcal{F}$, we obtain the following monotonicity result. Recall that $\lambda(f) = \lambda(\psi_f)$.

Theorem F. *Given $f, g \in \mathrm{PA}(m)$, we have*

$$\Phi(f) < \Phi(g) \iff \lambda(\psi_f) < \lambda(\psi_g).$$

Since the topological entropy of both f and ψ_f is $\log \lambda(\psi_f)$, one may interpret Theorem F as a monotonicity result for the entropy of the one-paramater family of pseudo-Anosovs generated by $\mathrm{PA}(m)$. Entropic monotonicity results exist for many natural families of interval maps (cf. for example [MT00] and [**?**]).

The proof of Theorem F is a chain of monotonicity statements. The key driver is Theorem 4.5.4, which presents a pair of transformation laws for the kneading data of $f = \Phi^{-1}(q)$ as $q \in \mathcal{F}$ varies. We treat $\mathcal{F}$ in Section 4.4 and describe how it models both $\mathrm{PA}(m)$ and $\Pi(m)$ in Section 4.5. We then prove Theorem F in Section 4.6.

Techniques from kneading theory also allow us to obtain best possible bounds on $\lambda(\psi_f)$.

Corollary 4.1.1. *Let $f \in \mathrm{PA}(m)$. Then,*

$$\inf\{\lambda(\psi_f) : f \in \mathrm{PA}(m)\} = \frac{m + 1 + \sqrt{(m+1)^2 - 8}}{2}$$

and

$$\sup\{\lambda(\psi_f) : f \in \mathrm{PA}(m)\} = m + 1.$$

Proof. Given $f \in \mathrm{PA}(m)$, $\lambda(f)$ is equal to the magnitude of the slope of the piecewise-linear map $f : I \to I$. The integer floor of $\lambda(f)$ is $\lfloor \lambda(f) \rfloor = m$. Thus, $\lambda(f)$ is bounded from below by m and from above by $m+1$.

For $n \geq 1$, define the maps $g_n, h_n \in \mathrm{PA}(m)$ by

$$g_n = \Phi^{-1}\left(\frac{1}{n}\right) \quad \text{and} \quad h_n = \Phi^{-1}\left(\frac{n-1}{n}\right).$$

Theorem F implies that the sequences $\{\lambda(g_n)\}_n$ and $\{\lambda(h_n)\}_n$ are monotonically decreasing and monotonically increasing, respectively. Since these sequences are bounded, it follows that they converge. Indeed, we must have

$$\lambda_0 := \lim_{n\to\infty} \lambda(g_n) = \inf\{\lambda(f) : f \in \mathrm{PA}(m)\}$$

and

$$\lambda_1 := \lim_{n\to\infty} \lambda(h_n) = \sup\{\lambda(f) : f \in \mathrm{PA}(m)\}.$$

The maps g_n converge uniformly to a λ_0-zig-zag map $\mathbf{0}_m$, and the maps h_n converge uniformly to a λ_1-zig-zag map $\mathbf{1}_m$ (cf. Definition 4.5.1). In the case of $\mathbf{1}_m$, it is not hard to see from the principal kneading sequence (cf. Definition 4.2.2) of h_n that $\mathbf{1}_m(1) = \lambda_1 - m$ is a fixed point of $\mathbf{1}_m$ in the subinterval $[m \cdot \lambda_1^{-1}, 1]$, on which $\mathbf{1}_m$ is defined by the linear map

$$\mathbf{1}_m(x) = \lambda_1 x - m \quad \text{for } x \in [m \cdot \lambda_1^{-1}, 1].$$

From this one determines that $\lambda_1 = m+1$. On the other hand, from the principal kneading sequence of g_n one sees that $\mathbf{0}_m(1) = \lambda_0 - m$ is a fixed point of $\mathbf{0}_m$ in the subinterval $[(m-2)\cdot\lambda_0^{-1}, (m-1)\cdot\lambda_0^{-1}]$. On this subinterval $\mathbf{0}_m$ is defined by the linear map

$$\mathbf{0}_m(x) = \lambda_0 x - (m-2) \quad \text{for } x \in [(m-2)\cdot\lambda_0^{-1}, (m-1)\cdot\lambda_0^{-1}].$$

Thus we find

$$\lambda_0 - m = \mathbf{0}_m(\lambda_0 - m) = \lambda_0^2 - m\lambda_0 - (m-2).$$

Solving for λ_0 gives

$$\lambda_0 = \frac{m+1+\sqrt{(m+1)^2-8}}{2}.$$

□

Just as Theorem E provides an interpretation of the first invariant $\Phi(f)$ from the perspective of ψ_f, we seek to understand the second invariant D_f using two- and three-dimensional techniques. Such an understanding provides immediate leverage over several topological invariants associated to ψ_f. Indeed, we will see that D_f will coincide or nearly coincide with these invariants.

The core idea in the following theorem is that, despite its original definition, D_f is the characteristic polynomial of several matrices naturally associated to f or to ψ_f.

Theorem G. *Let $f \in \mathrm{PA}(m)$ and set $n = |\,\mathrm{PC}(f)| = 1 + \deg(D_f)$. Set $\lambda = \lambda(f) = \lambda(\psi_f)$.*

1. *The digit polynomial determines the Artin-Mazur zeta function of f:*

$$\zeta_f(t) = \frac{1}{\mathcal{R}(D_f(t))} = \frac{1}{\mathcal{R}(\det(tI - W_f))}.$$

2. *The digit polynomial is equal to the strong Markov polynomial of f:*

$$D_f(t) = \chi_M(f;t).$$

3. *The digit polynomial determines the homology, symplectic, and puncture polynomials of ψ_f, defined by Birman-Brinkmann-Kawamuro ([BBK12]):*

$$D_f(t) = h(\psi_f;t) = \begin{cases} s(\psi_f;t) & \text{if } n \text{ is odd} \\ s(\psi_f;t)(t+1) & \text{if } n \text{ is even.} \end{cases}$$

4. *Let $\mathbb{S}$ be the surface obtained from the orientation double cover of ψ_f by filling in the lifts of the punctured 1-prong singularities of ψ_f. Denote by $\chi_+(t)$ (resp., $\chi_-(t)$) the characteristic polynomial of the lift ψ_+ (resp., ψ_-) acting on $H_1(\mathbb{S};\mathbb{Z})$. Then*

$$D_f(t) = \chi_+(t) \quad \text{and} \quad D_f(-t) = \chi_-(t).$$

5. *Let β_f be any n-braid representative of ψ_f obtained by ripping open $\infty \in S_{0,n+1}$ to a boundary circle. Let $\mathbb{B}(\beta_f, z)$ denote the reduced Burau matrix for β_f, and set $\chi(\beta_f;t) = \det(tI - \mathbb{B}(\beta_f, -1))$. Then*

$$\chi(\beta_f;t) = \begin{cases} D_f(t) & \text{if } \chi(\beta_f;t) \text{ has } \lambda \text{ as a root} \\ D_f(-t) & \text{if } \chi(\beta_f;t) \text{ has } -\lambda \text{ as a root.} \end{cases}$$

Note that since $D_f(t)$ is reciprocal, i.e. $\mathcal{R}(D_f(t)) = D_f(t)$, statement (1) implies that

$$\zeta_f(t) = \frac{1}{D_f(t)}.$$

We remark that any pseudo-Anosov that is the thickening of a PCF λ-expander is defined on a punctured sphere and has at most one singularity with 3 or more prongs. In a forthcoming paper with Karl Winsor we prove the converse statement: every such pseudo-Anosov is the thickening of a PCF λ-expander f. In this more general setting, the digit polynomial is replaced by the strong Markov polynomial $\chi_M(f;t)$, which is simply the characteristic polynomial of a minimal Markov partition associated to f.

Minor modifications to the proof of Theorem G will then prove analogous statements for the relevant invariants of ψ_f.

We will close with a brief discussion of these and other thoughts in Section 8, after proving Theorem G in Section 4.7.

4.2 Preliminaries

A central goal of the present paper is to argue that $\mathrm{PA}(m)$ in fact respects the combinatorial structure of a special graph $\mathcal{F}$ whose vertex set is $\mathbb{Q} \cap (0,1)$, the set paramterizing $\mathrm{PA}(m)$. We call this graph the *Farey tree*. See Section 4.4 for more.

From now on, we will use the notation $\mathcal{F}$ for $\mathbb{Q} \cap (0,1)$, in order to emphasize that we are endowing this set with the structure of the Farey tree.

We will spend most of this section synthesizing results from Chapter 3, before turning to the proof of Theorem E in Section 4.3.

4.2.1 The digit polynomial D_f

Let $f \in \mathrm{PA}(m)$ for $m \geq 2$ even, and write $\Phi(f) = \frac{a}{b}$. The point $x = 1$ is strictly periodic under f of minimal period $b+1$. Denote the elements of this orbit by $x_i = f^i(1)$ for $i = 0, \ldots, b+1$. Since f is a λ-zig-zag map, the x_i satisfy a collection of linear relations:

$$x_i = a_i \pm \lambda x_{i-1}, \qquad \text{for } i = 1, \ldots, b+1$$

where the indices are taken modulo $b+1$ and each a_i is an integer satisfying $|a_i| \leq m$. In fact, in Section 3.8 we proved

$$x_i = \begin{cases} \lambda x_{i-1} - m \text{ or } \lambda x_{i-1} - (m-2) & \text{if } 1 \leq i \leq b-1 \\ m - \lambda x_{i-1} & \text{if } i = b \\ 2 - \lambda x_{i-1} & \text{if } i = b+1. \end{cases} \tag{4.1}$$

Since $x_0 = x_{b+1} = 1$, composing these relations and solving for 0 produces a monic integral polynomial relation in λ, of the form

$$0 = D_f(\lambda) = \lambda^{b+1} + 1 - \sum_{i=1}^{b} c_i \lambda^{b+1-i}. \tag{4.2}$$

One checks that the c_i are integers satisfying

$$c_i = \begin{cases} \lambda x_{i-1} - x_i & \text{if } 1 \leq i \leq b-1 \\ \lambda x_{i-1} + x_i & \text{if } i = b \end{cases}. \tag{4.3}$$

If instead m is odd and $\Phi(f) = \frac{a}{b}$ then $x = 1$ is periodic of minimal period $b+2$. With notation as before, the corresponding restrictions are similar:

$$x_i = \begin{cases} \lambda x_{i-1} - m \text{ or } \lambda x_{i-1} - (m-2) & \text{if } 1 \leq i \leq b-1 \\ m - \lambda x_{i-1} & \text{if } i = b \\ \lambda x_{i-1} - 1 & \text{if } i = b+1 \\ 1 - \lambda x_{i-1} & \text{if } i = b+2. \end{cases} \tag{4.4}$$

Composing these relations and solving for 0 produces the equation

$$0 = \lambda D_f(\lambda),$$

where $D_f(\lambda)$ has the same form as in (4.2) and the c_i satisfy (4.3).

In either case, we make the following definition.

Definition 4.2.1. The polynomial $D_f(t) \in \mathbb{Z}[t]$ is the *digit polynomial* of f.

By definition, λ is a root of D_f. Less obvious is the fact that D_f is the characteristic polynomial of a matrix naturally associated to f, called the *strong Markov matrix*. This fact is crucial to the proof of Theorem G. See Section 4.7.2.

Definition 4.2.2. Let $f \in \mathrm{PA}(m)$ and write $\Phi(f) = \frac{a}{b}$. The *principal kneading sequence* $\nu(f)$ of f is the first period of $\mathrm{It}_f(1)$:

$$\nu(f) = \mathrm{Pre}_{b(m)}(\mathrm{It}_f(1)),$$

where

$$b(m) = \text{the minimal period of 1 under } f = \begin{cases} b+1 & \text{if } m \text{ is even} \\ b+2 & \text{if } m \text{ is odd.} \end{cases}$$

As with $\mathrm{It}_f(1)$, the sequence $\nu(f)$ is 0-indexed: for $0 \leq i \leq b(m)$,

$$\nu_i(f) = (\mathrm{It}_f(1))_i = A(f^i(x)).$$

The next proposition is immediate from Equations (4.1) and (4.4), and relates the coefficients of D_f to the entries of $\nu(f)$.

Proposition 4.2.3. *Let $f \in \mathrm{PA}(m)$ with digit polynomial*

$$D_f(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^{b+1-i}.$$

Then the principal kneading sequence of f is given by

$$\nu_i(f) = \begin{cases} c_{i+1} & \text{if } 0 \le i \le b-2 \\ m-1 & \text{if } i = b-1 \\ k_1 & \text{if } i = b \\ 0 & \text{if } m \text{ is odd and } i = b+1. \end{cases} \tag{4.5}$$

Together with Theorem C this statement implies the following simple description of the principle kneading sequence. We preserve notation.

Corollary 4.2.4. *The principal kneading sequence of $f \in \mathrm{PA}(m)$ satisfies*

$$\nu_i(f) = \begin{cases} m & \text{if } 0 \le i \le b-2 \text{ and } L(t) \in \mathbb{N} \text{ for some } t \in [i, i+1] \\ m-2 & \text{if } 0 \le i \le b-2 \text{ and } L(t) \notin \mathbb{N} \text{ for some } t \in [i, i+1] \\ m-1 & \text{if } i = b-1 \\ k_1 & \text{if } i = b \\ 0 & \text{if } m \text{ is odd and } i = b+1. \end{cases} \tag{4.6}$$

As these statements show, the following invariants of $f \in \mathrm{PA}(m)$ determine each other:

1. the fraction $\Phi(f) = \frac{a}{b}$,
2. the polynomial $D_f(t)$, and
3. the principal kneading sequence $\nu(f)$.

4.3 The proof of Theorem E

The definition of the map Φ assigning to the interval map $f \in \mathrm{PA}(m)$ a fraction in $\mathcal{F} = \mathbb{Q} \cap (0, 1)$ is opaque. In this section, we prove that $\Phi(f)$ has an interpretation from the perspective of the pseudo-Anosov ψ_f.

Theorem E. *Suppose* $f \in \mathrm{PA}(m)$*. Then*

$$\Phi(f) = 1 - \mathrm{rot}_\infty(\psi_f).$$

The proof of Theorem E proceeds as follows. We consider only the case when the number of critical points of f satisfies $m \geq 4$ even. The other two cases are completely analogous.

1. Show that the orbit on $\partial_H \Lambda_f$ has exactly one element on each of the connected components of the *lower horizontal boundary* $\partial_H^L \Lambda_f$.

2. Investigate the edge identifications of $\partial_H^L \Lambda_f$ in order to determine the counter-clockwise ordering of the prongs around ∞.

3. Show that ψ_f acts on the prongs by the permutation inverse to $\rho'(f)$ (cf. the notation in Proposition 3.7.7).

Proof. Cut the interval $I = [0,1]$ at the set $\mathrm{PC}(f) \setminus \{k_1\}$ to obtain the partition

$$I = \bigcup_{i=1}^{n-1} J_i.$$

Each Markov rectangle R_j has a horizontal boundary

$$\partial_H R_j = \partial_H^U R_j \sqcup \partial_H^L R_j,$$

where $\partial_H^L R_j$ is the *lower horizontal boundary of* R_j, i.e. the connected component of $\partial_H R_j$ that is at a lower height in $\mathbb{R}^2$. The *lower horizontal boundary of* Λ_f is the union

$$\partial_H^L \Lambda_f = \bigcup_{j=1}^{n} \partial_H^L R_j.$$

From our work in Section 3.5.1 we know that the connected components of $\partial_H^L \Lambda_f$ are precisely the sets

$$\widetilde{J}_i = \pi^{-1}(J_i) \cap \partial_H^L \Lambda_f, \text{ for } i = 1, \dots, n-1.$$

For each i the map $f_G^{-1} : \widetilde{J}_i \to f_G^{-1}(\widetilde{J}_i)$ is an affine contraction, with scaling factor λ^{-1}, and in fact is a homeomorphism onto its image. Moreover, we have

$$f_G^{-1}(\widetilde{J}_i) \subseteq \widetilde{J}_{\rho'(f)^{-1}(i)} = \widetilde{J}_{i-(n-k)}, \tag{4.7}$$

where $\Phi(f) = \frac{n-k}{n-1}$ and $\rho'(f)$ is the permutation ρ' in Proposition 3.7.7. Here we are interpreting the index $i - (n-k)$ modulo $n-1$.

It follows now that $f_G^{-(n-1)}$ sends each $\widetilde{J}_i$ into itself, implying the existence of an attracting fixed point $p_i \in \widetilde{J}_i$. The collection $\{p_1, \dots, p_{n-1}\}$ is necessarily the unique repelling orbit of f_G on $\partial_H \Lambda_f$. We have thus located the preimages of the singular point $\infty \in S^2$ under the quotient map that identifies segments of $\partial\Lambda_f$. It remains to understand how the neighborhoods of the points p_i are glued together.

We introduce a new piece of notation. Each p_i is contained in the interior of $\widetilde{J}_i$, since $\pi(\partial\widetilde{J}_i) \subseteq \mathrm{PC}(f)$ and $\pi(p_i) \notin \mathrm{PC}(f)$. Therefore, cutting $\widetilde{J}_i$ at p_i produces left and right subintervals $\widetilde{J}_i^l$ and $\widetilde{J}_i^r$, respectively.

We claim that for each index $i = 1, \dots, n-2$, the edge identifications glue $\widetilde{J}_i^r$ to $\widetilde{J}_{i+1}^l$ by an orientation-reversing isometry. Once we establish this fact, the conclusion of the theorem will follow. Indeed, the standard orientation on $\mathbb{R}^2$ descends to the orientation on S^2, and a positive turn around $\infty \in S^2$ lifts to a sequence of counterclockwise half-turns around the points p_i. Each such half-turn traverses the vertical line s_i emanating from the corresponding p_i, and s_i pushes down to a prong of the contracting foliation of ψ_f at ∞. Equation (4.7) implies that

$$f_G(p_i) = p_{i+(n-k)},$$

but since $\widetilde{J}_i^r$ is glued to $\widetilde{J}_{i+1}^l$ we require $n-1-(n-k) = k-1$ counterclockwise half-turns to reach $s_{i+(n-k)}$ from s_i. It follows that ψ_f rotates the prongs at ∞ by $k-1$ positive clicks, and hence

$$\mathrm{rot}_\infty(\psi_f) = \frac{k-1}{n-1} = 1 - \frac{n-k}{n-1} = 1 - \Phi(f).$$

It remains to show that $\widetilde{J}_i^r$ glues to $\widetilde{J}_{i+1}^l$ by an orientation-reversing isometry. This is a standard argument, and is part of a broader phenomenon described, for example, in Section 3.4 of [BH95], Bestvina and Handel's foundational paper on the subject of train tracks. In our case, one can see this by noting first that the segments of $\partial_H^U \Lambda_f$ adjacent to the unique periodic point in Λ_f projecting to $k_1 \in I$ are glued in this fashion. Applying f_G transports this gluing formation to all of the segments of $\partial_H^L \Lambda_f$. $\square$

4.4 The Farey tree

In this section we introduce the Farey tree and discuss its relevant properties. For treatments from the perspective of continued fractions, we recommend [Khi97]. For further reading on the Farey tree specifically, refer to [Hat22].

4.4.1 The Farey sum

All fractions $q = \frac{a}{b}$ are in lowest terms unless otherwise stated.

Definition 4.4.1. The *Farey sum* of two fractions $\frac{a}{b}$ and $\frac{c}{d}$ is the fraction

$$\frac{a}{b} \oplus \frac{c}{d} := \frac{a+c}{b+d},$$

where the fraction on the righthand side is not necessarily in lowest terms. Indeed, even if $\frac{a}{b}$ and $\frac{c}{d}$ are in lowest terms, their Farey sum need not be. If, however, $|ad - bc| = 1$, then it is not hard to show that $\frac{a+c}{b+d}$ is in lowest terms. For this reason, we introduce the following non-standard terminology.

Definition 4.4.2. We say that two fractions $\frac{a}{b}$ and $\frac{c}{d}$ are *compatible* if $|ad - bc| = 1$. Similarly, we say that two rationals p, q are *compatible* if they have compatible fractional representatives.

Note that if two fractions are compatible, then they are necessarily in lowest terms. From now on, we only consider the Farey sum of compatible fractions.

The next two propositions are exercises in arithmetic.

Proposition 4.4.3. *Let p, q be compatible rationals with $p < q$. Then $p < p \oplus q < q$.*

From now on, when we write $p \oplus q$ we will implicitly assume that $p < q$.

Proposition 4.4.4. *Let p, q be compatible rationals. Then p is compatible with $p \oplus q$, and $p \oplus q$ is compatible with q.*

4.4.2 Constructing the Farey tree

We are now ready to construct the Farey tree $\mathcal{F}$. We do so inductively, building a new level $\mathcal{F}_n$ on step at a time.

Set $\mathcal{F}_1 = \{\frac{1}{2}\}$, the root of $\mathcal{F}$. Suppose that the first $n-1$ levels of $\mathcal{F}$ have been constructed. The vertices in these levels are elements of $\mathbb{Q} \cap (0, 1)$: arrange them according to the usual linear order on $\mathbb{R}$, and include $\frac{0}{1}$ and $\frac{1}{1}$ on the far left and right, respectively. To construct $\mathcal{F}_n$, take the Farey sum of every element $q \in \mathcal{F}_{n-1}$ with each of its two neighbors. This produces left- and right children q_L, q_R of q, and the union of these children across all $q \in \mathcal{F}_{n-1}$ is defined to be $\mathcal{F}_n$. We furthermore include directed edges from each q to both of its children q_L and q_R.

