Algebraic degrees of pseudo-Anosov stretch factors

Balázs Strenner

November 4, 2017

Abstract

The motivation for this paper is to justify a remark of Thurston that the algebraic degree of stretch factors of pseudo-Anosov maps on a surface S can be as high as the dimension of the Teichmüller space of S. In addition to proving this, we completely determine the set of possible algebraic degrees of pseudo-Anosov stretch factors on almost all finite type surfaces. As a corollary, we find the possible degrees of the number fields that arise as trace fields of Veech groups of flat surfaces homeomorphic to closed orientable surfaces. Our construction also gives an algorithm for finding a pseudo-Anosov map on a given surface whose stretch factor has a prescribed degree. One ingredient of the proofs is a novel asymptotic irreducibility criterion for polynomials.

1 Introduction

Let S be a finite type surface. An element f of the mapping class group $\mathrm{Mod}(S)$ is *pseudo-Anosov* if there is a representative homeomorphism ψ, a number $\lambda > 1$ called the stretch factor (or dilatation), and a pair of transverse invariant singular measured foliations $\mathcal{F}^u$ and $\mathcal{F}^s$ such that $\psi(\mathcal{F}^u) = \lambda\mathcal{F}^u$ and $\psi(\mathcal{F}^s) = \lambda^{-1}\mathcal{F}^s$. The stretch factor λ is an algebraic integer. The goal of this paper is to determine the possible algebraic degrees of λ.

Studying the number λ is motivated by its connections to algebraic geometry, geometric topology and dynamics. For example, $\log(\lambda)$ is the translation length of f on Teichmüller space with the Teichmüller metric and hence the length of a closed geodesic in moduli space [FM12, Section 14.2.2]. The volume of the hyperbolic 3-manifold obtained as the mapping torus of f is also related to $\log(\lambda)$ [KM14]. Finally, the extension field $\mathbb{Q}(\lambda + \lambda^{-1})$ plays a role in Teichmüller dynamics [McM03].

Thurston's remark. Thurston announced the classification of elements of $\mathrm{Mod}(S)$ to finite order, reducible and pseudo-Anosov elements in his seminal bulletin paper [Thu88] (which had been circulating as a preprint much earlier). On pages 427–428, he provides a bound for the algebraic degree of pseudo-Anosov stretch factors λ. He denotes the dimension of the Teichmüller space of S by d and writes:

> *Therefore λ is an algebraic integer of degree $\leq d$. The examples of Theorem 7 show that this bound is sharp.*

Theorem 7, which describes a construction of pseudo-Anosov mapping classes using Dehn twists, definitely should produce examples realizing the degree d, but Thurston did not explain this in the paper, nor did anyone else since then. The intuition that supports Thurston's claim is that a random degree d polynomial is likely to be irreducible. However, it is not clear if the polynomials arising from Thurston's constuction are random in any sense. Another difficulty is the lack of irreducibility criteria that apply for defining polynomials of stretch factors. For example, the well-known Eisenstein's criterion does not apply since it requires the constant term of the polynomial to be divisible by a prime, whereas the polynomials in question always have constant term ± 1.

Even the fact that the degree can grow linearly with the genus is nontrivial. This is due to Arnoux and Yoccoz [AY81] who constructed a degree g stretch factor on each closed orientable surface S_g of genus $g \geq 3$. More recently Shin [Shi16] also realized the degree $2g$ on S_g. (For S_g, the dimension of Teichmüller space and hence the maximal degree predicted by Thurston is $6g-6$.)

The main result. In this paper, not only do we realize the theoretical maximum $6g-6$, but we completely answer the question of which degrees appear on S_g. Moreover, we also answer this question for most finite type surfaces, including all nonorientable surfaces, for which Thurston's construction does not apply.

Let $D(S)$ be the set of possible algebraic degrees of stretch factors of pseudo-Anosov elements of $\mathrm{Mod}(S)$. Let $D^+(S) \subset D(S)$ be the set of degrees arising from pseudo-Anosov maps with a transversely orientable invariant foliation. Finally, denote by $[a,b]_{\text{even}}$ and $[a,b]_{\text{odd}}$ the set of even and odd integers, respectively, in the interval $[a,b]$.

Theorem 1.1. *Let $g \geq 2$. We have*

$$D(S_g) = \big[2, 6g-6\big]_{\text{even}} \cup \big[3, 3g-3\big]_{\text{odd}}$$

and

$$D^+(S_g) = \big[2, 2g\big]_{\text{even}} \cup \big[3, g\big]_{\text{even}}.$$

In fact, we prove a more general result in Theorem 8.9, where we also determine $D^+(S)$ for any finite type surface, and $D(S)$ for

- nonorientable surfaces with any number of punctures and
- orientable surfaces with an even number of punctures.

We also almost completely determine $D(S)$ for orientable surfaces with an odd number of punctures. The reason we miss some cases is that these surfaces are not double covers of nonorientable surfaces. However, our construction relies on such a covering to realize the highest possible odd degree.

The fact that $D(S_g)$ and $D^+(S_g)$ cannot be larger than stated in Theorem 1.1 is well-known. We have $\min D(S_g) \geq 2$, because only 1 and -1 are algebraic units of degree 1. Thurston [Thu88] showed that $\max D(S_g) \leq 6g-6$ and $\max D^+(S_g) \leq 2g$, and Long [Lon85] showed that if $d \in D(S)$ is odd, then $d \leq 3g-3$. A similar argument shows that if $d \in D^+(S)$ is odd, then $d \leq g$. (Here and in what follows, we use d to denote any degree, not necessarily the maximal one.)

Irreducibility of polynomials via converging roots. In the work of Arnoux–Yoccoz [AY81] and Shin [Shi16], a construction of pseudo-Anosov maps is given such that λ is a root of a polynomial which can be shown to be irreducible[1]. The irreducibility criteria in both cases are specific to one particular sequence of polynomials and cannot easily be generalized to realize other degrees.

In this paper we use a novel irreducibility criterion. Since the statement is elementary and the proof is short, we include the criterion here together with the proof.

Lemma 1.2. *Let $u_k(x) \in \mathbb{Z}[x]$ be a sequence of integral polynomials and let $\lambda_k \in \mathbb{C}$ be a sequence of roots, that is, we assume $u_k(\lambda_k) = 0$ for all k. Suppose that there exists $v(x) \in \mathbb{C}[x]$ such that*

$$\lim_{k\to\infty} \frac{u_k(x)}{x - \lambda_k} = v(x). \tag{1.1}$$

Suppose $v(\theta) = 0$ for some $\theta \in \mathbb{C}$, and let $\theta_k \to \theta$ be a sequence with $u_k(\theta_k) = 0$ for all k. If $\theta_k \neq \theta$ for all but finitely many k, then θ_k and λ_k are roots of the same irreducible factor of $u_k(x)$ for all but finitely many k.

Proof. Factor the polynomials $u_k(x)$ to irreducible factors over $\mathbb{Z}$. Assume for a contradiction that θ_k and λ_k are in different irreducible factors for infinitely many k. By restricting to a subsequence, we may assume that θ_k and λ_k are in different irreducible factors for all k. For all k, let $w_k(x) \in \mathbb{Z}[x]$ be an irreducible factor with root θ_k. Since there is a uniform bound on the absolute values of the roots of $w_k(x)$, we can restrict further to a subsequence such that all $w_k(x)$ have the same degree and $w_k(x) \to w(x)$ for some $w(x)$. Since $w_k(x) \in \mathbb{Z}[x]$, we have $w(x) \in \mathbb{Z}[x]$ and $w_k(x) = w(x)$ if k is large enough. In particular, $\theta_k = \theta$ if k is large enough, which is a contradiction. ☐

The criteria are cleanest to state and prove for sequences of polynomials, but quantitative versions would also be possible to obtain using Newton's formulas for coefficients of polynomials in terms of the roots.

Finding stretch factors with prescribed degrees. To construct stretch factors of prescribed algebraic degrees, we use Penner's construction of pseudo-Anosov maps. In this construction one has two choices: the choice of a collection of curves and the choice of a product of Dehn twists about these curves. Interestingly, the algebraic degree of a stretch factor arising from this construction seems to depend primarily only on the choice of the collection of curves, and not the Dehn twist product. More specifically, computer experiments have led to the following observation.

> *The algebraic degree of a stretch factor arising from Penner's construction typically equals the rank of the intersection matrix of the collection of curves.*

However, the choice of the Dehn twist product also has some effect on the algebraic degree of the stretch factor. Unfortunate choices of Dehn twist products may result in a lower algebraic degree than the typical one.

On the other hand, we will show that for certain infinite sequences of Dehn twist products the above observation is guaranteed to hold asymptotically. This way we

[1] Irreducibility in this paper is always meant over $\mathbb{Z}$.

obtain a simple criterion (Theorem 5.4) stating that if a collection of curves with rank d intersection matrix exists on a surface S, then $d \in D(S)$. With this criterion in hand, the problem of realizing algebraic degrees reduces to the problem of constructing collections of curves on surfaces with certain properties. We construct collections of curves in Section 6.

Constructing sequences of polynomials with converging roots. In order to prove Theorem 5.4, we construct sequences $f_1, f_2, \dots$ of pseudo-Anosov mapping classes such that the defining polynomials $u_k(x)$ of the stretch factors satisfy the hypotheses of Lemma 1.2 and relate the rank of the intersection matrix to the number of disjoint sequences $\theta_k \to \theta$ where $\theta_k \neq \theta$ for all but finitely many k.

The sequences f_k are constructed as follows. Fixing a collection of curves in Penner's construction, we start with some product of Dehn twists and we modify this product for each k by replacing the Dehn twists with their kth power. It turns out that the defining polynomials of such sequences have the asymptotic behavior as in Lemma 1.2. We prove this in three parts, presented in Sections 3 to 5.

The fact that the defining polynomials converge in the sense of (1.1) is shown in Theorem 3.1. This has two main parts: showing that the left Perron–Frobenius eigenvectors of certain matrices associated to the Dehn twist products converge (Section 3.1) and showing that the left actions of the matrices on the orthogonal complements of these eigenvectors also converge (Section 3.3).

In Section 4, we show that the limit of the left actions in the previous paragraph turns out to be a composition of projections from hyperplanes in $\mathbb{R}^n$ to other hyperplanes in $\mathbb{R}^n$. Moreover, for appropriate choices of Dehn twist products, cancellations occur in this composition of projections, so the limit of the left actions *is* a projection and therefore we have $v(x) = x(x-1)^s$ for the limit polynomial in Lemma 1.2. Therefore we will use Lemma 1.2 with $\theta = 0$ or 1.

Finally, Proposition 5.1 relates the rank of the intersection matrix to the number of roots of $u_k(x)$ that are different from 1. The roots are also always different from 0. This gives a count for sequences $\theta_k \to \theta$ where $\theta_k \neq \theta$ for all k.

Realizing degrees algorithmically. Our approach also provides an algorithm for finding a pseudo-Anosov mapping class on a given surface whose stretch factor has a prescribed algebraic degree. Indeed, the pseudo-Anosov mapping classes f_k described above have stretch factors that eventually have the prescribed degree. Therefore one can iterate over $f_1, f_2, \dots$ to find a desired example in finite time.

The Dehn twist products for which the degree of the stretch factor is not the rank of the intersection matrix seem to be rare, so in practice f_1 is very often already a good example. However, it would be interesting to prove a bound on the smallest k such that f_k is guaranteed to have a stretch factor with the prescribed degree. Such a bound seems attainable by effectivizing Lemma 1.2 and estimating the rate of convergence in Theorem 3.1. Not only would such a bound give an estimate on the running time of the algorithm, but it would also allow one to give formulas for mapping classes whose stretch factors have prescribed algebraic degrees.

