ON DILATATIONS OF SURFACE AUTOMORPHISMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ahmad Rafiqi

August 2018

ON DILATATIONS OF SURFACE AUTOMORPHISMS

Ahmad Rafiqi, Ph.D.

Cornell University 2018

Suppose S is a compact topological surface without boundary, oriented and connected. In §1 we go over Thurston's classification of *automorphisms* of S: continuous functions $f : S \to S$ with continuous inverses. The classification provides an $f_\#$ homotopic to a given automorphism f, which is either: periodic ($f_\#^n$ is the identity for some $n \geq 1$); or $f_\#$ fixes a finite union of disjoint circles $\gamma_i : S^1 \hookrightarrow S$; or is what Thurston calls pseudo-Anosov (pA). In the pA case, there is a finite number of points on S whose complement carries two transverse foliations by 1-submanifolds each with a holonomy invariant measure on transverse arcs. Such a structure on S can be described in terms of a quadratic differential on a Riemann surface structure on it which we go over in §1.2. Further, there is a number $\lambda_f \geq 1$ (called the *dilatation* or *stretch factor*) such that the two measured foliations are stretched and shrunk respectively by λ_f under $f_\#$.

When the foliations of a pA map f can be oriented consistently in neighborhoods on S, the dilatation λ_f is an eigenvalue of the induced action f_* on the homology group $H_1(S;\mathbb{Z})$. The numbers λ_f thus satisfy monic polynomials with integer coefficients. Fried [10] showed that λ_f is a unit in $\mathbb{Z}[\lambda_f]$ and that the Galois conjugates of λ_f (excepts perhaps one of $\pm\lambda_f^{-1}$) lie in the open annulus $\{z \in \mathbb{C} : 1/\lambda_f < |z| < \lambda_f\}$. Numbers satisfying these properties are called *biPerron*. If the Galois conjugates of an algebraic integer $\lambda \geq 1$ are only in the disk $\{|z| < \lambda\}$, it is called a *Perron* number. We describe Fried's proof and some properties of

dilatations in §1.3. In [10, Problem 2], Fried asked whether some power of a biPerron unit is always a surface automorphism stretch factor. This thesis is an attempt to solve this problem.

In §2 we go over a well-known rectangular decomposition of a Riemann surface with a quadratic differential. When the associated foliations are orientable, the quadratic differential is the square of an Abelian differential and we study this latter structure on a Riemann surface in some detail.

A square matrix A of non-negative entries is called *mixing* if some power A^n has only positive entries. If only the sum $A + A^2 + ...$ is positive, A is *ergodic*. By the Perron-Frobenius theorem, an ergodic matrix has a real eigenvalue $\lambda > 0$ (called the *Perron root*) bigger than all its other eigenvalues in absolute value, and whose eigenspace is 1-dimensional. If the matrix has integer entries, the Perron root is a Perron number. Lind [20] showed the converse that all Perron numbers are eigenvalues of ergodic integer matrices. In Chapter 3 (see [3]), under additional hypotheses on an erdogic matrix A with entries in $\{0, 1\}$ we construct a closed orientable surface S of genus $g \leq \dim A/2$ with a self-homeomorphism f such that the dilatation λ_f is the Perron root of A (which is therefore biPerron).

On the other hand, Hamenstädt [12] showed that out of dilatations smaller than $R > 0$ on a surface of fixed genus g, the proportion of those that have only real Galois conjugates approaches 1 as $R \to \infty$. This suggests Fried's conjecture may be false and motivated our study of the asymptotic behavior of biPerron units which we present in chapter 4. Let $B_g(R)$ be the set of bi-Perron units no larger than R whose minimal polynomial has degree at most $2g$, and say $D_g(R)$

is the set of dilatations no larger than R of pseudo-Anosov maps with orientable invariant foliations on S_g.

Eskin-Mirzakhani-Rafi (2016) and Hamenstädt (2016) independently showed that the number of periodic orbits of length less than $\log(R)$ for the Teichmüller flow on the moduli space of area one Abelian differentials on S_g grows like $R^{4g-3}/\log(R)$ as $R \to \infty$. This is an upper bound for the number of dilatations, $|D_g(R)|$. We showed (with H. Baik and C. Wu) that $|B_g(R)|$ grows like $R^{g(g+1)/2}$ as $R \to \infty$. Since dilatations $D_g(R)$ form a subset of biPerron units $B_g(R)$, we get that for $g \geq 6$ the proportion of dilatations $D_g(R)$ inside biPerron units $B_g(R)$ approaches 0 as $R \to \infty$. Our result does not disprove Fried's conjecture.

For the lower genera (2, 3, 4 and 5) we get an interesting result as well. If there are exactly n closed geodesics in Moduli space of the same length $\log(\lambda)$, we say n is the *multiplicity* of each of these geodesics. For geodesics bounded in length, the proportion of those closed geodesics that have multiplicity greater than some positive integer k approaches 1 as the bound in length approaches ∞. That is, asymptotically almost all geodesics in Moduli spaces for these genera have arbitrarily high multiplicities.

BIOGRAPHICAL SKETCH

Ahmad Rafiqi was born in the town of Sopore in the region of Kashmir in India. After finishing high school in Srinagar, Kashmir, joined Stony Brook University in New York for a bachelors degree in mathematics followed by Cornell University for this PhD.

Results presented here were not ideas construed in a systematic or clever way, rather they were noticed; often after many months spent on the same page. The structure in these ideal settings *is* beautiful. I would dedicate my dissertation to the cause of this structure. To me that is The Real of Plato, the Dao of Lao-Tzu. So I begin in the name of God.

ACKNOWLEDGEMENTS

Professor John Hubbard has been extremely helpful, willing to lend an ear to my thoughts for long amounts of time and provide many helpful ideas. It has been a pleasure having him as my advisor.

Chapters 3 and 4 are based on papers published with Hyungryul Baik and Chenxi Wu and as such many of the ideas in these chapters are theirs, and I am indebted to them for many helpful discussions. In particular the construction of surfaces associated to odd-block matrices first appears in [3], written by the author in collaboration with them.

For helpful discussions and many helpful ideas I would also like to thank Dylan Thurston, Joshua P. Bowman, Chris Leininger, Sarah Koch, Eriko Hironaka, Curt McMullen and Balázs Strenner. I would also like to thank Professors Allen Hatcher and Jason Manning for their help in the writing and editing of this thesis.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

We will be concerned with spaces that are 2-dimensional at every point, like a bubble of water. We take a *surface* S to be a topological space whose every point has a neighborhood smoothly equivalent to an open set in the plane $\mathbb{R}^2$. We will assume S is Hausdorff (so distinct points have disjoint open neighborhoods), connected, and has a countable basis for its topology.

If for every collection of open sets covering S, some finite subcollection of the open sets also covers S, S is called *compact*. It is called *orientable* if a normal can be chosen to its tangent plane at every point, consistently all over the surface. Compact orientable surfaces are classified by genus and are the sphere S^2, the torus $S^1 \times S^1$ (surface of a donut), and higher genus surfaces, where a surface of genus g, denoted S_g, is a connect sum of g tori.

1.1 Surface automorphisms

Let $S = S_g$ be the surface of genus g, oriented. A continuous function $f : S \to S$ for which there is a continuous inverse f^{-1} is called a *homeomorphism* of S.

The *isotopy class* of f, denoted $[f]$, is the equivalence class of homeomorphisms $g : S \to S$ for which there is a continuous map

$$H : S \times [0, 1] \to S,$$

so that for each $t := [0, 1]$, the time t restriction $H\big|_{S\times\{t\}}$ is a homeomorphism, f and g being the time 0 and 1 restrictions respectively. We will give a brief description of these classes for genus 1, and a related positive quadratic unit invariant associated to each class before describing Thurston's classification of these isotopy classes for a higher genus surface.

1.1.1 Torus automorphisms

For the torus, $T = \mathbb{R}^2/\mathbb{Z}^2$, isotopy classes of orientation preserving automorphisms correspond to the special linear group $SL_2(\mathbb{Z})$: Every $A \in SL_2(\mathbb{Z})$ maps the integer lattice $\mathbb{Z}^2$ to itself, and as a result factors as a map of the torus. For the converse note that projections of the x, y-axes give a basis for the homology group $H_1(T;\mathbb{Z}) \cong \mathbb{Z}^2$. Now, a map $f : T \to T$ induces a homomorphism f_* of the first homology group $H_1(T;\mathbb{Z}) \cong \mathbb{Z}^2$. f_* is independent of the representative as projections of images under $f \simeq f'$ of the basis curves onto the x, y-axes are homotopic. f_* has inverse $(f^{-1})_*$, and $\det f_* = 1$ being orientation preserving, so it is an element of $SL_2(\mathbb{Z})$. Seen as a linear map of $\mathbb{R}^2$, f_* is homotopic via a straight line to a lift of the homeomorphism of the torus (assuming f fixes a point of T, but this can be arranged by isotopy) and this homotopy is evidently periodic so it descends to the torus. The set of isotopy classes of automorphisms of S forms a group under composition called the *mapping class group* of S, denoted $\mathcal{MCG}(S)$. Since $(f \circ g)_* = f_* \circ g_*$, we see that $\mathcal{MCG}(T) \cong SL_2(\mathbb{Z})$.

So one looks at the elements of $SL_2(\mathbb{Z})$ to classify torus automorphisms. For a square matrix A, let $tr(A)$ denote its trace, i.e. the sum of its diagonal entries. If $f_*^n = \mathrm{I}_2 = (f^n)_*$, for some $n > 0$, f^n is isotopic to the identity on T. This happens exactly when the integral matrix f_* satisfies $|tr(f_*)| \leq 1$. When $tr(f_*) = \pm 2$, there is an eigenvalue of ± 1 and thus a curve is mapped to itself up to isotopy. f is then isotopic to a power of the Dehn twist around the curve preserved. If the power is nonzero, f has infinite order.

When $|tr(f_*)| > 2$, f_* has two eigenspaces F^u, F^s of irrational slope, scaled by reciprocal eigenvalues $\lambda > 1 > 1/\lambda$. Translates of F^u, F^s descend to dense injectively immersed 1-submanifolds on T, transverse to each other. So T possesses two transverse 1-foliations one stretched by λ and one compressed by it, under the map $\overline{f_*}$. Such a map $\overline{f_*}$ is called *Anosov*, after D. V. Anosov. The eigenvalue $\lambda > 1$ is the *dilatation* of $[f]$.

When $|tr(f_*)| > 2$, λ is a root of $\chi_{f_*}(x) = x^2 - tr(f)x + 1$, so $tr(f) - \lambda$ is the inverse of λ, and they are both units in the ring of algebraic integers $\mathbb{Z}[\lambda] = \mathbb{Z}[x]/(\chi_{f_*}(x))$ contained in the index 2 field extension $\mathbb{Q}(\lambda)$ of $\mathbb{Q}$.

Suppose λ is a positive quadratic algebraic unit. I.e. it solves an irreducible quadratic equation $p(\lambda) = \lambda^2 + m\lambda + n = 0$, and has an inverse $a\lambda + b \in \mathbb{Z}[\lambda]$. Then $\lambda(a\lambda + b) = 1 \implies an = -1$. So the product of the roots of p is ± 1 and as λ is positive, it is bigger than 1. The linear map $C_\lambda = \left(\begin{smallmatrix} 0 & \pm 1 \\ 1 & -m \end{smallmatrix}\right)$ factors through the projection $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ as an invertible map of the torus, orientation reversing when $n = -1$. Thus every positive quadratic algebraic unit is the dilatation of an Anosov map of the torus.

1.1.2 Surfaces of higher genus

When $g > 1$, Thurston's classification gives a respresentative $\phi : S_g \to S_g$ of each isotopy class $[f]$ of automorphisms of S_g and these he classified as one of the following three types.

If ϕ^n is the identity for some $n > 0$, ϕ is called *periodic*.

If a collection $\Gamma = \coprod_{i=1}^{k} \gamma_i$ of disjoint embeddings $S^1 \overset{\gamma_i}{\hookrightarrow} S$ satisfies $\phi(\Gamma) = \Gamma$, ϕ is called *reducible*. This is because in this case some power ϕ^K fixes each component C_i of the complement of Γ in S, and $\phi\big|_{C_i}$ is reduced to the first or the third class below.

The final type is what Thurston called a *pseudo-Anosov* representative $\phi \in [f]$. The rest of this section is the definition.

A foliation on a surface of negative Euler characteristic similar to the foliations above on the torus would imply a nowhere vanishing vector field on the orientation double cover of the foliation contradicting $\chi(S) < 0$. The parallel notion for these surfaces is of a *singular foliation* $\mathcal{F}$, which is a decomposition into injectively immersed 1-submanifolds of $S \setminus N$, where $N \subset S$ is finite. Points $p \in N$ are called *singular* for $\mathcal{F}$ and the structure of the leaves around such a p is as suggested by figure 1.

Figure 1: A 3- and 4-pronged singularity. Singular foliations are k-pronged at points of $N, k \geq 3$.

In the topology inherited by the 1-submanifolds injectively immersed in the surface, maximal connected components are called *leaves* of the foliation. A leaf emanating from a point of N is called a *singular leaf*. A singular foliation with a holonomy invariant measure μ on the space of transversals to the foliation is called a *measured foliation*.

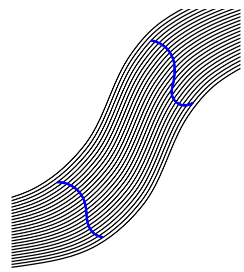


Figure 1: A holonomy of a transverse arc (blue).

That is, every arc that is transverse to the foliation is assigned a positive measure, and if the arc is moved by a homotopy that keeps it transverse to the foliation and each endpoint of the arc remains in the same leaf throughout the homotopy, then the measure is unchanged.