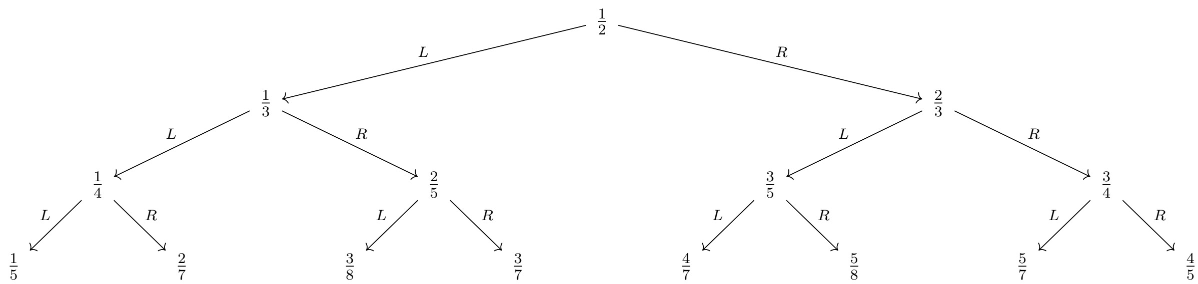


Figure 4.1: The first four levels of the Farey tree, with edges included.

Definition 4.4.5. The *Farey tree* is the directed graph $\mathcal{F}$ whose vertices are $\bigcup_{n\geq 1} \mathcal{F}_n$, and whose edges are precisely those described in the above procedure.

Remark 4.4.6. Some sources in the literature refer to $\mathcal{F}$ as the *Stern-Brocot tree*, while reserving the title of "Farey tree" for a larger tree whose vertices are $\mathbb{Q} \cup \{\infty\}$.

4.4.3 Properties of the Farey tree

A rational's position in $\mathcal{F}$ is intimately related to its *continued fraction expansion*. Indeed, one can use the theory of continued fractions to construct $\mathcal{F}$ and prove the following classical statement. See e.g. [Hat22] for more details.

Proposition 4.4.7. *The vertex set of $\mathcal{F}$ is precisely $\mathbb{Q} \cap (0,1)$. Moreover, for every vertex r there is exactly one pair of compatible rationals p, q such that $p \oplus q = r$.*

For this reason, we slightly abuse notation and write $\mathcal{F}$ for $\mathbb{Q} \cap (0,1)$. We will also make use of the larger set

$$\overline{\mathcal{F}} = \mathcal{F} \cup \left\{ \frac{0}{1}, \frac{1}{1} \right\}.$$

Definition 4.4.8. Given $p, q \in \mathcal{F}$, we say p is an *ancestor* of q if the directed path from $\frac{1}{2}$ to q contains p. We also say that q is a *descendant* of p.

Definition 4.4.9. Let $q \in \mathcal{F}_n$. The *left child* of q is the unique descendant $q_L$of q in $\mathcal{F}_{n+1}$ such that $q_L < q$. Similarly, the *right child* is the unique descendant q_R of q

in $\mathcal{F}_{n+1}$ such that $q < q_R$. The *left-* and *right parents* of q are the unique elements $q^L, q^R \in \overline{\mathcal{F}}$ such that $q^L \oplus q^R = q$.

Remark 4.4.10. The terminology of parents and children can be misleading at first. It is true that

$$(q_L)^R = q = (q_R)^L \text{ for all } q \in \mathcal{F}.$$

On the other hand, only one of $(q^L)_R$ and $(q^R)_L$ will be equal to q, since each $q \neq \frac{1}{2}$ has only one incoming edge. The other of the pair will be an ancestor of q.

Proposition 4.4.11. *Fix $q \in \mathcal{F}$. If $r \in \mathcal{F}$ satisfies $q^L < r < q^R$, then r is a descendant of q. In particular, the denominator of r is at least the denominator of q.*

Proof. If $r = q$ then there is nothing to show. Therefore, assume that $r \neq q$. Suppose $q \in \mathcal{F}_n$. According to the construction of $\mathcal{F}$, q is directly adjacent to its parents q^L and q^R in the set $\bigcup_{k \leq n} \mathcal{F}_k$. In particular, a rational r between q and q^L can only appear in $\mathcal{F}$ after taking the mediant of q and q^L, making r a descendant of q. Similarly, if $r \in \mathcal{F}$ is between q and q^R, then it can only appear after taking the mediant of q and q^R, making it once again a descendant of q. $\square$

Proposition 4.4.12. *Let $q \in \mathcal{F}_n$ for $n \geq 2$. Then exactly one of q^L and q^R is an element of $\mathcal{F}_{n-1}$. Moreover,*

1. *if $q^L \in \mathcal{F}_{n-1}$, then $q^R = (q^L)^R$, and*
2. *if $q^R \in \mathcal{F}_{n-1}$, then $q^L = (q^R)^L$.*

Proof. In the construction of $\mathcal{F}$, the two parents of $q \in \mathcal{F}_n$ are adjacent rationals in the first $n-1$ levels of $\mathcal{F}$. Thus, one of them must be the parent of the other. On the other hand, there is an edge of $\mathcal{F}$ pointing from some $r \in \mathcal{F}_{n-1}$ to q: this r must be a parent of q. Denote by s the other parent of q. As we noted above, s must be a parent of r: if $r = q^L$, then $r < q < s$ and it follows that $s = q^R$ and $s = r^R$. Similarly, if $r = q^R$ then $s < q < r$ and we have $s = q^L = r^L$. □

Definition 4.4.13. For $q \in \mathcal{F}$, we define $\mathcal{F}(q)$ to be the subtree of $\mathcal{F}$ with q as its root. Equivalently, $\mathcal{F}(q)$ is the collection of descendants of q, along with q itself.

Note that for each q there is a natural graph isomorphism from $\mathcal{F}(q)$ to $\mathcal{F}$ taking q to $\frac{1}{2}$ and preserving the order on $\mathbb{R}$. Thus we can speak of the levels $\mathcal{F}_n(q)$ of the subtree $\mathcal{F}(q)$.

Proposition 4.4.14. *Fix $q \in \mathcal{F}$. For any $s \in \mathcal{F}(q_L)$ (resp., $s \in \mathcal{F}(q_R)$) there exists a sequence*

$$s = s_0, \ldots, s_k = q$$

such that $s_i = (s_{i-1})^R$ for $i = 1, \ldots, k$ (resp., $s_i = (s_{i-1})^L$).

Proof. We consider only the case $s \in \mathcal{F}(q_L)$, since the other is analogous. We proceed by induction on the level n of $\mathcal{F}(q_L)$ containing s. If $n = 1$, then $s = q_L$ and the sequence

$$s_0 = q_L, s_1 = q$$

is as desired. Now suppose that the proposition has been proven for all s in the first $n-1$ levels of $\mathcal{F}(q_L)$, and let $s \in \mathcal{F}_n(q_L)$. By Proposition 4.4.12 exactly one of s^L and s^R is in $\mathcal{F}_{n-1}(q_L)$. If $s^R \in \mathcal{F}_{n-1}(q_L)$ then we are done by the inductive hypothesis: simply prepend s to the sequence of right parents connecting s^R to q. If instead $s^L \in \mathcal{F}_{n-1}(q_L)$, Proposition 4.4.12 says that $s^R = (s^L)^R$. The inductive hypothesis implies that we have a sequence of right parents connecting s^L to q, say

$$s^L, s^R, \ldots, q.$$

Then the desired sequence for s is identical, after substituting in s for s^L:

$$s, s^R, \ldots, q.$$

The proof is complete. $\square$

4.5 The Farey tree as a model for PA(m) and $\Pi(m)$

The vertices of $\mathcal{F}$ successively generate further vertices using the Farey sum operation. We transport this structure to both PA(m) and the family $\Pi(m)$ of pseudo-Anosovs they generate and extract dynamical consequences. The most important results from this section are Theorem 4.5.4 and Proposition 4.5.10.

4.5.1 The model for PA(m)

We derive a transformational law for the principal kneading sequence $\nu(f)$ as we travel down $\mathcal{F}$. Before stating this result, however, we make a few definitions.

Definition 4.5.1. Fix $m \geq 2$. We will denote by $\mathbf{0}_m$ the m-modal zig-zag map such that

1. the itinerary of $x = 1$ under $\mathbf{0}_m$ is $(m) \cdot (m-2)^\infty$, and

2. $\mathbf{0}_m$ is decreasing on I_{m-1}.

We also define $\mathbf{1}_m$ to be the m-modal zig-zag map such that

1. the itinerary of $x = 1$ under $\mathbf{1}_m$ is $(m)^\infty$, and

2. $\mathbf{1}_m$ is decreasing on I_{m-1}.

Set $\overline{\mathrm{PA}(m)} = \mathrm{PA}(m) \cup \{\mathbf{0}_m, \mathbf{1}_m\}$. We extend the map Φ to $\overline{\mathrm{PA}(m)}$ by declaring

$$\Phi(\mathbf{0}_m) = \frac{0}{1} \quad \text{and} \quad \Phi(\mathbf{1}_m) = \frac{1}{1}.$$

Definition 4.5.2. Fix $m \geq 2$. We say that $f, g \in \overline{\mathrm{PA}(m)}$ are *compatible* if $\Phi(f), \Phi(g) \in \overline{\mathcal{F}}$ are compatible. In this case, we define $f \oplus g \in \mathrm{PA}(m)$ to be the interval map such that

$$\Phi(f \oplus g) = \Phi(f) \oplus \Phi(g).$$

We also borrow the language of parents and children. If $f = \Phi^{-1}(q)$, then we define the following elements of $\overline{\mathrm{PA}(m)}$:

- $f^L = \Phi^{-1}(q^L)$,
- $f^R = \Phi^{-1}(q^R)$,
- $f_L = \Phi^{-1}(q_L)$, and
- $f_R = \Phi^{-1}(q_R)$.

As Corollary 4.2.4 shows, given $f \in \mathrm{PA}(m)$ we may write the principle kneading sequence $\nu(f)$ as the concatenation of sequences

$$\nu(f) = (m) \cdot \mathbf{w}(f) \cdot \mathbf{k}, \quad \text{where} \quad \mathbf{k} = \begin{cases} (m-1, k_1) & \text{if } m \text{ is even} \\ (m-1, k_1, 0) & \text{if } m \text{ is odd.} \end{cases} \tag{4.8}$$

Here $\mathbf{w}(f)$ is defined implicitly, and is the part of $\nu(f)$ determined by the intersection of a line segment of slope $\Phi(f)$ with horizontal and vertical integer lines in $\mathbb{R}^2$.

Definition 4.5.3. Fix $f \in \mathrm{PA}(m)$. We define the following three sequences by adusting the prefix or suffix of $\nu(f)$:

1. $\overline{\nu(f)} = (m) \cdot \mathbf{w}(f) \cdot (m)$
2. $\widehat{\nu(f)} = (m) \cdot \mathbf{w}(f) \cdot (m-2)$

3. $\underline{\nu(f)} = (m-2) \cdot \mathbf{w}(f) \cdot \mathbf{k}$.

Here are the transformation rules for $\nu(f)$.

Theorem 4.5.4. *Let $f, g \in \mathrm{PA}(m)$ be compatible with $\Phi(f) < \Phi(g)$. Then*

$$\nu(f \oplus g) = \overline{\nu(f)} \cdot \underline{\nu(g)}, \tag{4.9}$$

and

$$\nu(f \oplus g) = \widehat{\nu(g)} \cdot \nu(f). \tag{4.10}$$

Proof. Write $\Phi(f) = \frac{a}{b}$ and $\Phi(g) = \frac{c}{d}$. Consider the broken line segment parameterized as

$$\mathcal{L}(t) = \begin{cases} (t, \Phi(f) \cdot t) & \text{if } t \in [0, b] \\ (t, a + \Phi(g) \cdot (t - b)) & \text{if } t \in [b, b+d]. \end{cases}$$

This broken line consists of two straight line segments, one of slope $\Phi(f)$ and the second of slope $\Phi(g)$. The straight line segment $\mathcal{L}_q$ shares its endpoints with $\mathcal{L}$, but lies above it in the plane. Consider the isotopy $\mathcal{H}$ from $\mathcal{L}$ to $\mathcal{L}_q$ defined by dragging the non-smooth point of $\mathcal{L}$ vertically towards $\mathcal{L}_q$, while keeping this point connected by line segments to the endpoints $(0,0)$ and $(b+d, a+c)$. Pick a parameterization of this isotopy by $s \in [0,1]$ such that $\mathcal{H}_0(\mathcal{L}) = \mathcal{L}$ and $\mathcal{H}_1(\mathcal{L}) = \mathcal{L}_q$.

We can define the *kneading sequence* $\nu(s)$ of $\mathcal{H}_s(\mathcal{L})$ using the intersections with horizontal and vertical integer lines, going from left to right. Explicitly, we declare the following algorithm:

1. Initialize $\nu(s) = (m)$.

2. For each $1 \leq i \leq b + d - 2$:

(a) if the intersection of $\mathcal{H}_s(\mathcal{L})$ with $x = i + 1$ is an integer point, append $\mathbf{k}$ to $\nu(s)$;

(b) else if the intersection of $\mathcal{H}_s(\mathcal{L})$ with $x = i$ is an integer point, append m to $\nu(s)$;

(c) else if there is a horizontal intersection strictly between $x = i$ and $x = i+1$, append m to $\nu(s)$;

(d) else, append m to $\nu(s)$;

3. Append $\mathbf{k}$ to $\nu(s)$.

According to this algorithm, the initial and terminal kneading sequences of the isotopy satisfy

$$\nu(0) = \nu(f) \cdot \nu(g),$$
$$\nu(1) = \nu(f \oplus g).$$

Moreover, the value of $\nu(s)$ only changes when one of the line segments shifts across an integer point $(x, y) \in \mathbb{Z}^2$. Indeed, suppose the left line segment of $\mathcal{H}_{s_0}(\mathcal{L})$ passes through the integer point $(x, y) \neq (0, 0)$, for some $s_0 > 0$. Then this left line segment has slope y/x, which is equal to some rational r with denominator dividing x and therefore at most b, since $x \leq b$. But this rational r is in the interval $[\Phi(f), \Phi(g)]$, so Proposition 4.4.11 implies that x is at least as large as the denominator of $\Phi(f) \oplus \Phi(g)$, which is $b + d$. Since $d \geq 1$, we have a contradiction: the left line segment of $\mathcal{H}_{s_0}(\mathcal{L})$ cannot pass through a second integer point besides $(0, 0)$. An identical argument proves that the right line segment of $\mathcal{H}_{s_0}(\mathcal{L})$ cannot pass through a second integer point besides $(a + c, b + d)$.

Therefore $\nu(s) = \nu(f \oplus g)$ for all $s > 0$. The initial perturbation of $\mathcal{H}_0(\mathcal{L})$ replaces the suffix $\mathbf{k}$ of f with (m) and the prefix m of $\nu(g)$ with $m-2$, verifying (4.9). See Figure 4.3b.

To prove (4.10) we proceed by an identical argument, except that we define $\mathcal{L}$ to be the broken line whose first segment has slope $\Phi(g)$ and whose second segment has slope $\Phi(f)$. In this case, the initial isotopy replaces the suffix $\mathbf{k}$ of $\nu(g)$ with $(m-2)$, while it does not alter the prefix (m) of $\nu(f)$. See Figure 4.3a. □

Remark 4.5.5. The formulas are slightly different if $f = \mathbf{0}_m$ or $g = \mathbf{1}_m$. In the first case, $g = \Phi^{-1}(1/n)$ for some n, and so $\mathbf{w}(g) = (m-2)^{n-2}$. It is not hard to see then that

$$\mathbf{w}(\mathbf{0}_m \oplus g) = (m-2)^{n-1}.$$

In the second case, $f = \Phi^{-1}(1-1/n)$ and $\mathbf{w}(f) = (m)^{n-2}$. Therefore,

$$\mathbf{w}(f \oplus \mathbf{1}_m) = (m-1)^{n-1}.$$

From now on, when we write $\nu(q)$ for $q \in \mathcal{F}$ we will mean $\nu\left(\Phi^{-1}(q)\right)$. Theorem 4.5.4 allows us to compute $\nu(q)$ as we travel along $\mathcal{F}$ from $\frac{1}{2}$ to q.

Example 4.5.6. We compute $\nu(7/12)$ step by step. The path in $\mathcal{F}$ from $1/2$ to $7/12$ is pictured in Figure 4.2. Dotted arrows depict a parent-child relationship that is not an edge of $\mathcal{F}$.

We begin with $\nu(1/2) = (m) \cdot \mathbf{k}$. Remark 4.5.5 tells us that

$$\nu\left(\frac{2}{3}\right) = (m) \cdot (m) \cdot \mathbf{k}.$$

The left child of $2/3$ is $3/5$, which has left parent $1/2$, hence by (4.10)

$$\nu\left(\frac{3}{5}\right) = \widehat{\nu\left(\frac{2}{3}\right)} \cdot \nu\left(\frac{1}{2}\right) = (m) \cdot (m, m-2, m) \cdot \mathbf{k}.$$

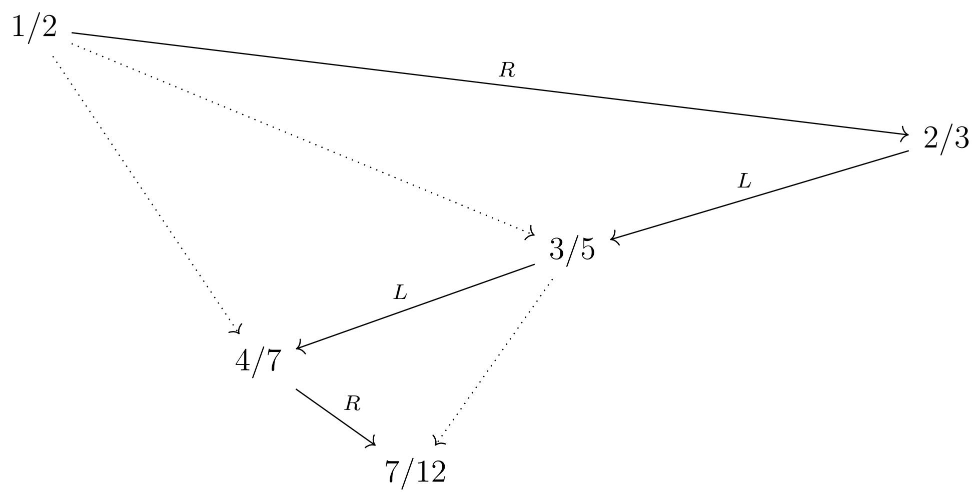


Figure 4.2: The unique path in the Farey tree from the root 1/2 to 7/12. Dotted arrows depict a parent-child relationship that is not an edge of $\mathcal{F}$.

Similarly, the left child of 3/5 is 4/7, whose left parent is again 1/2, giving

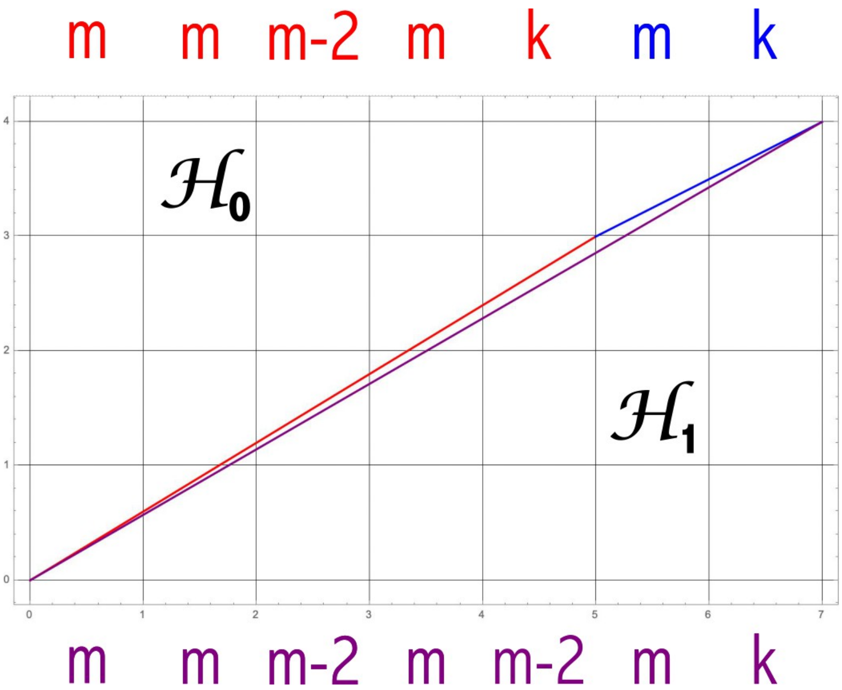
$$\nu\left(\frac{4}{7}\right) = \widehat{\nu\left(\frac{3}{5}\right)} \cdot \nu\left(\frac{1}{2}\right) = (m) \cdot (m, m-2, m, m-2, m) \cdot \mathbf{k}.$$

See Figure 4.3a. Finally we arrive at 7/12, whose left- and right parents are 4/7 and 3/5, respectively. Applying (4.9) gives

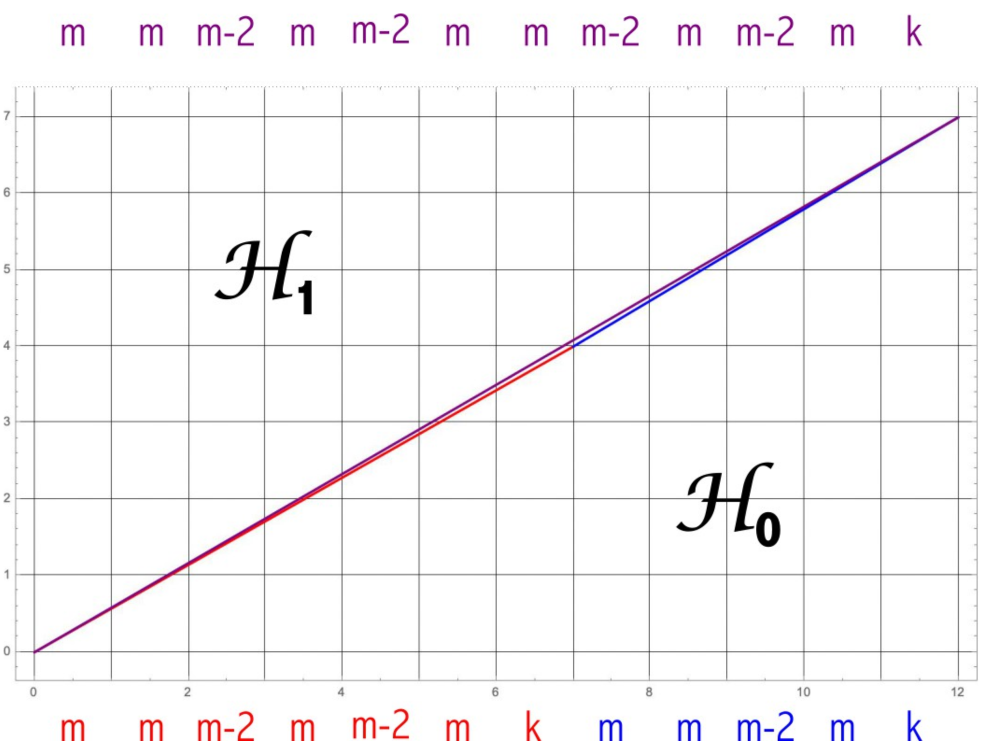
$$\begin{aligned} \nu\left(\frac{7}{12}\right) = \overline{\nu\left(\frac{4}{7}\right)} \cdot \underline{\nu\left(\frac{3}{5}\right)} &= (m) \cdot (m, m-2, m, m-2, m, m) \cdot (m-2, m, m-2, m) \cdot \mathbf{k} \\ &= (m) \cdot (m, m-2, m, m-2, m, m, m-2, m, m-2, m) \cdot \mathbf{k}. \end{aligned}$$

See Figure 4.3b.