Degrees of trace fields of Veech groups. Our second result concerns trace fields of Veech groups. A half-translation surface is a surface with a singular Euclidean structure with trivial or $\mathbb{Z}_2$-holonomy. Its Veech group is the group of its $\mathrm{PSL}(2,\mathbb{R})$-symmetries. Every Veech group is a Fuchsian group, and its trace field is a natural invariant of the half-translation surface. There is a half-translation surface associated with every pseudo-Anosov map f which is defined by the stable and unstable foliations of f. The trace field of the Veech group of this surface is $\mathbb{Q}(\lambda+\lambda^{-1})$, where λ is the stretch factor of f. The degree of the field extension $\mathbb{Q}(\lambda+\lambda^{-1}) : \mathbb{Q}$ is either the algebraic degree of λ or half of it. For more details, see [FM12, §11.3], [Zor06], [KS00, §7], [GJ00, §7] or [McM03, §9].

Which Fuchsian groups arise as Veech groups is an open question [HMSZ06, Problem 5]. Whether there is a cyclic Veech group generated by a hyperbolic element is also unknown [HMSZ06, Problem 6]. Also little is known about the number fields that arise as trace fields of Veech groups. As a corollary of our results on the algebraic degrees of stretch factors, we obtain the following.

Theorem 1.3. *The set of degrees of number fields that arise as trace fields of Veech groups of half-translation surfaces homeomorphic to S_g is $\{1,\dots,3g-3\}$.*

Pseudo-Anosov mapping classes that are not lifts. David Futer and Samuel Taylor has pointed out to us the following corollary of Theorem 1.1.

Corollary 1.4. *For any $g \geq 2$, there exists a pseudo-Anosov element of $\mathrm{Mod}(S_g)$ that has no power that arises by lifting a pseudo-Anosov mapping class on a lower genus surface by a branched covering.*

Proof. Any pseudo-Anosov mapping class whose stretch factor has degree $6g-6$ has the required property. This is because the degree is preserved under taking powers (Lemma 8.2) and it is clearly also preserved under lifts. However, $6g-6$ is not a possible degree on lower genus surfaces. $\square$

Bestvina and Fujiwara have also described a property of pseudo-Anosov maps such that if this property is satisfied, then no power of the map is a lift by a branched covering [BF17, Lemma 6.2]. In Example 6.4 of their paper, they build an explicit example in genus 3 satisfying with the above property. Earlier, Bestvina and Fujiwara also proved a result analogous to Corollary 1.4 for unbranched coverings instead of branched coverings [BF07, Proposition 4.2].

We remark that Corollary 1.4 can easily be generalized to punctured and nonorientable surfaces also using the more general Theorem 8.9 instead of Theorem 1.1.

Open questions. Our construction uses Penner's construction, not Thurston's construction, therefore Thurston's remark that the maximal degree arises from his construction is yet to be justified. We note that it is possible that his remark applies only to orientable surfaces, because we are not aware of a natural adaptation of his construction to nonorientable surfaces.

The algebraic degree of λ is an interesting measure of complexity of f. Franks and Rykken [FR99] (see also [GJ00, Theorem 5.5]) showed that if S is orientable and $\mathcal{F}^u$

and $\mathcal{F}^s$ are transversely orientable, then f is a lift of an Anosov mapping class of the torus by a branched covering if and only if the degree of λ is 2. Farb conjectured that this phenomenon generalizes to higher degrees.

Conjecture 1.5 (Farb)**.** *Given any d there exists $h(d)$ so that any pseudo-Anosov map with degree d stretch factor on a closed orientable surface arises by lifting a pseudo-Anosov map on some surface of genus at most $h(d)$ by a (branched or unbranched) cover.*

However, this generalization turns out to be false. Leininger and Reid [LR17] and independently Yazdi [Yaz17] have recently announced that they have counterexamples to Conjecture 1.5.

There are many other open questions about the degrees of stretch factors. Margalit asked what the possible algebraic degrees of stretch factors in the Torelli group are. Computer experiments suggest that the same degrees occur in the Torelli group as in the whole mapping class group. We wonder if the methods of this paper can be used to prove this. One can ask the same question for any other subgroup of $\mathrm{Mod}(S)$. For the point-pushing subgroup, it would be interesting to know if the degree of the stretch factor is related to some property of the corresponding element of the fundamental group.

It is also not known what degrees are generic in $\mathrm{Mod}(S)$, its subgroups or strata of the holomorphic quadratic differentials over the moduli space of S. We conjecture that the largest possible degree ($6g-6$ in the case of S_g) is generic in the whole mapping class group. In the non-maximal strata, the degree cannot reach $6g-6$, and the likely scenario is that the generic degree in each stratum is the maximal degree that can be realized in that stratum.

Finally, many of these questions have versions for outer automorphisms of free groups.

The structure of the paper. In Section 2 we review some basic facts about Penner's construction. In Sections 3 to 5 we prove the three parts contributing to the proof of Theorem 5.4 which reduces the problem of realizing degrees to realizing ranks of intersection matrices of collections of curves. We construct collections of curves in Section 6. In two cases, we fail to construct a collection of curves whose intersection matrix has the desired rank. In these two cases, we give explicit examples of pseudo-Anosov mapping classes in Section 7 without using Penner's construction. The proofs of the main theorems are given in Section 8.

Acknowledgements. This work is an improvement over a part of the Ph.D. thesis of the author. He thanks his advisor, Autumn Kent, for her guidance and Dan Margalit for suggesting the problem and for many valuable comments. He also appreciates helpful conversations with Joan Birman, Jordan Ellenberg, Benson Farb, Eriko Hironaka and Jean-Luc Thiffeault. Special thanks to the referees whose comments have greatly helped improve the clarity of the exposition.

The author was partially supported by the grant NSF DMS-1128155.

2 Background on Penner's construction

In this section we recall Penner's construction and some facts about the construction that will be used later in the paper.

2.1 Penner's construction

Consider the annulus $A = S^1 \times [0,1]$. We orient A via its embedding to the (θ, r)-plane (the plane parametrized in polar coordinates) by the map $(\theta, t) \mapsto (\theta, t+1)$. The orientation of A is defined to be consistent with the standard orientation of the plane. The standard Dehn twist $T : A \to A$ is defined by the formula $T(\theta, t) = (\theta + 2\pi t, t)$.

Let c be a two-sided simple closed curve on a surface S, and let $\phi : A \to S$ be a homeomorphism between A and a regular neighborhood of c. We refer to the pair (c, ϕ) as a *marked curve*. The *Dehn twist* about the marked curve (c, ϕ) is the homeomorphism $T_{c,\phi}$ defined by the formula

$$T_{c,\phi}(x) = \begin{cases} \phi \circ T \circ \phi^{-1}(x) & \text{if } x \in \phi(A) \\ x & \text{if } x \in S - \phi(A). \end{cases}$$

In the rest of the paper we will not distinguish between $T_{c,\phi}$ and its mapping class. Note that if S is oriented, then $T_{c,\phi}$ is the left Dehn twist T_c if ϕ is orientation-preserving and the right Dehn twist T_c^{-1} otherwise.

Let (c, ϕ_c) and (d, ϕ_d) be marked curves that intersect at a point p. We say that they are *marked inconsistently* at p if the pushforward of the orientation of A by ϕ_c and ϕ_d disagree near p.

Two simple closed curves on a surface are in *minimal position* if they realize the minimal intersection number in their homotopy classes. A collection of simple closed curves C on a surface is *filling* if the curves are in pairwise minimal position and the components of $S - C$ are disks or once-punctured disks.

Penner gave the following construction for pseudo-Anosov mapping classes [Pen88]. See also [Fat92] for a different proof.

Penner's Construction (General case)**.** *Let $C = \{(c_1, \phi_1), \ldots, (c_n, \phi_n)\}$ be a filling collection of marked curves on S. Suppose that they are marked inconsistently at every intersection. Then any product of the T_{c_i,ϕ_i} is pseudo-Anosov provided each twist is used at least once.*

When S is orientable, the hypotheses imply that C is a union of two multicurves A and B, and the statement takes the following more well-known form.

Penner's Construction (Orientable case)**.** *Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$ be a pair of filling multicurves on an orientable surface S. Then any product of T_{a_j} and $T_{b_k}^{-1}$ is pseudo-Anosov provided that each twist is used at least once.*

We remark that if C is a union of a pair of multicurves and the curves are marked inconsistently at every intersection, then the surface filled by C is necessarily orientable. Hence if S is nonorientable, then C in Penner's construction cannot be a union of two multicurves.

2.2 Oriented collections of marked curves

Let (c, ϕ) be a marked curve. We define its *left side* as $\phi(S^1 \times \{0\})$ and its *right side* as $\phi(S^1 \times \{1\})$. Note that the marking ϕ also induces an orientation of c from the standard (counterclockwise) orientation of S^1.

Suppose (c, ϕ_c) and (d, ϕ_d) are marked inconsistently at some $p \in c \cap d$. When we follow c in the direction of its orientation near p, we either cross from the left side of d to the right side of d or the other way around. In the first case, we call p a *left-to-right* crossing. In the second case, we call it *right-to-left* crossing. Note that the definition is symmetric in c and d: if c crosses from the left side of d to the right side of d, then d also must cross from the left side of c to the right side of c (Figure 2.1).

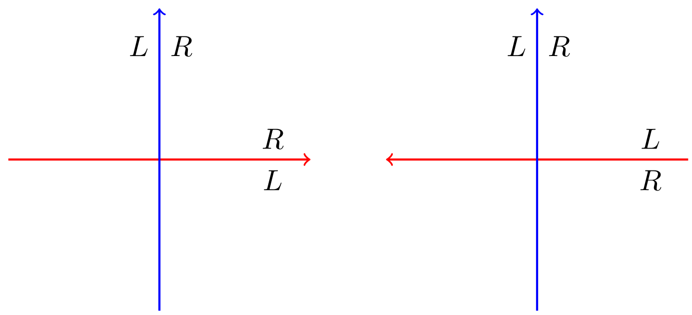

Figure 2.1: A left-to-right and a right-to-left crossing.

Let C be a collection of marked curves which are inconsistently marked at each intersection. Then C is called *completely left-to-right* if all crossings are left-to-right and *completely right-to-left* if all crossings are right-to-left.

We also recall that a singular foliation on a surface is *transversely orientable* if there is a continuous choice of vectors at the non-singular points of the foliation so that at each point the chosen vector is not tangent to the leaf of the foliation going through that point.

Proposition 2.1. *If C is completely left-to-right or completely right-to-left, then the pseudo-Anosov maps constructed from it by Penner's construction have a transversely orientable invariant foliation.*

Proof. Penner [Pen88, p. 188] observed that by smoothing out the intersections of C, one obtains a bigon track τ^+ invariant under the Dehn twists of C and hence under any pseudo-Anosov map ψ constructed from his construction. By choosing the smoothings differently at every intersection, we get a track τ^- invariant under ψ^{-1}. The unstable foliation is carried by τ^+, and the stable foliation is carried by τ^-.

When C is completely left-to-right, τ^- is transversely orientable: there is a continuously varying set of vectors transverse to τ^- such that the vectors point toward the right side of the curves of C (Figure 2.2). Hence the stable foliation is transversely orientable.

Similarly, when C is completely right-to-left, τ^+ is transversely orientable, and the unstable foliation is transversely orientable. $\square$

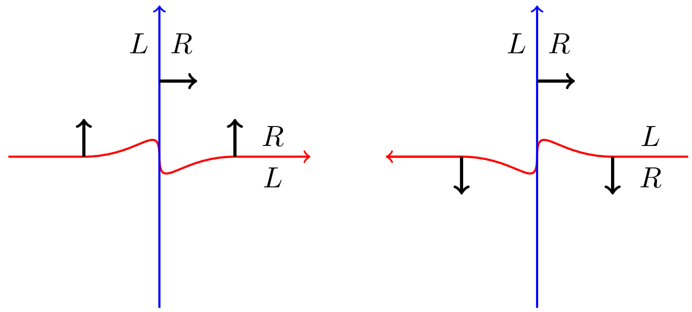

Figure 2.2: Transverse orientation in a neighborhood of a crossing. A left-to-right crossing and τ^- on the left. A right-to-left crossing and τ^+ on the right.