A measured foliation $(\mathcal{F}, \mu)$ is *scaled* by λ under $\phi : S \to S$ if ϕ takes the singular set of $\mathcal{F}$ to itself, leaves are stretched by λ and sent to other leaves, and for every arc α transverse to $\mathcal{F}$, $\lambda\mu(\phi(\alpha)) = \mu(\alpha)$.

Finally, ϕ is called ***pseudo-Anosov*** if there is a number $\lambda > 1$ and S possesses two measured foliations $(\mathcal{F}^s, \mu^s), (\mathcal{F}^u, \mu^u)$, each with the same singular set N, and with leaves transverse to each other, so that under ϕ, $(\mathcal{F}^u, \mu^u)$ is scaled by λ while $(\mathcal{F}^s, \mu^s)$ is scaled by $1/\lambda$. The number λ appearing here is called the *dilatation* of ϕ. Thurston proved the remarkable theorem:

Theorem 1 (Thurston)**.** *Any isotopy class of maps $f : S_g \to S_g$, $g > 1$ contains a representative that is either periodic, reducible, or pseudo-Anosov. Further, a pseudo-Anosov class cannot contain a periodic or reducible representative.*

1.2 Quadratic differentials

"I have in the first place, therefore, set myself the task of constructing the notion of a multiply extended magnitude out of the general notions of magnitude. It will follow from this that a multiply extended magnitude is capable of different measure-relations, and consequently that space is only a particular case of a triply extended magnitude.

...I restrict myself, therefore, to those manifoldnesses in which the line element is expressed as the square root of a quadric differential expression."

Bernhard Riemann (Translated by W. K. Clifford)
On the Hypotheses which lie at the Bases of Geometry

Another way to look at this is in terms of quadratic differentials (defined below) which encode the rigid almost-flat structure on an orientable surface provided by the two transverse measured foliations of Thurston's pseudo-Anosov representatives, whenever defined by them. First, X is a *Riemann surface* if for some open cover $\{U_i\}$ defining the topology of X, there are homeomorphisms $z_i : U_i \to \mathbb{C}$ (called *coordinates*), such that the transition functions $\psi_{ij} = z_i \circ z_j^{-1}$ are complex differentiable (*conformal*). This is to say on each $U_{ij} := U_i \cap U_j$, ψ_{ij} is analytic as a map of $z_j(U_{ij}) \subset \mathbb{R}^2$ and for $p \in U_{ij}$ the derivative $D\psi_{ij}\Big|_{z_j(p)} : T_{z_j(p)}\mathbb{C} \to T_{z_i(p)}\mathbb{C}$ is complex linear, (so of the form $v \mapsto cv$ for all v and a fixed c in $T_{z_j(p)}\mathbb{C} \cong \mathbb{C}$).

Circles $\{|z| < \epsilon\}$ in the tangent space at $z_j(p)$ are sent to circles in the tangent space at $z_i(p)$ and scaled by $\left|\frac{d\psi_{ij}}{dz_j}\right| = |c| \neq 0$. In particular, there is a well-defined notion of angle of intersection of two smooth intersecting curves on a Riemann surface X. Since conformal maps between domains in $\mathbb{C}$ can change the length of a curve, there is no metric chosen on X as such. A quadratic differential q equips X with a rigid 'distance' between points (and area) which is flat (isometric to $\mathbb{R}^2$ with the Euclidean metric $dx^2 + dy^2$) everywhere in the complement of a discrete set $N \subset X$. Even though the curvature is constant outside a finite set, this is not the hyperbolic metric one puts on S by writing it as a quotient of the hyperbolic plane $\mathbb{H}^2$ by a Fuchsian group.

At a point $p \in X$, a quadratic differential q_p is defined as a symmetric 2-form on $T_pX \cong \mathbb{C}$, i.e q_p is a $\mathbb{C}$-bilinear map $T_pX \times T_pX \to \mathbb{C}$, with $q_p(u, v) = q_p(v, u)$. Positive-definite symmetric 2-forms are called *Riemannian metrics* on real manifolds, (the dot product in $\mathbb{R}^n$ for example). Under an identification of T_pX with $\mathbb{C}$, bilinearity gives $q_p(v, w) = q_p(1, 1) \cdot vw$, so each q_p is of the form $c(p)\,dz^2$. A *holomorphic (resp. meromorphic) quadratic differential* q on X is a choice of a 2-form $q_{z(p)} = c(z(p))\,dz^2$ for every p in every coordinate $z : U \to \mathbb{C}$ of a conformal open cover $\mathfrak{U}$, so that each $c : z(U) \to \mathbb{C}$ is holomorphic (resp. meromorphic) and the 2-form $c(z)\,dz^2$ is invariant under coordinate changes: i.e. if ψ is a transition map of conformal coordinates $z = \psi(z')$ then $c'(z')\,dz'^2$ equals the pullback $\psi^*(c(z)\,dz^2) = c\,(\psi(z'))\left(\frac{d\psi}{dz'}\right)^2 dz'^2$. Write $c(z)$ as $q(z)$ and for tangent vectors v define $q(v) := q(v, v)$, with slight abuse of notation. The coefficient functions $q(z)$, $q'(z')$ in conformal coordinates z, z' are related by

$$q'(z') = q\,(z)\left(\frac{d\psi}{dz'}\right)^2 .$$

Note that the forms q are $\mathbb{C}$-valued, which is not a metric on the underlying $\mathbb{R}^2$ structure, but its absolute value $|q(z)|\,dx\,dy$ (denoted $|q|$) is invariant under coordinate changes and does provide an area element. This follows since the local coefficients $q(z)$ of q are related by the Jacobian of the transition maps. In terms of the $\mathbb{R}^2$ structure of the coordinate $z : U \to \mathbb{C}$, $|q|$ corresponds to the singular Riemannian metric $ds^2 = |q(z)|(dx^2 + dy^2)$.

As $\frac{d\psi}{dz'} \neq 0$, the discrete set $N \subset X$ of the zeros and poles of the meromorphic q is well-defined. N is finite in our case, being a discrete subset of a compact space. The ratio of two quadratic differentials is a meromorphic function $X \to \mathbb{CP}^1$, which if not constant has as many zeros as poles. It follows that each holomorphic q not identically zero has the same number of zeros counted with multiplicity. This number is $4g - 4$ on a surface of genus g, (see e.g. [15, §5.3]).

Example 2 (General quadratic differentials)**.** *Non-compact Riemann surfaces have many quadratic differentials, e.g. on a Riemann surface $U \subset \mathbb{C}$ open, any meromorphic function $q : U \to \mathbb{C}$ defines a quadratic differential $q(z)\,dz^2$.*

On $X = \mathbb{CP}^1$, let $q = dz^2$ in the coordinate z around 0. Then for the coordinate around $w = \frac{1}{z} = 0$, q is given by $\left(\frac{-1}{w^2}\right)^2 dw^2 = \frac{dw^2}{w^4}$, and has a pole of order 4 at $w = 0$. q is a meromorphic quadratic differential on $\mathbb{CP}^1$, though it is not integrable*: $\int_X |q|$ is not finite (in this case it is $\int_{\mathbb{R}^2} dx\,dy$, the area of the flat plane). We will construct some integrable quadratic differentials on $\mathbb{CP}^1$ in chapter 3.*

On a torus $X = \mathbb{C}/(\mathbb{Z} \oplus \tau\mathbb{Z})$, $\mathtt{Im}(\tau) \neq 0$, any holomorphic function lifts to a bounded entire function, so it is constant. Thus the only holomorphic quadratic differentials on X are given by $c\,dz^2$. All of these are squares of global holomorphic 1-forms ($\pm\sqrt{c}\,dz$), with no zeros. □

For $p \in X$ with $q_p \neq 0$, one can by translation of a given coordinate choose a simply connected conformal coordinate $z : U \to \mathbb{C}$ with $0 \notin z(U)$. Now choose a branch of $\sqrt{q(z)}$ on $z(U)$. Then in the new coordinate $w = \int_{z(p)}^{w} \sqrt{q(z)}\,dz$ we have $q(w) = 1$ so $q = dw^2$. Such a coordinate w (called *preferred*) is unique up to translation and sign: if $dw^2 = dz^2$ then $z = \pm w + c$.

Consider a smooth curve $\gamma : [0, 1] \to X$ with zeros of q only at the ends if any. By definition of q, $v_t := q_{\gamma(t)}(\gamma'(t)) \in \mathbb{C}$ is well-defined for $t \in (0, 1)$. In a preferred coordinate z, it is computed as $v_t = dz^2((z \circ \gamma)'(t)) = ((z \circ \gamma)'(t))^2$. So we see $v_t \in \mathbb{R}_{>0} \iff \gamma'(t)$ is real (i.e. horizontal), and $v_t \in \mathbb{R}_{<0} \iff \gamma'(t)$ is purely imaginary (i.e. vertical) in preferred coordinates. So call a curve $\gamma : I \to X$ with no zeros of q in the interior *horizontal* (resp. *vertical*) *for* q if for all t in the interior, $q_{\gamma(t)}(\gamma'(t)) > 0$ (resp. < 0).

Every point $p \in X \setminus N$ has a preferred coordinate z in which $q = dz^2$. Moreover, the transition function for the overlap of two preferred coordinates is of the form $z \mapsto \pm z + c$, for some $c \in \mathbb{C}$. So we can define the *horizontal measured foliation of* q on $X \setminus N$ by defining it at p as the horizontal foliation in some preferred coordinate $z = x + iy$ at p, and setting $\mu^u := |dy|$. This definition doesn't depend on the choice of the preferred coordinate used, as horizontal lines and the measure $|dy|$ are invariant under transition maps between preferred coordinates. Similarly the *vertical measured foliation of* q may be defined using the vertical foliations of preferred coordinates z, and their transverse measures $|dx|$.

In a conformal coordinate around a point $p \in N$, q vanishes to some order k. In this case *preferred coordinates* can be similarly chosen so that $q = z^k\, dz^2$. Then $z(p) = 0$ and z is uniquely determined up to multiplication by a $(k+2)^{nd}$ root of unity. Here $k \geq 1$ for q holomorphic, but can be $k = -1$ for q *integrable* meromorphic, (i.e. with $\int_X |q|$ finite). The local structure of the leaves around p looks like a $(k+2)$-pronged singularity. In this way a quadratic differential q encodes the structure of two transverse measured foliations on X with the zeros and simple poles of q forming the singular sets of both. The leaves of a foliation on $X \setminus N$ are not orientable when q has a zero of odd order or a simple pole.

Note that for any $p, p' \notin N$, a preferred coordinate z at p can be continued along any continuous injective path $\gamma : I \to X \setminus N$ connecting p to p'. $z(p')$ will depend on the path γ in general but stays constant when γ is changed by an isotopy in $X \setminus N$ fixing its endpoints at p, p'. In particular, at any $p \notin N$, the domain of a preferred coordinate can be assumed to contain $f(p)$

Now consider a homeomorphism f_0 of the oriented surface S of genus $g > 1$, f_0 neither homotopic to a periodic nor a reducible map. Thurston's classification provides for f_0: a Riemann surface structure (marked say $\sigma : S \to X$); a homeomorphism (*pseudo-Anosov*) $f : X \to X$ isotopic to $\sigma \circ f_0 \circ \sigma^{-1}$; and a real number $\lambda > 1$ (the *dilatation*); and a meromorphic quadratic differential q on X with a finite set N of zeros and simple poles such that $f(N) = N$ (preserving orders of singularities) and for every $p \notin N$, in preferred coordinates for q, $z = x + iy$ centered at p (i.e. with $z(p) = 0$) and $w = u + iv$ centered at $f(p)$, f is affine of the form $(u, v) = f(x, y) = (\lambda x, y/\lambda)$. The horizontal foliation of q is scaled by λ under f, and the vertical by $1/\lambda$. Proofs of these statements can be found in Hubbard [16]. If a preferred coordinate z centered at p is continued along some path in $X \setminus N$ that connects p to $f(p)$, f takes the form $f(z) = \lambda \mathtt{Re}(z) + i \frac{\mathtt{Im}(z)}{\lambda} + z(f(p))$.

Such a map is not conformal since it takes infinitesimal circles to ellipses. It is quasiconformal (short qc) of constant $K = \lambda^2$. This constant is a measure of how far $\phi \circ f \circ \phi^{-1}$ is from being conformal. Loosely speaking, a *quasiconformal structure* on a surface S is a homeomorphism $\sigma : S \to X$ to a Riemann surface, where another such map $\sigma' : S \to X'$ defines the same qc-structure if there exists a map $\alpha : X \to X'$ with $\alpha \circ \sigma$ homotopic to σ' and with the property that α takes infinitesimal circles to infinitesimal ellipses of eccentricities globally bounded by some M. A *quasiconformal map* g of the qc surface S is a homeomorphism $f : S \to S$ so that for some qc structure $\sigma : S \to X$, under the map $\sigma \circ f \circ \sigma^{-1} : X \to X$ the eccentricities of the images of infinitesimal circles are bounded by some M almost everywhere on X. The infimum of the bounds M, taken over maps in the class $[f]$ is called *quasiconformal dilatation* of $[f]$. A pseudo-Anosov map f

is an extremal map, in the sense of Teichmüller [27], in that its qc-dilatation is the least element in the set of qc-dilatations of qc maps homotopic to f under all possible qc-structures $\sigma' : S \to X$, (see Bers [4]).

1.3 Some properties of dilatations

Thurston showed that the algebraic degree of a dilatation λ is bounded by $6g-6$ for a map of a closed surface S_g. If the associated measured foliations are orientable, then one doesn't need to take a branched double cover, and λ is the root of a polynomial of degree bounded by $2g = \dim(H_1)$. Arnoux-Yoccoz (1981) and Ivanov (1990) showed that for a fixed surface S_g of genus g, the set of dilatations smaller than some finite number c is finite. Thus there is a least dilatation $l_g > 1$ if we fix the genus. Penner (1991) showed that the least dilatation l_g is less than $11^{1/g}$, by way of constructing specific examples. Better estimates have been made since.