Remark 4.5.7. Before moving on to $\Pi(m)$, we pause to reflect on the ramifications of Theorem 4.5.4 and Remark 4.5.5. We may interpret these as saying that making a turn in $\mathcal{F}$ at $\Phi(f)$ amounts to *perturbing* the point $f^{b-1}(1) \in I_{m-1}$. Indeed, suppose that $\Phi(f) = \frac{a}{b}$. Then for any left child g and any right child h we have



(a) Computing $\nu(4/7)$ from $\nu(3/5)$ and $\nu(1/2)$.



(b) Computing $\nu(7/12)$ from $\nu(4/7)$ and $\nu(3/5)$.

Figure 4.3: The last two steps in computing $\nu(7/12)$.

$$
\begin{aligned}
\mathrm{Pre}_b(\mathrm{It}_g(1)) &= \widehat{\nu(f)} &&= (m)\cdot \mathbf{w}(f)\cdot(m-2)\\
\mathrm{Pre}_b(\mathrm{It}_f(1)) &= \mathrm{Pre}_b(\nu(f)) &&= (m)\cdot \mathbf{w}(f)\cdot(m-1)\\
\mathrm{Pre}_b(\mathrm{It}_h(1)) &= \overline{\nu(f)} &&= (m)\cdot \mathbf{w}(f)\cdot(m)
\end{aligned}
\tag{4.11}
$$

Descending to the left (resp. right) in $\mathcal{F}$ perturbs $f^{b-1}(1)$ to the left (resp., right). As we will see in Section 4.6, it is this behavior that causes the stretch factor $\lambda(f)$ to change monotonically with $\Phi(f)$.

4.5.2 The model for $\Pi(m)$

Theorem 4.5.4 and Remark 4.5.5 have implications for how the Markov partition and transition matrix of $f = \Phi^{-1}(q) \in \mathrm{PA}(m)$ transform as $q \in \mathcal{F}$ varies. These transformation laws are reincarnated in the Markov partition and transition matrix of $\psi(f) \in \Pi(m)$. For brevity, however, we spend this subsection describing a dynamico-topological transformation law that is more meaningful from the perspective of $\Pi(m)$.

Definition 4.5.8. If $f, g \in \mathrm{PA}(m)$ are compatible, then we say that the pseudo-Anosovs $\psi_f, \psi_g \in \Pi(m)$ are *compatible*, as well. In this case, define $\psi_f \oplus \psi_g$ to be the element $\Psi(f \oplus g) \in \Pi(m)$.

Definition 4.5.9. Given $f \in \mathrm{PA}(m)$, denote by $n_\infty(\psi_f)$ the number of prongs of the invariant foliations of ψ_f at the singularity ∞.

Note that if $\Phi(f) = \frac{a}{b}$ then $n_\infty(\psi_f) = b$.

Proposition 4.5.10. *Fix $m \geq 2$, and let $\phi, \psi \in \Pi(m)$ be compatible. Then*

$$\mathrm{rot}_\infty(\phi \oplus \psi) = \mathrm{rot}_\infty(\phi) \oplus \mathrm{rot}_\infty(\psi). \tag{4.12}$$

Moreover,

$$n_\infty(\phi \oplus \psi) = n_\infty(\phi) + n_\infty(\psi). \tag{4.13}$$

Proof. These relations follow directly from our definitions and Theorem E. Let $f, g \in \mathrm{PA}(m)$ be the compatible maps generating ϕ and ψ, respectively. Write $\Phi(f) = \frac{a}{b}$ and $\Phi(g) = \frac{c}{d}$. Then

$$\begin{aligned}
\mathrm{rot}_\infty(\phi \oplus \psi) &= 1 - \left(\frac{a}{b} \oplus \frac{c}{d}\right) \\
&= \frac{b + d - (a + c)}{b + d} \\
&= \frac{(b - a) + (d - c)}{b + d} \\
&= \frac{b - a}{b} \oplus \frac{d - c}{d} \\
&= \left(1 - \frac{a}{b}\right) \oplus \left(1 - \frac{c}{d}\right) \\
&= \mathrm{rot}_\infty(\phi) \oplus \mathrm{rot}_\infty(\psi).
\end{aligned}$$

This proves Equation (4.12). Since for any $h \in \mathrm{PA}(m)$ the fraction $\Phi(h)$ is always in lowest terms, and since the number of prongs of ψ_h at infinity is equal to the denominator of $\Phi(h)$, Equation (4.13) now follows. □

4.6 Monotonicity of entropy

Recall that the topological entropy of a λ-zig-zag map f is $\mathrm{h}(f) = \log(\lambda)$. In the study of interval dynamics, many one-parameter families of maps have been shown to exhibit *monotonicity of entropy*: that is, entropy monotonically increases or decreases with the parameter. In this section we prove Theorem F, which can be viewed as another example of this phenomenon. Here, the parameter is the anti-rotation number $\Phi : \mathrm{PA}(m) \to \mathcal{F}$.

Theorem F. *Let $f, g \in \mathrm{PA}(m)$. Then*

$$\Phi(f) < \Phi(g) \iff \lambda(\psi_f) < \lambda(\psi_g).$$

Proof. The result follows from the chain of equivalences:

$$\begin{aligned}
\Phi(f) < \Phi(g) &\iff \mathrm{It}_f(1) <_E \mathrm{It}_g(1) && (4.6.4)\\
&\iff \mathcal{K}_f \ll \mathcal{K}_g && (4.6.5)\\
&\iff \lambda(f) < \lambda(g) && (4.6.6)\\
&\iff \lambda(\psi_f) < \lambda(\psi_g).
\end{aligned}$$

The last equivalence is because $\lambda(f) = \lambda(\psi_f)$. □

In Section 4.6.1 we introduce the *twisted lexicographic order* $\leq_E$, and in Section 4.6.2 we prove Proposition 4.6.4 and Corollary 4.6.6.

4.6.1 Ordering kneading sequences

In this section we prove a few basic facts about the sign function and cumulative sign vector of $f \in \mathrm{PA}(m)$ (cf. Section 2.4).

Lemma 4.6.1. *Suppose $f \in \mathrm{PA}(m)$ with sign vector E. Then for $0 \leq j \leq m$,*

$$E(j) = \begin{cases} (-1)^j & \textit{if } m \textit{ is even} \\ (-1)^{j+1} & \textit{if } m \textit{ is odd.} \end{cases}$$

In particular, $E(m) = E(m-2) = +1$ always.

Proof. This follows directly from Corollary 3.6.6 and Definition 3.6.8. If $f \in \mathrm{PA}(m)$, then f is positive if and only if m is even. □

Lemma 4.6.2. *Suppose $f \in \mathrm{PA}(m)$ and write $\Phi(f) = \frac{a}{b}$. Then the cumulative sign vector of $\mathrm{It}_f(1)$ satisfies*

$$s_i(\mathrm{It}_f(1)) = +1 \text{ for all } 0 \leq i \leq b-1. \tag{4.14}$$

Proof. From Corollary 4.2.4 we know that the entries of the principal kneading sequence of f satisfy

$$\nu_i(f) \in \{m-2, m\} \text{ for all } 0 \leq i \leq b-2.$$

As Lemma 4.6.1 shows, $E(m) = E(m-2) = +1$ for all values of m. The proof is complete. □

4.6.2 Completing the proof of Theorem F

In this section we synthesize our work from Sections 4.4 and 4.5 to prove that the stretch factor $\lambda(f)$ grows monotonically in the anti-rotation number $\Phi(f)$.

Proposition 4.6.3. *For all $f \in \mathrm{PA}(m)$ we have*

$$\mathrm{It}_{f^L}(1) <_E \mathrm{It}_f(1) <_E \mathrm{It}_{f^R}(1).$$

Proof. In the case that both f^L and f^R are in $\mathcal{F}$, these inequalities follow directly from Equations (4.11) and (4.14). For the special case where $f^L = \mathbf{0}_m$ or $f^R = \mathbf{1}_m$, we instead appeal to the formulas in Remark 4.5.5. □

We are now ready to compare the kneading data of elements of $\mathrm{PA}(m)$.

Proposition 4.6.4. *Fix $m \geq 2$, and let f and g be distinct elements of $\mathrm{PA}(m)$. Then*

$$\Phi(f) < \Phi(g) \iff \mathrm{It}_f(1) <_E \mathrm{It}_g(1).$$

Proof. Let r be the latest common ancestor of $\Phi(f)$ and $\Phi(g)$, and set $h = \Phi^{-1}(r) \in \mathrm{PA}(m)$.

Suppose that $\Phi(f) < \Phi(g)$. In the general case that r is distinct from $\Phi(f)$ and $\Phi(g)$, we have $\Phi(f) \in \mathcal{F}(r_L)$, implying by Proposition 4.4.14 that there exists a sequence of right parents

$$s_0 = \Phi(f), s_1, \ldots, s_k = r$$

connecting $\Phi(f)$ to r. Setting $f_i = \Phi^{-1}(s_i)$, we have $f_i^R = f_{i+1}$ for $i = 0, \ldots, k-1$. Proposition 4.6.3 now implies that

$$\mathrm{It}_f(f) <_E \mathrm{It}_{f_1}(1) <_E \cdots <_E \mathrm{It}_h(1).$$

A similar argument involving left parents shows that $\mathrm{It}_h(1) <_E \mathrm{It}_g(1)$.

In the special case that $r = \Phi(f)$ or $\Phi(g)$, we have $f = h$ or $g = h$, respectively. From here we apply the same argument to obtain the same conclusion.

Suppose instead that $\Phi(f) \geq \Phi(g)$. Then in fact $\Phi(f) > \Phi(g)$, and repeating our above argument shows that $\mathrm{It}_f(1) >_E \mathrm{It}_g(1)$. □

Proposition 4.6.5. *For any $m \geq 2$, let f and g be distinct elements of* $\mathrm{PA}(m)$. *Then*

$$\mathcal{K}_f \ll \mathcal{K}_g \iff \mathrm{It}_f(1) <_E \mathrm{It}_g(1).$$

Proof. Suppose first that m is even. Then $E(j) = (-1)^j$, so the conditions in (2.1) become

$$\mathcal{K}_f \ll \mathcal{K}_g \iff \begin{cases} \mathrm{It}_f(1) \leq_E \mathrm{It}_g(1) & \text{for } j \text{ odd, and} \\ 0^\infty \geq_E 0^\infty & \text{for } j \text{ even.} \end{cases}$$

Since f and g are distinct elements of $\mathrm{PA}(m)$ they cannot have the same itinerary, and so the conclusion holds in this case.

Suppose instead that m is odd. Then $E(j) = (-1)^{j+1}$, so the conditions in (2.1) are now

$$\mathcal{K}_f \ll \mathcal{K}_g \iff \begin{cases} \mathrm{It}_f(1) \leq_E \mathrm{It}_g(1) & \text{for } j \text{ odd, and} \\ 0 \cdot \mathrm{It}_f(1) \geq_E 0 \cdot \mathrm{It}_g(1) & \text{for } j \text{ even.} \end{cases}$$

Since $E(0) = -1$, the conclusion again holds. $\square$

Corollary 4.6.6. *For any $m \geq 2$, let f and g be distinct elements of* $\mathrm{PA}(m)$*. Then*

$$\lambda(f) < \lambda(g) \iff \mathrm{It}_f(1) <_E \mathrm{It}_g(1).$$

Proof. If $\mathrm{It}_f(1) <_E \mathrm{It}_g(1)$, then $\mathcal{K}_f \ll \mathcal{K}_g$ by Proposition 4.6.5, hence by Proposition 2.4.8 we have

$$\log \lambda(f) = \mathrm{h}(f) \leq h(g) = \log \lambda(g).$$

This inequality must in fact be strict, since elements of $\mathrm{PA}(m)$ have distinct slopes and hence distinct entropies. If $\mathrm{It}_f(1) \geq_E \mathrm{It}_g(1)$, then in fact $\mathrm{It}_f(1) > \mathrm{It}_g(1)$, and the same arguments imply $\lambda(f) > \lambda(g)$. $\square$

4.7 The digit polynomial as other invariants

We conclude by showing that the digit polynomial of a map $f \in \mathrm{PA}(m)$ recovers several well-known invariants. These include:

1. the *strong Markov polynomial* $\chi_M(f;t)$ of f (Section 4.7.2),
2. the *Artin-Mazur zeta function* $\zeta_f(t)$ of f (Section 4.7.2),
3. the *homology and symplectic polynomials* of ψ_f (Section 4.7.3), and

4. the *symplectic Burau polynomial* $\chi_\beta(-1,t)$ of any braid representative β for ψ_f (Section 4.7.5).

Importantly, all of these polynomials are essentially the characteristic polynomial of the strong Markov matrix for f, which may be understood as the transition matrix for ψ_f acting on a particular invariant train track.

Theorem G. *Let $f \in \mathrm{PA}(m)$ and set $n = |\,\mathrm{PC}(f)| = 1 + \deg(D_f)$. Set $\lambda = \lambda(f) = \lambda(\psi_f)$.*

1. *The digit polynomial determines the Artin-Mazur zeta function of f:*

$$\zeta_f(t) = \frac{1}{\mathcal{R}(D_f(t))} = \frac{1}{\mathcal{R}(\det(tI - W_f))}.$$

2. *The digit polynomial is equal to the strong Markov polynomial of f:*

$$D_f(t) = \chi_M(f;t).$$

3. *The digit polynomial determines the homology, symplectic, and puncture polynomials of ψ_f:*

$$D_f(t) = h(\psi_f;t) = \begin{cases} s(\psi_f;t) & \text{if } n \text{ is odd} \\ s(\psi_f;t)(t+1) & \text{if } n \text{ is even.} \end{cases}$$

4. *Let $\mathbb{S}$ be the surface obtained from the orientation cover of ψ_f by filling in the lifts of the punctured 1-prong singularities of ψ_f. Denote by $\chi_+(t)$ (resp., $\chi_-(t)$) the characteristic polynomial of ψ_+ (resp., ψ_-) acting on $H_1(\mathbb{S};\mathbb{Z})$. Then*

$$D_f(t) = \chi_+(t) \quad \text{and} \quad D_f(-t) = \chi_-(t).$$

5. *Let β_f be any n-braid representative of ψ_f obtained by ripping open $\infty \in S_{0,n+1}$ to a boundary circle. Let $\mathbb{B}(\beta_f, z)$ denote the reduced Burau matrix for β_f, and set $\chi(\beta_f;t) = \det(tI - \mathbb{B}(\beta_f, -1))$. Then*

$$\chi(\beta_f;t) = \begin{cases} D_f(t) & \textit{if } \chi(\beta_f;t) \textit{ has } \lambda \textit{ as a root} \\ D_f(-t) & \textit{if } \chi(\beta_f;t) \textit{ has } -\lambda \textit{ as a root.} \end{cases}$$

4.7.1 The reverse of a polynomial

The *reverse* of a polynomial $p(t) \in \mathbb{C}[t]$ is

$$\mathcal{R}(p(t)) = t^{\deg(p)} p(t^{-1}).$$

The polynomial $p(t)$ is *reciprocal* or *symmetric* if $\mathcal{R}(p(t)) = p(t)$.

Remark 4.7.1. Note that $D_f(t)$ is reciprocal, by Theorem C.

We will need the following fact.

Proposition 4.7.2. *For any $f, g \in \mathbb{C}[t]$ we have*

$$\mathcal{R}(f(t)) = \mathcal{R}(g(t)) \iff f(t) = t^{\deg(f)-\deg(g)} g(t).$$

4.7.2 Zeta functions of zig-zag maps

In this section we prove statements (1) and (2) of Theorem G.

Suppose that $f : X \to X$ is a dynamical system such that the sets

$$\mathrm{Fix}(f^i) = \{x \in X : f^i(x) = x\}$$

are finite for each $i \geq 1$. Then the *Artin-Mazur zeta function* of f is the formal power series

$$\zeta_f(t) = \exp\left(\sum_{i=1}^{\infty} \frac{|\operatorname{Fix}(f^i)|}{i} \cdot t^i\right).$$

Let f be a λ-zig-zag map with m critical points. We say f is *simple* if for some minimal $N \geq 1$, $f^N(1)$ is a critical point:

$$f^N(1) = \frac{k}{\lambda} \qquad \text{for some } 1 \leq k \leq m.$$

Remark 4.7.3. Note that each $f \in \mathrm{PA}(m)$ is simple with $k = 1$.

If f is simple, then the *coefficients* of f are the integers c_i for $0 \leq i \leq N-1$ such that

$$c_i(f) = \begin{cases} A(f^i(1)) & \text{if } E(A(f^i(1))) = +1 \\ A(f^i(1)) + 1 & \text{if } E(A(f^i(1))) = -1. \end{cases}$$

Furthermore, we set $s_i(f) = s_i(\mathrm{It}_f(1))$, the cumulative signs of the itinerary of f. Finally, define the polynomials

$$\rho_f(t) = k \cdot s_N(f) \cdot t^N + \sum_{i=0}^{N-1} s_i(f) c_i(f) t^i$$

and

$$\phi_f(t) = t \cdot \rho_f(t).$$

Proposition 4.7.4. *Suppose $f \in \mathrm{PA}(m)$. Then*

$$1 - \phi_f(t) = \mathcal{R}(D_f)(t).$$

Proof. Write $\Phi(f) = \frac{a}{b}$, and $\lambda = \lambda(f)$. Then $N = b$, with $f^b(1) = \frac{1}{\lambda}$. Furthermore, Remark 4.6.2 shows that $s_j(f) = +1$ for $0 \leq j \leq b-1$, whereas $s_b(f) = s_b(\mathrm{It}_f(1)) =$

-1. Therefore,

$$1 - \phi_f(t) = 1 - t\rho_f(t) = 1 + t^{b+1} - \sum_{i=1}^{b} c_{i-1}(f)t^i.$$

On the other hand, we know that

$$D_f(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^{b+1-i},$$

where the c_i satisfy the relations in (4.3). Taking the reverse gives

$$\mathcal{R}(D_f)(t) = t^{b+1} + 1 - \sum_{i=1}^{b} c_i t^i,$$

so it remains to show that $c_i = c_{i-1}(f)$ for $1 \leq i \leq b$. Indeed, for $1 \leq i \leq b-1$ we have

$$c_i = \nu_{i-1}(f) = A(f^{i-1}(1)) = c_{i-1}(f),$$

while Theorem C tells us that

$$c_b = m = (m-1) + 1 = c_{b-1}(f).$$

□

Combining Proposition 4.7.4 above and Theorem 1.1 of [Suz17] gives the following:

Proposition 4.7.5. *Fix $f \in \mathrm{PA}(m)$. Then $\zeta_f(t)$ converges absolutely in $|t| < \lambda(f)^{-1}$. In addition, for $|t| < \lambda(f)^{-1}$ we have*

$$\zeta_f(t) = \frac{1}{\mathcal{R}(D_f(t))} = \frac{1}{\det(I - tW_f)}. \tag{4.15}$$

In particular,

$$\chi_W(f;t) = t^{m-1} D_f(t). \tag{4.16}$$

Proof. Since f is continuous, Suzuki's argument in the proof of Theorem 1.1 in [Suz17] verifies (4.15). Therefore,

$$\mathcal{R}(D_f(t)) = \mathcal{R}(\det(tI - W_f)) = \mathcal{R}(\chi_W(f;t)).$$

Since $\deg(D_f) = b+1$ and $\dim(W_f) = b+1+(m-1)$, Proposition 4.7.2 implies formula (4.16). □

Corollary 4.7.6. *Fix $f \in \mathrm{PA}(m)$. Then $\chi_M(f;t) = D_f(t)$.*

Proof. Proposition 2.2.5 implies that $\chi_W(f;t) = t^{m-1}\chi_M(f;t)$. The statement now follows from Proposition 4.7.5. □

4.7.3 D_f is the homology polynomial of ψ_f

In [BBK12], Birman-Brinkmann-Kawamuro introduce several related polynomial invariants of a pseudo-Anosov ψ on a closed, connected, orientable surface S with punctures, treated as marked points. These are:

- the *homology polynomial* $h(\psi;t)$, which records the action of ψ on a vector space W of weight functions for an invariant train track τ which satisfy the *switch conditions*.