2.3 Description Penner's construction by linear algebra

Denote by $i(a, b)$ the geometric intersection number of the simple closed curves a and b. For two collections of curves $A = \{a_j\}$ and $B = \{b_k\}$, the intersection matrix $i(A, B)$ is a matrix whose (j, k)-entry is $i(a_j, b_k)$.

Suppose we apply Penner's construction with the collection $C = \{(c_1, \phi_1), \ldots, (c_n, \phi_n)\}$. Let

$$\Omega = i(C, C)$$

and

$$Q_i = I + D_i\Omega \quad (1 \leq i \leq n), \tag{2.1}$$

where I denotes the $n \times n$ identity matrix, and D_i denotes the $n \times n$ matrix whose ith entry on the diagonal is 1 and whose other entries are zero. When a product of the twists T_{c_i,ϕ_i} satisfies the hypotheses of Penner's construction, the corresponding product of the Q_i is *Perron–Frobenius*, that is, it has nonnegative entries and some power of it has strictly positive entries [Fat92, Proposition 5.3]. Moreover, the stretch factor of the resulting pseudo-Anosov mapping class equals the Perron–Frobenius eigenvalue (the unique largest real eigenvalue) of the corresponding Perron–Frobenius matrix [Fat92, Théorème 5.4]. Our matrices Q_i are the transposes of $M(T_j)$ and $M(S_j^{-1})$ in Section 5 of [Fat92].

3 Sequences of pseudo-Anosov mapping classes

In this section we study the asymptotic behavior of certain sequences of pseudo-Anosov mapping classes arising from Penner's construction. We show that the defining polynomials of the stretch factors converge in the sense of (1.1). Before we state the theorem more precisely, we need to introduce some notation.

Let Ω be the intersection matrix of a filling collection of curves C. Let Z_i be the orthogonal complement of the ith row of Ω. Since each curve in C intersects some other curve in C, all rows of Ω are nonzero, and the Z_i are hyperplanes. Let

$$p_{i \leftarrow j} : \mathbb{R}^n \to Z_i \tag{3.1}$$

be the (not necessarily orthogonal) projection onto the hyperplane Z_i in the direction of $\mathbf{e}_j$, the jth standard basis vector in $\mathbb{R}^n$. This projection is defined if and only if $\mathbf{e}_j$ is not contained in Z_i, which is in turn equivalent to the statement that the (i,j)-entry of Ω is positive.

Let $\mathbf{G}(\Omega)$ be the graph on the vertex set $\{1,\dots,n\}$ where i and j are connected if the (i,j)-entry of Ω is positive. For a closed path

$$\gamma = (i_1 \cdots i_K i_1)$$

in $\mathbf{G}(\Omega)$, define the linear map $f_\gamma : Z_{i_1} \to Z_{i_1}$ by the formula

$$f_\gamma = (p_{i_1 \leftarrow i_K} \circ \cdots \circ p_{i_2 \leftarrow i_1})|_{Z_{i_1}}. \tag{3.2}$$

In words, f_γ is a composition of projections: first from Z_{i_1} to Z_{i_2}, then from Z_{i_2} to Z_{i_3}, and so on, and finally from Z_{i_K} back to Z_{i_1}.

Theorem 3.1. *Let Ω be the intersection matrix of a collection of curves satisfying the hypotheses of Penner's construction. Let $\gamma = (i_1 \dots i_K i_1)$ be a closed path in $\mathbf{G}(\Omega)$ visiting each vertex at least once and let*

$$M_{\gamma,k} = Q_{i_K}^k \cdots Q_{i_1}^k.$$

Let λ_k be the Perron–Frobenius eigenvalue of $M_{\gamma,k}$ and denote by $u_k(x)$ and $v(x)$ the characteristic polynomials $\chi(M_{\gamma,k})$ and $\chi(f_\gamma)$, respectively. Then we have

$$\lim_{k\to\infty} \frac{u_k(x)}{x-\lambda_k} = v(x).$$

The rest of the section is devoted to the proof of this statement.

3.1 Estimating the left Perron–Frobenius eigenvectors

In this section we study the asymptotic behavior of the left Perron–Frobenius eigenvectors of the matrices $M_{\gamma,k}$ in Theorem 3.1. First we introduce a few conventions.

We follow the convention that vectors denoted by bold lowercase letters are column vectors. Row vectors are written as transposes of column vectors. Note that the ith row of a matrix M can be written as $\mathbf{e}_i^T M$. We will write ω_{ij} for the (i,j)-entry of Ω.

We recall that the Perron–Frobenius eigenvectors of a Perron–Frobenius matrix are the eigenvectors corresponding to the Perron–Frobenius eigenvalue whose coordinates are positive. The Perron–Frobenius eigenvectors form a ray emanating from the origin. When we say that the Perron–Frobenius eigenvector can be chosen to have some property, we mean that there is an eigenvector on this ray satisfying that property.

The result of this section is the following.

Proposition 3.2. *Let Ω be the intersection matrix of a collection of curves satisfying the hypotheses of Penner's construction. Let $\gamma = (i_1 \dots i_K i_1)$ be a closed path in $\mathbf{G}(\Omega)$ visiting each vertex at least once and let*

$$M_{\gamma,k} = Q_{i_K}^k \cdots Q_{i_1}^k.$$

Then the left Perron–Frobenius eigenvector $\mathbf{w}_k^T$ of $M_{\gamma,k}$ can be chosen for all k so that

$$\lim_{k\to\infty}\left[\mathbf{w}_k^T - \left(k\mathbf{e}_{i_1}^T\Omega + \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega\right)\right] = 0. \tag{3.3}$$

Proof. Note that we can assume that γ visits every vertex of $\mathbf{G}(\Omega)$ at least twice. This is because the square of $M_{\gamma,k}$ is associated to a path visiting every vertex at least twice, has the same Perron–Frobenius eigenvectors as $M_{\gamma,k}$, and the quantities in (3.3) are the same for $M_{\gamma,k}$ and its square.

Let $||\Omega||_\infty$ be the maximum of the entries of Ω and let $||\Omega||_{\min,\gamma}$ be the minimum of those entries of Ω that are at a position (i,j) such that ij is an edge in γ. We will show that

$$||\mathbf{w}_k^T - k\mathbf{e}_{i_1}^T\Omega - \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega||_\infty \le \frac{2^K||\Omega||_\infty^{K-1}}{k||\Omega||_{\min,\gamma}^K}, \tag{3.4}$$

which clearly implies the statement of the proposition. We break the proof of this inequality to two steps.

Step 1. *The left Perron–Frobenius eigenvectors of $M_{\gamma,k}$ are contained in the cone generated by the rows of $\Omega M_{\gamma,k}$.*

Proof. First we recall a geometric proof of the fact that every Perron–Frobenius matrix has a (left) eigenvector in the positive cone $\mathbb{R}^n_{\ge 0}$. The key idea is that the right action of the matrix fixes the positive cone and it maps rays to rays. So we have a continuous action on the space of rays in the positive cone. This space is homeomorphic to a disk, so by the Brouwer fixed point theorem there is a fixed ray. This fixed ray corresponds to an eigenvector [BT92].

We are going to run the same argument not for the positive cone but the smaller cone generated by the rows of $\Omega M_{\gamma,k}$. For this, we need to prove that this cone is invariant under the right action of $M_{\gamma,k}$. With that in hand, we obtain that $M_{\gamma,k}$ has an eigenvector in this smaller cone. It is well-known that the only eigenvectors in the positive cone are the Perron–Frobenius eigenvectors, so this implies the statement to be proven.

Now we turn to showing that the cone C_k generated by the rows of $\Omega M_{\gamma,k}$ is invariant under the right action of $M_{\gamma,k}$. Let

$$C = \{\mathbf{v}^T\Omega : \mathbf{v} \ge 0\}$$

be the cone generated by the rows of Ω. Note that $C_k = CM_{\gamma,k}$. It suffices to show that C is invariant under the right action of $M_{\gamma,k}$ for all k, that is, that $CM_{\gamma,k} \subset C$. This is because multiplying both sides by $M_{\gamma,k}$ gives $CM_{\gamma,k}^2 \subset CM_{\gamma,k}$, and substituting $C_k = CM_{\gamma,k}$ yields $C_k M_{\gamma,k} \subset C_k$.

To show that C is invariant under the right action of $M_{\gamma,k}$ for all k, it suffices to show that C is invariant under the right action of the generators Q_i. For this, observe that

$$\Omega Q_i = \Omega(I + D_i\Omega) = (I + \Omega D_i)\Omega = Q_i^T\Omega$$

for $1 \le i \le n$. Finally note that for any $\mathbf{v} \ge 0$ and $\mathbf{v}^T\Omega \in C$, we have $(\mathbf{v}^T\Omega)Q_i = (\mathbf{v}^T Q_i^T)\Omega \in C$, since $\mathbf{v} \ge 0$ implies $\mathbf{v}^T Q_i^T \ge 0$. $\square$

Step 2. *For all k, every row of $\Omega M_{\gamma,k}$ can be normalized to a vector $\mathbf{u}^T$ (that depends on k and the row) that satisfies*

$$||\mathbf{u}^T - k\mathbf{e}_{i_1}^T\Omega - \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega||_\infty \le \frac{2^K||\Omega||_\infty^{K-1}}{k||\Omega||_{\min,\gamma}^K}.$$

Proof. By expanding $M_{\gamma,k}$ and the Q_i out using their definitions, we obtain

$$\begin{aligned}\Omega M_{\gamma,k} = \Omega Q_{i_K}^k \cdots Q_{i_1}^k &= \Omega(I + kD_{i_K}\Omega)\cdots(I + kD_{i_1}\Omega) = \\ &= \Omega + \sum_{1\le t_1<\ldots<t_\ell\le K} k^\ell \Omega D_{i_{t_\ell}}\Omega\cdots\Omega D_{i_{t_1}}\Omega = \\ &= \Omega + \sum_{i=1}^n \sum_{\substack{1\le t_1<\ldots<t_\ell\le K\\ i_{t_\ell}=i}} k^\ell \Omega D_{i_{t_\ell}}\Omega\cdots\Omega D_{i_{t_1}}\Omega.\end{aligned}$$

So the i'th row of $\Omega M_{\gamma,k}$ is

$$\mathbf{e}_{i'}^T\Omega M_{\gamma,k} = \mathbf{e}_{i'}^T\Omega + \sum_{i=1}^n H(i,i'), \tag{3.5}$$

where

$$H(i,i') = \sum_{\substack{1\le t_1<\ldots<t_\ell\le K\\ i_{t_\ell}=i}} k^\ell \mathbf{e}_{i'}^T\Omega D_{i_{t_\ell}}\Omega\cdots\Omega D_{i_{t_1}}\Omega. \tag{3.6}$$

Right multiplication by D_j zeroes out all columns except the jth column, therefore the following identity holds:

$$\mathbf{e}_i^T\Omega D_j = \omega_{ij}\mathbf{e}_j^T.$$

Repetitively applying this identity for the terms in (3.6), we obtain

$$H(i,i') = \sum_{\substack{1\le t_1<\ldots<t_\ell\le K\\ i_{t_\ell}=i}} k^\ell \omega_{i' i_{t_\ell}}\omega_{i_{t_\ell} i_{t_{\ell-1}}}\cdots\omega_{i_{t_2} i_{t_1}}\mathbf{e}_{i_{t_1}}^T\Omega. \tag{3.7}$$

A summand on the right hand side is nonzero if and only if the path $(i_{t_1}\ldots i_{t_\ell} i')$ is contained $\mathbf{G}(\Omega)$. In particular, if $ii' \notin \mathbf{G}(\Omega)$, then all summands vanish and $H(i,i') = 0$.