Since f^n is pseudo-Anosov with dilatation λ^n if f is pseudo-Anosov with dilatation λ, we see that the set of all dilatations is a dense subset of $[1, \infty)$. Even though positive quadratic algebraic units are exactly the dilatations in the case of the torus, B. Strenner [26] showed that quadratic units can also arise as dilatations on surfaces of arbitrarily high genus. In fact, he shows that algebraic units of every even degree bounded by $6g-6$ arise as dilatations of pseudo-Anosov maps on S_g, (bounded by $2g$ if we only allow transversely orientable invariant foliations).

Computing dilatations:

Let $f : (X, q) \to (X, q)$ be the pseudo-Anosov representative of its mapping class, with N the finite set of zeros of the quadratic differential q of f. The dilatation $\lambda(f)$ can be computed using the action of f_* on the first cohomology group of X provided the leaves of the horizontal foliation of q on $X \setminus N$ can be oriented so that every point has an open neighborhood in which the leaves all have the same orientation. However, one can always construct the *orientation double cover* of the foliation where this is the case: first take the double sheeted covering space $\pi : \tilde{X} \to X \setminus N$ where $\tilde{X}$ consists of two points $(p, \pm 1)$ for each $p \in X \setminus N$, such that the second coordinate ± 1 is locally constant. The ± 1 at p records a choice of orientation of the leaf of the foliation through p and the lifted foliation on $\tilde{X}$ can be oriented by it.

Consider a closed loop $\gamma : I \to X \setminus N$ going once around a k-pronged singular point $p \in N$, enclosing no other points of N. Lifting γ starting at $(\gamma(0), 1) \in \tilde{X}$ one arrives at $(\gamma(1), (-1)^k)$, as can be seen in a local picture. The endpoint of the lift is its starting point iff k is even. For k odd, the endpoint of the lift is the same point of X but with the opposite orientation of the leaf through it, while the lift of 2γ is a closed loop. Let U be the component of $X \setminus \gamma(I)$ containing p. Then if k is even, $\pi^{-1}(U \setminus \{p\})$ has two disjoint components with foliations isomorphic to $U \setminus \{p\}$, each k-pronged. If k is odd, $\pi^{-1}(U \setminus \{p\})$ is a single punctured disk, with a $2k$-pronged foliation. $\tilde{X}$ can thus be completed by including these punctures and π extends to a *branched* double cover $\pi : \tilde{X} \to X$, branched at the inverse image of each odd pronged singularity. f lifts to a homeomorphism $\tilde{f}$ of $\tilde{X}$ having the same affine structure with respect to the lifted foliation on $\tilde{X}$ as f has on X. Thus it has the same dilatation, $\lambda(\tilde{f}) = \lambda(f)$.

Consider the pullback $\pi^* q$ of q to $\tilde{X}$ with the lifted horizontal foliation oriented. The simple poles of q, if any, become non-singular points of $\pi^* q$. Since the lifted foliation on $\tilde{X}$ only has even-pronged singularities, at an $x \in \tilde{X}$, in preferred coordinates $\pi^* q$ is of the form $z^{2k}\,dz^2$, for some integer $k = k(x) \geq 0$. The 1-forms $\pm z^k\,dz$ both square to $\pi^* q$ and provide a choice of a square root ω in these coordinates. A consistent choice can be made globally on X if ω is chosen to yield positive values on positively oriented tangent curves to the horizontal leaves of $\pi^* q$. The real and imaginary parts of ω define elements of $H^1(\tilde{X}; \mathbb{R})$. These real and imaginary parts are scaled by $\lambda^{\pm 1}$ under $\tilde{f}^*$, so $\lambda^{\pm 1}$ are eigenvalues of the induced action on the first cohomology group of $\tilde{S}$.

The matrix $\tilde{f}^*$ preserves the integer lattice $H^1(\tilde{X}; \mathbb{Z}) \subset H^1(\tilde{X}; \mathbb{R})$. λ & $1/\lambda$ therefore satisfy monic polynomials over the integers, which is to say λ is an algebraic unit. Fried [10] showed this and further that dilatations are the growth rate of the induced homomorphism $f_* : \pi_1(S) \to \pi_1(S)$.

An algebraic integer (i.e. the root of a polynomial over the integers with leading coefficient 1) is called *weak Perron* if it is positive and greater or equal than all its Galois conjugates in absolute value. If it is strictly greater in absolute value, it is called *(strong) Perron*. Thurston proved that dilatations are Perron numbers using a Markov partition for the homeomorphism with an associated non-negative incidence matrix A [9]. Since pseudo-Anosov maps are topologically mixing, A is *mixing*: some power A^n is strictly positive. If just the sum of positive powers of a non-negative matrix A is positive, it is called *ergodic*. The classic Perron-Frobenius theorem tells us that every ergodic matrix has a simple positive eigenvalue, λ (called the *Perron root*), and a corresponding positive eigenvector. Further, $\lambda > |\mu|$ for any other eigenvalue μ of A. The Perron root λ lies between the min and max of the L^1-norms of the columns (or rows) of A.

David Fried (1985) proved that pseudo-Anosov dilatations are *biPerron* algebraic units: these are algebraic units λ all of whose Galois conjugates $\alpha \in \mathbb{C}$ lie in the annulus $1/\lambda < |\alpha| < \lambda$, (except perhaps one of $\pm\lambda^{-1}$). This follows since $H^1(\tilde{X}; \mathbb{R})$ is naturally a symplectic vector space under the cup product, preserved under homeomorphisms, so the induced map on cohomology is symplectic. Since $1/\alpha$ is an eigenvalue of a symplectic matrix whenever α is, a Perron root of symplectic matrix is biPerron. For details on this see [16, Sec 8.6]. Fried asked (see Problem 2 in [10]) if for every biPerron algebraic unit λ, some power λ^β is a pseudo-Anosov dilatation, but this is not known to be true.

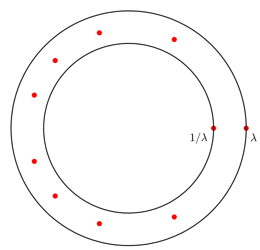


Figure 2: Galois conjugates of Lehmer's number, a biPerron unit with minimal polynomial $x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$.

CHAPTER 2

RECTANGULAR STRUCTURE OF ABELIAN DIFFERENTIALS

As is well-known, a holomorphic quadratic differential q on a compact Riemann surface X provides many decompositions of X into finite collections of rectangles and cylinders identified along their boundary edges, which are horizontal and vertical leaves of q. The following proposition is copied from *Teichmüller Theory - Volume 1*.

Proposition 3. *[15, Prop. 5.3.4] Given a holomorphic quadratic differential q on a compact Riemann surface X, and $L > 0$, there exists a finite collection of cylinders;*

$$A_i := \{z \in \mathbb{C} : \ 0 < \mathtt{Im}(z) < b_i\} \Big/ a_i\mathbb{Z},$$

and a finite collection of rectangles

$$R_j := \{z \in \mathbb{C} : \ 0 < \mathtt{Re}(z) < c_j, \ 0 < \mathtt{Im}(z) < d_j \ \text{ and } \ c_j > L\}$$

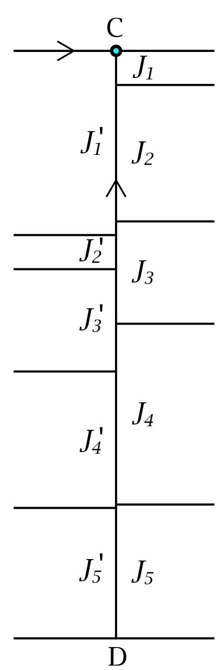
with holomorphic injections $\phi_i : A_i \to X$, $\psi_j : R_j \to X$ with disjoint images whose closures cover X, such that ${\psi_j}^* q = dz^2$ on R_j (in the coordinate z in the definition of R_j), and similarly $\phi_i^* q = dz^2$ on A_i.

We will follow the proof in [15] in our discussion of this proposition, in the case that X is a closed oriented surface of genus g. We will assume q is subordinate to a pseudo-Anosov map $f : (X, q) \to (X, q)$. We will further assume the foliations of q to be oriented (having taken a branched double cover if necessary), so each zero of q is of even multiplicity. As we saw above, with this assumption, $q = \omega^2$ for a holomorphic 1-form ω on X, called an *Abelian differential*: every point $p \in X$ has a simply-connected neighborhood U and a conformal coordinate $z : U \to \mathbb{C}$ on which $\omega = z^k\, dz$, $k \in \mathbb{Z}_{\geq 0}$, ($k > 0 \iff p \in N \subset X$ finite).

Note that for $p \in X$ where locally $\omega = dz$, the coordinate z is unique if we require $z(p) = 0 \in \mathbb{C}$. When $k > 1$, $z(p)$ is 0, but z is unique up to a $(k+1)^{st}$ root-of-unity since $(\zeta z)^k\, d(\zeta z) = z^k\, dz$ when $\zeta^{k+1} = 1$.

When there is a pseudo-Anosov representative $f : (X, q) \to (X, q)$, it takes zeros of q to zeros of the same order, taking horizontal leaves to horizontal leaves, vertical to vertical. Suppose L were a compact horizontal segment of q, of length l, connecting two zeros. Then $f^{-n}(L)$ would be horizontal and of length l/λ^n also connecting zeros of q. Since q has finitely many zeros and $\lambda > 1$ this is impossible. So such a quadratic differential has no compact horizontal segments connecting zeros. A similar argument can be made about compact critical vertical segments and the decomposition in the proposition above has no cylinders, given f.

Pick a point $C \in X$ where $\omega_C = 0$ and draw one segment J of an incoming vertical leaf to C. The rectangular decomposition of (X, ω^2) depends on J. Consider a horizontal trajectory with one endpoint a zero of ω (possibly C itself). It can be uniquely continued in X away from the zero indefinitely since it neither meets another zero nor crosses itself. The closure of this trajectory is all of X: otherwise the compact boundary of the closure is scaled by λ^{-n} under f^{-n}, while the lengths of closed curves are bounded below on (X, ω^2). Thus each horizontal trajectory eventually crosses J, and we can draw every horizontal trajectory from every zero of ω until it first intersects J. Among the finite points on J marked by these intersections, suppose $D \in J$ is furthest from C. Shorten J to this point, i.e. redefine J to be the vertical segment from the point D



to the zero C, of reduced length $L > 0$. Now continue the criti-
cal horizontal trajectory through D on the other side of J until it meets J again. Each horizontal segment thus drawn is critical. The segments leaving J (i.e. on the *right* of J) partition J into (closed) subintervals $J_1, \ldots, J_n$. Denote the interior of J_r by J_r^o.

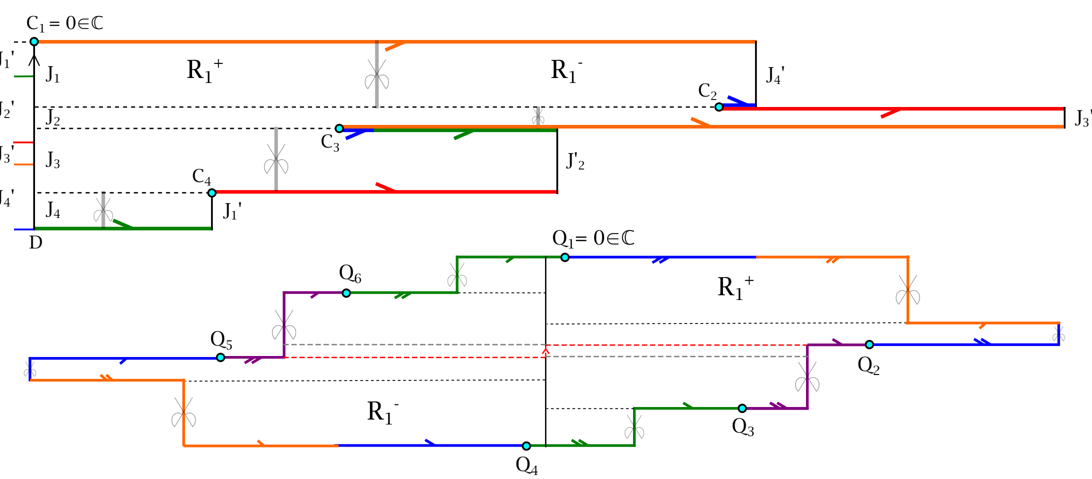
Identify the oriented J with the segment of length L from $-iL \in \mathbb{C}$ to 0: (i.e. set $z(p) = \int_C^p \omega$ for $p \in J$). By construction, the horizontal trajectory leaving a point $y \in J_r^o$ does not meet a zero of ω before intersecting J again (on the left). Define $\psi_r : [0, \delta] \times J_r^o \to X$ by sending (x, y) to the point on the horizontal leaf leaving $y \in J_r^o$ and at a distance x from y. ψ_r is well-defined until the trajectory meets a zero of ω, which doesn't happen before it intersects J. Pick $y \in J_r^o$ and the smallest $x_y > 0$ such that $\psi_r(x_y, y) \in J$. Since $\psi_r(\{x\} \times J_r^o)$ consists of points all at a distance x measured from the vertical J_r^o along parallel horizontal leaves, $\psi_r(\{x\} \times J_r^o)$ is vertical, and x_y is constant for $y \in J_r^o$, call it c_r. So the rectangle $R_r := (0, c_r) \times J_r^o$ isometrically embeds in X under ψ_r as a rectangle consisting of horizontal curves leaving J on the right and arriving at J on the left, bounded by critical trajectories above and below.