- the *puncture polynomial* $p(\psi;t)$, which records the action of ψ on the degenerate subspace $Z \subseteq W$ of a certain skew-symmetric bilinear form. This polynomial is a product of cyclotomics, and may be reinterpreted using the permutation action of ψ on a certain subset of the punctures of S.

- the *symplectic polynomial* $s(\psi;t)$, which records the action of ψ on W/Z.

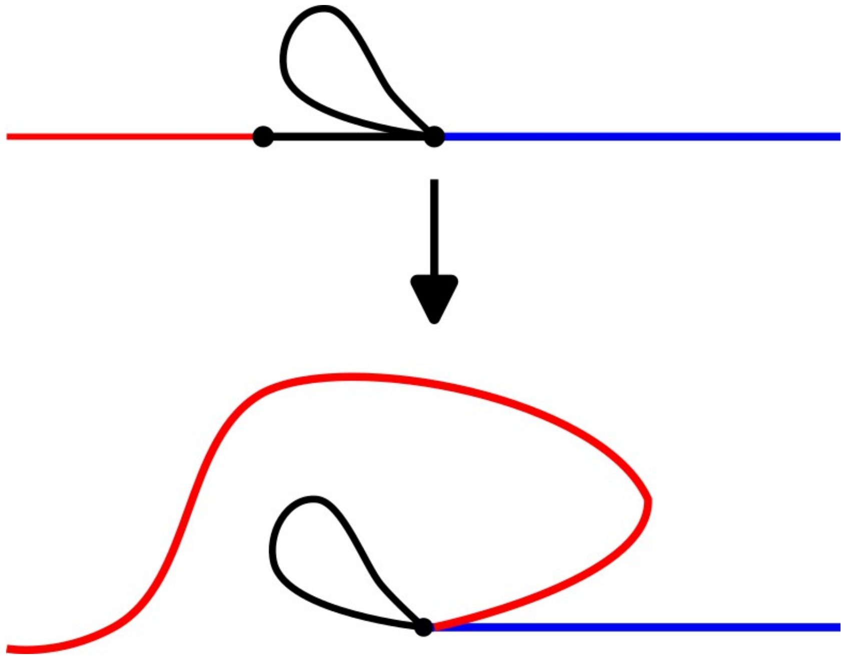


Figure 4.4: The folding move in the proof of Theorem 4.7.7. The loop and the middle edge are *infinitesimal edges*, while the other two edges are expanding. The two infinitesimal edges have the same image under f. Thus we may identify them to obtain a new graph τ' with one fewer edge. Applying this move to every non-extremal loop in τ produces a Bestvina-Handel train track that satisfies the requirements of [BBK12]. Importantly, the transition matrix of the new train track map is identical to that of the original, when restricting to the expanding edges.

Note that $h(\psi;t) = s(\psi;t) \cdot p(\psi;t)$. The symplectic polynomial, and hence the homology polynomial, has the stretch factor $\lambda(\psi)$ as its root of maximal modulus.

In this section we prove statement (3) of Theorem G.

Theorem 4.7.7. *Fix $f \in \mathrm{PA}(m)$ and ψ_f the pseudo-Anosov it generates. Set $n = |\mathrm{PC}(f)|$. Then*

$$D_f(t) = h(\psi_f;t) = \begin{cases} s(\psi_f;t) & \text{if } n \text{ is odd} \\ s(\psi_f;t)(t+1) & \text{if } n \text{ is even.} \end{cases}$$

Proof. Recall that the homology polynomial $h(\psi_f;t)$ is equal to the characteristic polynomial of the action of ψ_f on a space of weights on an underlying graph that satisfy the *switch conditions* (cf. Theorem 2.1 in [BBK12].) We first compute the dimension of this weight space, denoted $W(G, f)$, and show that it is equal to $\deg(D_f) = n - 1$.

We make minor adjustments to our train track, which is not in quite the same form as that assumed by Birman-Brinkmann-Kawamuro. We refer the reader to Section 3.6 for a description of the general form of our train tracks. Recall that each loop γ of τ corresponds to an element of $\mathrm{PC}(f)$. The non-extremal loops of τ are those that do not correspond to $x = 0, 1$ in I. For each non-extremal loop, we apply a *folding move* as in Figure 4.4. The resulting train track τ' is still invariant for ψ_f, and has the same transition matrix M_f, but is now a Bestvina-Handel track. In particular, it satisfies the hypotheses in [BBK12].

The underlying graph G of this new track τ' has $n-1$ expanding edges, n infinitesimal loops, and n vertices. All of the vertices are *partial*, in the terminology of Birman-Brinkmann-Kawamuro, and hence we have the count

$$\begin{aligned}\deg(h(\psi_f;t)) = \dim W(G,f) &= \#\text{edges of } G \text{ - } \#\text{non-odd vertices of } G \\ &= (n-1+n) - n \\ &= n-1.\end{aligned}$$

To show that $h(\psi_f;t) = D_f(t)$, it remains to find a basis for $W(G, f)$. There is a natural such basis: each expanding edge e_i is adjacent to two infinitesimal loops, γ_{i-1} and γ_i. Define $\eta_i \in W(G, f)$ to be the weight function such that

$$\eta_i(e_i) = 1, \quad \eta_i(\gamma_{i-1}) = \eta_i(\gamma_i) = 1/2,$$

and $\eta_i = 0$ outside of these three edges. The functions $\{\eta_i\}_{i=1}^{n-1}$ are linearly independent, since exactly one of them assigns non-zero weight to any expanding edge.

Therefore, these η_i form a basis for $W(G, f)$. Moreover, by the construction of the train track the action of ψ_f on the expanding edges of τ' is precisely the action of f on the subintervals of the strong Markov partition. Thus,

$$h(\psi_f; t) = \chi_M(f; t) = D_f(t).$$

Next we compute the puncture polynomial $p(\psi_f; t)$. Recall that this polynomial is the characteristic polynomial of the action of ψ_f on a certain degenerate subspace Z of $W(G, f)$. The dimension of Z is equal to the number of punctures in $S_{0,n+1}$ represented by a loop with an even number of corners in τ' (cf. Proposition 3.3 in [BBK12]). The number of corners in such a loop is precisely equal to the number of prongs of the singularity at the puncture. Therefore, the puncture at infinity is the only puncture that could possibly contribute to the dimension count of Z, and it does so precisely when it has an even number of prongs, i.e. when $n-2$ is even. Thus,

$$\deg(p(\psi_f; t)) = \dim(Z) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even.} \end{cases}$$

In particular, since $p(\psi_f; t)$ is a product of cyclotomics and since $D_f(1) \neq 0$ we have

$$p(\psi_f; t) = \begin{cases} 1 & \text{if } n \text{ is odd} \\ t+1 & \text{if } n \text{ is even.} \end{cases}$$

Therefore, the symplectic polynomial $s(\psi_f; t)$ satisfies

$$s(\psi_f; t) = \frac{h(\psi_f; t)}{p(\psi_f; t)} = \begin{cases} D_f(t) & \text{if } n \text{ is odd} \\ D_f(t)/(t+1) & \text{if } n \text{ is even.} \end{cases}$$

$\square$

4.7.4 The orientation double cover for ψ_f

In this section we prove statement (4) of Theorem G.

A singular foliation $\mathcal{F}$ of a surface S is *orientable* if there exists a vector field X on S which is zero at the singularities of $\mathcal{F}$ and everywhere else is nonzero and tangent to $\mathcal{F}$. Otherwise, we say $\mathcal{F}$ is *non-orientable*. Similarly, we may speak of a train track $\tau \subseteq S$ being orientable: say τ is *orientable* if there is a choice of orientation for each edge of τ such that, if there is a smooth path from e_1 to e_2 then the orientation of e_2 agrees with the orientation of e_1 flowed along the path to e_2.

Since the literature on this subject is vast and since there is some confusion on the relation between these notions of orientability, we include here a brief summary of relevant properties.

Proposition 4.7.8. *Let $\psi : S \to S$ be a pseudo-Anosov of the compact, connected, oriented surface S. Then the following are equivalent:*

1. *The unstable foliation of ψ is orientable.*
2. *The stable foliation of ψ is orientable.*
3. *Any invariant train track for ψ is orientable.*
4. *The spectral radius of the map $\psi_* : H_1(S;\mathbb{Z}) \to H_1(S;\mathbb{Z})$ is the stretch factor λ of ψ, and in fact exactly one of λ and $-\lambda$ is an eigenvalue of ψ_*.*

Proof. The proof of Lemma 4.3 in [BB07] provides the equivalences

$$(1) \iff (2) \iff (4).$$

It remains to show that (1) is equivalent to (3). If the unstable foliation $\mathcal{F}$ of ψ is orientable, then any invariant train track τ of ψ carries $\mathcal{F}$, and in particular inherits an orientation from that of $\mathcal{F}$. Conversely, an orientable train track τ carries only orientable invariant foliations (cf. Lemma 2.6 in [Los10]). □

Definition 4.7.9. We say a pseudo-Anosov ψ is *orientable* if it satisfies any of the equivalent conditions of Proposition 4.7.8.

Thus the orientability of a pseudo-Anosov is an important invariant. There is a standard construction for producing a pseudo-Anosov from a non-orientable one. This construction is called the *orientation double cover*, and underlies the techniques in [BB07], [BBK12], and [Los10]. Briefly, one takes the branched double cover $p : \tilde{S} \to S$, branching along the singularities of ψ that have an odd number of prongs. Denoting by $\iota : \tilde{S} \to \tilde{S}$ the unique non-trivial deck transformation for this cover, we see that $\iota^2 = 1$. Hence, we have the decomposition

$$H_1(\tilde{S}; \mathbb{Z}) = V_+ \oplus V_-,$$

where $V_\pm$ is the eigenspace of ι_* for the eigenvalue ± 1. The pseudo-Anosov ψ lifts to two maps $\psi_1, \psi_2 : \tilde{S} \to \tilde{S}$ that both commute with ι and satisfy $\psi_1 = \iota \circ \psi_2$.

The following proposition combines the work of [BB07] and [BBK12].

Proposition 4.7.10. *Let $\psi : S \to S$ be a non-orientable pseudo-Anosov with stretch factor $\lambda > 1$, and let $p : \tilde{S} \to S$ be the orientation double cover of ψ. Let $\iota : \tilde{S} \to \tilde{S}$ be the unique non-trivial deck transformation for this cover. Then:*

1. *Both ψ_1 and ψ_2 are orientable pseudo-Anosovs with stretch factor λ.*

2. *The kernel of p_* is precisely* $\ker(p_*) = V_- \subseteq H_1(\tilde{S}; \mathbb{Z})$.

3. *The actions of $(\psi_1)_*$ and $(\psi_2)_*$ on V_+ are each conjugate to the action of ψ_* on $H_1(S; \mathbb{Z})$.*

4. *Exactly one of $(\psi_1)_*$ and $(\psi_2)_*$ has λ as an eigenvalue, while the other has $-\lambda$ as an eigenvalue. Denote the corresponding homeomorphisms of $\tilde{S}$ by ψ_+ and ψ_-, respectively.*

5. *The characteristic polynomials of $(\psi_+)_*$ and $(\psi_-)_*$ acting on V_- satisfy*

$$\det(tI - (\psi_+)_*|_{V_-}) = h(\psi; t) \qquad \textit{and} \qquad \det(tI - (\psi_-)_*|_{V_-}) = h(\psi; -t).$$

Consider $f \in \mathrm{PA}(m)$ with $n = |\,\mathrm{PC}(f)|$. The pseudo-Anosov ψ_f is defined on $S = S_{0,n+1}$, and its set of singularities coincides with the puncture set of S: each of the points of $\mathrm{PC}(f)$ corresponds to a punctured 1-pronged singularity, and the final puncture is a singularity with $n-2$ prongs. The orientation double cover of ψ_f is the surface

$$\widetilde{S} \cong \begin{cases} S_{\frac{n-1}{2},\ n+1} & \text{if } n \text{ is odd} \\ S_{\frac{n-2}{2},\ n+2} & \text{if } n \text{ is even.} \end{cases}$$

In $\widetilde{S}$ there are n punctured 2-prong singularities lying over the 1-prong singularities in S. The remaining punctures (there are either one or two) project to the final puncture of S. In either case, Proposition 4.7.10 implies that

$$H_1(\widetilde{S}; \mathbb{Z}) \cong H_1(S; \mathbb{Z}) \oplus V_- \cong \mathbb{Z}^n \oplus \mathbb{Z}^{n-1}, \tag{4.17}$$

and that the actions of $(\psi_f)_+$ and $(\psi_f)_-$ preserve this splitting.

We are now ready to prove statement (4) of Theorem G.

Theorem 4.7.11. *Let $\mathbb{S}$ be the surface obtained from $\widetilde{S}$ by filling in the lifts of the punctured 1-prong singularities of ψ_f. Denote by $\chi_+(t)$ (resp., $\chi_-(t)$) the characteristic polynomial of ψ_+ (resp., ψ_-) acting on $H_1(\mathbb{S}; \mathbb{Z})$. Then*

$$D_f(t) = \chi_+(t) \quad \textit{and} \quad D_f(-t) = \chi_-(t).$$

Proof. We prove that $D_f(t) = \chi_+(t)$, since the other statement follows immediately.

Filling in the 2-prong singularities of $\widetilde{S}$ kills the factor of $H_1(S;\mathbb{Z})$ in (4.17). Thus, the action of $(\psi_+)_*$ on $H_1(\mathbb{S};\mathbb{Z})$ is conjugate to the action of $(\psi_+)_*$ on V_-. By statement (5) of Proposition 4.7.10 we find

$$\chi_+(t) = h(\psi;t) = D_f(t),$$

where the final equality was proved in Theorem 4.7.7. □

4.7.5 The Burau representation

In this section we complete the proof of Theorem G.

Fix $n \geq 2$, and denote by B_n the n-stranded braid group. The *reduced Burau representation in dimension* n is a homomorphism

$$\mathbb{B}(\cdot, z) : B_n \to \mathrm{GL}(n-1, \mathbb{Z}[z, z^{-1}]).$$

Thus, the image $\mathbb{B}(\beta, z)$ of an n-braid β is an $(n-1)\times(n-1)$ matrix whose entries are Laurent polynomials in the variable z. There are explicit formulas for the Burau representation in terms of the Artin generators for B_n; cf. for example Chapter 3 of [Bir75].

Definition 4.7.12. The *symplectic representation* is the homomorphism

$$\mathbb{B}(\cdot, -1) : B_n \to \mathrm{GL}(n-1, \mathbb{Z})$$

obtained by specializing $\mathbb{B}(\cdot, z)$ at $z = -1$. We define the *symplectic Burau polynomial* of $\beta \in B_n$ to be the characteristic polynomial of $\mathbb{B}(\beta, -1)$:

$$\chi(\beta;t) = \det(tI - \mathbb{B}(\beta, -1)).$$

Note that $\chi(\beta;t)$ has degree $n-1$.

It is well-known that the mapping class group of the n-punctured disc D_n, relative to its boundary circle ∂D_n, is

$$\operatorname{Mod}(D_n, \partial D_n) \cong B_n.$$

Moreover, the center of B_n is generated by the *full twist* Δ_n^2, the Dehn twist around ∂D_n. Capping off this boundary component by a punctured disc kills this element of B_n, and we obtain the isomorphism

$$B_n/\Delta_n^2 \cong \operatorname{Mod}(S_{0,n+1}, p),$$

where the group on the right is the subgroup of $\operatorname{Mod}(S_{0,n+1})$ fixing the puncture p in the capping disc. Thus, given $\psi \in \operatorname{Mod}(S_{0,n+1}, p)$ we may associate to it a braid $\beta \in B_n$, which is well-defined up to multiplication by a power of Δ_n^2.

Remark 4.7.13. One can show that $\mathbb{B}(\Delta_n^2, -1) = -I_{n-1}$. Therefore,

$$\chi(\Delta_n^2\beta; t) = \chi(\beta; -t).$$

In particular, the symplectic Burau polynomial of a braid representative of a pseudo-Anosov $\psi \in \operatorname{Mod}(S_{0,n+1}, p)$ is nearly an invariant of ψ itself: the only ambiguity is a possible change of variable $t \mapsto -t$.

Given $f \in \mathrm{PA}(m)$ with $|\operatorname{PC}(f)| = n$ the pseudo-Ansov $\psi_f : S_{0,n+1} \to S_{0,n+1}$ has a unique singularity ∞ with $n-2$ prongs, which is necessarily a fixed point. Ripping open this puncture to a boundary component, we obtain an n-braid β_f defined up to multiplication by Δ_n^2. Regardless of our choice for β_f, however, it will always be pseudo-Anosov with a 1-prong singularity at each puncture of D_n and an $(n-2)$-prong singularity at ∂D_n. Moreover, the stretch factor of β_f is equal to the stretch factor λ of ψ_f.

With the remark in mind, we state the main result of this section.

Theorem 4.7.14. *Fix $f \in \mathrm{PA}(m)$, and let β_f be any n-braid representative of ψ_f obtained by ripping open $\infty \in S_{0,n+1}$ to a boundary circle. Then*

$$\chi(\beta_f;t) = \begin{cases} D_f(t) & \text{if } \chi(\beta_f;t) \text{ has } \lambda \text{ as a root} \\ D_f(-t) & \text{if } \chi(\beta_f;t) \text{ has } -\lambda \text{ as a root.} \end{cases}$$

Our argument relies heavily on the work of Band-Boyland in [BB07]. We direct the reader to sections 2 and 3 of that paper for details. The reduced Burau matrix $\mathbb{B}(\beta,z)$ describes the action on first homology of a preferred lift $\tilde{h}$ of β to a certain cover of D_n, denoted $D_n^{(\infty)}$. The deck group of the cover $\tilde{p} : D_n^{(\infty)} \to D_n$ is isomorphic to $\mathbb{Z}$, and thus for each $k \geq 1$ we obtain a cover $p_k : D_n^{(k)} \to D_n$ via the quotient map $\xi_k : \mathbb{Z} \to \mathbb{Z}/k\mathbb{Z}$, as well as a preferred lift $h^{(k)} : D_n^{(k)} \to D_n^{(k)}$ of β.

Importantly, one obtains information about $h_*^{(k)} : H_1(D_n^{(k)}) \to H_1(D_n^{(k)})$ by specializing $\mathbb{B}(\beta,z)$ at the kth roots of unity.

Proposition 4.7.15 (Theorem 3.4 in [BB07])**.** *For $k \geq 1$ set ζ_k to be a primitive kth root of unity. There is an invariant subspace $S_{\mathbb{C}}^{(k)}$ of $H_1(D_n^{(k)};\mathbb{C})$ such that the action of $h_*^{(k)}$ on this subspace is by*

$$\mathbb{B}(\beta,1) \oplus \mathbb{B}(\beta,\zeta_k) \oplus \cdots \oplus \mathbb{B}(\beta,\zeta_k^{k-1}). \tag{4.18}$$

Moreover, set T to be the generator of the deck group $\mathbb{Z}/k\mathbb{Z}$ for $p_k : D_n^{(k)} \to D_n$. Then in (4.18), the matrix $\mathbb{B}(\beta,\zeta_k^j)$ is the action of $h_*^{(k)}$ on the eigenspace of T with eigenvalue ζ_k^j.

Finally, any eigenvector of $h_*^{(k)}$ not lying in $S_{\mathbb{C}}^{(k)}$ has as its eigenvalue a root of unity.

Recall that, by Proposition 4.7.8, a pseudo-Anosov ψ with stretch factor λ is orientable if and only if λ or $-\lambda$ is an eigenvalue of ψ_*. Band-Boyland use this fact to conclude the following.

Theorem 4.7.16 (Theorem 5.1 in [BB07]). *Suppose $\beta \in B_n$ is a pseudo-Anosov braid having stretch factor $\lambda > 1$. Then the following are equivalent:*

1. *The spectral radius of $\mathbb{B}(\beta, -1)$ is λ, and -1 is the only root of unity for which this occurs.*
2. *The invariant foliations $\mathcal{F}^u$ and $\mathcal{F}^s$ have an odd-order singularity at each puncture of D_n (not including the boundary), and all other singularities of $\mathcal{F}^u$ and $\mathcal{F}^s$ in the interior of D_n have even order.*
3. *$D_n^{(2)}$ is the orientation double cover of β.*

We are now ready to prove Theorem 4.7.14. This will conclude the proof of Theorem G.

Proof of Theorem 4.7.14. Let β_f be any pseudo-Anosov braid obtained from ψ_f by ripping open $\infty \in S_{0,n+1}$ to the boundary circle of D_n. Then β_f has the following singularity data:

- a 1-prong singularity at each puncture of D_n, and
- a singularity with $n-2$ prongs on ∂D_n.

Therefore, β_f satisfies condition (2) of Theorem 4.7.16. It follows that $D_n^{(2)}$ is the orientation double cover of β_f. In this setting, the deck group of $p_2 : D_n^{(2)} \to D_n$ is generated by $T = \iota$, the hyperelliptic involution of $D_n^{(2)}$.

On the other hand, one may take the orientation cover of $\psi_f : S_{0,n+1} \to S_{0,n+1}$, obtaining the surface $\widetilde{S}$ with its own hyperelliptic involution ι'. As we have noted before, ψ_f has a pair of lifts, ψ_+ and $\psi_- = \iota' \circ \psi_+$.

We will relate the following two actions:

1. the action of the preferred lift $\beta_f^{(2)}$ on $V_-(\iota)$
2. the action of a lift $\widetilde{\psi}$ of ψ on $V_-(\iota')$.