Now suppose that $ii' \in \mathbf{G}(\Omega)$. Let t be the largest element of $\{1,\ldots,K\}$ such that $i_t = i$. Since γ visits i at least twice, such t exists and $t \ge 3$. Since $(i_1\ldots i_t i')$ and $(i_2\ldots i_t i')$ are paths in $\mathbf{G}(\Omega)$, the right hand side of (3.7) contains the nonzero summands

$$k^t\omega_{i' i_t}\cdots\omega_{i_2 i_1}\mathbf{e}_{i_1}^T\Omega,$$

$$k^{t-1}\omega_{i' i_t}\cdots\omega_{i_3 i_2}\mathbf{e}_{i_2}^T\Omega.$$

Moreover, these two paths are the unique longest and second longest paths of the form $(i_{t_1}\ldots i_{t_\ell} i')$ in $\mathbf{G}(\Omega)$ satisfying $i_{t_\ell} = i$. So we have

$$H(i,i') = c(i,i')\left(\mathbf{e}_{i_1}^T\Omega + k^{-1}\omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega + R(i,i')\right), \tag{3.8}$$

where $c(i,i') = k^t \omega_{i' i_t} \cdots \omega_{i_2 i_1}$ and $R(i,i')$ is a sum of expressions of the form

$$\frac{\prod k\omega_{ij}}{\prod k\omega_{i'j'}} \mathbf{e}_{i''}^T \Omega \tag{3.9}$$

such that the number of multiplicands in the denominator is at least two more than the number of multiplicands in the numerator. The trivial estimate yields that the supremum norm of each expression of the form (3.9) is bounded from above by

$$\frac{||\Omega||_\infty^{K-2}}{k^2 ||\Omega||_{\min,\gamma}^K} ||\Omega||_\infty .$$

A trivial upper bound on the number of terms in the sum $R(i,i')$ is $2^K - 1$, the number of nonempty subsets of $\{i_1, \ldots, i_K\}$, hence

$$||R(i,i')||_\infty \leq (2^K - 1) \frac{||\Omega||_\infty^{K-1}}{k^2 ||\Omega||_{\min,\gamma}^K}. \tag{3.10}$$

We remark that (3.8) and (3.10) hold also when ii' is a non-edge in $\mathbf{G}(\Omega)$ provided $c(i,i')$ and $R(i,i')$ are defined to be zero.

By substituting (3.8) into (3.5) and rescaling each side we obtain the equation

$$k \frac{\mathbf{e}_{i'}^T \Omega M_{\gamma,k}}{\sum_{i=1}^n c(i,i')} = k\mathbf{e}_{i_1}^T \Omega + \omega_{i_2 i_1}^{-1} \mathbf{e}_{i_2}^T \Omega + k \left(\frac{\sum_{i=1}^n c(i,i') R(i,i')}{\sum_{i=1}^n c(i,i')} + \frac{\mathbf{e}_{i'}^T \Omega}{\sum_{i=1}^n c(i,i')} \right).$$

Note that the expression is the i'th row of $\Omega M_{\gamma,k}$, renormalized. The first term inside the parentheses is a convex combination of the $R(i,i')$, so the same bound holds for it as in (3.10). To obtain a bound for the second term, note that $c(i,i') \geq k^2 ||\Omega||_{\min,\gamma}^2$ whenever $ii' \in \mathbf{G}(\Omega)$, because the number of multiplicands ω_{ij} in the definition of $c(i,i')$ is always at least two. Hence

$$\left\| \frac{\mathbf{e}_{i'}^T \Omega}{\sum_{i=1}^n c(i,i')} \right\|_\infty \leq \frac{||\Omega||_\infty}{k^2 ||\Omega||_{\min,\gamma}^2} \leq \frac{||\Omega||_\infty^{K-1}}{k^2 ||\Omega||_{\min,\gamma}^K}.$$

By combining the estimates for the two terms inside the parentheses, we obtain the desired inequality. □

The inequality (3.4) immediately follows from Step 1 and 2. This completes the proof of Proposition 3.2. □

The intuition behind Proposition 3.2 is the following. The cone C_k generated by the rows of $\Omega M_{\gamma,k}$ is the image of the cone C generated by the rows of Ω under the right action of $Q_{i_K}^k$, …, $Q_{i_1}^k$ in this order. Where C ends up after these actions is determined for the most part by the last action. The right action of $Q_{i_1}^k$ is trivial on all standard basis vectors except one which is translated by $k\mathbf{e}_{i_1}^T \Omega$. Thus we can think of the right action of $Q_{i_1}^k$ as a map sending the cone C towards the direction of $\mathbf{e}_{i_1}^T \Omega$. This is why $\mathbf{e}_{i_1}^T \Omega$ appears in (3.3) and why it appears with a large weight k.

The second to last action is the action of $Q_{i_2}^k$, sending the C towards the direction of $\mathbf{e}_{i_2}^T \Omega$. This action has a smaller effect than the last action, but a more significant one than all the actions before. So $\mathbf{e}_{i_2}^T \Omega$ still appears in (3.3), but with a smaller weight than $\mathbf{e}_{i_1}^T \Omega_k$. The rest of the actions turn out to be negligible as $k \to \infty$.

3.2 Projections to hyperplanes

The goal of this section is to describe the projections $p_{i\leftarrow j}$ defined in (3.1) by matrices.

Introduce the definition

$$Q_{i\leftarrow j} = I - \omega_{ij}^{-1} T_{ji} \Omega \tag{3.11}$$

where T_{ji} is the $n \times n$ matrix whose (j,i)-entry is 1 and whose other entries are zero. The matrix $Q_{i\leftarrow j}$ is defined if $\omega_{ij} > 0$, in other words, if ij is an edge of $\mathbf{G}(\Omega)$.

Note that multiplication by T_{ji} on the left has the effect of zeroing out all rows except the ith row and moving the ith row to the jth row. So in words, $Q_{i\leftarrow j}$ is calculated in the following way. Zero out all rows of Ω except the ith row, move the ith row to the jth row, and then normalize it so that the entry on the diagonal is 1. Subtracting this matrix from the identity matrix gives $Q_{i\leftarrow j}$. Note that the jth column of $Q_{i\leftarrow j}$ is zero.

Lemma 3.3. *If ij is an edge of $\mathbf{G}(\Omega)$, then the left action of $Q_{i\leftarrow j}$ is the projection $p_{i\leftarrow j}$.*

Proof. Recall that $p_{i\leftarrow j}$ is the projection to the hyperplane Z_i in the direction of $\mathbf{e}_j$ where Z_i is the orthogonal complement of the ith row of Ω. To show that the left action of $Q_{i\leftarrow j}$ is the same transformation, it suffices to show that $Q_{i\leftarrow j}\mathbf{e}_j = 0$ and $Q_{i\leftarrow j}\mathbf{v} = \mathbf{v}$ whenever $\mathbf{e}_i^T \Omega \mathbf{v} = 0$.

The first statement is clear, since the jth column of $Q_{i\leftarrow j}$ is zero.

For the second statement, note that the equation $Q_{i\leftarrow j}\mathbf{v} = \mathbf{v}$ is equivalent to $T_{ji}\Omega\mathbf{v} = 0$, which is in turn equivalent to saying that the ith coordinate of the vector $\Omega\mathbf{v}$ is zero. But this is equivalent to $\mathbf{e}_i^T\Omega\mathbf{v} = 0$, so we are done. ☐

3.3 Convergence of linear maps

In this section, we show that the matrices $M_{\gamma,k}$ in Theorem 3.1 act on certain codimension 1 subspaces by left multiplication and these left actions asymptotically stabilize as k goes to infinity. We will use this fact to prove Theorem 3.1 at the end of the section.

Proposition 3.4. *Let Ω be the intersection matrix of a collection of curves satisfying the hypotheses of Penner's construction. Let $\gamma = (i_1 \dots i_K i_1)$ be a closed path in $\mathbf{G}(\Omega)$ visiting each vertex at least once and let*

$$M_{\gamma,k} = Q_{i_K}^k \cdots Q_{i_1}^k.$$

Let $\mathbf{w}_k^T$ be a left Perron–Frobenius eigenvector of $M_{\gamma,k}$. If a sequence of vectors $\mathbf{v}_k$ converges to some vector $\mathbf{v}^*$ and $\mathbf{w}_k^T\mathbf{v}_k = 0$ for all k, then

$$\lim_{k\to\infty} M_{\gamma,k}\mathbf{v}_k = Q_{i_1\leftarrow i_K} \cdots Q_{i_3\leftarrow i_2} Q_{i_2\leftarrow i_1}\mathbf{v}^*.$$

Proof. We show the following convergences by induction:

$$\lim_{k\to\infty} Q_{i_1}^k \mathbf{v}_k = Q_{i_2\leftarrow i_1}\mathbf{v}^* \tag{3.12}$$

$$\lim_{k\to\infty} Q_{i_2}^k Q_{i_1}^k \mathbf{v}_k = Q_{i_3\leftarrow i_2} Q_{i_2\leftarrow i_1}\mathbf{v}^* \tag{3.13}$$

$$\vdots$$

$$\lim_{k\to\infty} Q_{i_K}^k \cdots Q_{i_1}^k \mathbf{v}_k = Q_{i_1\leftarrow i_K} \cdots Q_{i_2\leftarrow i_1}\mathbf{v}^* \tag{3.14}$$

First we prove the base case (3.12). Since

$$\lim_{k\to\infty} Q_{i_2\leftarrow i_1}\mathbf{v}_k = Q_{i_2\leftarrow i_1}\mathbf{v}^*,$$

it suffices to show that

$$\lim_{k\to\infty} \left[Q_{i_1}^k \mathbf{v}_k - Q_{i_2\leftarrow i_1}\mathbf{v}_k\right] = 0.$$

Suppose that the Perron–Frobenius eigenvectors $\mathbf{w}_k^T$ are chosen as guaranteed by Proposition 3.2. Then we have

$$\lim_{k\to\infty} \left|\left(k\mathbf{e}_{i_1}^T\Omega + \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega\right)\mathbf{v}_k\right| = \lim_{k\to\infty} \left|\left(\mathbf{w}_k^T - \left(k\mathbf{e}_{i_1}^T\Omega + \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega\right)\right)\mathbf{v}_k\right| = 0$$

by (3.3). Therefore

$$\lim_{k\to\infty} \left(Q_{i_1}^k - Q_{i_2\leftarrow i_1}\right)\mathbf{v}_k = \lim_{k\to\infty} \left(kD_{i_1}\Omega + \omega_{i_2 i_1}^{-1} T_{i_1 i_2}\Omega\right)\mathbf{v}_k = 0,$$

since $kD_{i_1}\Omega + \omega_{i_2 i_1}^{-1} T_{i_1 i_2}\Omega$ is a matrix whose i_1st row equals $k\mathbf{e}_{i_1}^T\Omega + \omega_{i_2 i_1}^{-1}\mathbf{e}_{i_2}^T\Omega$ and whose other rows are zero. This completes the proof of the base case (3.12).

Next, we describe how to obtain (3.13) from (3.12). The remaining inductive steps are analogous.

Let $M_{\gamma',k}$ be the cyclic permutation $Q_{i_1}^k Q_{i_K}^k \cdots Q_{i_2}^k$ of the product $M_{\gamma,k}$. If $\mathbf{w}_k^T$ is a left Perron–Frobenius eigenvector of $M_{\gamma,k}$, then

$$\mathbf{u}_k^T = \mathbf{w}_k^T Q_{i_K}^k \cdots Q_{i_2}^k$$

is a left Perron–Frobenius eigenvector of $M_{\gamma',k}$. On the other hand we have

$$\mathbf{u}_k^T Q_{i_1}^k \mathbf{v}_k = \mathbf{w}_k^T M_{\gamma,k}\mathbf{v}_k = \lambda_k \mathbf{w}_k^T \mathbf{v}_k = 0.$$

In words, $Q_{i_1}^k \mathbf{v}_k$ is orthogonal to the left Perron–Frobenius eigenspace of $M_{\gamma',k}$. Therefore we can apply (3.12) for $Q_{i_2}^k$ instead of $Q_{i_1}^k$ and for $Q_{i_1}^k \mathbf{v}_k$ instead of $\mathbf{v}_k$ to obtain (3.13). □

Proof of Theorem 3.1. Let $u_k(x) = (x - \lambda_k)v_k(x)$ and denote by W_k the orthogonal complement of the left Perron–Frobenius eigenspace of $M_{\gamma,k}$. Note that W_k is invariant under the left action of $M_{\gamma,k}$ and the induced linear transformation has characteristic polynomial $v_k(x)$.