J is also partitioned by incoming critical horizontal curves on the left, say as $J'_1, \ldots, J'_{n'}$. Then each J'_s is the vertical boundary of a rectangle with other vertical boundary some J_r, so $J_{\sigma(r)} \cong J'_r$ for some permutation σ of $\{1, ..., n\}$. Note that the inverse of each $\psi_r : R_r \to X$ is a preferred coordinate $z_r : \psi_r(R_r) \to R_r \subset \mathbb{C}$, since at a point p on a horizontal trajectory leaving $y \in J_r^o$, it agrees with the integral of ω along the vertical path from 0 to y followed by the horizontal from y to p. We will identify the rectangle R_r with its image in X under ψ_r. The z_r have disjoint images in $\mathbb{C}$ and can be considered as a single coordinate $z = z_J$ on $\cup\psi_r(R_r)$.

Each horizontal boundary of a rectangle contains a zero of ω (except perhaps the one through D which might have its zero to the left of J). z can be extended continuously to the intersection of adjacent rectangles from $y = J_r \cap J_{r+1}$ only until the zero of ω on $R_r \cap R_{r+1}$, call this zero C_r.

The union of the closures of the rectangles $R_1, \ldots, R_n$ in X is a closed submanifold of X whose boundary comprises of closed critical horizontal leaves. Since there is a pseudo-Anosov map $f : (X, \omega^2) \to (X, \omega^2)$, there are no closed horizontal leaves, so the closure $\overline{\cup_r R_r}$ is X. This gives us a decomposition of X into rectangles whose boundary edges are critical horizontal and vertical segments of trajectories of $q = \omega^2$.

Example 4. *Below is a decomposition of a Riemann surface X of genus 2 corresponding to a holomorphic 1-form ω on it, depicted as a polygon P (with slits) in $\mathbb{C}$ with identifications on the boundary (including the slits). It corresponds to the pA map of example 12, page 34 below. There is a single zero $C \in X$ where in preferred coordinates $\omega = z^2\, dz$. Each of the C_i on the horizontal critical edges correspond to the same C in this example.*



We have rearranged the picture above by cutting each rectangle in half between the two zeros on it's boundary, and placing the right half on the left of J. The pA map in this case sends the L-shaped purple edge to the one in green, green to blue, and blue to the red through the middle; for details see example 12, page 34 below.

Lemma 5. *Given an Abelian differential on a closed Riemann surface, one can find a single chart centered on a vertical segment J' with rectangles $R'_1, ...R'_n$ glued on the right and congruent rectangles $R''_1, ..., R''_n$ glued in a different order on the left of J. The horizontal boundaries of the rectangles are found using segments of critical horizontal segments starting from the critical points and drawn until they intersect J'. In particular, a segment of <u>every</u> critical horizontal trajectory is drawn in such a chart until it intersects the vertical J'.*

Proof. Take the two horizontal edges of ∂R_1. Each contains a zero at a unique point along the trajectory ($C_1 = 0$ and say C_2 with $\mathrm{Re}(C_2) > 0$). Consider the vertical segment in R_1 (depicted gray in the example above) at the midpoint of the diagonal straight line between C_1 and C_2 (i.e. $\{\frac{\mathrm{Re}(C_2)}{2}\} \times J_1$). We cut the part of R_1 to the right of this vertical edge (R_1^-) and re-place it on the left side of J (where its right edge is identified). Similarly for $1 < j < n$, R_j contains a unique vertical segment midway between the two zeros on its two horizontal boundaries. Cut the part of R_j past this vertical edge and place it to the left of J. If the horizontal edge through D has a zero on the right of J, repeat as above for R_n. If not, then by construction the zero on the horizontal edge through D is on the left of J, and therefore on the bottom edge of one of the other R_j (with $j < n$) and has already been placed on the left of D, say a distance l to its left. In this case cut R_n along the vertical edge with real part $|J_n| - l/2$ and place what's to its right on the left of J. To complete the symmetry, we also move the vertical J under a holonomy along the horizontal foliation to be between the zeros on the top and bottom edges. Note that under a small horizontal translation of J, say to the left, the critical trajectories incident to it on the left are shortened, while those on its right are extended but the structure of the decomposition doesn't change much. A similar argument works if it is moved to the right.

In this way, for any critical vertical segment J'' of ω, one can find a shorter segment J of positive length, and a single coordinate $z = z_J$ on a set of full measure on X, with image in the shape of a rectangular polygon in $\mathbb{C}$. Moreover, if J is moved slightly to be between the zeros on the top and bottom edge of this polygon, a symmetric chart is found around the moved J' (symmetric in the sense that for each rectangle on the right, there is a congruent rectangle on the left, but the order in which rectangles are glued on the left may be arbitrary). □

This rectangular structure provides further justification for why the procedure in §3 below works, where we construct Riemann surfaces with differentials and pseudo-Anosov maps subordinate to the differentials starting from a matrix of 0s and 1s.

CHAPTER 3

CONSTRUCTING SURFACE AUTOMORPHISMS

This chapter is an elaboration of the ideas in [3] written in collaboration with Hyungryul Baik and Chenxi Wu. There a surface map was constructed for extended incidence matrices (see [29]) of 0's and 1's under additional hypotheses. In this way, an infinite subset of biPerron numbers was shown to be pseudo-Anosov dilatations.

Before constructing the surface, in §3.1 below, we recall Thurston's construction of a self-map of an interval $f : I \to I$ with $e^{H(f)} = \lambda$ for any given weak Perron λ. For an interval map $f : I \to I$, we call a point $x \in I$ *critical* if $f(x)$ is a local extremum. The set of all future images of the critical points is called the *post-critical set*. If this set is finite, the map f is called *post-critically finite* (shortened *pcf*).

3.1 Odd-block matrices and pcf interval maps

Thurston proved [29, Thm 2] that logarithms of weak Perron numbers are exactly the topological entropies of post-critically finite self-maps of intervals. Moreover, for a Perron number λ, he builds an ergodic non-negative integer matrix A (i.e. the sum of positive powers of A is a matrix with only positive entries) with λ as the Perron root (thus proving a theorem of Lind [20]). Thurston's matrix is the incidence matrix of a continuous piecewise linear (PL) map, $h = h_A$ of the interval $I = [0, 1]$, with respect to some partition $\mathscr{P}$. The partition is Markov with respect to h (i.e. $h(\mathscr{P}) \subset \mathscr{P}$ and {critical images} $\subset \mathscr{P}$) so h is pcf. Incidence matrices of post-critically finite self-maps of intervals Thurston recognizes as having properties (i) & (ii) below. (See Example below.)

Theorem 6. *[29, Thm. 1.4, Prop 3.1] Every Perron [in particular biPerron] number, λ, is the leading eigenvalue of an ergodic non-negative integer matrix, $A_{n\times n}$, satisfying:*

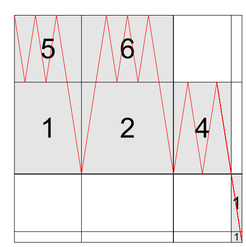
(i) In each column of A, the non-zero entries form one consecutive block; and

(ii) There is a map $\phi : \{0, 1, ..., n\} \to \{0, 1, ..., n\}$ such that the odd entries of the j^{th} column of A are in rows between $\phi(j-1)$ and $\phi(j)$, excluding the smaller of the two and including the bigger. (I.e. $\min\{\phi(j-1), \phi(j)\} < i \leq \max\{\phi(j-1), \phi(j)\}$).

Essentially, the continuity of the PL self-map of the interval I is encoded by ϕ into the data of the images of the points of the partition of $I = [0, 1]$.

Example 7.

$$\begin{pmatrix} 5 & 6 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 4 & 0 \\ 3 & 5 & 1 & 0 & 1 & 2 & 1 \\ 8 & 4 & 1 & 0 & 7 & 4 & 1 \\ 2 & 0 & 1 & 3 & 0 & 6 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



Thurston's example of an odd-block matrix and the associated interval map of constant slope in absolute value. $\lambda \sim 6.37$ satisfies $\lambda^2 - 7\lambda + 4 = 0$. The matrix is singular and not ergodic and some intervals in the partition collapse to points. Here $(\phi_i)_{i=0}^{7} = (0, 3, 2, 5, 4, 2, 2, 7)$.

Note that the odd entries in every column are also consecutive and these odd blocks can be connected end-to-end via ϕ (depicted by the blue dashes in the example above). We will call a non-negative integer matrix satisfying properties (i) and (ii) above an *odd-block* matrix. Thurston calls them *extended incidence matrices*, however we would like to say odd-block.

To construct the map $h = h_A$ out of such a matrix A, partition the interval $I = [0, 1]$ into n subintervals as $\mathscr{P} = \{0 = x_0 \leq x_1 \leq ... \leq x_n = 1\}$. x_i is mapped to $x_{\phi(i)}$. Note that in folding an interval an even number of times across another interval, both endpoints of the domain are sent to a single endpoint of the image. Placing it an odd number of times forces the boundary of the domain to be sent to distinct endpoints of the image. Now, starting from $x_{\phi(i-1)}$, the interval $I_i := (x_{i-1}, x_i)$ is placed piecewise linearly over the other subintervals of the partition according to the i^{th}-column A_i of A: first cover the intervals that need to be covered an even number of times on one side of $x_{\phi(i-1)}$, returning to it; then cover the intervals of the odd block of A_i, finishing at $x_{\phi(i)}$; then cover the intervals corresponding to even entries on the other side of $x_{\phi(i)}$ (the side yet uncovered by I_i), returning back to $x_{\phi(i)}$. Note this process is not unique.

The piecewise linear parts will have random slopes. However, Thurston shows they can be made to have constant slope λ in absolute value if the partition is chosen according to the non-negative λ-eigenvector $\overrightarrow{w}$ of the transpose A^t, i.e. set $x_i - x_{i-1} = w_i$. A non-negative eigenvector is guaranteed to exist by the Perron-Frobenius theorem since the matrix is non-negative. If the eigenvector is not strictly positive, some subintervals of the partition collapse. This doesn't happen though if the matrix A is *ergodic* (the sum of its positive powers is strictly positive) in which case a strict positive eigenvector $\overrightarrow{w}$ is guaranteed.

To see why we can guarantee constant slope of the PL map, let the columns of A be $\overrightarrow{A_i}$. The λ-eigenvector $\overrightarrow{w}$ of A^t satisfies $\overrightarrow{w} \cdot \overrightarrow{A_i} = \lambda w_i$. This equation is delineated in the graph over the i^{th}-subinterval I_i of the partition: I_i of width w_i is stretched by λ and folded over other subintervals according to the i^{th}-column of A. This is a total length of $\overrightarrow{w} \cdot \overrightarrow{A_i}$.

The topological entropy $H(f)$ of an interval map f with finitely many critical points is

$$H(f) = \lim_{n\to\infty} \frac{1}{n} \log(Var(f^n))$$

by a theorem of Misiurewics and Szlenk [24], where $Var(f)$ is the arclength of f considered as a path in I. $Var(h_A^n)$ is a positive linear combination of the entries of A^n. The entropy of h_A is the exponent of growth of total variation, so it is the log of the largest eigenvalue of A, $\log(\lambda)$.

3.2 Odd-block matrices consisting of 0′s and 1′s

Let M be a fixed odd-block matrix of size n - the incidence matrix of a PL interval map h with respect to a partition $\mathscr{P} = \{x_i\}_{i=0}^n \subset I$. Denote by $\mathbb{N}_n$ the set $\{0, 1, ..., n\}$ and recall that an odd-block matrix comes equipped with $\phi : \mathbb{N}_n \to \mathbb{N}_n$, which encodes the images of the partition $\mathscr{P} = \{x_i\}_{i=0}^n$ under the associated interval map: $h(x_i) = x_{\phi(i)}$.

If the partition is extended by including every critical point of h, the incidence matrix with respect to the bigger partition is larger and composed entirely of 0′s and 1′s. It still has the same leading eigenvalue λ, as h has entropy $\log \lambda$. Whether M is ergodic or mixing doesn't change under this extension, since that is a reflection of the property of the continuous interval map h. However, the bigger matrix of 0′s and 1′s may have determinant zero now even if the smaller was non-singular.

Let $\mathscr{OB}_n$ be the set of odd-block matrices of size n consisting only of 0′s and 1′s. We will develop the theory for these. Given $\phi \in \operatorname{Hom}(\mathbb{N}_n, \mathbb{N}_n)$, $M_\phi \in \mathscr{OB}_n$ is determined: $M_{ij} = 1$ iff $(\phi(j) < i \leq \phi(j+1)$ or $\phi(j+1) < i \leq \phi(j))$. This map is surjective since every odd-block matrix has a map ϕ associated to it by definition. However, it is not injective: for instance both $(\phi(i))_{i=0}^{2} = (2,0,2)$ and $(0,2,0)$ yield the mixing but singular matrix $M_\phi = \left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)$ with Perron root $\lambda = 2$, (not a unit). Thus $|\mathscr{OB}_n| \leq |\operatorname{Hom}(\mathbb{N}_n, \mathbb{N}_n)| = (n+1)^{n+1}$. The non-singular $M \in \mathscr{OB}_n$ are precisely those for which ϕ is a permutation:

Proposition 8. *Determinants of matrices $M \in \mathscr{OB}_n$ lie in the set $\{0, \pm 1\}$. $\det\left(M_\phi\right) = \pm 1$ if ϕ is a permutation and zero otherwise.*

Proof. If ϕ is not a permutation, some $i \in \mathbb{N}_n$ will not be in the image of ϕ. If $i = 0$ or n, the first or the last row of M is $\overrightarrow{0}$. If $0 < i < n$, no block of 1's will begin or end between the i^{th} and $(i+1)^{st}$ row of M, so these two rows will be the same. Either way, $\det(M) = 0$.