From Proposition 4.7.15 we know that action (1) has characteristic polynomial

$$\det\left(tI - \left(\beta_f^{(2)}\right)_*\Big|_{V_-(\iota)}\right) = \det(tI - \mathbb{B}(\beta_f, -1)) = \chi(\beta_f; t).$$

For action (2), the work of Birman-Brinkmann-Kawamuro tells us that

$$\det\left(tI - \left(\widetilde{\psi}\right)_*\Big|_{V_-(\iota')}\right) = \begin{cases} h(\psi_f; t) & \text{if } \widetilde{\psi} = \psi_+ \\ h(\psi_f; -t) & \text{if } \widetilde{\psi} = \psi_-. \end{cases}$$

There is a natural isomorphism $\pi : H_1(D_n^{(2)}; \mathbb{Z}) \to H_1(\widetilde{S}; \mathbb{Z})$ obtained by capping off each boundary component of $D_n^{(2)}$ with a punctured disc. Moreover, the following diagram commutes:

$$\begin{tikzcd}
H_1(D_n^{(2)}; \mathbb{Z}) \arrow[r, "\pi"] & H_1(\widetilde{S}; \mathbb{Z}) \\
V_-(\iota) \arrow[u, hook] \arrow[r, "\pi"] & V_-(\iota') \arrow[u, hook] \\
V_-(\iota) \arrow[u, "(\beta_f^{(2)})_*"] \arrow[r, "\pi"] & V_-(\iota') \arrow[u, "\widetilde{\psi}_*"']
\end{tikzcd}$$

Here, $\widetilde{\psi}$ is the unique lift of ψ_f that we obtain by capping off $\beta_f^{(2)}$. We now see that actions (1) and (2) are conjugate, and hence their characteristic polynomials coincide. To finish the proof, we apply Theorem 4.7.7 to conclude that

$$\chi(\beta_f; t) = \begin{cases} D_f(t) & \text{if } \widetilde{\psi} = \psi_+ \\ D_f(-t) & \text{if } \widetilde{\psi} = \psi_-. \end{cases}$$

□

Remark 4.7.17. Theorem 4.7.14 is a special case of a result proved in the thesis of Warren Michael Shultz, cf. Theorem 1.0.1 in [Shi16]. There is a slight error in the statement of this theorem: namely, item (1) does not include the possible sign ambiguity that we noted in Remark 4.7.13.

4.8 Future directions

Over the course of Chapters 3 and 4 we have undertaken the study of a particular family $\mathrm{PA}(m)$ of interval maps that may be realized as the train track map for a pseudo-Anosov. In our case, $\mathrm{PA}(m)$ is a one-parameter family, determined by a single critical orbit. Moreover, we have shown that this family exhibits the structure of the Farey tree $\mathcal{F}$, in that the critical orbit for the map associated to $q_1 \oplus q_2$ is essentially the concatenation of the critical orbits for the maps associated to q_1 and q_2.

4.8.1 Tree-like families

We strongly suspect that analogous results hold in the following more general context. Let $f : I \to I$ be a PCF interval map of pseudo-Anosov type. Since ψ permutes the set of 1-prong singularities, so too does f permute the elements of $\mathrm{PC}(f)$. We conjecture that, fixing all but one postcritical orbit and allowing the final one to vary, a tree-like family of interval maps of pseudo-Anosov type will result. The branching is due to the different choices of how to modify the orbit. In the case of $\mathrm{PA}(m)$, our only choices are to perturb the orbit of $x = 1$ to the right or to the left, determining whether we descend to the left or to the right in $\mathcal{F}$, respectively (cf. Remark 4.5.7).

It is conceivable that some generalization of the invariant $\Phi(f)$ will allow one to track their progress through this tree-like family of pseudo-Anosovs, but the details of such behavior are unclear to us. For one thing, it seems possible that the same rotation number might appear for different maps, unlike what happens for $\mathrm{PA}(m)$.

4.8.2 Generalizing Theorem G

What is more clear is how to generalize the statements in Theorem G to any interval map of pseudo-Anosov type. For a general such map, the natural replacement for D_f is the strong Markov polynomial $\chi_M(f;t)$. Because the only critical values of a zig-zag map are $x = 0$ and $x = 1$, and because $x = 0$ is either a fixed point or in the orbit of $x = 1$, we can extract $\chi_M(f;t)$ purely from the itinerary of $x = 1$. Thus we cannot define $D_f(t)$ in general.

An analysis of the proof of Theorem G shows, however, that apart from proving that $D_f(t) = \chi_M(f;t)$, the other statements do not rely on any special characteristic of zig-zag maps. The essential fact is that $\chi_M(f;t)$ is the characteristic polynomial of a train track map for ψ_f. Statement 3 has to be slightly altered, since it is possible for a factor of $t-1$ to appear if $n = \deg(\chi_M(f;t)) + 1$ is odd for general f, but statements 1, 4, and 5 hold verbatim after replacing $D_f(t)$ with $\chi_M(f;t)$.

4.8.3 Interval maps from pseudo-Anosovs

A PCF interval map f of pseudo-Anosov type with $|\,\mathrm{PC}(f)| = p$ generates a pseudo-Anosov $\psi_f : S_{0,p+1} \to S_{0,p+1}$ with at most one singularity having two or more prongs. Call a pseudo-Anosov with only one such singularity *hyperpolar*, since each 1-prong singularity is a *pole* for the associated quadratic differential.

Furthermore, call a pseudo-Anosov *interval-like* if it admits an invariant train track τ consisting of a chain of expanding edges with an infinitesimal loop at each vertex. Thus, every interval-like pseudo-Anosov is hyperpolar. As we have seen, the dynamics of the PCF interval map underlying $f : \tau \to \tau$ determine those of the interval-like pseudo-Anosov. It is natural, then, to ask the following question.

Question 4.8.1. Which hyperpolar pseudo-Anosovs are interval-like?

In the next chapter we prove that the two sets of pseudo-Anosovs coincide (cf. Theorem H). This result is joint work with Karl Winsor. Thus, hyperpolar pseudo-Anosovs may be said to have the dynamics of PCF interval maps. In order to truly leverage this connection, however, it would be useful to have a better understanding of when a PCF interval map is of pseudo-Anosov type. In Chapter 3 we achieved this explicitly for the case of zig-zag maps from combinatorial data. In general, one would hope for a set of criteria relying on the kneading data of a given interval map to determine if it is of pseudo-Anosov type.

Question 4.8.2. Given a PCF interval map f, is there a set of criteria, depending on the kneading data of f, that is equivalent to f being of pseudo-Anosov type?

Chapter 5

Interval maps from pseudo-Anosovs

In Chapters 3 and 4 we investigated the properties of pseudo-Anosovs having interval-type train tracks. In this chapter we characterize pseudo-Anosovs with interval-type train tracks as precisely those whose singularity data is of the form $(1^p, p-2)$.

Definition 5.0.1. A pseudo-Anosov $\psi : S_{0,p+1} \to S_{0,p+1}$ is *interval-like* if it admits an invariant train track τ consisting of a chain of expanding edges with a single infinitesimal loop attached to each vertex.

As we have seen, every interval-like pseudo-Anosov has singularity data $(1^p, p-2)$ where $p = |\operatorname{PC}(f)|$ and $f : I \to I$ is the interval map induced by the action of ψ on the expanding edges of τ.

Note that if ψ has singularity data $(1^p, p-2)$ then ψ is necessarily defined on a marked sphere, by the Euler-Poincaré formula (cf. Proposition 2.5.4).

The following theorem is joint work with Karl Winsor.

Theorem H. *If ψ is a pseudo-Anosov with singularity data $(1^p, p-2)$ for some $p \geq 3$, then ψ is interval-like.*

The consequences of this theorem are manifold. First and foremost, it implies that the dynamics of a pseudo-Anosov ψ with singularity data $(1^p, p-2)$, such as its periodic orbits and its topological entropy, are essentially the dynamics of a PCF interval map f with $|\operatorname{PC}(f)| = p$. Indeed, the two systems have strong Markov partitions with the same Markov matrix. With interval dynamics being so well-understood, it is tantalizing to imagine the applications of such a connection.

One such application is a satisfying new proof of a result of Boissy-Lanneau. In the following theorem, a translation surface (X, ω) is in a *hyperelliptic connected component* if

1. (X, ω) is *hyperelliptic*, i.e. there is an involution $\iota : X \to X$ preserving the translation structure of (X, ω),
2. the abelian differential ω has at most two zeros, and
3. if ω has two zeros, then ι exchanges them.

Theorem 5.0.2 (Theorem 1.1 in [BL12])**.** *Let $\widetilde{\psi}$ be an affine pseudo-Anosov homeomorphism on a genus $g \geq 1$ translation surface (X, ω) in a hyperelliptic connected component. Then the stretch factor of $\widetilde{\psi}$ satisfies*

$$\lambda\left(\widetilde{\psi}\right) > \sqrt{2}.$$

As currently stated, it is not immediately clear how Theorem 5.0.2 is related to Theorem H. We give an equivalent formulation of the theorem to illuminate this connection. Since ι preserves with the translation structure of (X, ω) it commutes with $\widetilde{\psi}$, producing a pseudo-Anosov ψ in the quotient:

$$\begin{tikzcd} (X,\omega) \arrow[r, "\widetilde{\psi}"] \arrow[d] & (X,\omega) \arrow[d] \\ S \arrow[r, "\psi"] & S \end{tikzcd}$$

We have seen this picture before: the system $\widetilde{\psi} : X \to X$ is the *orientation double cover* of $\psi : S \to S$ (cf. Section 4.7.4).

Proposition 5.0.3. *S is a topological sphere. Moreover, ψ has the same stretch factor as $\widetilde{\psi}$, and its singularity data is as follows:*

1. *$(1^{2g+1}, 2g-1)$ if ω has one zero, or*

2. *$(1^{2g+2}, 2g)$ if ω has two zeros.*

Proof. S is a topological sphere because ι is the hyperelliptic involution of the hyperelliptic Riemann surface (X, ω). The branched covering $X \to S$ has degree 2 and $2g+2$ branch points in S, by Riemann-Hurwitz. If ω has a single zero, then ι necessarily fixes that point, which is a zero of order $2g-2$ and hence has $4g-2$ prongs. In the quotient it becomes a singularity with $2g-1$ prongs, half as many. The remaining $2g+1$ branch points of ι in this case come from regular points upstairs, and so are 1-prong singularities. If instead ω has two zeros, which ι swaps by assumption, then these are both zeros of order $g-1$, and hence have $2g$ prongs. In the quotient they are identified to a single $2g$-prong singularity. This point is not a branch point, so all $2g+2$ branch points are 1-prong singularities. $\square$

Here is the equivalent formulation of Theorem 5.0.2 that we will prove using Theorem H. The proof is in Section 5.4.

Theorem 5.0.4 (Reformulation of Theorem 5.0.2)**.** *Let $\psi : S_{0,p+1} \to S_{0,p+1}$ be a pseudo-Anosov with singularity data $(1^p, p-2)$. Then the stretch factor of ψ satisfies*

$$\lambda(\psi) > \sqrt{2}.$$

Theorem H also suggests a potential avenue to answering Fried's Question (Question 2.5.8) in the affirmative. W. Thurston proved in [Thu14] that the set of entropies of PCF interval maps $f : I \to I$ is the set of *weak Perron numbers*,

$$\mathrm{WP} = \{\log \lambda : \lambda^k \text{ is a Perron number for some } k \geq 1\}.$$

All bi-Perron units are weak Perron. Thus, given a bi-Perron unit λ, one can find a PCF interval map f having entropy $\log \lambda$. If we can further guarantee that there is such an f of pseudo-Anosov type, then we will have answered Fried's Question, since the corresponding pseudo-Anosov will have stretch factor λ. This final task, however, is very combinatorially challenging, and would require new insights that we do not provide here.

The proof of Theorem H relies on the theory of *tight splitting*, developed in [FRW22] by the author along with Braeden Reinoso and Luya Wang. We introduce the necessary background for train tracks in Section 5.1, and treat tight splitting in Section 5.2. The figures in these sections are due to Luya Wang. We prove Theorem H in Section 5.3.

5.1 Fibered surfaces and train tracks

In this chapter we will denote by S' the surface S with its marked points deleted, and by $\widehat{S}$ the closed surface obtained by capping-off the boundary components of S with disks and marking a point in the interior of each disk. We will also assume that the surface S' has negative Euler characteristic.

In [BH95], Bestvina and Handel prove that one may associate to any geometric ψ a *fibered surface* $F \subseteq S'$. This fibered surface is decomposed into *strips* and *junctions*, where the strips are foliated by intervals, i.e. *leaves*. See Figure 5.1. Together, the leaves and junctions of F are called *decomposition elements*, and $\psi(F) \subseteq F$, sending decomposition elements into decomposition elements and, in particular, junctions into junctions. Collapsing each decomposition element to a point produces a graph G with a graph map $g : G \to G$. The vertices of G correspond to the junctions of F, and the

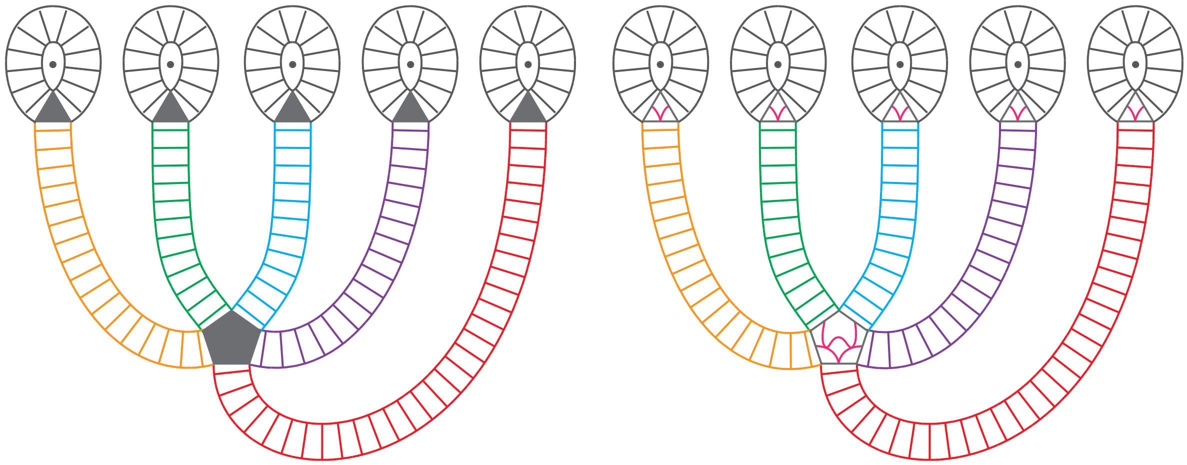


Figure 5.1: Left: the fibered surface for some geometric ψ on $S^1_{0,5}$. The shaded regions are the junctions, and the striped bands connecting them are the strips. Right: following Bestvina-Handel, one inserts additional, “infinitesimal” edges into the junctions. These will also be inserted into the graph G that one obtains by collapsing all of the decomposition elements. Their inclusion will produce the smooth analog τ of G, which is a train track. See Figure 5.2 below.

edges of G correspond to the strips of F.

Roughly speaking, a graph map g is *efficient* if the image of no edge backtracks under any power of g. After adjusting F so that g is efficient, Bestvina and Handel construct a “smoothed” version of G as follows. Within each junction $J \subseteq F$, one inserts additional edges that smoothly connect the strips of F and encode how images of strips under ψ pass through J.

In this way, we obtain a new graph τ smoothly embedded in S', called a *train track*. At each vertex s of τ, called a *switch*, there is a well-defined tangent line. Two arcs a, b of τ are *tangent* at s if $a(0) = b(0) = s$ and $a'(0) = b'(0)$. A *cusp* is the data of a pair (a, b) of adjacent arcs tangent at s. See Figure 5.2 for an example.

The following proposition appears as Proposition 3.3.5 in [BH95].

Proposition 5.1.1. *Suppose ψ is pseudo-Anosov. Then each component of $\widehat{S} \setminus \tau$ is either:*

1. *a disk with $k \geq 3$ cusps on its boundary, or*

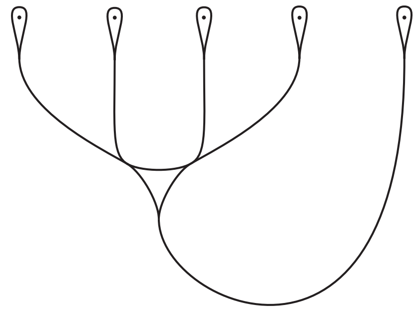


Figure 5.2: A train track τ on the five-punctured disk D_5. The components of the complement of τ consist of: five once-punctured *monogons*, i.e. disks with a single boundary cusp; a *trigon*, i.e. a disk with three boundary cusps; and an exterior once-punctured bigon. Pseudo-Anosovs carried by this track lie in the stratum $(2; 1^5; 3)$.

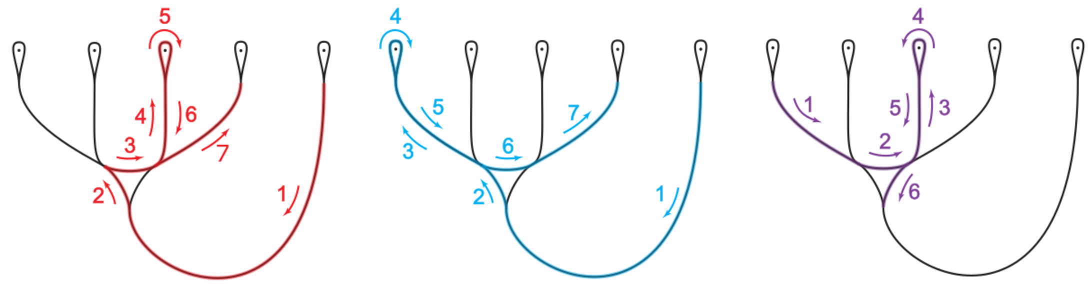


Figure 5.3: Three edge-paths on the train track τ from Figure 5.2. **Left:** an edge path of length 7 which is not a train path, since it makes several sharp turns. **Middle:** a train path of length 7 which can be "pushed off" of τ into a small neighborhood so that it does not intersect itself. **Right:** a train path of length 6 which cannot be "pushed off" of τ so that it becomes injective.

2. *a disk with a single marked point in its interior and $k \geq 1$ cusps on its boundary.*

Remark 5.1.2. From this perspective, the cusps of a component $C \subset \widehat{S} \backslash \tau$ correspond precisely to the prongs of a singularity $p \in C$ of the invariant foliations of ψ.

Definition 5.1.3. An *edge path* in τ is a map $e : I \to \tau$ such that $e(0)$ and $e(1)$ are switches. A *train path* is an edge path that is also a smooth immersion. The *length* of a train path e is defined to be the number of edges traversed by $e(I)$, counting with multiplicity. Let $e(I) = e_1 \cdots e_k$ denote a train path whose directed image traverses first e_1, then e_2, etc. See Figure 5.3 for examples.

Definition 5.1.4. A *train track map* is a map $f : \tau \to \tau$ such that for any train path $g : I \to \tau$ the composition $f \circ g : I \to \tau$ is a train path.

Remark 5.1.5. Note that if $f : \tau \to \tau$ is a train track map, then $f(e)$ is a train path for each edge e of τ. Indeed, from this it follows that $f^k(e)$ is a train path for each $k \geq 1$, and hence f^k is a train track map for all $k \geq 1$.

The map $\psi : F \to F$, or equivalently the graph map $g : G \to G$ corresponding to ψ and F, defines a map $f : \tau \to \tau$. The fact that g is efficient implies that f is a train track map. In this case, we say that the train track τ *carries* the map ψ, and the map ψ *induces* the train track map f. The *data* of a geometric map will then be a triple (τ, ψ, f) in a commutative diagram:

$$\begin{tikzcd} \tau \arrow[r, "\psi"] \arrow[dr, "f"'] & \psi(\tau) \arrow[d, "\text{collapse}"] \\ & \tau \end{tikzcd}$$

Here, one should imagine τ being mapped forward by ψ into F, meeting the leaves of F transversely. The map f is then defined by collapsing each leaf of F to a point, while inside each junction the arcs of $\psi(\tau)$ are collapsed onto the appropriate edges of τ. See Figure 5.4.