Let

$$\mathcal{B}^* = \{\mathbf{b}_1^*, \ldots, \mathbf{b}_{n-1}^*\}$$

be a basis for Z_{i_1}, the orthogonal complement of the i_1st row of Ω. For each k, choose a basis

$$\mathcal{B}_k = \{\mathbf{b}_1^k, \ldots, \mathbf{b}_{n-1}^k\}$$

for W_k such that $\mathbf{b}_i^k \to \mathbf{b}_i^*$ for $i = 1, \ldots, n-1$. This is possible, since the subspaces W_k converge to Z_{i_1} by Proposition 3.2.

By Lemma 3.3, the left action of $Q_{i_1 \leftarrow i_K} \cdots Q_{i_3 \leftarrow i_2} Q_{i_2 \leftarrow i_1}$, restricted to the hyperplane Z_{i_1} equals f_γ. Let A^* be the matrix describing this left action in the basis $\mathcal{B}^*$. Let A_k be the matrix describing the left action of $M_{\gamma,k}$ on W_k in the basis $\mathcal{B}_k$.

By Proposition 3.4, we have

$$\lim_{k\to\infty} M_{\gamma,k}\mathbf{b}_i^k = Q_{i_1 \leftarrow i_K} \cdots Q_{i_3 \leftarrow i_2} Q_{i_2 \leftarrow i_1} \mathbf{b}_i^*$$

for $i = 1, \ldots, n-1$, therefore $A_k \to A$ and $\chi(A_k) \to \chi(A)$. Since $\chi(A_k) = u_k(x)$ and $\chi(A) = v(x)$, this completes the proof. □

4 Homotopy invariance

The goal of this section is to show that the eigenvalues of the linear transformation f_γ defined in (3.2) are invariant under homotopy of γ. As a result, we will be able to determine the eigenvalues of f_γ for any contractible γ. It turns out that the only eigenvalues in these cases are 0 and 1.

We say that the closed paths γ and γ' are homotopic in $\mathbf{G}(\Omega)$ if the naturally associated maps from S^1 to $\mathbf{G}(\Omega)$ are homotopic. It is easy to see that two closed paths are homotopic if and only if they are connected by a sequence of the insertions and removals of backtrackings (paths of the form (iji)) and cyclic permutations of the vertices. Since f_γ is not defined for a closed path of length 0, we require that all closed paths appearing in such a sequence have length at least two.

Proposition 4.1. *If γ and γ' are homotopic closed paths in $\mathbf{G}(\Omega)$, then the characteristic polynomials of f_γ and $f_{\gamma'}$ are equal.*

Proof. First, we will show that removing or inserting a backtracking to γ does not change f_γ as long as the last edge of γ is not changed. It is easy to see that it is necessary to require at least that the starting vertex of γ is fixed, since the domain of f_γ depends on the first vertex of γ.

Since inserting a backtracking is the inverse operation of removing one, it suffices to show the statement for removals. For this, let $\gamma = (i_1 \ldots i_K i_1)$, let $2 \le k \le K-1$ and suppose that $i_{k-1} = i_{k+1} = i$. We will to show that the removal of the backtracking $(i_{k-1} i_k i_{k+1})$ leaves f_γ unchanged.

If $k \ge 3$, then the composition

$$p_{i_{k+2} \leftarrow i} \circ p_{i \leftarrow i_k} \circ p_{i_k \leftarrow i} \circ p_{i \leftarrow i_{k-2}}$$

appears in the formula (3.2). We are using the convention that $i_{K+1} = i_1$. However, this is the same as $p_{i_{k+2} \leftarrow i} \circ p_{i \leftarrow i_{k-2}}$ for the following reasons. The image of $p_{i \leftarrow i_{k-2}}$ is the hyperplane Z_i, the orthogonal complement of the ith row of Ω. The subsequent projection, $p_{i_k \leftarrow i}$ is in the direction of $\mathbf{e}_i$. However, we have $\mathbf{e}_i \in Z_i$, because the diagonal entries of Ω are zero, so $\mathbf{e}_i^T \Omega \mathbf{e}_i = 0$. Hence the image of $p_{i_k \leftarrow i} \circ p_{i \leftarrow i_{k-2}}$ is still inside Z_i. As a consequence, the next projection, $p_{i \leftarrow i_k}$, which also projects onto Z_i, does not have any effect. This shows that

$$p_{i_{k+2} \leftarrow i} \circ p_{i \leftarrow i_k} \circ p_{i_k \leftarrow i} \circ p_{i \leftarrow i_{k-2}} = p_{i_{k+2} \leftarrow i} \circ p_{i_k \leftarrow i} \circ p_{i \leftarrow i_{k-2}}.$$

But now the subsequent projections $p_{i_{k+2} \leftarrow i}$ and $p_{i_k \leftarrow i}$ are both projections in the direction of $\mathbf{e}_i$, hence $p_{i_{k+2} \leftarrow i} \circ p_{i_k \leftarrow i} = p_{i_{k+2} \leftarrow i}$. So we indeed end up with shorter the composition $p_{i_{k+2} \leftarrow i} \circ p_{i \leftarrow i_{k-2}}$.

If $k = 2$, this argument needs to be slightly modified. We now have

$$f_\gamma = (p_{i_1 \leftarrow i_K} \circ \cdots \circ p_{i_4 \leftarrow i} \circ p_{i \leftarrow i_2} \circ p_{i_2 \leftarrow i})|_{Z_i}.$$

The image of Z_i under $p_{i_2 \leftarrow i}$ is contained in Z_i, since $\mathbf{e}_i \in Z_i$. Now the same arguments as above show that $p_{i \leftarrow i_2}$ and then $p_{i_2 \leftarrow i}$ can be omitted from the composition. This completes the proof of the fact that f_γ is invariant under homotopy of γ rel the last edge of γ.

Another fact we need is that cyclic permutation of the vertices of γ does not change the characteristic polynomial of f_γ. One can see this directly from the formula (3.2), because cyclic permutation of the vertices of γ changes f_γ by conjugation.

We can now give the proof of the proposition. Suppose that γ and γ' are homotopic and therefore connected by a sequence of insertions and removals of backtrackings and cyclic permutations of the vertices. Since our paths have length at least 2, we can always permute the vertices before an insertion of removal of a backtracking so that the last edge is unchanged. So in each step f_γ either does not change or it changes by conjugation. Either way, the characteristic polynomial does not change. □

Corollary 4.2. *If the matrix Ω has size $n \times n$ and the path γ is contractible, then the characteristic polynomial of f_γ takes the form*

$$\chi(f_\gamma) = x(x-1)^{n-2}.$$

Proof. By Proposition 4.1, the characteristic polynomial of f_γ is the same as that of $f_{\gamma'}$ where $\gamma' = (iji)$ is a path of length two. Writing out the definition of $f_{\gamma'}$, we have

$$f_{(iji)} = (p_{i \leftarrow j} \circ p_{j \leftarrow i})|_{Z_i}.$$

As we have observed in the proof of Proposition 4.1, we have $\mathbf{e}_i \in Z_i$, so the image of Z_i under $p_{j \leftarrow i}$ is contained in Z_i. Therefore $p_{i \leftarrow j}$ can be omitted from the formula and $f_{(iji)}$ is just the projection $p_{j \leftarrow i}|_{Z_i}$ projecting Z_i to a codimension 1 subspace of Z_i. Hence $\chi(f_\gamma) = x(x-1)^{n-2}$ as stated. □

5 A simple criterion for realizing degrees

In this section, we give a simple way to certify that a given degree can be realized on a given surface.

We will need the following fact relating the rank of the intersection matrix Ω to the 1-eigenspaces of products of the matrices Q_i.

Proposition 5.1. *Let ψ be a pseudo-Anosov map arising from Penner's construction using a collection of curves with an $n \times n$ intersection matrix Ω. If M is the product of the Q_i describing ψ, then 1 is an eigenvalue of M with multiplicity $n-r$ where $r = \operatorname{rank}(\Omega)$. In particular, the characteristic polynomial of M takes the form $(x-1)^{n-r}p(x)$ where $\deg(p) = r$ and $p(1) \neq 0$.*

Proof. First we show that the multiplicity is at least $n-r$. This is because the left action of every $Q_i = I + D_i\Omega$ is the identity on the null space $\operatorname{Nul}(\Omega)$ of Ω, so the left action of M on $\operatorname{Nul}(\Omega)$ is also the identity. Moreover, the dimension of $\operatorname{Nul}(\Omega)$ is $n-r$.

Now we turn to showing that the multiplicity is at most $n-r$. Since M acts on $\operatorname{Nul}(\Omega)$ as the identity, it has a well-defined left action

$$\ell_M : \widehat{V} \to \widehat{V}$$

on the quotient space $\widehat{V} = \mathbb{R}^n / \operatorname{Nul}(\Omega)$. It suffices to show that 1 is not an eigenvalue of ℓ_M.

In other to show this, we will consider the quadratic form h on $\mathbb{R}^n$ defined by the formula

$$h(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T \Omega \mathbf{v}.$$

The function h can be thought of as a height function on $\mathbb{R}^n$. It was shown in Proposition 2.1 of [SS15] that

$$h(Q_i\mathbf{v}) - h(\mathbf{v}) = ||Q_i\mathbf{v} - \mathbf{v}||^2 \tag{5.1}$$

for all $i = 1, \ldots, n$ and $\mathbf{v} \in \mathbb{R}^n$. In words, the matrices Q_i act on $\mathbb{R}^n$ by not decreasing the height. Moreover, there is no increase in the height if and only if $\mathbf{v}$ is fixed by Q_i.

Now suppose that $\ell_M(\widehat{\mathbf{v}}) = \widehat{\mathbf{v}}$ for some $\widehat{\mathbf{v}} \in \widehat{V}$; we want to show that $\widehat{\mathbf{v}} = 0$. Then there is some $\mathbf{v} \in \mathbb{R}^n$ such that $M\mathbf{v} - \mathbf{v} \in \operatorname{Nul}(\Omega)$, so $\Omega M\mathbf{v} = \Omega\mathbf{v}$. Hence

$$h(M\mathbf{v}) = \frac{1}{2}\mathbf{v}^T M^T \Omega M \mathbf{v} = \frac{1}{2}\mathbf{v}^T \Omega \mathbf{v} = h(\mathbf{v}).$$

By (5.1), this is only possible if $Q_i\mathbf{v} = \mathbf{v}$ for all $i = 1, \ldots, n$, since each Q_i appears in M. It follows that $D_i\Omega\mathbf{v} = 0$ for all $i = 1, \ldots, n$. Since $\sum_{i=1}^n D_i = I$, we have $\Omega\mathbf{v} = 0$, hence $\widehat{\mathbf{v}}$ is the zero vector of $\widehat{V}$. Therefore 1 is indeed not an eigenvalue of ℓ_M. $\square$

Corollary 5.2. *Let λ be a pseudo-Anosov stretch factor arising from Penner's construction using a collection of curves with an $n \times n$ intersection matrix Ω. Then $\deg(\lambda) \leq \operatorname{rank}(\Omega)$.*

Proof. The number λ is an eigenvalue of a matrix M and hence a root of the polynomial $p(x)$ in Proposition 5.1. $\square$

The following theorem gives a recipe for constructing a stretch factor with a specified algebraic degree.