If ϕ is a permutation, 0 is in the image of ϕ, so there is at least one and at most two (adjacent) columns that begin with 1. Say the rows and columns of M are R_i and C_i. If there is only one C_j beginning with 1, permute the columns as $(C_j, C_1, ...C_{j-1}, C_{j+1}, ..., C_n)$. If there are two choices, C_j, C_k, one is shorter since ϕ is a permutation, and we move the shorter one to the first column as above. Now subtract the shorter column from the longer one. Follow by operations of the form $R_i \to R_i - R_1$ until the first column is just $e_1 = (1, 0, ..., 0)$. We are then left with a matrix of size $(n-1)$ where we have not introduced any negative signs, and is still in fact odd-block. The process can then be repeated until we get to the identity matrix. Since the determinant was at most multiplied by -1 under these changes, we see that $\det(M) = \pm 1$. □

Note that the determinant of M_ϕ is not the sign of the permutation $\sigma(\phi)$ as one might expect. For instance $(\phi(i))_{i=0}^{4} = (1,3,0,4,2)$ with sign $\sigma(\phi) = 1$ yields $M_\phi = \left(\begin{smallmatrix}0&1&1&0\\1&1&1&0\\1&1&1&1\\0&0&1&1\end{smallmatrix}\right)$ with determinant 1, whereas $(\phi(i))_{i=0}^{4} = (1,3,4,0,2)$ with sign

$\sigma(\phi) = -1$ yields $M_\phi = \begin{pmatrix} 0&0&1&1 \\ 1&0&1&1 \\ 1&0&1&0 \\ 0&1&1&0 \end{pmatrix}$ with determinant 1. (Both these matrices produce pseudo-Anosov maps of S_2, orientation reversing and preserving respectively, see below.)

There are $(n+1)!$ matrices in $\mathcal{OB}_n$ coming from permutations, and in fact the map $\mathcal{S}_{n+1} \to \mathcal{OB}_n$ is injective for $n > 1$. This is no homomorphism of course since $\mathcal{S}_{n+1}$ is a finite group and typical elements of $\mathcal{OB}_n$ have infinite order. However, some structure can be transferred.

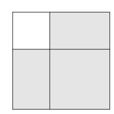
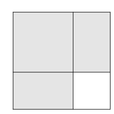
For $\phi, \phi' \in \mathcal{S}_{n+1}$, cut I into n equal parts and define interval maps $h_\phi, h_{\phi'}$ whose incidence matrices are $M_\phi, M_{\phi'} \in \mathcal{OB}_n$. These will not have constant slope, but $N = M_\phi \cdot M_{\phi'}$ then is the incidence matrix of $h_\phi \circ h_{\phi'}$. This composite map is continuous so the product matrix is odd-block (it may have entries > 1, so not in $\mathcal{OB}_n$ neccesarily). However, the permutation corresponding to the odd-block matrix N is just the composite permutation $\phi \circ \phi'$ as an endpoints x_i gets sent to $x_{\phi(\phi'(i))}$ under $h_\phi \circ h_{\phi'}$. Thus we see, for instance:

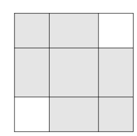
Proposition 9. *$M_\phi \cdot M_{\phi'} \equiv M_{\phi\circ\phi'}$ (mod 2) and in particular, $M_\phi \cdot M_{\phi^{-1}} \equiv I_n$ (mod 2).*

Ergodicity of M_ϕ can also be translated into properties of the permutation. M is ergodic iff for every pair of subintervals I_i, I_j of the partition, some iterate $h_M^N(I_i)$ contains I_j. A subinterval I_j corresponds to a pair $\{j-1, j\}$ in the domain of ϕ. So ergodicity of M_ϕ becomes: $\forall\ 0 < i, j \leq n\ \exists\ N$ so that the set $\{(j-1), j\}$ lies between $\phi^N(i-1), \phi^N(i)$. i.e. either $\phi^N(i-1) < j \leq \phi^N(i)$ or $\phi^N(i) < j \leq \phi^N(i-1)$. This can be guaranteed, for instance, if the cycle decomposition of ϕ has only one cycle, or if it has two cycles where one is singleton.

3.3 Surface map for $M \in \mathscr{OB}_n$

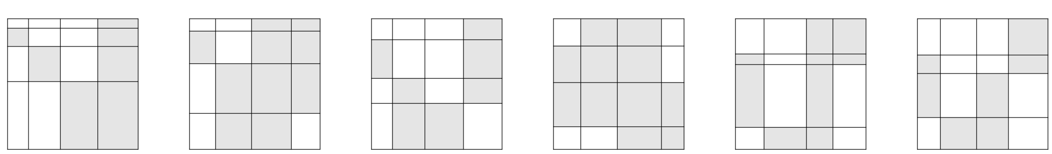
For a matrix $M \in \mathscr{OB}_n$, one can associate a polygon in the plane and a self map of it as follows (this was shown in an earlier paper [3], written by the author in collaboration with Hyungryul Baik and Chenxi Wu): The matrix M and its transpose M^t both have non-negative λ–eigenvectors, $\overrightarrow{v}$ and $\overrightarrow{w}$, (normalized so $\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} v_i = 1$). We partition $[0,1] \times [0,1]$ into an $n \times n$ grid, with widths of subintervals given by the entries of $\overrightarrow{v}$ (going left to right), and heights given by the entries of $\overrightarrow{w}$ (going from top to bottom). Shade the $\{i, j\}^{th}$ rectangle of the partition iff $M_{ij} = 1$ and let the closed shaded region be denoted P. With slight abuse of notation, we will call the part of P in the i^{th}-column of M as C_i. The part of P in the i^{th}-row of M is often disconnected, but we call the union of the rectangles in the i^{th}-row as R_i.

Example 10. *For $n = 2$, only $\phi_i = (1, 2, 0)$ and $(2, 0, 1)$ have eigenvalues $\lambda > 1$, and yield matrices $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, with associated polygons $P =$*   *$\lambda^2 - \lambda - 1 = 0$ for both.*

An ergodic example for $n = 3$ is $\phi_i = (2, 0, 3, 1)$, $M_\phi = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $P =$ 

Some ergodic examples for $n = 4$ are:

$\begin{pmatrix} 0&0&0&1 \\ 1&0&0&1 \\ 0&1&0&1 \\ 0&0&1&1 \end{pmatrix}$ $\begin{pmatrix} 0&0&1&1 \\ 1&0&1&1 \\ 0&1&1&1 \\ 0&1&1&0 \end{pmatrix}$ $\begin{pmatrix} 0&0&0&1 \\ 1&0&0&1 \\ 0&1&0&1 \\ 0&1&1&0 \end{pmatrix}$ $\begin{pmatrix} 0&1&1&0 \\ 1&1&1&0 \\ 1&1&1&1 \\ 0&0&1&1 \end{pmatrix}$ $\begin{pmatrix} 0&0&1&1 \\ 1&0&1&1 \\ 1&0&1&0 \\ 0&1&1&0 \end{pmatrix}$ $\begin{pmatrix} 0&0&0&1 \\ 1&0&0&1 \\ 1&0&1&0 \\ 0&1&1&0 \end{pmatrix}$

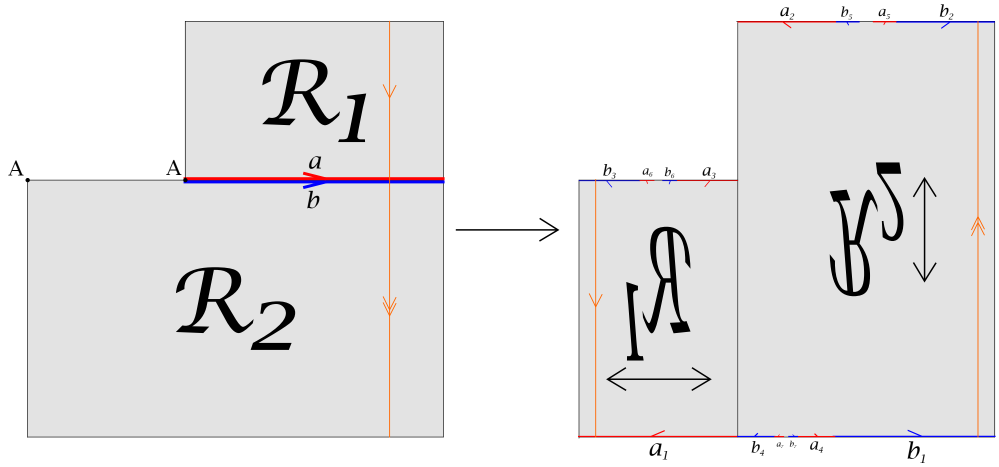


The widths and heights of each row and column are chosen so that if the union of rectangles in the i^{th} row, $R_i \subset P$ is squeezed horizontally by λ and stretched vertically by λ, the horizontal widths add up to the width of $C_i \subset P$ and the heights are the height of C_i. Thus one is tempted to define a map on this object by sending R_i to C_i, stretching vertically and shrinking horizontally by λ. If identifications can be found on the boundaries of the rectangles in P, identifying only via translations or central symmetries, so that the identifications are compatible with this row i to column i map, we would have found a Teichmüller extremal map on a surface with a quadratic differential, (the pull back of dz^2 from a choice of a $\mathbb{C}$ structure on the plane). If there are finitely many intervals on the boundaries of the rectangles in P identified in this way, we would have found a pseudo-Anosov map on a compact surface. If there are infinitely many cone points there is still a meromorphic quadratic differential on a surface of infinite type, and a pair of transverse measured foliations, infinitely singular, one stretched by λ and one compressed by it, but we will not be looking at these in this dissertation.

We will study the first of the examples above, for which this procedure works, in some detail below. Thurston describes this example in [29], where he obtains the surface as the ω-limit set of a PL map defined in the plane. In the setup of this paper it is associated to the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$.

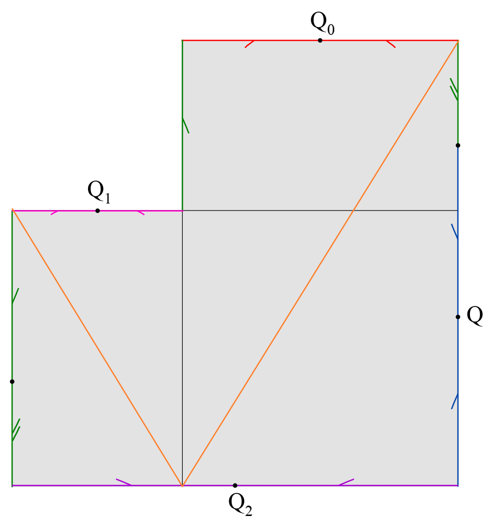
Example 11. *Let $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Define a function f_0 on the interiors of the rows R_1, R_2 in the associated polygon P of M by stretching each row by a factor of λ in the y-direction and shrinking each rectangle in the x-direction by λ. The rectangles thus obtained are isometric via translations to the first and second columns of P respectively. In this example we place the first of these reflected horizontally in the first column, and the second reflected vertically.*

A vertical segment going through the rows (orange) is depicted with its image on the right. The projection of its image back onto it (or to the y-axis) is the pcf interval map h_M defined in §2.1 above since the heights of rows were chosen according to the λ-eigenvector $\overrightarrow{w}$ of M^t, also used to define h_M. In particular, $R_i \cap R_{i+1}$ lies on the horizontal line $y = x_i$, for all $1 \leq i < n$. Recall x_i were the points of the partition chosen so $x_i - x_{i-1} = w_i$.



In the figure above, a_i and b_i are the i^{th} iterates of the edge $a = b = R_1 \cap R_2$ under continuous extensions of f_0 to ∂R_1 and ∂R_2 respectively. If f_0 is to extend to a continuous function on $R_1 \cup R_2$, all future iterates $a_i \sim b_i$ are to be identified as labelled. In the current example this can be arranged by folding each horizontal segment of $\partial(R_1 \cup R_2)$ in half, forming a cone point of angle π in the middle.

Continuity at $a = b$ suffices to show that there is only one cone point (of angle π) on the vertical segments as the following argument (of Chenxi Wu) shows: the vertical segments of $\partial(R_1 \cup R_2)$ form a circle after each horizontal segment is folded in half. This circle is stretched by $\lambda > 1$ and vertical segments going in opposite directions (up and down) from the point P are identified. This can be thought of as a map of a circle where an arc is compressed to a point and the rest of the circle is stretched uniformly to preserve circumference. This identifies the circle in half to itself from the source to the sink (for *bigger matrices, when there are more vertical edges that are mapped into the interior of P, there are more sinks for the dynamics on the circle formed by the vertical edges, and there are as many sinks as sources on this circle, so there are then $2(n-1)$ intervals on this circle that define the vertical identifications).*



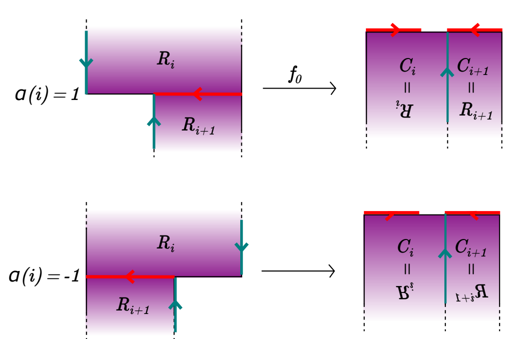
After these identifications, f_0 extents to the intersections of rows, and descends to a homeomorphism f of the quotient space $P/\sim$ which is the "pillowcase": A sphere with 4 cone points of angle π. Or with a quadratic differential having 4 simple poles. Note that the map f in the vertical direction follows the interval map h_M. The projection of the orange curve above (the image of a vertical segment going through the rows) to the y-axis is the interval map h_M. f is of the form $(x, y) \mapsto (\pm\frac{x}{\lambda} + c, h_M(y))$ for some $c \in \mathbb{R}$ on each R_i.