The edges of G (other than those loops peripheral to marked points/punctures of S) are in bijection with a subset of the edges of τ, which we call the *real* edges. All other edges of τ are *infinitesimal*. In particular, all edges of τ contained in a junction of F are infinitesimal. Enumerate the edges of τ so that $e_1, \ldots, e_k$ are the real edges and $e_{k+1}, \ldots, e_n$ are the infinitesimal edges. For each pair (i, j) with $1 \leq i, j, \leq n$ define the integer

$$m_{i,j} = \text{the number of times the train path } f(e_j) \text{ traverses } e_i.$$

The *extended transition matrix* of f is the matrix $\widetilde{M}$ whose (i, j)-entry is the integer

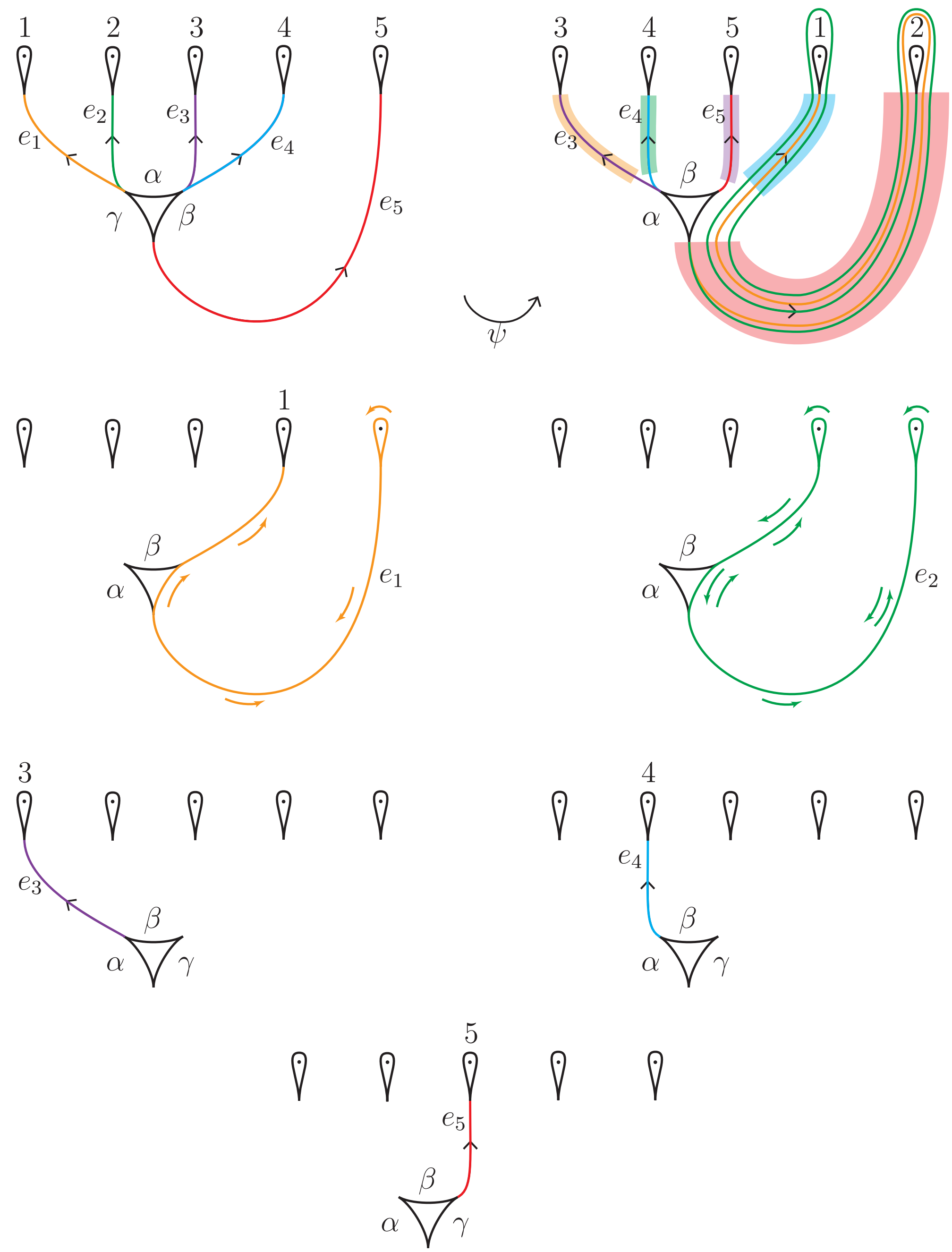


Figure 5.4: **Top row:** the train track τ from Figure 5.2 and the action of a pseudo-Anosov ψ that it carries. The real edges of τ are labeled $e_1, \ldots, e_5$. The shaded regions on the right denote the neighborhoods that deformation retract onto these edges. **Bottom three rows:** The action of f = (collapse $\circ\, \psi$) on each edge of τ, depicted separately.

$m_{i,j}$. The *transition matrix* of f is the submatrix $M \subset \tilde{M}$ recording the transitions between real edges of τ: in other words,

$$M = (m_{i,j}) \text{ where } 1 \leq i, j \leq k.$$

The next theorem follows from work of Bestvina-Handel in [BH95].

Theorem 5.1.6. *Let (τ, ψ, f) be the data of a geometric map, where τ satisfies the conclusion of Proposition 5.1.1. Let M be the transition matrix of f. Then*

$$\psi \textit{ is pseudo-Anosov} \iff M \textit{ is primitive.}$$

The dominant eigenvalue λ of M is called the *dilatation* of ψ. By Theorem 2.1.3 there is a unique right λ-eigenvector w of M, up to scale, and its entries w_i for $i = 1 \ldots, k$ define *transverse weights* on the real edges e_i of M.

Remark 5.1.7. If a train track map F is induced by some pseudo-Anosov ψ, then such a ψ is unique up to conjugacy in $\mathrm{Mod}(\widehat{S})$. Indeed, Bestvina–Handel in [BH95] provide an algorithm to determine the measured foliations preserved by ψ.

5.2 The tight splitting

This section is devoted to developing a tool which will be integral to the proof of Theorem H: a specialized form of "splitting," which will allow us to algorithmically alter a train track until it has a desired form.

5.2.1 Splitting standardly embedded tracks

We first describe a particular class of train tracks on the punctured disk, called *standardly embedded* tracks, which will aid in the description of our splitting procedure.

Standardly embedded train tracks have previously appeared in the work of Ko–Los–Song ([KLS02]), Cho–Ham ([CH08]), and Ham–Song ([HS07]), who used them to study pseudo-Anosovs on a marked disc $S^1_{0,n}$ for small n.

Definition 5.2.1. An *infinitesimal polygon* of a train track τ is a connected component of $S^1_{0,n} \setminus \tau$ whose boundary consists of finitely many infinitesimal edges of τ. A train track τ on $S^1_{0,n}$ is *standardly embedded* if the following conditions hold:

1. Every component of $S^1_{0,n} \setminus \tau$ is an infinitesimal polygon, except for the one containing $\partial S^1_{0,n}$.
2. If two edges of τ are tangent at a switch, then either both are real or both are infinitesimal.
3. Cusps only occur at vertices of infinitesimal polygons.

Figure 5.2 is an example of a standardly embedded track, and Figure 5.4 shows a pseudo-Anosov carried by this track, as well as the induced train track map. Every train track may be adjusted to a standarly embedded one, and this adjustment does not affect which pseudo-Anosovs the track carries. So, we have:

Proposition 5.2.2. *Every pseudo-Anosov on $S^1_{0,n}$ is carried by a standardly embedded train track.*

We adapt the following definition from Ham–Song's notion of an elementary folding map [HS07].

Definition 5.2.3. Let $\tau, \tau_1 \hookrightarrow S^1_{0,n}$ be standardly embedded train tracks. A *Markov map* is a graph map $p : \tau_1 \to \tau$ that maps vertices to vertices, and is locally injective away from the preimages of vertices. An *elementary folding map* is a smooth Markov map such that for exactly one real edge α, the image $p(\alpha)$ has word length 2, while the images of all other edges have word length 1. We require that the distinguished edge α belong to a cusp (α, β) of τ_1, and that $p(\alpha)$ be of the form

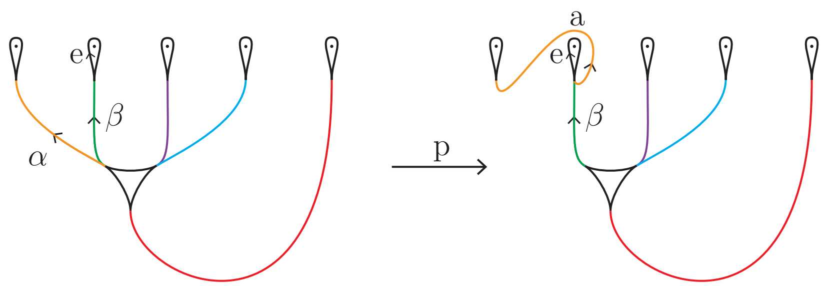


Figure 5.5: An example of an elementary folding map. The map p is the identity except at the edge α, which is mapped as a directed path to $\beta \cdot e \cdot a$.

$$p(\alpha) = p(\beta) \cdot a,$$

where a is a real edge joined to $p(\beta)$ by an infinitesimal edge.

For the purposes of this paper, an elementary folding map $p : \tau_1 \to \tau$ will be the identity map away from the distinguished real edge α. See Figure 5.5.

Remark 5.2.4. An elementary folding map in our terminology is the composition of two elementary moves in Ham-Song's terminology [HS07].

Suppose now that (τ, ψ, f) is the data of a pseudo-Anosov ψ on $S^1_{0,n}$ carried by the standardly embedded τ:

$$\begin{tikzcd}
 & \tau \\
\tau \arrow[ur, "f"] \arrow[r, "\psi"'] & \psi(\tau) \arrow[u, "\text{collapse}"']
\end{tikzcd}$$

Suppose further that $\tau_1 \hookrightarrow S^1_{0,n}$ is another standardly embedded train track such that there exists an elementary folding map $p : \tau_1 \to \tau$. Then there is a well-defined elementary folding map $p_\psi : \psi(\tau_1) \to \psi(\tau)$ such that the following diagram commutes:

$$\begin{tikzcd}
 & \tau \\
\tau \arrow[ur, "f"] \arrow[r, "\psi"'] & \psi(\tau) \arrow[u, "\text{collapse}"'] \\
\tau_1 \arrow[u, "p"] \arrow[r, "\psi"] & \psi(\tau_1) \arrow[u, "p_\psi"']
\end{tikzcd}$$

In general, we cannot expect τ_1 to carry ψ: if it did, we would then be able to complete the above commutative diagram as follows:

$$\begin{tikzcd}
 & \tau \\
\tau \arrow[ur, "f"] \arrow[r, "\psi"'] & \psi(\tau) \arrow[u, "\text{collapse}"'] \\
\tau_1 \arrow[u, "p"] \arrow[r, "\psi"] \arrow[dr, "f_1"'] & \psi(\tau_1) \arrow[u, "p_\psi"'] \arrow[d, "\text{collapse}"] \\
 & \tau_1
\end{tikzcd}$$

In this section we will discuss how to find such a τ_1. The process of producing the data (τ_1, ψ, f_1) from (τ, ψ, f) is called *tight splitting*, or *t-splitting* for short.

Let $\tau \hookrightarrow S^1_{0,n}$ be standardly embedded, and let $v \in \tau$ be a switch. The *link* of v is the collection $\mathrm{Lk}(v)$ of edges of τ incident to v. The elements of $\mathrm{Lk}(v)$ inherit a natural counterclockwise cyclic order $e_1, \ldots, e_k$. A subset $C \subseteq \mathrm{Lk}(v)$ is *connected* if whenever $e_i, e_j \in C$ and $i < j$, then either

1. $e_{i+1}, \ldots, e_{j-1} \in C$, or
2. $e_{j+1}, \ldots, e_k, e_1, \ldots, e_{i-1} \in C$.

The collections

$$R(v) = \{\text{real edges in } \mathrm{Lk}(v)\}, \quad I(v) = \{\text{infinitesimal edges in } \mathrm{Lk}(v)\}$$

are connected. We index the elements of $\mathrm{Lk}(v)$ so that the real edges are $e_1, \ldots, e_m$

under the cyclic order. In other words, from the perspective of v facing its real edges, e_1 is the real edge furthest to the right and e_m is the edge furthest to the left.

Definition 5.2.5. The *right extremal edge* of v is $r(v) = e_1$, and the *left extremal edge* is $l(v) = e_m$. If $R(v) = \{e\}$ is a singleton, then we set $e = l(v) = r(v)$.

If v is a switch at an infinitesimal loop of τ, we treat each end of the loop as a distinct element of $\mathrm{Lk}(v)$. Hence $I(v)$ always consists of two elements, i_l and i_r. These are defined so that, under the cyclic order, we have

$$l(v) < i_l < i_r < r(v).$$

Definition 5.2.6. We denote by v_l the switch of τ at the other end of i_l from v. Similarly, we denote by v_r the switch of τ at the other end of i_r from v. In the case that v is at a loop of τ, we set $v_l = v_r = v$.

From now on, we set the convention that, for a given switch v of τ, all edges in $R(v)$ are oriented into v as paths.

Definition 5.2.7. Let $\tau \hookrightarrow S^1_{0,n}$ be a standardly embedded train track. Let v be a switch of τ. Fix a train track map $f : \tau \to \tau$. We say that v *splits tightly to the left* or *l-splits* if for every real edge $x \subseteq \tau$ the following two conditions hold:

1. Whenever $l(v)$ appears in the train path $f(x)$, it is followed by $\overline{r(v_l)}$, and
2. whenever $\overline{l(v)}$ appears in the train path $f(x)$, it is preceded by $r(v_l)$.

Similarly, we say that v *splits tightly to the right* or *r-splits* if for every real edge $x \subseteq \tau$ the following two conditions hold:

1. Whenever $r(v)$ appears in the train path $f(x)$, it is followed by $\overline{l(v_r)}$, and
2. whenever $\overline{r(v)}$ appears in the train path $f(x)$, it is preceded by $l(v_r)$.

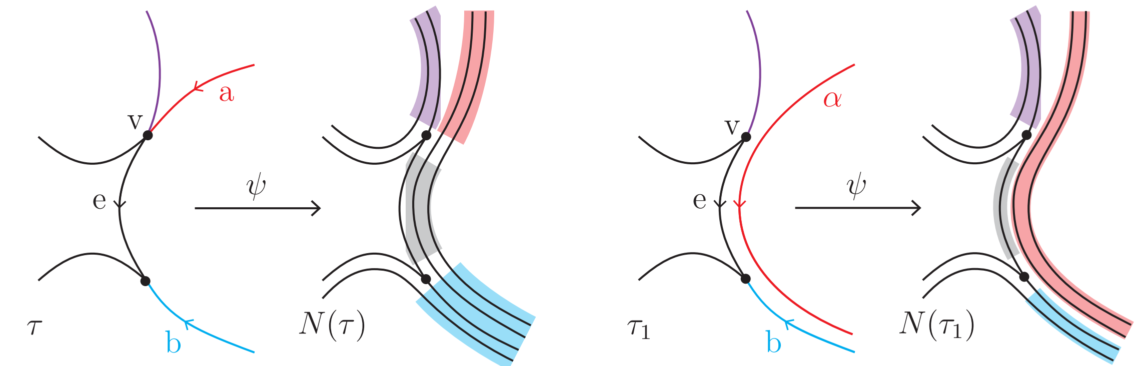


Figure 5.6: **Left:** part of a train track τ and the image of a pseudo-Anosov ψ carried by τ. Here ψ induces a train track map $f : \tau \to \tau$ for which v splits tightly to the right. **Right:** The train track τ_1 after r-splitting v, and the action of ψ on τ_1. Note in particular that ψ has not changed, only τ and its fibered neighborhood $N(\tau)$. In each row, the highlighted regions are collapsed by a deformation retraction onto the corresponding edges.

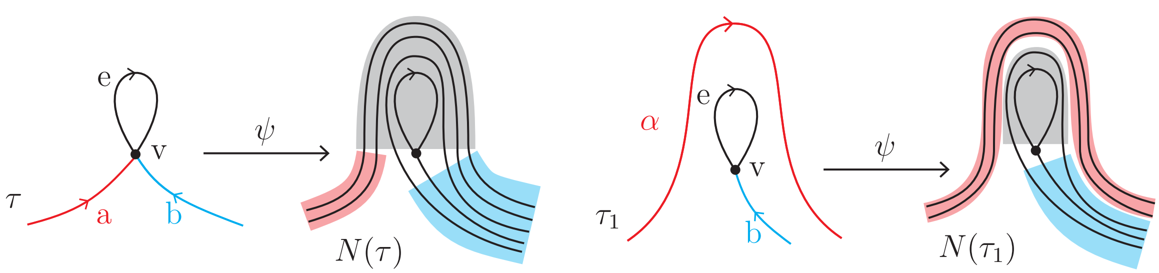


Figure 5.7: **Left:** another train track τ and map $f : \tau \to \tau$ for which v splits tightly to the right. **Right:** the train track τ_1 after r-splitting v.

In either case, we say that v *splits tightly*, or that v *t-splits*. See Figures 5.6 and 5.7.

If v splits tightly, we define a new train track that maps to τ by an elementary folding map. In this way, we view splitting as an inverse operation to folding. In what follows we will restrict our attention to the case that v tightly splits to the left: all definitions are analogous if v splits tightly to the right. To obtain these analogous statements and proofs, one need only replace all l's with r's and vice versa.

Suppose v l-splits. Define τ_v^l to be the standardly embedded train track obtained by deleting $l(v)$ and replacing it with a real edge α such that

1. As a directed edge, $\alpha(0) = l(v)(0)$ and $\alpha(1) = r(v_l)(1)$.

2. The edge α forms a bigon (i.e. a two-cusped disk) with the train path $l(v) \cdot \overline{r(v_l)}$, and there is an isotopy rel the punctures of $S^1_{0,n}$ so that α lies transverse to the leaves of the fibered neighborhood of τ.

The standardly embedded track τ_v^l comes equipped with a natural elementary folding map $p : \tau_v^l \to \tau$, defined by

$$p(x) = \begin{cases} x & \text{if } x \neq \alpha \\ l(v) \cdot \overline{r(v_l)} & \text{if } x = \alpha. \end{cases}$$

Definition 5.2.8. If v splits tightly to the left, then the map $p : \tau_v^l \to \tau$ is called a *tight left split* or an *l-split* of τ. We analogously define the *tight right split* or *r-split* $p : \tau_v^r \to \tau$.

Proposition 5.2.9. *Suppose (τ, ψ, f) is the data of a pseudo-Anosov carried by the standardly embedded train track τ:*

$$\begin{tikzcd} & \tau \\ \tau \arrow[ur, "f"] \arrow[r, "\psi"'] & \psi(\tau) \arrow[u, "collapse"'] \end{tikzcd}$$

If v l-splits, then τ_v^l carries ψ. Hence the above diagram may be completed to the commutative diagram

$$\begin{tikzcd} & \tau \\ \tau \arrow[ur, "f"] \arrow[r, "\psi"'] & \psi(\tau) \arrow[u, "collapse"'] \\ \tau_v^l \arrow[u, "p"] \arrow[r, "\psi"] \arrow[dr, "f_v^l"'] & \psi(\tau_v^l) \arrow[u, "p_\psi"'] \arrow[d, "collapse"] \\ & \tau_v^l \end{tikzcd}$$

where f_v^l is a train track map.

Proof. Let $F \subseteq S^1_{0,n}$ be a fibered surface for ψ from which the Bestvina-Handel algorithm produces τ. Let L, I, and R denote the strips of F collapsing to the (unoriented) edges $l(v), i_l$, and $r(v_l)$ of τ, respectively. Deleting L and replacing it with a strip A collapsing to α produces a new fibered surface F' from which the algorithm produces τ_v^l. The fact that F' is a fibered surface for ψ follows from the fact that v l-splits: any strip of $\psi(F)$ passing through L in fact passes through all three of L, I, and R in order, and hence after an isotopy we may arrange for the strip to pass through A instead. Furthermore, since α is isotopic to $l(v) \cdot i_l \cdot \overline{r(l_v)}$ and $\psi(L)$, $\psi(I)$, and $\psi(R)$ may be isotoped into F', it follows that $\psi(A)$ may be isotoped into F' as well. $\square$

Proposition 5.2.10. *Suppose that v l-splits and let M and M_v be the transition matrices of $f : \tau \to \tau$ and $f_v^l : \tau_v^l \to \tau_v^l$, respectively. Then*

$$M_v = P^{-1} M P,$$

where P is the transition matrix of the elementary folding map $p : \tau_v^l \to \tau$: that is, if $l(v)$ is the jth edge and $r(v_l)$ is the ith edge, then we have

$$P = D_{i,j},$$

where $D_{i,j}$ is the square matrix with 1's along the diagonal and in the (i, j)-entry, and 0's elsewhere.

Proof. We will argue that we have the following commutative diagram:

$$\begin{array}{ccc} \tau_v^l & \xrightarrow{f_v^l} & \tau_v^l \\ \big\downarrow{\scriptstyle p} & & \big\downarrow{\scriptstyle p} \\ \tau & \xrightarrow{f} & \tau \end{array}$$

From this the claim will follow, since each of the arrows is a Markov map, and so upon passing to transition matrices we obtain the relation

$$PM_v = MP.$$

Suppose x is an edge of τ_v^l. By the definition of p we have

$$(f \circ p)(x) = \begin{cases} f(x) & \text{if } x \neq \alpha \\ f(l(v)) \cdot f\left(\overline{r(v_l)}\right) & \text{if } x = \alpha. \end{cases}$$

On the other hand, we must understand the map $f_v^l : \tau_v^l \to \tau_v^l$ in order to analyze the composition $p \circ f_v^l$. For any edge $y \in \tau$, define $f'(y)$ to be the word obtained from the train path $f(y)$ by replacing each instance of $l(v) \cdot \overline{r(v_l)}$ with α and each instance of $r(v_l) \cdot \overline{l(v)}$ with $\overline{\alpha}$. In other words, $f'(x)$ is the unique word such that

$$p(f'(x)) = f(x).$$

If $x \neq \alpha$ is an edge of τ_v^l, then $f_v^l(x) = f'(x)$. If $x = \alpha$, then $f_v^l(x) = f_v^l(\alpha) = f'(l(v)) \cdot f'\left(\overline{r(v_l)}\right)$. In either case, we obtain the formula

$$(p \circ f_v^l)(x) = \begin{cases} f(x) & \text{if } x \neq \alpha \\ f(l(v)) \cdot f\left(\overline{r(v_l)}\right) & \text{if } x = \alpha. \end{cases}$$

This agrees with the formula for $f \circ p$, so the proof is complete. □

Recall that by the Perron-Frobenius theorem (Theorem 2.1.3), the dilatation of ψ is a simple eigenvalue of the transition matrix M, and there exists a positive right λ-eigenvector μ of M. For a fixed choice of μ we will denote by $\mu(x)$ the entry of μ corresponding to the real edge x.