Theorem 5.3. *Let Ω be the intersection matrix of a collection of curves C satisfying the hypotheses of Penner's construction such that* $\operatorname{rank}(\Omega) = r$. *Let $\gamma = (i_1 \dots i_K i_1)$ be a closed path in $\mathbf{G}(\Omega)$ visiting each vertex at least once and let*

$$M_{\gamma,k} = Q_{i_K}^k \cdots Q_{i_1}^k.$$

Let f_k be the pseudo-Anosov mapping class described by the matrix $M_{\gamma,k}$ and let λ_k be its stretch factor. Then $\deg(\lambda_k) = r$ for all but finitely many k.

Proof. Let $u_k(x)$ be the characteristic polynomial of $M_{\gamma,k}$. Proposition 5.1 shows that $u_k(x) = (x-1)^{n-r} p_k(x)$ where $p_k(1) \neq 0$ and the degree of $p_k(x)$ is r. Since λ_k is a root of $p_k(x)$, it suffices to show that $p_k(x)$ irreducible if k is large enough.

By Theorem 3.1 and Corollary 4.2, we have

$$\lim_{k\to\infty} \frac{u_k(x)}{x - \lambda_k} = x(x-1)^{n-2}.$$

So all roots of $p_k(x)$ except for λ_k converge to either 0 or 1. Note that $p_k(0) \neq 0$, since the matrices Q_i are invertible, hence so is $M_{\gamma,k}$. Therefore all roots are different from their limits for all k and Lemma 1.2 implies that $p_k(x)$ is indeed irreducible if k is large enough. □

As an immediate corollary of Theorem 5.3 and Proposition 2.1, we have the following.

Theorem 5.4. *If the surface S admits a filling collection of curves C with inconsistent markings such that* $\operatorname{rank}(i(C,C)) = r$, *then $r \in D(S)$. In addition, if C is completely left-to-right or completely right-to-left, then $r \in D^+(S)$.*

6 Collections of curves

In this section, we construct filling collections of curves on various surfaces. By Theorem 5.4, the integers that arise are ranks of the intersection matrices also arise as algebraic degrees of stretch factors.

We consider both orientable and nonorientable surfaces. In the orientable case, the constructions are fairly straightforward. The nonorientable case is also not difficult, but we will need to do more case-by-case analysis for surfaces with small Euler characteristic.

6.1 Orientable surfaces

Recall the definition of completely left-to-right and completely right-to-left collections of curves from Section 2.2. Let $\overline{S}$ denote the closed surface obtained from S by filling in the punctures.

Proposition 6.1. *Let S be an orientable surface. For all $1 \leq r \leq \frac{1}{2}\dim(\mathrm{Teich}(S))$ there is a filling collection of curves C with inconsistent markings on S such that* $\operatorname{rank}(i(C,C)) = 2r$.

Moreover, for $1 \leq r \leq \frac{1}{2}\dim(H_1(\overline{S}))$, the collection C can be chosen to be completely left-to-right or completely right-to-left.

Proof. Since S is orientable, C is necessarily a union of two multicurves A and B, where the curves of A are marked consistently with the orientation of S, and the curves of B are marked inconsistently with the orientation of S. Note that $\operatorname{rank}(i(C,C)) = 2\operatorname{rank}(i(A,B))$.

Let $S_{g,n}$ be the orientable surface of genus g with n punctures. In the special case $(g,n) = (4,3)$, Figure 6.1 shows a pair of filling multicurves A and B on $S_{g,n}$ with $\frac{1}{2}\dim(\mathrm{Teich}(S)) = 3g-3+n$ simple closed curves in each multicurve. This construction generalizes for all $S_{g,n}$ (where $g \geq 2$) in the following way. The separating curves of B shown on Figure 6.1 divide $S_{4,3}$ to two once-punctured tori on the left and right, and two twice-punctured tori in the middle. To draw the analogous picture for $S_{g,3}$, change the number of twice-punctured tori in the middle from two to $g-2$. Then, to get the curve systems on $S_{g,n}$ for arbitrary n, change the number of punctures in the once-punctured torus on the right, and change the number of parallel curves of A and B around the punctures accordingly.

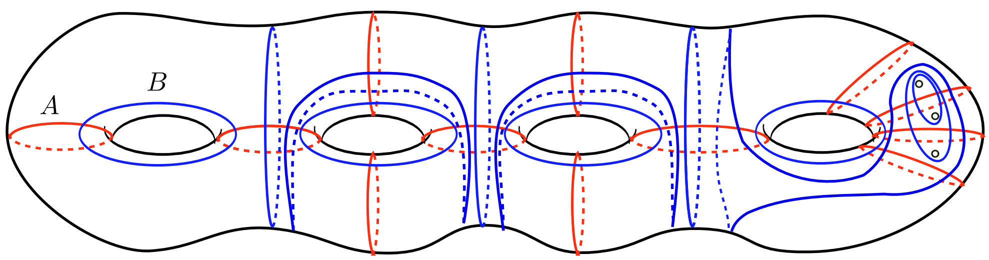

Figure 6.1: A maximal pair of filling multicurves on $S_{4,3}$.

By numbering the curves in each multicurve left-to-right and top-to-bottom, $i(A,B)$ takes the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \end{pmatrix}. \tag{6.1}$$

From the pattern, it is not hard to see that $i(A,B)$ has nonzero determinant for all $S_{g,n}$ where $g \geq 2$.

Note that A and the multicurve consisting of the g curves of B around the holes still fill the surface. Therefore A and any submulticurve of B that contains those g curves also fill. This gives examples for pairs of filling multicurves with intersection matrices of rank r for $g \leq r \leq 3g-3+n$.

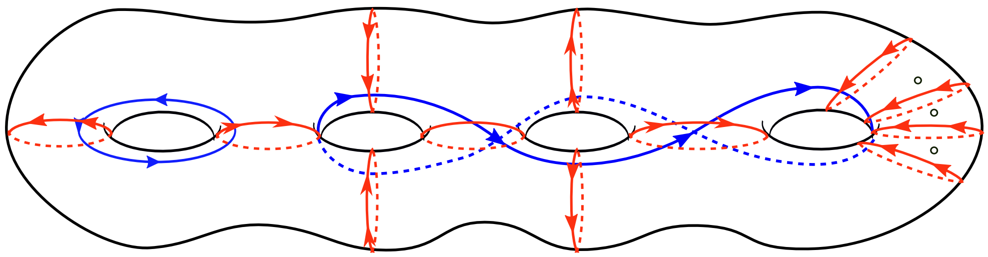

Figure 6.2: Pairs of multicurves realizing ranks $1 \leq r \leq g$.

To obtain examples for all ranks $1 \leq r \leq g-1$, link together the g curves around the holes one by one as on Figure 6.2, resulting in multicurves B' consisting of fewer and fewer curves. These multicurves still fill with A, and the columns of $i(A, B')$ are linearly independent, because for every curve in B' there is a curve in A that intersects only that curve. Note that when $1 \leq r \leq g$, the pairs of multicurves in the examples shown on Figure 6.2 are completely left-to-right if oriented as on the figure, because the red curves always cross the blue curves from left to right (cf. Figure 2.1). This completes the case $g \geq 2$.

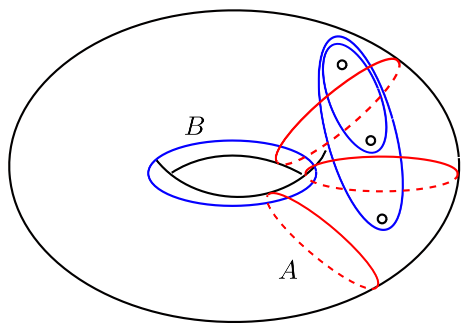

Figure 6.3: The case $g = 1$.

Figures 6.3 and 6.4 show examples with $\mathrm{rank}(i(A, B)) = 3g - 3 + n$ in the cases $g = 1, n \geq 1$ and $g = 0, n \geq 4$, respectively. In all these cases, there is a curve in B that intersects all curves of A and which alone fills the surface with A. Hence once again we can drop curves from B preserving the filling property and decreasing the rank. The rank 1 example for $g = 1$ is completely left-to-right (with the appropriate markings), so this proves the second part of the proposition when $g = 1$, while for $g = 0$ there is nothing to prove since $\dim(H_1(\overline{S_{0,n}})) = 0$ for all n.

In the remaining cases $g = 1$, $n = 0$ and $g = 0$, $n < 4$, the formula $\dim(\mathrm{Teich}(S_{g,n})) = 6g - 6 + 2n$ does not hold. The case $(g, n) = (1, 0)$ is straightforward to check as we have $\dim(\mathrm{Teich}(S_1)) = 2$. In the cases $g = 0$, $n < 4$, we have $\dim(\mathrm{Teich}(S_{g,n})) = 0$, so there is nothing to check. $\square$

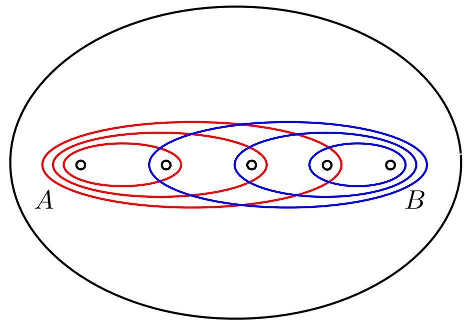

Figure 6.4: The case $g = 0$.

6.2 Nonorientable surfaces

Let $N_{g,n}$ be the nonorientable surface of genus g—the connected sum of g copies of the projective plane—with n punctures. We also use the term *crosscap* for the projective plane. Analogously to the orientable case, we abbreviate $N_{g,0}$ as N_g.

One way to obtain a nonorientable surface is to cut an open disk out of a surface and glue the resulting boundary component to the boundary of a Möbius strip. In other words, the resulting surface is the connected sum of the original surface and a crosscap. We refer to this process as *attaching a crosscap*. On figures, it is common to mark the location of the above surgery by a cross inside a disk. For example, Figure 6.8 on the left shows a surface obtained from the sphere by attaching five crosscaps, while Figure 6.9 shows a surface obtained from S_2 by attaching one crosscap.

A slightly different way of thinking about attaching a crosscap is by cutting an open disk out and identifying antipodal points of the resulting boundary component. That means that if a curve enters an attached crosscap, then it exits the crosscap at the antipodal point.

On many figures in this section, we represent a nonorientable surface as an orientable surface with crosscaps attached. We fix an orientation on the complement of the crosscaps. We color parts of a marked curve (cf. Section 2.1) on such a surface using two different colors depending on whether the embedding of the regular neighborhood of the curve is orientation-preserving or orientation-reversing on that part. Note that the color of the curve changes when it goes through a crosscap.

The result of this section is the following.

Proposition 6.2. *If $g \geq 3$ and $g + n \geq 5$ or $1 \leq g \leq 2$ and $g + n \geq 4$, then for all $3 \leq r \leq \dim(\mathrm{Teich}(N_{g,n})) = 3g + 2n - 6$ there is a filling inconsistently marked collection of curves C on $N_{g,n}$ such that $\mathrm{rank}(i(C, C)) = r$.*

If $(g, n) = (4, 0)$ or $(3, 1)$, then for all $3 \leq r \leq \dim(\mathrm{Teich}(N_{g,n})) - 1$ there is a filling inconsistently marked collection of curves C on $N_{g,n}$ such that $\mathrm{rank}(i(C, C)) = r$.

Moreover, when $3 \leq r \leq \dim(H_1(N_g, \mathbb{R})) = g - 1$, the collection C on $N_{g,n}$ can be chosen to be completely left-to-right or completely right-to-left.

See Table 6.1 for a summary of the surfaces with small Euler characteristic.

The proof is based on Proposition 6.1 and two lemmas that we discuss next.

n\g	1	2	3	4	5
0	∅	∅	∅	-1	E
1	∅	∅	-1	E	E
2	∅	E	E	E	E
3	E	E	E	E	E

Table 6.1: The surfaces for which every number between 3 and $\dim(\mathrm{Teich}(N_{g,n}))$ are realized as $\mathrm{rank}(i(C,C))$ are marked by E. The surfaces for which all these numbers except the maximum $\dim(\mathrm{Teich}(N_{g,n}))$ are realized are marked by -1. The surfaces that do not admit pseudo-Anosov maps are marked by $\emptyset$.