For the polygon P in the example above, any combination of rotations or reflection on the rectangles other than the one given above or its inverse ($C_i \mapsto R_i$) yield identifications on the boundary with infinitely many cone points or cone points of infinite angle. We will try to see when $M = M_\phi$ in $\mathcal{OB}_n$ with leading eigenvalue $\lambda > 1$ can lead to a surface of finite type. Let P be the associated closed polygon, and H_i be the horizontal edge of the boundary of P at $y = x_i$, ($H_i := \{(x, y) \in \partial P; y = x_i\}$. $f_0^{n+1}(H_i) \subset H_i$ since $\phi \in \mathcal{S}_{n+1}$, and it is uniformly compressed by λ^{n+1}. If identifications lead to a surface of finite type, there can be at most one cone point on H_i (at a point of a periodic orbit, a cycle of the permutation ϕ) otherwise the set of images of a cone point accumulating to a point of the attracting periodic orbit will consist only of cone points and the surface will be of infinite type. In the example above, the image of $R_1 \cap R_2$ under both $R_1 \mapsto C_1$ and $R_2 \mapsto C_2$ falls on the same horizontal edge $H_{\phi(1)}$ of ∂P. Note that under the inverse $C_i \mapsto R_i$, the vertical edge $C_1 \cap C_2$ is sent to vertical edges with distinct x-coordinates. This is since $R_1 \cap R_2$ was sent to the extremes of $H_{\phi(i)}$ under f_0. This is also necessary since if the vertical images had the same x-coordinate, there would be a cone point of angle π at an endpoint of H_i, which would lead to an infinite type surface.

So we try to use the continuous interval map h_M to define the y-coordinate of f_0. For every $1 \leq i < n$, for which x_i is a critical point of h_M, the two columns C_i, C_{i+1} of P are aligned at the horizontal line $\{y = x_{\phi(i)}\}$ since M is odd-block. The first problem we run into is that $R_i \cap R_{i+1}$ may not be connected. In this case there are at least two points of angle $3\frac{\pi}{2}$ on $H_{\phi(i)}$ which after identifications lead to > 1 cone points. To avoid this, the multiple rectangles in each row can be merged by moving the columns of P_0 horizontally, and presumably in some order so each row R_i becomes a single rectangle, without breaking the continuity of the columns. The interval map h_M sometimes provides a way to consistently move the columns horizontally so the adjacent pairs of rows could be aligned either on the right or the left, but this seems generally impossible for a critical x_i if there is the shape of a 'W' in the graph of h_M centered at x_i, (where the inner $\wedge$ of 'W' is shorter than each of the outer edges), and if the adjacent columns there have different widths.

The first restriction we therefore impose is that we be able to align each pair of adjacent rows on the left or the right without breaking columns. This can be guaranteed if, for instance, no horizontal line intersecting the graph of h_M at $h_M(x_i) = x_{\phi(i)}$ intersects the graph on both sides of $x = x_i$. In this case h_M gives a well defined solid column for every broken column to fall on. We will say in this case that there *exists an alignment function* α on the rows ($\alpha : \{1, \ldots, (n-1)\} \to \{\pm 1\}$ can record if R_i and R_{i+1} are aligned on the right $(+1)$ or left (-1)).

We are allowed at most one cone point on $H_{\phi(i)}$ if we want a finite type surface. There are two systematic ways to do this once the rows are aligned: We can let the y-coordinate of f_0 be h_M. Then if x_i is critical for h_M, the intersection of the rows R_i, R_{i+1} is sent to the edge where columns C_i, C_{i+1} are aligned. We must place the two images of $R_i \cap R_{i+1}$ at the extreme ends of $H_{\phi(i)}$ pointing opposite to each other so as to not introduce multiple cone points. We can scale by $\frac{1}{\lambda}$ in the x direction for one of the rows and $-\frac{1}{\lambda}$ for the other so the two images of $R_i \cap R_{i+1}$ point opposite, and there is a choice to do this such that $H_i \subset \partial P$ is sent to the middle of $H_{\phi(i)}$. When done this way, the two images of the edge $C_i \cap C_{i+1}$ under the inverse ($C_i \mapsto R_i$) have different x-coordinates, and induce identifications on the unaligned vertical edges of R_i , R_{i+1}. Starting from the two ends of H_i the two vertical edges of the rows are identified for the length of the shorter of the rows.



The figure above illustrates the case when columns C_i and C_{i+1} are aligned at their top edges, and x_i is a critical point. When the rows are aligned on the right, only an orientation preserving map gives the right identification. When aligned on the left, an orientation reversing map.

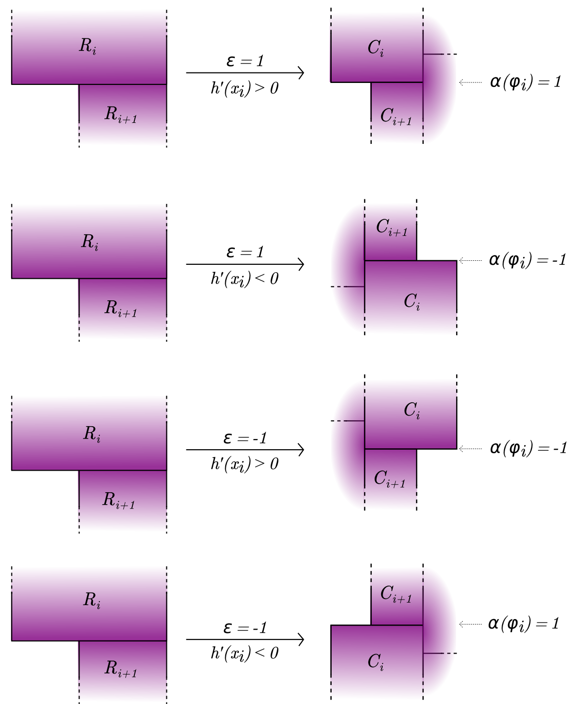
Doing so, we have either (*i*) reflected one of R_i, R_{i+1} horizontally and the other vertically before scaling or, (*ii*) rotated one row by π and the other by 0. (*i*) is orientation reversing and (*ii*) preserving. One of these choices is forced upon us for each *critical* $i \in \{1, \ldots, (n-1)\}$. Since a homeomorphism of a connected oriented surface couldn't change from being orientation preserving to orientation reversing, these alignments must be compatible with each other. In other words, we can choose $\epsilon = \pm 1$ to record whether we are trying to construct an orientation preserving ($\epsilon = 1$), or reversing ($\epsilon = -1$) map. However, once ϵ is chosen, alignments of rows R_i, R_{i+1} and columns C_i, C_{i+1} determine how f_0 maps between them, for i critical.

Remark: We could use the opposite of h_M: if h_M is increasing on I_i, we make f_0 decreasing on R_i, and vice-versa. However, anytime this procedure works, the inverse map $C_i \mapsto R_i$ corresponds to using a continuous map $h_{M'}$ on the transpose of M.

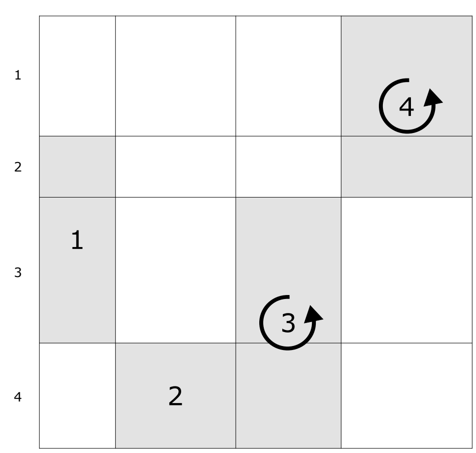
For x_i of the parition $\mathscr{P}$ that are not critical for h_M, the union $R_i \cup R_{i+1}$ is mapped homeomorphically by f_0 to columns $C_i \cup C_{i+1}$, and we need the alignments to be compatible. Putting these conditions together places the following *alignment conditions* on the alignment function α of the rows:

(*i*) If x_i is critical, $\alpha(i) = \epsilon$ (resp. $-\epsilon$) if x_i is a local minimum (resp. maximum).

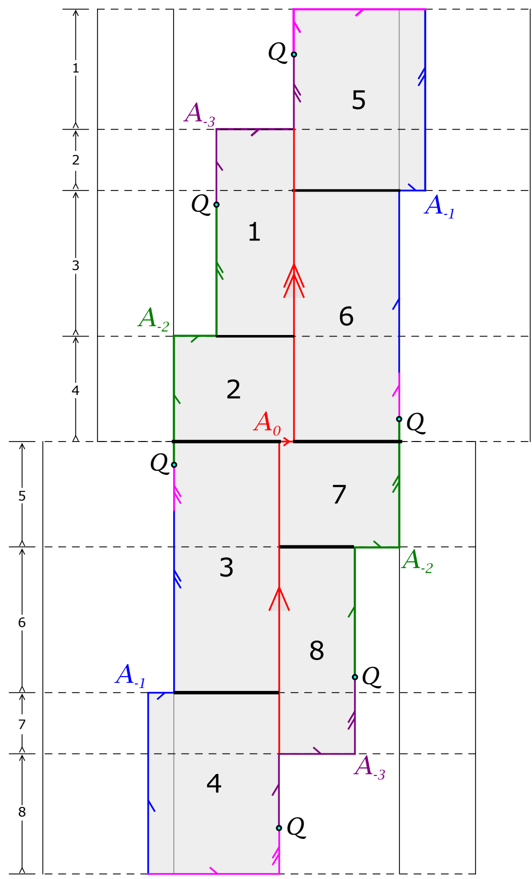
(*ii*) If x_i is not critical, when h_M is increasing at x_i, $\alpha(i) = \epsilon\alpha(\phi(i))$. When h_M is decreasing, $\alpha(i) = -\epsilon\alpha(\phi(i))$. This condition is illustrated in the figure below.



Example 12. *For* $\phi_i = (1, 3, 4, 2, 0)$ *with* $\lambda \sim 1.72$ *satisfying* $\lambda^4 - \lambda^3 - \lambda^2 - \lambda + 1 = 0$. $M_\phi = \begin{pmatrix} 0&0&0&1\\ 1&0&0&1\\ 1&0&1&0\\ 0&1&1&0 \end{pmatrix}$. *Below on the left is the polygon* P_ϕ. f_0 *sends rows* 1 *and* 2 *(labelled on the left) to columns* 1 *and* 2 *scaled by* $(\frac{1}{\lambda}, \lambda)$. *Rows* 3 *and* 4 *are also rotated by* π *after being scaled, and placed in columns* 3 *and* 4 *respectively.*



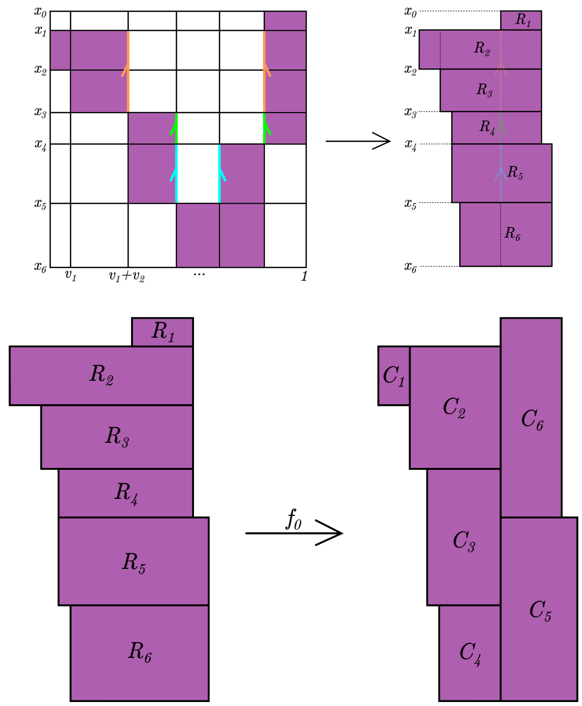
Below we show the branched double cover of P_ϕ *with a lift of the map* f_0 *shown above.*



For the double cover, take two copies of the polygon P_ϕ. We have rotated one copy on P_ϕ by π in the figure above and moved the columns horizontally until each (shaded) row is connected. Rows $R_1, \ldots, R_8$ of the double cover are labelled on the left as $1, \ldots, 8$. Each row R_i is multiplied by $(\frac{1}{\lambda}, \lambda)$ and placed in C_i labelled as i in the interior of P_ϕ (without any rotation in this example).

The red interior edge A_0 (quasigeodesic) has two preimages under continuous extensions of f_0: the two blue edges A_{-1}, which are identified as shown. These two have preimages in the green curves A_{-2}, which come from the purple A_{-3}'s, whose preimage are the pink curves. Under the identifications in $S = P/\sim$ there is only one cone point Q of angle 6π. Thus $\chi(S) = -2$ and S has genus 2.

Example 13. *For* $(\phi_i)_{i=0}^{6} = (2, 1, 3, 5, 6, 4, 0)$ *we get the matrix below, which we align* $(\alpha_i)_{i=1}^{5} = (1, 1, 1, -1, 1)$*, as shown below. This alignment satisfies the conditions and* f_0 *gives a homeomorphism of a surface with a quadratic differential of type* $Q(1, 1, 1, 1, 1, 1, 1, 5)$.