Corollary 5.2.11. *Let (τ, ψ, f) be the data of a pseudo-Anosov carried by a standardly embedded train track. Let M be the transition matrix for $f : \tau \to \tau$, and let λ be the dilatation of ψ. Fix a positive right λ-eigenvector μ of M. If v l-splits then $\mu_v = P^{-1}\mu$ is a positive right λ-eigenvector of M_v. Consequently,*

$$\mu(l(v)) < \mu(r(v_l)).$$

Proof. Since $M_v = P^{-1}MP$, it immediately follows that $\mu_v = P^{-1}\mu$ is a right λ-eigenvector of M_v. At least one entry of μ_v is positive, since $\mu_v(\alpha) = \mu(l(v)) > 0$. Therefore μ_v is positive, since Theorem 2.1.3 states that λ is a simple eigenvalue of M_v and has a positive eigenvector.

To see that $\mu(l(v)) < \mu(r(v_l))$, observe that

$$0 < \mu_v(r(v_l)) = \mu(r(v_l)) - \mu(l(v)).$$

□

Example 5.2.12. Here is an extended example of a sequence of t-splits. Let (τ, ψ, f) be the data of the pseudo-Anosov represented in Figure 5.8. The transition matrix for $f : \tau \to \tau$ is

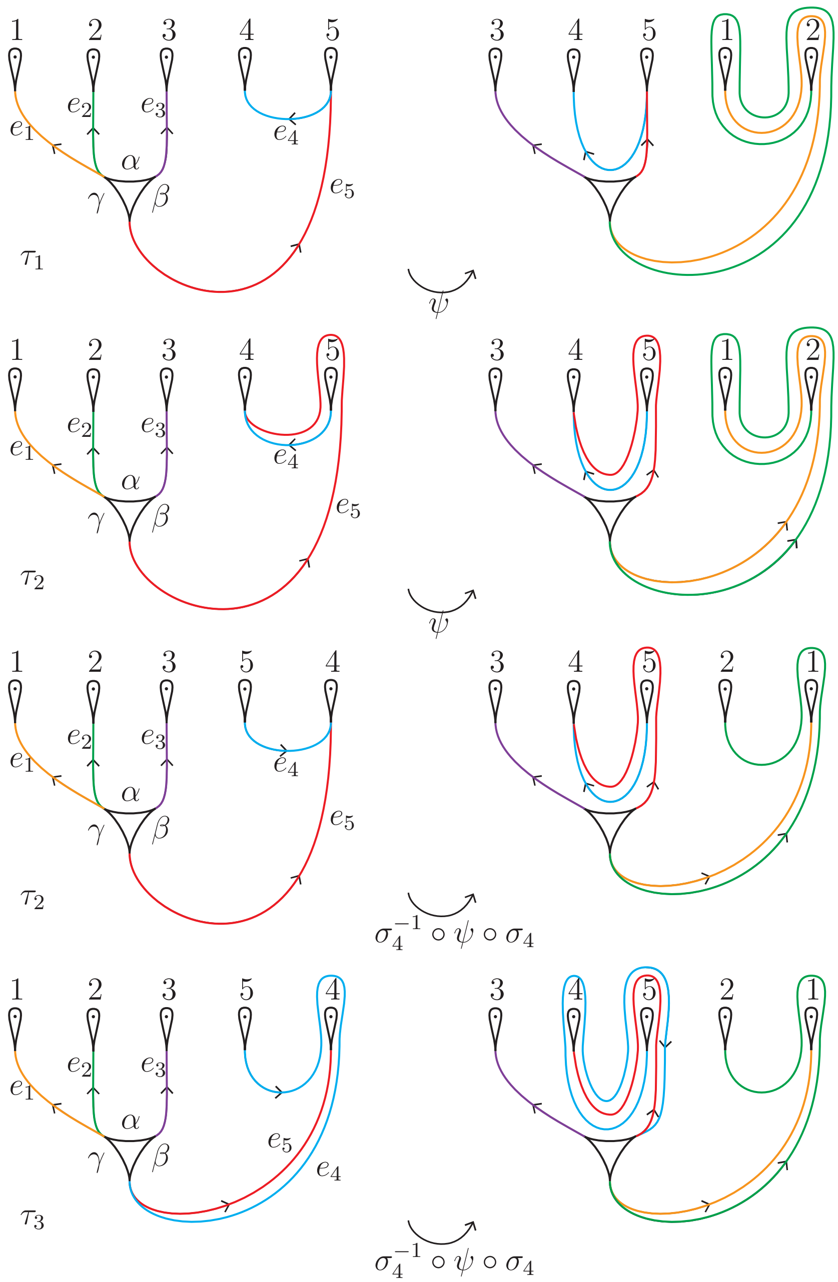


Figure 5.8: The track τ_1, τ_2 carries ψ. The track $\tau_2' = \sigma_4^{-1}(\tau_2)$ carries $\sigma_4^{-1} \circ \psi \circ \sigma_4$. The track τ_3 carries $\sigma_4^{-1} \circ \psi \circ \sigma_4$.

$$M_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The characteristic polynomial of M_1 is $\chi(t) = (t+1)(t^4-t^3-t^2-t+1)$. The dilatation of ψ is the root λ of this polynomial with largest absolute value, as in subsection 5.1. A positive right λ-eigenvector for M_1 is

$$\mu_1 = \begin{pmatrix} \mu_1(e_1) \\ \mu_1(e_2) \\ \mu_1(e_3) \\ \mu_1(e_4) \\ \mu_1(e_5) \end{pmatrix} = \begin{pmatrix} 2+5\lambda-\lambda^2-\lambda^3 \\ -2-2\lambda+\lambda^2+\lambda^3 \\ 1+\lambda+4\lambda^2-2\lambda^3 \\ -1-\lambda-\lambda^2+2\lambda^3 \\ 3 \end{pmatrix} = \begin{pmatrix} 2.537... \\ 2.628... \\ 4.370... \\ 4.526... \\ 3 \end{pmatrix}$$

One can see that the vertex at loop 5 splits tightly to the left. Performing this l-split produces the track τ_2, which also carries ψ. See Figure 5.8. The transition matrix of the l-split $p_1 : \tau_2 \to \tau_1$ is

$$P_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = D_{4,5}$$

and the transition matrix for $f_2 : \tau_2 \to \tau_2$ is

$$M_2 = P_1^{-1} M_1 P_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

which has right λ-eigenvector

$$\mu_2 = P_1^{-1}\mu_1 = \begin{pmatrix} \mu_2(e_1) \\ \mu_2(e_2) \\ \mu_2(e_3) \\ \mu_2(e_4) \\ \mu_2(e_5) \end{pmatrix} = \begin{pmatrix} \mu_1(e_1) \\ \mu_1(e_2) \\ \mu_1(e_3) \\ \mu_1(e_4) - \mu_1(e_5) \\ \mu_1(e_5) \end{pmatrix} = \begin{pmatrix} 2.537... \\ 2.628... \\ 4.370... \\ 1.526... \\ 3 \end{pmatrix}$$

We may conjugate by σ_4^{-1} to obtain the track τ_2', which is slightly easier to read. See Figure 5.8. This move is a standardizing braid move in the language of [KLS02]. It is not a t-split and is purely cosmetic. It does not alter the transition matrix or any other relevant dynamical information.

We can now see that the switch at loop 4 splits tightly to the right. Performing this r-split produces the track τ_3, which also carries $\sigma_4^{-1} \circ \psi \circ \sigma_4$. See Figure 5.8. The transition matrix of the r-split $p_2 : \tau_3 \to \tau_2$ is

$$P_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = D_{5,4}$$

and the transition matrix for $f_3 : \tau_3 \to \tau_3$ is

$$M_3 = P_2^{-1} M_2 P_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which has right λ-eigenvector

$$\mu_3 = P_2^{-1}\mu_2 = \begin{pmatrix} \mu_2(e_1) \\ \mu_2(e_2) \\ \mu_2(e_3) \\ \mu_2(e_4) \\ \mu_2(e_5) - \mu_2(e_4) \end{pmatrix} = \begin{pmatrix} \mu_1(e_1) \\ \mu_1(e_2) \\ \mu_1(e_3) \\ \mu_1(e_4) - \mu_1(e_5) \\ 2\mu_1(e_4) - \mu_1(e_4) \end{pmatrix} = \begin{pmatrix} 2.537... \\ 2.638... \\ 4.370... \\ 1.526... \\ 1.473... \end{pmatrix}.$$

5.2.2 Switch rigidity

In this section we investigate when a t-split is possible at a given switch, identifying the essential obstruction. We call this obstruction *switch rigidity* and show that it is uncommon. Indeed, the orbit of every switch contains a switch that is t-splittable (cf. Proposition 5.2.18).

Let v be a switch of the train track τ. Recall that $\mathrm{Lk}(v)$ is the set of edges of τ incident to v. A Markov map $f : \tau \to \tau$ induces a map $Df : \mathrm{Lk}(v) \to \mathrm{Lk}(f(v))$ as follows. Orient all edges in $\mathrm{Lk}(v)$ and $\mathrm{Lk}(f(v))$ away from v and $f(v)$, respectively. Then

$$Df(a) = b \text{ if } f(a) \text{ begins with } b.$$

As a consequence of the Bestvina-Handel algorithm, all elements of $R(v)$ belong to the same *gate*: that is, there exists an integer $k \geq 1$ such that $(Df)^k = D(f^k)$ is

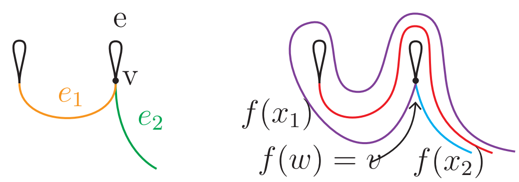


Figure 5.9: An example of a rigid switch. On the left is the switch, on the right the image of the map near the switch.

constant on $R(v)$.

Definition 5.2.13. Let $\tau \hookrightarrow S^1_{0,n}$ be standardly embedded, and let $f : \tau \to \tau$ be a train track map. Let v be a switch of τ such that $R(v)$ is not a singleton, and set $R(v) = \{e_1, \ldots, e_k\}$. Let w be the switch of τ such that $f(w) = v$. We say that v is *rigid* if there exist $x_1, \ldots, x_k \in R(w)$ such that

$$Df(x_i) = e_i \text{ for all } i.$$

Lemma 5.2.14. *Let (τ, ψ, f) be the data of the pseudo-Anosov ψ on $S^1_{0,n}$ carried by the standardly embedded τ. Let w be a switch of τ. Write $\alpha = r(w)$, $\beta = l(w)$, and $v = f(w)$. For any $c \in R(v)$ between $Df(\alpha)$ and $Df(\beta)$, there exists $\gamma \in R(w)$ such that $Df(\gamma) = c$. In other words, the set $Df(\mathrm{Lk}(w)) \subseteq \mathrm{Lk}(v)$ is connected.*

Proof. Suppose $c \in R(v)$ is between $Df(\alpha)$ and $Df(\beta)$. Since ψ is pseudo-Anosov, f is surjective. Hence there exists a real edge γ such that $f(\gamma)$ collapses onto c. But since ψ is a homeomorphism, $\psi(\gamma)$ cannot intersect $\psi(\alpha \cup \beta)$, so γ must be incident to w. In other words, $c = Df(\gamma)$. $\square$

Definition 5.2.15. We say a switch v of τ is a *loop switch* if it is incident to an infinitesimal loop.

The next lemma says that switch rigidity is the only barrier to the existence of a t-split at a loop switch. Note that if v is a loops switch, then $v_l = v_r = v$.

Lemma 5.2.16. *Let (τ, ψ, f) be the data of a pseudo-Anosov ψ on $S^1_{0,n}$ carried by the standardly embedded τ. Let v be a loop switch, and suppose that $R(v)$ is not a singleton. Then exactly one of the following three possibilities is true.*

1. *The switch v splits tightly to the left.*

2. *The switch v splits tightly to the right.*

3. *The switch v is rigid.*

Proof. Let w be the loop switch of τ such that $f(w) = v$. If either (1) or (2) holds then v cannot be rigid: for example, if v l-splits then there does not exist $x \in R(w)$ such that $Df(x) = l(v)$. On the other hand, if v is not rigid then Lemma 5.2.14 implies that at least one of $l(v), r(v)$ is not in the image $Df(\mathrm{Lk}(w))$.

Assume without loss of generality that $l(v) \notin Df(\mathrm{Lk}(w))$. Then any appearance of $l(v)$ in an image train path is in fact an appearance of $l(v) \cdot \overline{x}$, up to orientation. Here x is some edge in $R(v)$ that might vary. If x is always $r(v)$ then v l-splits. Otherwise, we claim that v r-splits.

Indeed, suppose that there exists a real edge $y \subseteq \tau$ such that $f(y)$ contains $l(v) \cdot \overline{x}$, up to orientation, for some real edge $x \neq r(v)$. Lemma 5.2.14 implies that $Df(\mathrm{Lk}(w))$ is a subset of the real edges between $l(v)$ and x. In particular, $r(v) \notin Df(\mathrm{Lk}(w))$. Let z be a real edge such that $f(z)$ contains $r(v)$, up to orientation. Since ψ is a homeomorphism and $f(z)$ is a train path, the appearance of $r(v)$ in $f(z)$ must be followed by $\overline{l(v)}$, due to the existence of $\psi(y)$. In other words, v r-splits.

Thus we have established that (1) or (2) holds if and only if (3) does not hold. It remains to show that (1) and (2) are mutually exclusive. Corollary 5.2.11 says that if v l-splits then $\mu(l(v)) < \mu(r(v))$. It follows that if (1) holds then (2) cannot. The proof is complete. $\square$

Lemma 5.2.16 says that if we cannot split at a particular switch v, then it is rigid. The natural next step is to consider the preimage switch v_1 causing v to be rigid.

If v_1 is also rigid, we look at its preimage v_2. It might happen that we never find a splittable switch. In this case, the periodic orbit of v consists of a cycle of rigid switches.

Definition 5.2.17. A *rigid cycle* of length k is a collection of rigid switches $v_1, \ldots, v_k \in \tau$ such that $f(v_j) = v_{j-1}$ for all j, where the indices are taken modulo k.

Proposition 5.2.18. *Rigid cycles do not exist.*

Proof. Let $v \in \tau$ be a switch. Since τ is standardly embedded, every element of $R(v)$ belongs to the same gate of v, hence there exists $k \geq 1$ such that $(Df)^k$ is constant on $R(v)$. In fact, for all $n \geq k$ we have that $(Df)^n$ is constant on $R(v)$. On the other hand, if v belonged to a rigid cycle of length n then $(Df)^n : R(v) \to R(v)$ would be the identity map, a contradiction. □

Corollary 5.2.19. *Let $v \in \tau$ be a switch such that $R(v)$ is not a singleton. Then some iterated preimage switch w of v is not rigid.*

It is well-known that if (τ, ψ, f) is the data of a pseudo-Anosov, then f permutes the infinitesimal k-gons for each k (cf. [BH95]). We obtain the following corollary, which will be of central importance in the following section. The *real valence* of a switch v is the cardinality of $R(v)$.

Corollary 5.2.20. *Let n_k denote the maximal real valence of a switch at an infinitesimal k-gon of τ, where $k \geq 1$. If $n_k > 1$ then there exists a switch of valence n_k at such a k-gon which is not rigid.*

Proof. The infinitesimal k-gons are permuted by f. If every such maximal valence switch is rigid, then they must form a rigid cycle, since real valence cannot decrease when passing to the preimage of a rigid switch. This is impossible, since rigid cycles do not exist. □

5.3 Train Tracks of Interval Type

While we have been discussing standardly embedded train tracks as graphs embedded in a punctured disc $S^1_{0,p}$, we may equivalently think of them as embedded in the sphere $S_{0,p+1}$ with a distinguished marked point coming from the collapsed boundary component. That is the perspective we will take in this chapter.

Definition 5.3.1. Let $\tau \hookrightarrow S_{0,p+1}$ be standardly embedded. We say that τ is *interval-like* if there are exactly 2 switches with real valence 1. Equivalently, after replacing each loop edge in τ around a puncture with a point, we get an interval.

In this section we prove Theorem H.

Theorem H. *Suppose ψ is a pseudo-Anosov with singularity data $(1^p, p-2)$. Then ψ has an invariant interval-like train track.*

Proof. Choose a standardly embedded invariant train track $\tau \hookrightarrow S_{0,p+1}$ for ψ. The connected components of $S_{0,p+1} \setminus \tau$ consist of p once-punctured monogons and a single $(p-2)$-gon, which we may consider punctured as well. Let E be the set of real edges of τ and let M be the transition matrix of ψ acting on τ. The matrix M is a $(p-1)$-dimensional primitive matrix with integer entries, and has a Perron-Frobenius eigenvector μ with strictly positive components $\mu(e) > 0$, $e \in E$.

Let $N(\tau)$ be the number of switches in τ of real valence 1. By Corollary 5.2.20, there is a switch $v \in \tau$ of maximal valence that admits a t-split. We may assume that τ is not interval-like. Then $N(\tau) \geq 3$ and the valence of v is at least 3. By Proposition 5.2.9, applying a t-split at v produces another train track τ' that carries ψ. The valence of v decreases by 1, and the valence of some adjacent switch increases by 1. Since v has valence at least 3, $N(\tau') \leq N(\tau)$.

Recall that the transition matrix associated to the pair (ψ, τ') is given by $M' = D_{i,j} M D_{i,j}^{-1}$ for some elementary matrix $D_{i,j}$, and the primitive eigenvector μ' of M' is given by $\mu'_i = \mu_i - \mu_j$ and $\mu'_k = \mu_k$ for all $k \neq i$. In particular, M' is another primitive matrix with integer entries and the same dominant eigenvalue as M, we have $\mu'_k \leq \mu_k$ for all k, and $\mu'_k < \mu_k$ for exactly one k.

Continue this procedure, and let τ_a, M_a, μ_a be the train track, transition matrix, and vector of edge weights obtained at step a. By Proposition 2.1.6, there exists $a > b$ such that $M_a = M_b$. Moreover, $\mu_a = r\mu_b$ for some $0 < r < 1$. Since the entries of μ_a are all strictly positive, this means that every edge's weight has been reduced by a t-split. But when we t-split over an edge adjacent to a switch of real valence 1, the number of such switches decreases by 1. Moreover, since we only split at switches of maximal valence, the number of such switches will never increase.

Thus τ_a has fewer switches of real valence 1 than τ_b. We repeat this argument until we find a train track with only two such switches. $\square$

Example 5.3.2. Here is an extended example of the algorithm used in the previous proof. We begin with the pseudo-Anosov $\psi : S_{0,6} \to S_{0,6}$ acting on an invariant train track τ_0. See Figure 5.10. In the notation of the proof, τ_0 has $N(\tau_0) = 4$ switches of valence 1, and $m = 4$ real edges. Labeling the edges 1 to 4 in counterclockwise order, we compute the following Markov matrix for $f_0 : \tau_0 \to \tau_0$:

$$M_0 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

The characteristic polynomial of M is $\chi_M(t) = t^4 - 3t^3 - t^2 - 3t + 1$. Thus the dominant root of this polynomial is the stretch factor of ψ: namely,

$$\lambda = \frac{1}{4}\left(3 + \sqrt{21} + \sqrt{14 + 6\sqrt{21}}\right) = 3.51... \tag{5.1}$$

Step 1. Corollary 5.2.20 tells us that the switch of order 4 in τ_0 is t-splittable. Indeed, after some inspection one sees that we this switch splits tightly to the right, i.e. it r-splits. Equivalently, edge 1 (red in the figure) splits over edge 4 (purple). From this operation we obtain a new invariant train track τ_1 carrying ψ, and a new train track map $f_1 : \tau_1 \to \tau_1$. See Figure 5.11.

Step 2. We can see that $N(\tau_1) = 3$, down one from before. Moreover, the switch of maximal valence in τ_1 has valence 3. Appealing to Corollary 5.2.20 again, we know that this switch is not rigid. It again splits to the right: edge 2 (orange in the figure) splits over edge 4 (purple). Applying this split gives us the new system $f_2 : \tau_2 \to \tau_2$. For visibility, we conjugate ψ by a normalizing braid β. See Figure 5.12. This does not change any topological properties of τ_2, nor any dynamical properties of ψ or f_2.

Step 3. We still have $N(\tau_2) = 3$, so we split again. Another r-split is in order, this time with edge 1 (red) splitting over edge 4 (purple). In Figure 5.13 we again conjugate by a braid to simplify the picture.

Step 4. We are nearly there. $N(\tau_3) = 3$, so we split once more. This time we apply an l-split, with edge 4 (purple) splitting over edge 3 (blue), producing the system $f_4 : \tau_4 \to \tau_4$. See Figure 5.14. This time $N(\tau_4) = 2$, and we have found an interval-type track carrying ψ.

To see the resulting interval map, pick an orientation for the spine (i.e., the union of the real edges) of τ_4. In Figure 5.15 we chose to begin at the second loop from the left. The pictured interval map is a uniform λ-expander, where λ is the stretch factor of ψ in Equation (5.1).

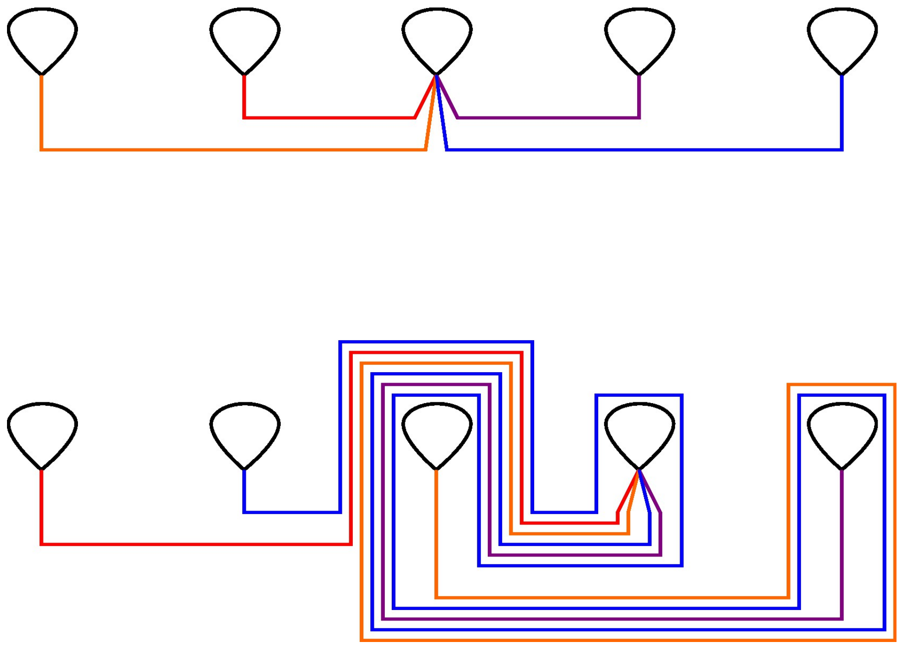


Figure 5.10: **Above:** An invariant train track $\tau = \tau_0$ for ψ. **Below:** The action of ψ on τ.