Suppose we have a filling collection of curves C on a surface. Provided C satisfies certain conditions, Lemma 6.3 says that attaching a crosscap to the surface allows extending C to a filling collection on the new surface in a way that $\mathrm{rank}(i(C,C))$ increases by $0, 2$ or 3. Lemma 6.4 says that adding a puncture allows increasing $\mathrm{rank}(i(C,C))$ by 0 or 2.

In the following statement, let $\mathcal{N}$ denote a small regular open neighborhood.

Lemma 6.3. *Let $C = \{c_i\}$ be a collection of inconsistently marked simple closed curves on a surface S which are in pairwise minimal position. Let R be a component of the complement of $\mathcal{N}(C)$. Note that ∂R is a union of arcs a_j, each of which lies on the boundary of some $\overline{\mathcal{N}(c_{i_j})}$.*

Let a_1 and a_2 be two arcs such that

- *c_{i_1} and c_{i_2} are non-isotopic and disjoint,*
- *there exists an arc b inside R connecting a_1 and a_2 such that the markings of c_{i_1} and c_{i_2} induce different orientations on $\mathcal{N}(c_{i_1}) \cup \mathcal{N}(b) \cup \mathcal{N}(c_{i_2}) \cong S_{0,3}$.*

Let S' be the surface obtained by attaching a crosscap to S inside R. Consider the curves d_1, d_2 and e on S' illustrated on Figure 6.5. The curves d_1 and d_2 are obtained from c_{i_1} and c_{i_2} by replacing the arcs a_1 and a_2 with arcs going around the crosscap in R. The curve e is obtained from c_{i_1} and c_{i_2} by replacing the arcs a_1 and a_2 by two arcs going through the crosscap in R. Note that e is simple, since c_1 and c_2 are disjoint. We endow d_1, d_2 and e with markings illustrated on Figure 6.5.

Finally, we introduce notations for the following three collections of curves on S':

$$C' = C \cup \{e\}$$
$$C'' = C \cup \{e, d_1\}$$
$$C''' = C \cup \{e, d_1, d_2\}.$$

The statement of the lemma is that the following hold for C', C'' and C'':

(i) *C', C'' and C''' are inconsistently marked;*

(ii) *C', C'' and C''' are in pairwise minimal position;*

(iii) $\mathrm{rank}(i(C,C)) = \mathrm{rank}(i(C',C')) = \mathrm{rank}(i(C'',C'')) - 2 = \mathrm{rank}(i(C''',C''')) - 3$*;*

(iv) *if C fills S, then C', C'' and C''' fill S'.*

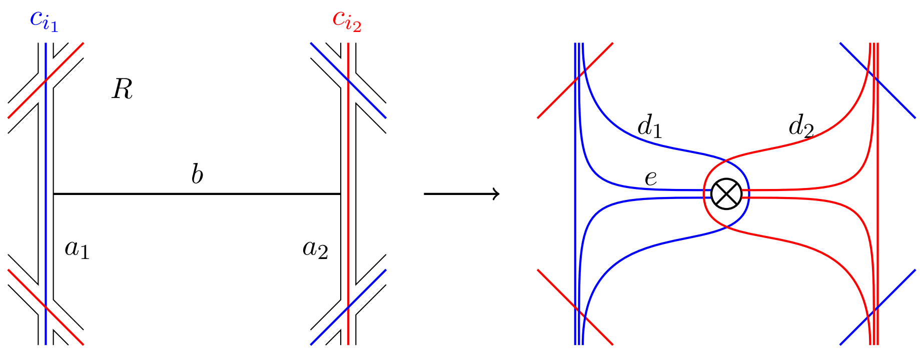

Figure 6.5: Creating three new curves by attaching a crosscap.

Proof. The statements (i) and (iv) are clear.

Next, we describe how the statement (ii) can be verified using the bigon criterion. The bigon criterion says that two curves are in minimal position if an only if they do not form a bigon: an embedded disk whose boundary is a union of an arc of one curve and an arc of the other and whose interior is disjoint from the two curves [FM12, Proposition 1.7].

First we show that d_1 does not form a bigon with the curves of C. (By symmetry, the same will be true for d_2.) Assume for contradiction that d_1 does form a bigon with some curve f of C. Note that in S', the curves d_1 and c_{i_1} bound an annulus $\mathcal{A}$ with an attached crosscap. So there are two possibilities: either the bigon is on the side of d_1 contained in $\mathcal{A}$ or the other side. In the first scenario the bigon has to be entirely contained in $\mathcal{A}$, otherwise there would be a bigon between f and c_{i_1}. But this cannot happen, since every curve of C entering $\mathcal{A}$ through d_1 exits it through c_{i_1}. Using this last observation, one can see that the second scenario is also not plausible, since we could obtain a bigon between f and c_{i_1} in the *original* surface S by extending the bigon between f and d_1 into the annulus $\mathcal{A}$ in S.

Next, we show that there is no bigon between e and the curves of C. For this, note that c_{i_1}, c_{i_2} and e bound a pair of pants $\mathcal{P}$. If there was a bigon between e and some curve f of C, then there are again two possibilities: the bigon is on the side of e contained in $\mathcal{P}$ or the other side. The first scenario is impossible for reasons similar to above: every curve of C entering $\mathcal{P}$ through e exits it either through c_{i_1} or through c_{i_2}. In the second scenario, consider the arc AB of f that forms a bigon with e. Following f from this arc in both directions into $\mathcal{P}$ until f exits $\mathcal{P}$ yields a longer arc $A'B'$ of f (Figure 6.6). One endpoint of this longer arc is on c_{i_1}, the other is on c_{i_2}, otherwise c_{i_1} or c_{i_2} forms a bigon with f. Note that the arc AB of f is part of the original surface S so it does not intersect the core curve of the attached crosscap. However, from the fact that A' and B' cannot both lie on c_{i_1} or c_{i_2}, it is easy to see that the arc AB of e forming the other side of the bigon goes through the crosscap and hence intersects its core curve once. So the arc AB of e and the arc AB of f cannot be homotopic rel A and B, a contradiction.

To show that d_1 and d_2 do not form a bigon, note that they intersect twice, so there are only four regions that are candidates for bigons. One of them contains the

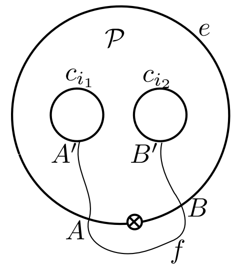

Figure 6.6: Ruling out the possibility of a bigon between e and a curve $f \in C$.

crosscap, so it is not a bigon. Two other regions are ruled out because c_{i_1} and c_{i_2} are not nullhomotopic. The fourth region is a bigon if any only if c_{i_1} and c_{i_2} are isotopic in S, but our assumption is that they are not isotopic. Hence d_1 and d_2 are in minimal position.

Finally, the pairs (d_1, e) and (d_2, e) are checked similarly, again using the fact that c_{i_1} and c_{i_2} are not nullhomotopic and not isotopic. This finishes the proof of (ii).

Let $C_0 = C - \{c_{i_1}, c_{i_2}\}$. We can write $i(C''', C''')$ in the following block form.

$$
\begin{aligned}
i(C''', C''') &= \\
&= \begin{pmatrix}
0 & i(C_0, c_{i_1}) & i(C_0, c_{i_2}) & i(C_0, d_1) & i(C_0, d_2) & i(C_0, e) \\
i(c_{i_1}, C_0) & 0 & i(c_{i_1}, c_{i_2}) & i(c_{i_1}, d_1) & i(c_{i_1}, d_2) & i(c_{i_1}, e) \\
i(c_{i_2}, C_0) & i(c_{i_2}, c_{i_1}) & 0 & i(c_{i_2}, d_1) & i(c_{i_2}, d_2) & i(c_{i_2}, e) \\
i(d_1, C_0) & i(d_1, c_{i_1}) & i(d_1, c_{i_2}) & 0 & i(d_1, d_2) & i(d_1, e) \\
i(d_2, C_0) & i(d_2, c_{i_1}) & i(d_2, c_{i_2}) & i(d_2, d_1) & 0 & i(d_2, e) \\
i(e, C_0) & i(e, c_{i_1}) & i(e, c_{i_2}) & i(e, d_1) & i(e, d_2) & 0
\end{pmatrix} = \\
&= \begin{pmatrix}
0 & X & Y & X & Y & X+Y \\
X^T & 0 & 0 & 0 & 0 & 0 \\
Y^T & 0 & 0 & 0 & 0 & 0 \\
X^T & 0 & 0 & 0 & 2 & 2 \\
Y^T & 0 & 0 & 2 & 0 & 2 \\
X^T + Y^T & 0 & 0 & 2 & 2 & 0
\end{pmatrix} \sim \begin{pmatrix}
0 & X & Y & 0 & 0 & 0 \\
X^T & 0 & 0 & 0 & 0 & 0 \\
Y^T & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 2 & 0 & 2 \\
0 & 0 & 0 & 2 & 2 & 0
\end{pmatrix},
\end{aligned}
$$

where $X = i(C_0, c_{i_1})$ and $Y = i(C_0, c_{i_2})$ and where the last relation is the equivalence under column and row operations. The upper left 3×3 block is $i(C, C)$, and the lower right 3×3 block is invertible. Hence $\operatorname{rank}(i(C''', C''')) = \operatorname{rank}(i(C, C)) + 3$.

The calculation is analogous for $i(C'', C'')$ and $i(C', C')$. In the first case, we have the invertible matrix $\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$ in the lower right corner. In the second case, the lower right corner is a single zero entry, hence $\operatorname{rank}(i(C', C')) = \operatorname{rank}(i(C, C))$. $\square$

Lemma 6.4. *Let C be a filling collection of inconsistently marked simple closed curves on a surface S with at least one puncture. Suppose that the curves of C are in pairwise minimal position.*

Then there is a point $p \in S - C$ and marked simple closed curves d and e on $S - \{p\}$ such that $C' = C \cup \{d\}$ and $C'' = C \cup \{d, e\}$ are filling inconsistently marked collec-

tions on $S - \{p\}$, and the curves in each collection are in pairwise minimal position. Moreover, we have

$$\operatorname{rank}(i(C, C)) = \operatorname{rank}(i(C', C')) = \operatorname{rank}(i(C'', C'')) - 2.$$

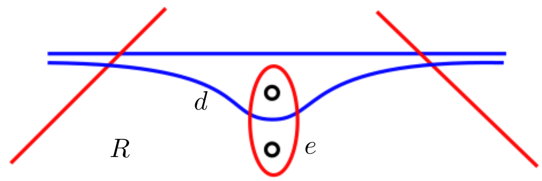

Figure 6.7: Creating two new curves by duplicating a puncture.

Proof. Let R be a component of $S - C$ which is a once-punctured disk. Let $p \in R$. Let e be a curve surrounding the puncture and p inside R. Let d be a curve c obtained from a curve on the boundary of R by pulling it over p. (See Figure 6.7.) The properties of inconsistent marking, filling and minimal position are easy to verify. In addition, we have

$$i(C'', C'') = \begin{pmatrix} i(C_0, C_0) & \mathbf{x} & \mathbf{x} & 0 \\ \mathbf{x}^T & 0 & 0 & 0 \\ \mathbf{x}^T & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix},$$

where $C_0 = C - c$ and $\mathbf{x} = i(C_0, c)$. Note that the $i(C', C')$ is the submatrix obtained by deleting the last row and last column, and $i(C, C)$ is the submatrix obtained by deleting the last two rows and the last two columns. This proves the equation about the ranks. □

We are now ready to give the proof of Proposition 6.2.

Proof of Proposition 6.2. Since the constructions in the different cases have different flavors, we divide the proof into three parts. First we give examples for completely left-to-right collections of curves, then for unrestricted collections of curves when $g \geq 5$ and $g \leq 4$, respectively.