Note we may assume without loss of generality that each cycle of the permutation ϕ contains a critical point of $h_{/m}$. This is because if there was a cycle consisting entirely of non-critical points, we could simply remove them from our partition of $[0, 1]$ and get a smaller matrix with the same leading eigenvalue. The smaller matrix would still consist of only 0s and 1s, as these points are not critical. Finally, as this process doesn't change h, the smaller matrix would also satisfy the alignment conditions. With this assumption, $X = \{x_0, ..., x_n\}$ is in fact the postcritical set of h. Since we don't lose any eigenvalues λ this way, we can make this assumption in what follows if needed. We can then show (see [3, Thm 5] for an earlier formulation of this result):

Theorem 14. *For* $M_\phi \in \mathscr{O}\mathscr{B}_n$ *with* ϕ *a permutation of* $\mathbb{N}_n$ *let* P *be the associated poly-*

gon. If the columns of P can be moved horizontally until the rows of P are connected, and if there exists a function α satisfying the alignment condition aligning the rows of P, then:

(i) *f_0 defined as $R_i \mapsto C_i$ defines a pseudo-Anosov map of a compact surface S_g;*

(ii) *When M is ergodic and every cycle of ϕ contains a critical point of h_M, the genus g equals $\lfloor n/2 \rfloor$, and otherwise $g \leq n/2$;*

(iii) *ψ is orientation preserving (resp. reversing) when ϵ is 1 (resp. -1); and*

(iv) *When $g = \frac{1}{2}\dim(M)$ a basis for $H_1(S_g)$ may be chosen so that the action of ψ on $H_1(S_g)$ is given by M. In this case $\chi(M)$ is palindromic or anti-palindromic depending on whether ϵ is 1 or -1.*

Proof. Assume the pair (M, ϵ) satisfies the alignment conditions.

If x_i is a noncritical point of h, the interior of $R_i \cup R_{i+1}$ is mapped homeomorphically to its image: part (*ii*) of the alignment condition ensures that f_0 continuous on the intersection $R_i \cap R_{i+1}$.

If x_i is a critical point of h, the midpoint of H_i is sent to the midpoint of $H_\phi(i)$. $R_i \cap R_{i+1}$ is sent to extreme parts of $H_{\phi(i)}$ under f_0 inducing a cone point of angle π in the middle. After folding each horizontal edge H_i in half, f_0 is well-defined and continuous.

We may assume each H_i has positive width: If two adjacent rows R_i and R_{i+1} had the same width, then $x_{\phi^{-1}(i)}$ could not have been a critical point of h. So $R_{\phi^{-1}(i)} \cup R_{\phi^{-1}(i)+1}$ is mapped homeomorphically to $R_i \cup R_{i+1}$, and hence $R_{\phi^{-1}(i)}$ and $R_{\phi^{-1}(i)+1}$ have the same width, so $x_{\phi^{-2}(i)}$ must be noncritical, and so on. Since ϕ is a permutation, this implies all iterates of x_i must be noncritical. But we may assume WLOG (see end of section 3) that all cycles in the permutation ϕ contain critical points.

Now we turn to the vertical edges of ∂P. These are stretched by λ under f_0, and sent to other vertical edges or to the interior of P. Since the horizontal edges of ∂P are folded in half, the vertical edges are connected end-to-end and form a circle, say $\boldsymbol{S}$. i.e. S is the quotient of ∂P obtained by collapsing each H_i to a point. Denote the image of H_i in S as $\boldsymbol{a_i}$, $i = 0, \ldots, n$.

For x_i critical, $1 \leq i \leq n-1$, for some $c_i > 0$, the two arcs starting at a_i, say $\boldsymbol{J_i :=} (a_i - c_i, a_i)$ and $\boldsymbol{J'_i :=} (a_i, a_i + c_i)$ are identified and sent to the interior of P, (see figure 7). The points $a_i - c_i$ and $a_i + c_i$ are sent to the same point $a_{\phi(i)}$.

For x_i noncritical, f_0 does not induce any identifications under one iterate. The vertical edges at a_i are sent to $a_{\phi(i)}$. However, some iterate of x_i is a critical point, and as a result, similar identifications are induced at noncritical x_i (except at x_0 and x_n).

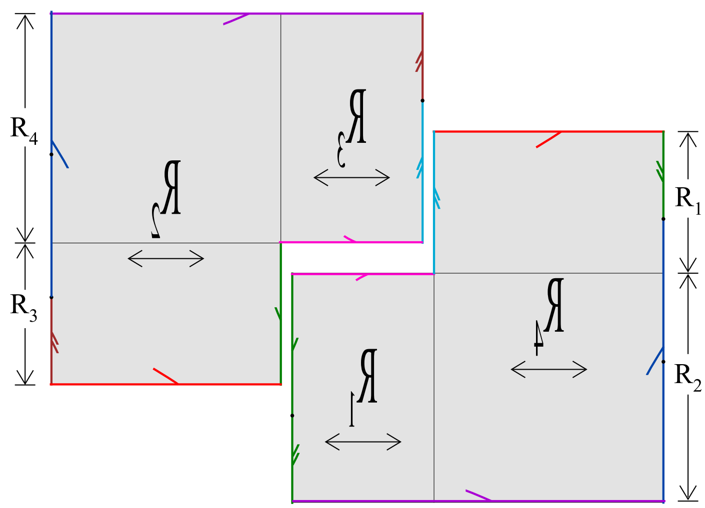
All gluings on vertical edges are induced this way. So the parts that are not identified under any forward iterate form $n-1$ closed intervals. Since f_0 stretches by $\lambda > 1$, these intervals have to be single points. Moreover, each component of the complement of these $n-1$ points is folded in half, so the $n-1$ points form a single point under identification, of angle $(n-1)\pi$, call it $\boldsymbol{Q}$.

All other points on ∂P are regular points (angle = 2π) by inspection. So we have $n+1$ points, Q_i, of angle π and a single point, Q, of angle $(n-1)\pi$. Thus the Euler characteristic of $P = P/\sim$ is 2, i.e. P is a sphere.

Taking the double cover of P, (ramified at Q and the Q_i, when n is even, and only ramified at the Q_i when n is odd), gives us a surface S_g of genus $g = n/2$ when n is even, and $g = (n-1)/2$ when n is odd. Lift the map f to a map $\psi : S_g \to S_g$, and lift the horizontal and vertical foliations from the plane, to get two transverse measured foliations on S_g, invariant under ψ.

Below we show the lift of the map of example 9 to its branched double cover. In general we pair the singularities and cut along paths on the boundary of P between singularities in each pair (i.e. take quasigeodesic paths). Now take two copies of this setup (P, P') and identify along the corresponding cuts between singularities, alternating the sheets at each cut.

Example 15.



In the figure above, we rotate P' by π and label the rows of P' as R_3, R_4. Identifications are simpler: each horizontal edge of ∂P is identitified to the corresponding edge in $\partial P'$ with opposite orientation (i.e. by a translation to the rotated copy). If the horizontal edge is not the top or the bottom, the adjacent vertical edges in the double are identified for a length equal to the smaller of the heights of the two adjacent rows. Since all identifications are by translations we obtain an abelian differential on this double cover, (with no zeros in this case). The lift can be chosen to either reflect every rectangle horizontally *or vertically; we choose horizontally in the figure above. The resulting surface is a torus and the map corresponds to the element* $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ *of determinant* -1.

Each curve between Q_i and Q_{i+1} in P lifts to two curves in the double cover. Taking one forwards, and the other backwards we get a closed loop. When n is even, these loops, for $i = 0, ..., n-1$, form a basis for $H_1(S_g)$. Under this basis, the induced map in homology ψ_* is given by the matrix $\pm M$, (depending on the choice of the lift of f to the double cover). □

CHAPTER 4

ASYMPTOTICS OF BIPERRON NUMBERS

This chapter is based on a paper written in collaboration with Huyngryul Baik and Chenxi Wu. We show that asymptotically almost all bi-Perron algebraic unit whose characteristic polynomial has degree at most $2n$ do not correspond to dilatations of pseudo-Anosov maps on a closed orientable surface of genus n for $n \geq 10$. As an application of the argument, we also obtain a statement on the number of closed geodesics of the same length in the moduli space of area one abelian differentials for low genus cases.

Let $n \geq 2$ be fixed, and R be any positive real number. Define $\mathcal{B}_n(R)$ to be the set of bi-Perron algebraic units no larger than R whose characteristic polynomial has degree at most $2n$. Here, by the characteristic polynomial of a bi-Perron algebraic unit λ, we mean the monic palindromic integral polynomial whose leading root is λ and has the lowest degree among all such polynomials. And let $\mathcal{D}_n(R)$ be the set of dilatations no larger than R of pseudo-Anosov maps with orientable invariant foliations on a closed orientable surface S_n of genus n. Similarly, let $\mathcal{D}'_n(R)$ be the set of dilatations no larger than R of pseudo-Anosov maps, not necessarily with orientable invariant foliation, on a closed orientable surface with genus n.

We remark that $\mathcal{D}_n(R)$ is contained in $\mathcal{B}_n(R)$, and $\mathcal{D}'_n(R)$ is contained in $\mathcal{B}_{3n-3}(R)$. A pseudo-Anosov dilatation λ on a surface of genus n is a root of an integral palindromic polynomial of degree at most $2n$ if its invariant foliations are orientable. This is because λ is the leading eigenvalue of the induced symplectic action on the homology group of the surface, which is $\mathbb{Z}^{2n}$. If we do not require the invariant foliation to be orientable, the upper bound on degree is $6n-6$: we can reduce this case to the case of orientable foliation by taking a double cover of the surface, and this bound follows from the fact that a quadratic differential on a surface of genus n has at most $2n$ zeros which is due to Gauss-Bonnet together with the Riemann-Hurwitz formula.

Note that Fried's conjecture is equivalent to $\mathcal{B}_n(R)$ being contained in $\mathcal{D}_m(R)$ for some large enough m. But a priori, m could be arbitrarily large and we do not know how to prove or disprove the claim. Instead we show the following.

Theorem 16. *Let $\mathcal{B}_n(R)$, $\mathcal{D}'_n(R)$ and $\mathcal{D}_n(R)$ be as above. Then*

1.
$$\lim_{R\to\infty} \frac{|\mathcal{D}_m(R) \cap \mathcal{B}_n(R)|}{|\mathcal{B}_n(R)|} = 0\,,$$
where $4m - 3 \leq n(n+1)/2$. *In particular,* $\lim_{R\to\infty} \frac{|\mathcal{D}_n(R)|}{|\mathcal{B}_n(R)|} = 0, \forall n \geq 6$.

2.
$$\lim_{R\to\infty} \frac{|\mathcal{D}'_m(R) \cap \mathcal{B}_n(R)|}{|\mathcal{B}_n(R)|} = 0\,,$$
where $6m - 6 \leq n(n+1)/2$. *In particular,* $\lim_{R\to\infty} \frac{|\mathcal{D}'_n(R)\cap\mathcal{B}_n(R)|}{|\mathcal{B}_n(R)|} = 0, \forall n \geq 10$.

Here $|A|$ means the cardinality of A for a finite set A.

Theorem 16 says that asymptotically almost all bi-Perron algebraic units whose characteristic polynomial has degree at most $2n$ do not correspond to dilatations of pseudo-Anosov maps on a surface of genus n for all $n \geq 10$, and for $n \geq 6$ if the invariant foliation is further assumed to be orientable. It would be interesting to see if the statement still holds for lower genera.

Let Γ_n be the set of all periodic orbits for the Teichmüller flow on the moduli space of area one abelian differentials on S_n. Then there exists a surjective map

$\Gamma_n \to \cup_{R>0}\mathcal{D}_n(R)$ defined by $\gamma \mapsto e^{\ell(\gamma)}$ where $\ell(\gamma)$ is the length of the orbit $\gamma \in \Gamma_n$. Let $\Gamma_n(R)$ be the preimage of $\mathcal{D}_n(R)$ of this map.

By $f \sim g$ we mean $\exists C$ such that $\frac{1}{C}f(x) \leq g(x) \leq Cf(x)$ when $x \gg 0$. By $f \lesssim g$ we mean $f = O(g)$ when $x \to \infty$. At our best knowledge, the following theorem is independently due to Eskin-Mirzakhani-Rafi [7] and Hamenstädt [11] (which is rephrased for our purpose).

Theorem 17 ([7], [11])**.** $|\Gamma_n(R)| \sim \frac{R^{4n-3}}{(4n-3)\log R}$.

We give a brief explanation of the above statement. In [7] and [11], it was stated that when restricted to each connected component of the strata of the moduli space of area one abelian differentials on S_n,

$$\begin{aligned}|\Gamma_n(R)| &= |\{\gamma \in \Gamma_n : e^{\ell(\gamma)} \leq R\}| \\ &= |\{\gamma \in \Gamma_n : \ell(\gamma) \leq \log R\}| \\ &\sim \frac{e^{(2n+\ell-1)\log R}}{(2n+\ell-1)\log R} = \frac{R^{2n+\ell-1}}{(2n+\ell-1)\log R}.\end{aligned}$$

Here ℓ is the maximum number of zeros of an area one abelian differential on S_n, so it is $2n-2$. See [22], [31], [6] and [7] for the relevant background of this theorem.

As a result, we have $|\Gamma_n(R)| \sim \frac{R^{4n-3}}{(4n-3)\log R}$ on each connected component of the strata. But for fixed n, there exists only finite number of such components. Therefore we get $|\Gamma_n(R)| \sim \frac{R^{4n-3}}{(4n-3)\log R}$ without restricting to the components. Note that since n is fixed, we can just say $|\Gamma_n(R)| \sim \frac{R^{4n-3}}{\log R}$.

As a direct corollary, we have:

Corollary 18. $|\mathcal{D}_n(R)| \lesssim \frac{R^{4n-3}}{\log R}$.