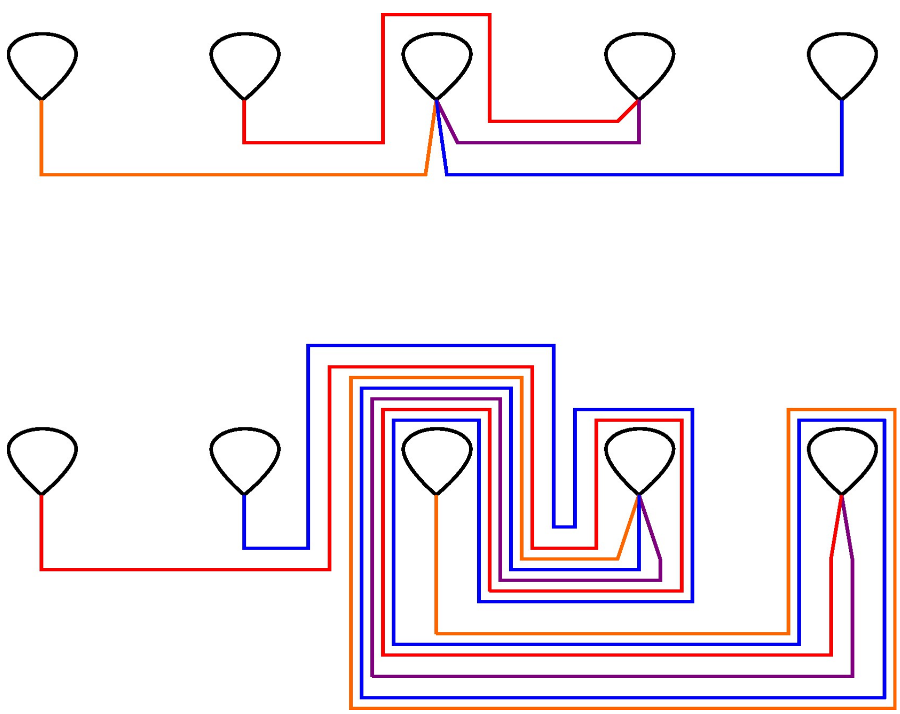


Figure 5.11: The action of ψ on τ_1, after the first t-split. The number of switches of valence 1 has decreased to $N(\tau_1) = 3$.

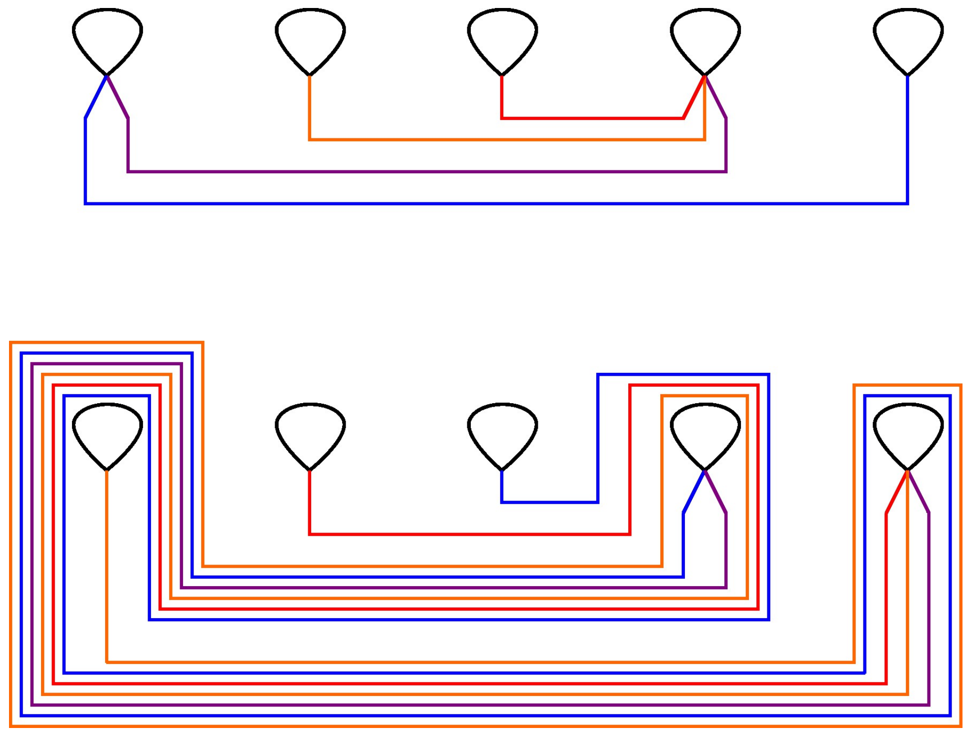


Figure 5.12: The action of ψ on τ_2, after the second t-split and conjugation by a normalizing braid β. The number of switches of valence 1 remains at 3.

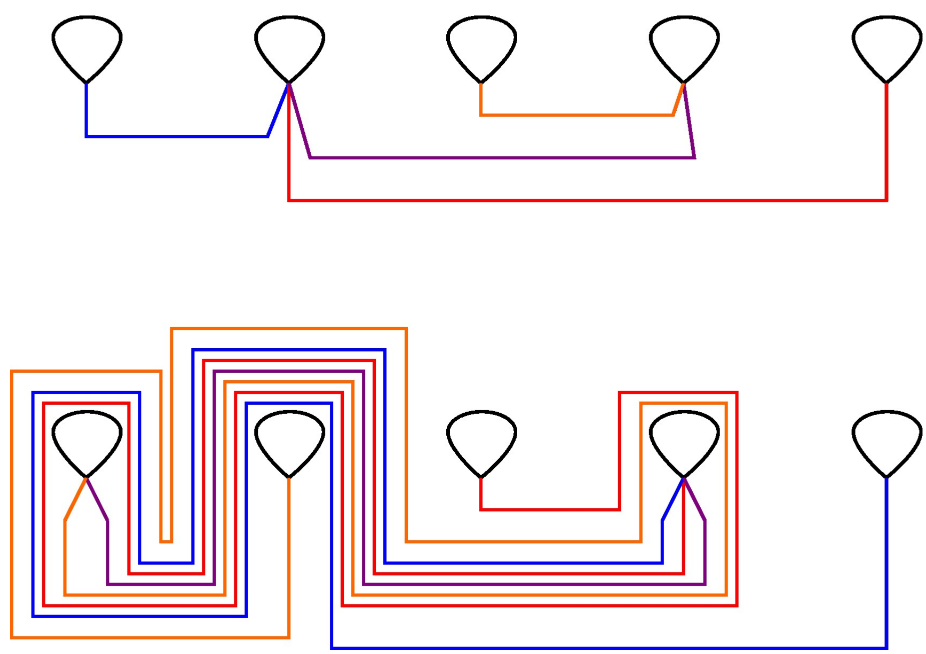


Figure 5.13: The action of ψ on τ_3, after the third t-split and conjugation by a normalizing braid. The number of switches of valence 1 remains at 3.

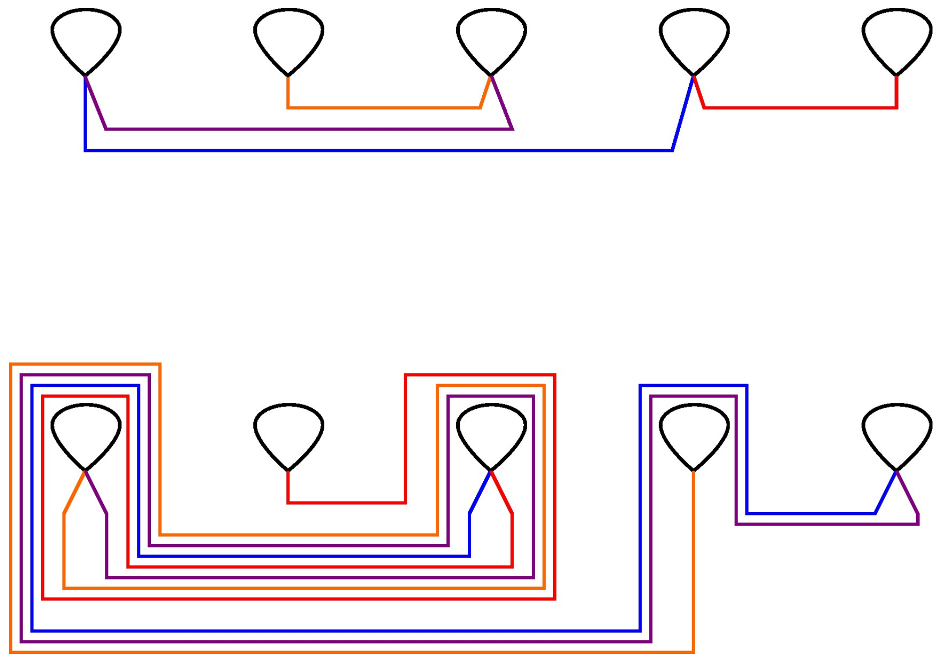


Figure 5.14: The action of ψ on τ_4, after the final t-split and conjugation by a braid word.

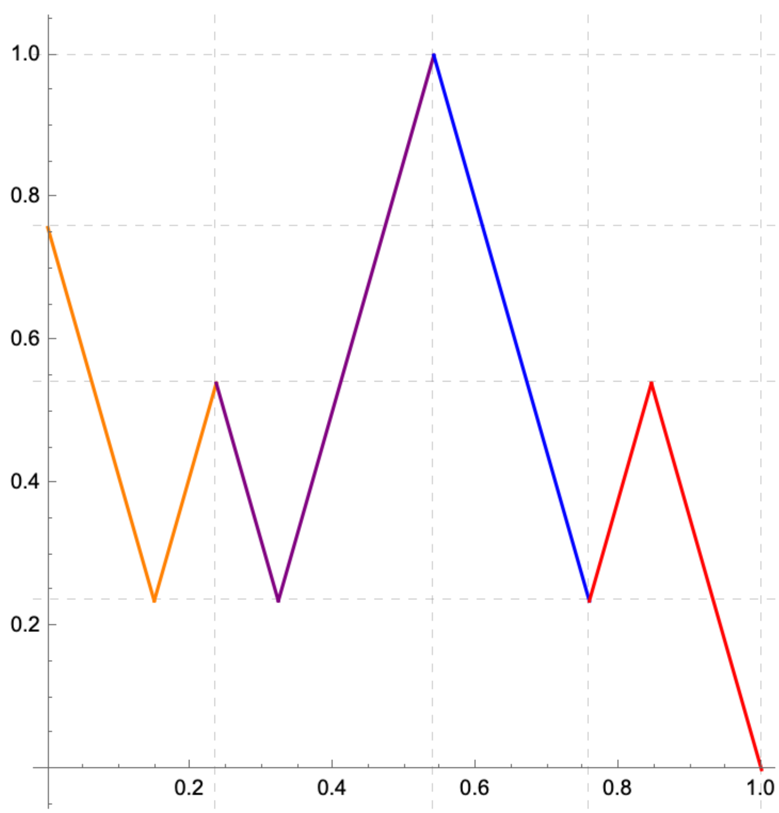


Figure 5.15: The interval map $f : I \to I$ induced by ψ. We chose 0 to correspond to the second loop from the left in Figure 5.14.

5.4 A lower bound on stretch factors

Let $\psi : S_{0,p+1} \to S_{0,p+1}$ be a pseudo-Anosov homeomorphism carried by an interval-like train track τ. Let $f : \tau \to \tau$ be the train track map induced by ψ. By removing the interiors of the infinitesimal edges of τ, we obtain a continuous map $f : I \to I$ where I is an interval. The decomposition of τ into real edges induces a decomposition

$$I = J_1 \cup \cdots \cup J_{p-1}$$

into subintervals with disjoint interiors, in order from left to right. Since the edges of τ arise from a Markov partition of ψ with weak Markov matrix M, each image $f(J_i)$ traverses finitely many of the subintervals J_j, making turns only at endpoints of subintervals. In fact, we have an almost-everywhere defined measurable semiconjugacy

$$\begin{tikzcd} S_{0,p+1} \arrow[r, "\psi"] \arrow[d, "\pi"] & S_{0,p+1} \arrow[d, "\pi"] \\ I \arrow[r, "f"] & I \end{tikzcd}$$

given by forgetting the vertical coordinate in each rectangle in the Markov partition for ψ. It is a fact that ψ is ergodic with respect to the measure μ on $S_{0,p+1}$ that restricts to the Lebesgue measure on each rectangle. The pushforward $\nu = \pi_*\mu$ is an invariant ergodic measure for f, and on each J_i is given by a nonzero multiple of Lebesgue measure determined by the height of the associated rectangle. In particular, the measure ν is equivalent to Lebesgue measure on I.

Let $J \subset I$ be a nonempty open interval. Since $\nu(J) > 0$ and f is ergodic with respect to ν, the set $\bigcup_{n=1}^{\infty} f^{-n}(J)$ has full ν-measure. This set is also the set of points in I whose forward orbit under f meets J. By letting J range over the subintervals of I with rational endpoints, we obtain a countable collection of subsets of full ν-measure. The intersection of these countably many sets is a full ν-measure subset of

I consisting of points whose forward f-orbit is dense. In particular, f has a dense orbit.

On the other hand, the existence of a dense orbit provides a lower bound on the topological entropy of f. The following statement is classical, although its earliest proofs are difficult to track down. It seems to have first been stated in [Blo82].

Proposition 5.4.1 (Proposition 4.70 in [Rue17])**.** *Let $f : I \to I$ be m-modal. If f has a dense orbit, then the topological entropy of f satisfies*

$$\mathrm{h}(f) \geq \log \sqrt{2}.$$

In particular, this bound is satisfies for f continuous and ergodic.

With this perspective in hand, we now reprove the theorem of Boissy-Lanneau. Recall

Proof of Theorem 5.0.4. The map $\psi : S \to S$ is a pseudo-Anosov with the same stretch factor λ as $\widetilde{\psi} : X \to X$, and has singularity data either $(1^{2g+2}, 2g)$ or $(1^{2g+1}, 2g-1)$. In particular, we may apply Theorem H to find an invariant interval-like track τ for ψ. From the discussion at the beginning of this section, we obtain an ergodic interval map $f : I \to I$ whose entropy is equal to $\log \lambda$. Proposition 5.4.1 now implies that

$$\log \lambda = \mathrm{h}(f) \geq \log \sqrt{2},$$

and hence $\lambda \geq \sqrt{2}$. This inequality must in fact be strict, since $\sqrt{2}$ is not a bi-Perron unit and hence cannot be the stretch factor of a pseudo-Anosov. $\square$

Bibliography

[Api18] Paul Apisa. $\mathrm{GL}_2\mathbb{R}$ orbit closures in hyperelliptic components of strata. *Duke Math. J.*, 167(4):679–742, 2018.

[BB07] Gavin Band and Philip Boyland. The Burau estimate for the entropy of a braid. *Algebr. Geom. Topol.*, 7:1345–1378, 2007.

[BBK12] Joan Birman, Peter Brinkmann, and Keiko Kawamuro. Polynomial invariants of pseudo-Anosov maps. *J. Topol. Anal.*, 4(1):13–47, 2012.

[BH95] M. Bestvina and M. Handel. Train-tracks for surface homeomorphisms. *Topology*, 34(1):109–140, 1995.

[Bir75] Joan S. Birman. *Erratum: "Braids, links, and mapping class groups" (Ann. of Math. Studies, No. 82, Princeton Univ. Press, Princeton, N. J., 1974).* Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1975. Based on lecture notes by James Cannon.

[BL12] Corentin Boissy and Erwan Lanneau. Pseudo-Anosov homeomorphisms on translation surfaces in hyperelliptic components have large entropy. *Geom. Funct. Anal.*, 22(1):74–106, 2012.

[Blo82] A. M. Blokh. Sensitive mappings of an interval. *Uspekhi Mat. Nauk*, 37(2(224)):189–190, 1982.

[BRW16] Hyungryul Baik, Ahmad Rafiqi, and Chenxi Wu. Constructing pseudo-Anosov maps with given dilatations. *Geom. Dedicata*, 180:39–48, 2016.

[CH08] Jin-Hwan Cho and Ji-Young Ham. The minimal dilatation of a genus-two surface. *Experimental Mathematics*, 17(3):257 – 267, 2008.

[dC05] André de Carvalho. Extensions, quotients and generalized pseudo-Anosov maps. In *Graphs and patterns in mathematics and theoretical physics*, volume 73 of *Proc. Sympos. Pure Math.*, pages 315–338. Amer. Math. Soc., Providence, RI, 2005.

[dCH04] André de Carvalho and Toby Hall. Unimodal generalized pseudo-Anosov maps. *Geom. Topol.*, 8:1127–1188, 2004.

[Far21] Ethan Farber. Constructing pseudo-Anosovs from expanding interval maps. *arXiv:2101.01721*, 2021.

[Fri85] David Fried. Growth rate of surface homeomorphisms and flow equivalence. *Ergodic Theory Dynam. Systems*, 5(4):539–563, 1985.

[FRW22] Ethan Farber, Braeden Reinoso, and Luya Wang. Fixed point-free pseudo-anosovs and the cinquefoil, 2022.

[Gan59] F. R. Gantmacher. *The theory of matrices. Vols. 1, 2.* Chelsea Publishing Co., New York, 1959. Translated by K. A. Hirsch.

[Hal94] Toby Hall. The creation of horseshoes. *Nonlinearity*, 7(3):861–924, 1994.

[Hat22] Allen Hatcher. *Topology of Numbers.* American Mathematical Society, Providence, Rhode Island, 2022.

[HS07] Ji-Young Ham and Won Taek Song. The minimum dilatation of pseudo-ansonov 5-braids. *Experimental Mathematics*, 16(2):167 – 180, 2007.

[Khi97] A. Ya. Khinchin. *Continued fractions.* Dover Publications, Inc., Mineola, NY, russian edition, 1997. With a preface by B. V. Gnedenko, Reprint of the 1964 translation.

[KLS02] Ki Hyoung Ko, Jérôme E. Los, and Won Taek Song. Entropies of braids. *Journal of Knot Theory and Its Ramifications*, 11(04):647–666, 2002.

[Lin92] Douglas Lind. Matrices of Perron numbers. *J. Number Theory*, 40(2):211–217, 1992.

[Los10] Jérôme Los. Infinite sequence of fixed-point free pseudo-Anosov homeomorphisms. *Ergodic Theory Dynam. Systems*, 30(6):1739–1755, 2010.

[M0̈8] Martin Möller. Finiteness results for Teichmüller curves. *Ann. Inst. Fourier (Grenoble)*, 58(1):63–83, 2008.

[M1̈3] Martin Möller. Teichmüller curves, mainly from the viewpoint of algebraic geometry. In *Moduli spaces of Riemann surfaces*, volume 20 of *IAS/Park City Math. Ser.*, pages 267–318. Amer. Math. Soc., Providence, RI, 2013.

[MS80] M. Misiurewicz and W. Szlenk. Entropy of piecewise monotone mappings. *Studia Math.*, 67(1):45–63, 1980.

[MT88] John Milnor and William Thurston. On iterated maps of the interval. In *Dynamical systems (College Park, MD, 1986–87)*, volume 1342 of *Lecture Notes in Math.*, pages 465–563. Springer, Berlin, 1988.

[MT00] John Milnor and Charles Tresser. On entropy and monotonicity for real cubic maps. *Comm. Math. Phys.*, 209(1):123–178, 2000. With an appendix by Adrien Douady and Pierrette Sentenac.

[Par64] William Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.*, 112:55–66, 1964.

[Par66] William Parry. Symbolic dynamics and transformations of the unit interval. *Trans. Amer. Math. Soc.*, 122:368–378, 1966.

[Rue17] Sylvie Ruette. *Chaos on the interval*, volume 67 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2017.

[Sal83] Raphaël Salem. Algebraic numbers and Fourier analysis. In *Selected reprints*, Wadsworth Math. Ser., pages iii+68. Wadsworth, Belmont, CA, 1983.

[Shi16] Hyunshik Shin. Algebraic degrees of stretch factors in mapping class groups. *Algebr. Geom. Topol.*, 16(3):1567–1584, 2016.

[Suz17] Shintaro Suzuki. Artin-Mazur zeta functions of generalized beta-transformations. *Kyushu J. Math.*, 71(1):85–103, 2017.

[Tho16] Daniel J Thompson. Generalized beta-transformations and the entropy of unimodal maps. *arXiv preprint arXiv:1602.03518*, 2016.

[Thu88] William P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2):417–431, 1988.

[Thu14] William P. Thurston. Entropy in dimension one. In *Frontiers in complex dynamics*, volume 51 of *Princeton Math. Ser.*, pages 339–384. Princeton Univ. Press, Princeton, NJ, 2014.

[Wil75] Keith M. Wilkinson. Ergodic properties of a class of piecewise linear transformations. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:303–328, 1974/75.