In the exactly same way, we can obtain an analogue of Corollary 18 for $|\mathcal{D}'_n(R)|$ from the following theorem which is due to Eskin and Mirzakhani [6]:

Theorem 19 (Theorem 1.1, [6]). *The number of geodesics in the moduli space of genus n surface of length at most* $\log(R) \sim \frac{R^{6n-6}}{(6n-6)\log(R)}$.

And as a direct corollary, just like Corollary 18, we have:

Corollary 20. $|\mathcal{D}'_n(R)| \lesssim \frac{R^{6n-6}}{\log R}$.

We remark that this does not imply $|\mathcal{D}_n(R)| \sim \frac{R^{4n-3}}{\log R}$ or $|\mathcal{D}'_n(R)| \sim \frac{R^{6n-6}}{\log R}$, since each element in $\mathcal{D}_n(R)$ or $\mathcal{D}'_n(R)$ may correspond to a lot of different closed geodesics in the moduli space.

Now we study how $|\mathcal{B}_n(R)|$ grows. Let $P_n(R)$ be the set of Perron polynomials of degree n with roots no larger than R, ($x > 1$ is Perron if it is the root of a monic irreducible polynomial with integer coefficients, so that the other roots of the polynomial are (strictly) less than x in absolute value).

Lemma 21. $|P_n(R)| \sim R^{n(n+1)/2}$.

Proof. $|P_n(R)| \lesssim R^{n(n+1)/2}$: Due to Vieta's formula the absolute value of the coefficient of x^k of a monic, degree n polynomial with all roots no larger than R is $\lesssim R^{n-k}$. Hence, the total number of such polynomials must be $\lesssim \prod_k R^{n-k} = R^{n(n+1)/2}$.

$R^{n(n+1)/2} \lesssim |P_n(R)|$: Let a_k be the coefficient of x^k in a degree n monic polynimial (so $a_n = 1$). By Rouché's theorem, when

$$1 > |\frac{a_0}{R^n}| + |\frac{a_1}{R^{n-1}}| + \cdots + |\frac{a_{n-1}}{R}| \tag{4.1}$$

$$\left(\frac{1}{2}\right)^{n-1}|\frac{a_{n-1}}{R}| > |\frac{a_0}{R^n}| + \cdots + |\frac{a_{n-2}}{R^2}|\left(\frac{1}{2}\right)^{n-2} + \left(\frac{1}{2}\right)^n \tag{4.2}$$

and

$$\left(\frac{1}{3}\right)^{n-1}|\frac{a_{n-1}}{R}| > |\frac{a_0}{R^n}| + \cdots + |\frac{a_{n-2}}{R^2}|\left(\frac{1}{3}\right)^{n-2} + \left(\frac{1}{3}\right)^n \tag{4.3}$$

one root λ of this polynomial has magnitude between $R/2$ and R while all other roots have magnitude smaller than $R/3$. Hence λ must be real. If $\lambda < 0$ one can multiply $(-1)^{n-k}$ to a_k to get a polynomial with a root $-\lambda$. Hence, half of those polynomials have a leading positive real root, so they are in $P_n(R)$.

The inequalities (4.1), (4.2) and (4.3) are satisfied if and only if the point $(a_0/R^n, \ldots, a_{n-1}/R)$ lies in a non-empty open subset $U \subset [-1,1]^n$. The number of such points as $R \to \infty$ converges to the volume of this open subset divided by the co-volume of the lattice $\mathbb{Z}/R^n \times \mathbb{Z}/R^{n-1} \times \ldots \times \mathbb{Z}/R$, and the co-volume of this lattice is $R^{-n(n+1)/2}$.

□

Lemma 22. $\lim_{R\to\infty} \frac{|\{reducible\ elements\ in\ P_n(R)\}|}{|P_n(R)|} = 0$

Proof. Because any reducible monic integer polynomial can be written as the product of two monic integer polynomials of lower degree, we have:

$$|\{\text{reducible elements in } P_n(R)\}| \leq \sum_k |b_k(R)||b_{n-k}(R)|$$

where $b_k(R)$ is the set of monic polynomials with roots bounded by R. The first part of the proof of the previous lemma implies that $|b_k(R)| \lesssim R^{k(k+1)/2}$, hence $|\{\text{reducible elements in } P_n(R)\}| \sim o(R^{n(n+1)/2})$. □

Let $\mathcal{P}_n(R)$ be the set of Perron numbers of degree n no larger than R. They have one-one correspondence with irreducible elements in $P_n(R)$ which by

Lemma 22 constitute almost all of $P_n(R)$ asymptotically. Hence by Lemma 21, $|\mathcal{P}_n(R)| \sim R^{n(n+1)/2}$.

Lemma 23. $|\mathcal{B}_n(R)| \sim R^{n(n+1)/2}$.

Proof. When $x > 1$, $1/x < 1$. Hence, $x \in \mathcal{B}_n(R)$ implies that $x + 1/x \in \mathcal{P}_n(R+1)$, hence $|\mathcal{B}_n(R)| \lesssim (R+1)^{n(n+1)/2} \sim R^{n(n+1)/2}$.

On the other hand, from the proof of Lemma 21 and the discussion above, the number of Perron numbers of degree n which lie between R and $R/2$ and have all conjugates smaller than $R/3$ is $\sim R^{n(n+1)/2}$. When R is sufficiently large, each such Perron number y correspond to a bi-Perron number x by the relation $x + 1/x = y$. And in fact x is an algebraic unit (so it is in $\mathcal{B}_n(R)$). Note that both roots of the polynomial $x^2 - yx + 1$ are algebraic integers, since it is a monic polynomial with algebraic integer coefficients. Furthermore, the product of these roots is 1, so they must be algebraic units. Hence $\mathcal{B}_n(R) \gtrsim R^{n(n+1)/2}$. □

Now we are ready to prove our main theorem.

Proof of Theorem 16. By Corollary 18 and Lemma 23, we get

$$\frac{|\mathcal{D}_m(R) \cap \mathcal{B}_n(R)|}{|\mathcal{B}_n(R)|} \lesssim \frac{R^{4m-3}}{\log R \cdot R^{n(n+1)/2}}.$$

The right-hand side goes to 0 as R goes to $+\infty$ as long as we have $4m - 3 \leq n(n+1)/2$. When $m = n$, this inequality is satisfied if and only if $n \geq 6$ (recall that n is always assumed to be at least 2), which proved part (1). Part (2) follows from the same argument above but using Corollary 20 instead of Corollary 18. □

Recall that Γ_n is the set of all closed geodesics in the moduli space of area one abelian differentials on the surface S_n, and $\Gamma_n(R)$ is the subset of Γ_n which consists of the closed geodesics of length no larger than $\log R$. For each $\gamma \in \Gamma_n$, let m_γ be the number $|\{\gamma' \in \Gamma_n : \ell(\gamma') = \ell(\gamma)\}|$.

We remark that if m_γ were uniformly bounded, one could have obtained Theorem 16 (1) for $n \geq 2$ instead of $n \geq 6$ using Theorem 1 (1) of [13]. But at least in the low genus cases, this is not true. As an application of our argument, we obtain the following theorem.

Theorem 24. *Suppose $n \leq 5$. For any positive integer k, the set $\{\gamma \in \Gamma_n : m_\gamma \geq k\}$ is typical, i.e.,*

$$\lim_{R\to\infty} \frac{|\{\gamma \in \Gamma_n(R) : m_\gamma \geq k\}|}{|\Gamma_n(R)|} \to 1.$$

Proof. Suppose not. Then for some k, we have

$$\limsup \frac{|\{\gamma \in \Gamma_n(R) : m_\gamma < k\}|}{|\Gamma_n(R)|} > 0.$$

But this implies that

$$\limsup \frac{|\mathcal{D}_n(R)|}{|\Gamma_n(R)|} \geq \limsup \frac{\frac{1}{k}|\{\gamma \in \Gamma_n(R) : m_\gamma < k\}|}{|\Gamma_n(R)|} > 0.$$

On the other hand, we know that $\mathcal{D}_n(R) \subset \mathcal{B}_n(R)$. As a consequence, $\lim \frac{|\mathcal{B}_n(R)|}{|\Gamma_n(R)|} \geq \limsup \frac{|\mathcal{D}_n(R)|}{|\Gamma_n(R)|}$. By Corollary 18 and Lemma 23, we get

$$\frac{|\mathcal{B}_n(R)|}{|\Gamma_n(R)|} \sim \frac{R^{n(n+1)/2}\log R}{R^{4n-3}} \to 0, \text{ for } n \leq 5,$$

a contradiction.

□

BIBLIOGRAPHY

[1] Pierre Arnoux and Jean-Christophe Yoccoz. Construction de difféomorphismes pseudo-anosov. *CR Acad. Sci. Paris Sér. I Math*, 292(1):75–78, 1981.

[2] Artur Avila and Marcelo Viana. Simplicity of lyapunov spectra: proof of the zorich-kontsevich conjecture. *Acta mathematica*, 198(1):1–56, 2007.

[3] Hyungryul Baik, Ahmad Rafiqi, and Chenxi Wu. Constructing pseudo-anosov maps with given dilatations. *Geometriae Dedicata*, 180(1):39–48, 2016.

[4] Lipman Bers. An extremal problem for quasiconformal mappings and a theorem by thurston. *Acta Mathematica*, 141(1):73–98, 1978.

[5] André de Carvalho and Toby Hall. Unimodal generalized pseudo-anosov maps. *Geometry & Topology*, 8(3):1127–1188, 2004.

[6] Alex Eskin and Maryam Mirzakhani. Counting closed geodesics in moduli space. *arXiv preprint arXiv:0811.2362*, 2008.

[7] Alex Eskin, Maryam Mirzakhani, and Kasra Rafi. Counting closed geodesics in strata. *arXiv preprint arXiv:1206.5574*, 2012.

[8] Benson Farb and Dan Margalit. A primer on mapping class groups, volume 49 of princeton mathematical series. 2012.

[9] Albert Fathi, François Laudenbach, and Valentin Poénaru. Travaux de thurston sur les surfaces, volume 66 of astérisque. société mathématique de france, paris, 1979. séminaire orsay. *With an English summary*, 332, 1979.

[10] David Fried. Growth rate of surface homeomorphisms and flow equivalence. *Ergodic Theory and Dynamical Systems*, 5(4):539–563, 1985.

[11] Ursula Hamenstaedt. Bowen's construction for the teichmueller flow. *arXiv preprint arXiv:1007.2289*, 2010.

[12] Ursula Hamenstaedt. Symbolic dynamics for the teichmüller flow. *arXiv preprint arXiv:1112.6107*, 2011.

[13] Ursula Hamenstaedt. Typical properties of periodic teichmueller geodesics. *arXiv preprint arXiv:1409.5978*, 2014.

[14] Eriko Hironaka. Small dilatation mapping classes coming from the simplest hyperbolic braid. *Algebraic & Geometric Topology*, 10(4):2041–2060, 2010.

[15] John Hamal Hubbard. Teichmüller theory and applications to geometry, topology, and dynamics, volume i: Teichmüller theory. *Matrix Editions*, 2006.

[16] John Hamal Hubbard. Teichmüller theory and applications to geometry, topology, and dynamics, volume 2: Surface homeomorphisms and rational functions. *Matrix Editions*, 2016.

[17] Erwan Lanneau and Jean-Luc Thiffeault. On the minimum dilatation of braids on punctured discs. *Geometriae Dedicata*, 152(1):165–182, 2011.

[18] Erwan Lanneau and Jean-Luc Thiffeault. On the minimum dilatation of pseudo-anosov homeromorphisms on surfaces of small genus (dilatation minimales des homéomorphismes de type pseudo-anosov sur des surfaces de petit genres). In *Annales de l'institut Fourier*, volume 61, pages 105–144, 2011.

[19] Christopher J Leininger. On groups generated by two positive multi-twists: Teichmüller curves and lehmers number. *Geometry & Topology*, 8(3):1301–1359, 2004.

[20] D. A. Lind. The entropies of topological markov shifts and a related class of algebraic integers. *Ergodic Theory Dynam. Systems*, 4(2):283–300, 1984.

[21] DD Long. Constructing pseudo-anosov maps. In *Knot theory and manifolds*, pages 108–114. Springer, 1985.

[22] Howard Masur. Interval exchange transformations and measured foliations. *Annals of Mathematics*, 115(1):169–200, 1982.

[23] Curtis T McMullen. Polynomial invariants for fibered 3-manifolds and teichmüller geodesics for foliations. In *Annales scientifiques de lEcole normale supérieure*, volume 33, pages 519–560. Elsevier, 2000.

[24] Michal Misiurewicz and Wieslaw Szlenk. Entropy of piecewise monotone mappings. *Studia Mathematica*, 67(1):45–63, 1980.

[25] Robert C Penner. A construction of pseudo-anosov homeomorphisms. *Transactions of the American Mathematical Society*, 310(1):179–197, 1988.

[26] Balázs Strenner. Algebraic degrees of pseudo-anosov stretch factors. *Geometric and Functional Analysis*, 27(6):1497–1539, 2017.

[27] Oswald Teichmüller. *Extremale quasikonforme Abbildungen und quadratische Differentiale*. Number 22. Verlag der Akademie der Wissenschaften, in Kommission bei W. de Gruyter u. Company, 1940.

[28] William P Thurston. A norm for the homology of 3-manifolds. *Mem. Amer. Math. Soc.*, 339:99–130, 1986.

[29] William P. Thurston. Entropy in dimension one. *Frontiers in Complex Dynamics: In Celebration of John Milnor's 80th Birthday*, pages 339–384, 2014.

[30] William P Thurston et al. On the geometry and dynamics of diffeomorphisms of surfaces. *Bulletin (new series) of the american mathematical society*, 19(2):417–431, 1988.

[31] William A Veech. The teichmuller geodesic flow. *Annals of Mathematics*, 124(3):441–530, 1986.