Part 1 (Completely left-to-right collections)**.** *For any $g \geq 4$, $n \geq 0$ and $3 \leq r \leq g - 1$, we need to construct a filling and inconsistently marked completely left-to-right collection of curves on $N_{g,n}$ whose intersection matrix has rank r.*

When $r = g-1$ and $n = 0$, we think about N_g and a sphere with g crosscaps attached. We arrange r curves around a central crosscap as shown on the left on Figure 6.8. When the curves are marked as shown on Figure 6.8, we obtain a completely left-to-right collection, because red arcs always cross the blue arcs from left to right (see the paragraphs before Proposition 6.2 for the explanation of the coloring convention). The intersection matrix is the $r \times r$ square matrix whose off-diagonal entries are 1 and whose diagonal entries are 0. Note that this matrix is invertible for all r: the inverse is the matrix whose off-diagonal entries are $1/(r - 1)$ and whose diagonal entries are $-(r - 2)/(r - 1)$.

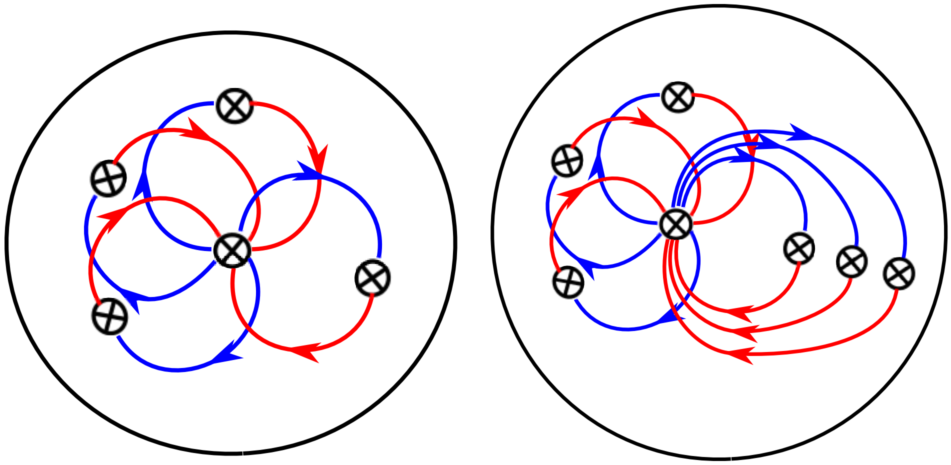

Figure 6.8: Rank 4 left-to-right collections on N_5 and N_7. Red arcs always cross blue arcs from the left to right.

For the cases $r < g-1$ and $n = 0$, we modify the previous arrangement by attaching crosscaps and replacing one of the curves with a set of new curves as on the right of Figure 6.8. A bit of care should be taken here as the number of new curves is less than the number of attached crosscaps, whereas it may seem first to the eyes used to orientable surfaces that the two numbers are equal. This is because the curves that connect at the central crosscap are the antipodal ones. In the particular case shown by Figure 6.8, the new set of curves consists of two curves, not three.

We claim that this modification does not change the rank of the intersection matrix. For this, note that the intersection number of the new curves with the rest of the curves are proportional, so the corresponding columns and rows in the intersection matrix are scalar multiples of each other. Hence the rank of the intersection matrix is the same for the two examples on Figure 6.8.

When $n > 0$, we consider the collection C already constructed in the case $n = 0$, add n disjoint isotopic copies of one of the curves in C, and arrange the n punctures between the $n+1$ isotopic curves so that no two of them are isotopic in the punctured surface. These $n+1$ curves have the same intersection numbers with the other curves, hence the rank of the intersection matrix remains the same as in the case $n = 0$. It is also clear that the new collection of curves remains completely left-to-right if the duplicated curves inherit their marking from the original curve.

Part 2 (Unrestricted case, $g \geq 5$)**.** *Note that attaching a crosscap to the orientable surface $S_{g',n}$ yields the nonorientable surface $N_{2g'+1,n}$. Hence we can obtain nonorientable surfaces of odd (resp. even) genus by attaching one (resp. two) crosscap(s) to an orientable surface.*

During the proof of Proposition 6.1, we constructed for many g', n, r' a collection of curves $C_{g',n,2r'}$ on $S_{g',n}$ such that $\operatorname{rank}(i(C_{g',n,2r'}, C_{g',n,2r'})) = 2r'$. If $g' \geq 2$ and $r' \geq 2$, then $C_{g',n,2r'}$ has at least two complementary regions that allow applying Lemma 6.3: both of the two left-most regions work on Figures 6.1 and 6.2. In fact, Lemma 6.3 can *be applied subsequently for the two regions. That is, after applying it for the first region, the hypotheses of the lemma still hold for the other region. This gives examples for all triples (g, n, r) where $g \geq 5$, $r \neq 3, 5$ and n is arbitrary.*

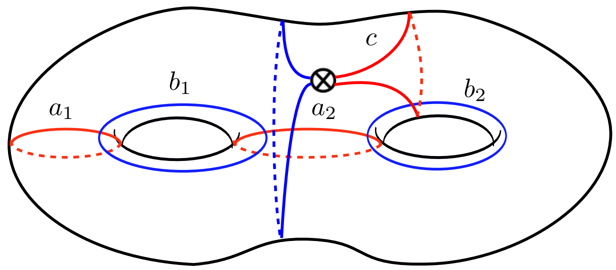

Figure 6.9: A rank 5 collection on N_5.

Examples for the cases

- $g \geq 5$, $r = 3$, *n is arbitrary,*
- $g \geq 6$, $r = 5$, *n is arbitrary,*

have been given in Part 1. The only case that remains is $g = 5$, $r = 5$. Figure 6.9 gives an example when $n = 0$. The intersection matrix is

$$i(C, C) = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 \end{pmatrix}$$

and it is invertible. When $n > 0$, we add parallel curves separating the punctures as in the end of Part 1.

Part 3 (Unrestricted case, $g \leq 4$)**.** *Table 6.2 summarizes the ranks we need to realize on the surfaces $N_{g,n}$ when $g \leq 4$.*

n\g	1	2	3	4
0	∅	∅	∅	3–5 (3,4,5)
1	∅	∅	3–4 (3,4)	3–8 (8)
2	∅	3–4 (3,4)	3–7 (7)	3–10
3	3 (3)	3–6	3–9	3–12
4	3–5 (4)	3–8	3–11	3–14
5	3–7	3–10	3-13	3–16

Table 6.2: Ranks to realize on nonorientable surfaces of genus at most 4. We give explicit examples of curve collections realizing the ranks shown in the parentheses and construct the other examples of using Lemmas 6.3 and 6.4.

We construct only finitely many examples (shown in parentheses in Table 6.2). When $g \leq 3$, these examples and Lemma 6.4 take care of all cases. What is different in the case $g = 4$ is that the surface with the smallest number of punctures (N_4) is closed, so we cannot apply Lemma 6.4 to construct examples on $N_{4,1}$ from the examples on N_4. However, the rank 3, 4 and 5 collections we will construct on N_4 still fill when a puncture is added, so the same collections realize the ranks 3, 4 and 5 on $N_{4,1}$. To realize 6 and 7, we apply Lemma 6.3 for our rank 4 collections on $N_{3,1}$. Finally, we describe a rank 8 example explicitly. Lemma 6.4 can then be used to complete the construction for all $N_{g,n}$ with $g = 4$ and $n > 1$.

To complete the proof, we now give the examples for the cases listed in the parentheses in Table 6.2. We recall that other than the filling property, the collections of curves also need to be marked inconsistently. In other words, a blue and red arc should meet at every intersection.

Case I ($g = 1$)**.** *Figure 6.10 shows collections with*

$$i(C,C) = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix} \quad \textit{and} \quad i(C,C) = \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 0 \\ 2 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \end{pmatrix}.$$

It is easy to see that both matrices are invertible.

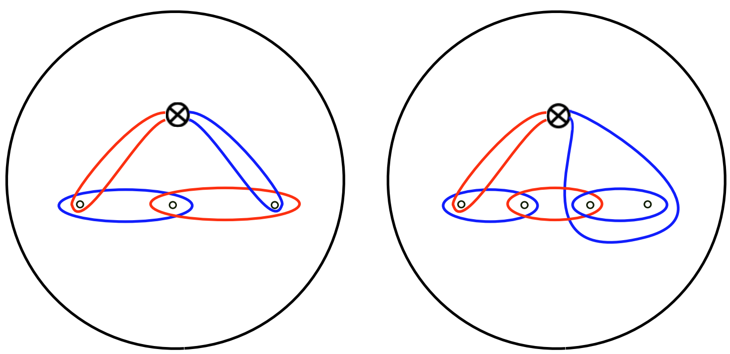

Figure 6.10: A rank 3 collection on $N_{1,3}$ and a rank 4 collection on $N_{1,4}$. The surfaces are represented as spheres with added crosscaps and punctures.

Case II ($g = 2$)**.** *Figure 6.11 shows collections with*

$$i(C,C) = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 4 \\ 2 & 4 & 0 \end{pmatrix} \quad \textit{and} \quad i(C,C) = \begin{pmatrix} 0 & 2 & 2 & 2 \\ 2 & 0 & 2 & 2 \\ 2 & 2 & 0 & 4 \\ 2 & 2 & 4 & 0 \end{pmatrix}.$$

Again, one can check that both matrices are invertible.

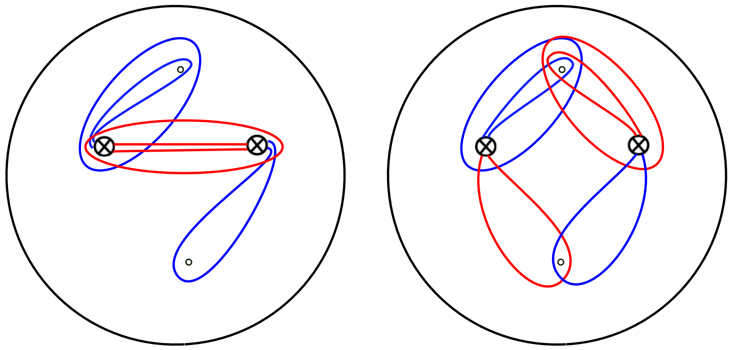

Figure 6.11: A rank 3 and rank 4 collection on $N_{2,2}$.

Case III $(g = 3)$**.** *Figure 6.12 shows a filling collection with intersection matrix*

$$i(C, C) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{pmatrix}, \tag{6.2}$$

which has rank 4. Note that there is a complementary region with two disjoint curves on its boundary, hence Lemma 6.3 indeed applies to yield rank 4, 6 and 7 collections on $N_{4,1}$*.*

By dropping the curve surrounding the hole which intersects only one other curve, the remaining three curves still fill and the intersection matrix is the lower right 3×3 *submatrix of (6.2), which has rank 3.*

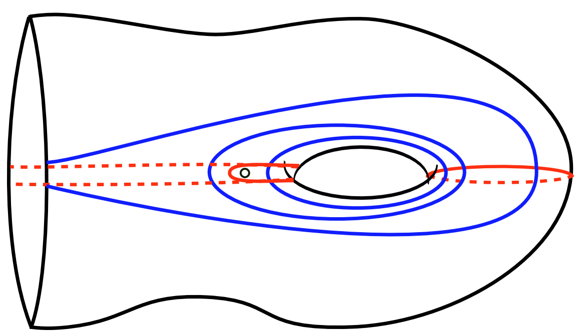

Figure 6.12: A rank 4 collection on $N_{3,1}$. Here the surface $N_{3,1}$ is represented by identifying antipodal points of the boundary component of the pictured orientable surface.

On the left, Figure 6.13 shows a filling collection of curves on $N_{3,2}$ *with intersection* *matrix*

$$i(C,C)=\begin{pmatrix}0&2&2&2&2&2&4\\2&0&0&2&4&4&4\\2&0&0&2&4&2&2\\2&2&2&0&2&2&4\\2&4&4&2&0&0&4\\2&4&2&2&0&0&2\\4&4&2&4&4&2&0\end{pmatrix},$$

which has rank 7.

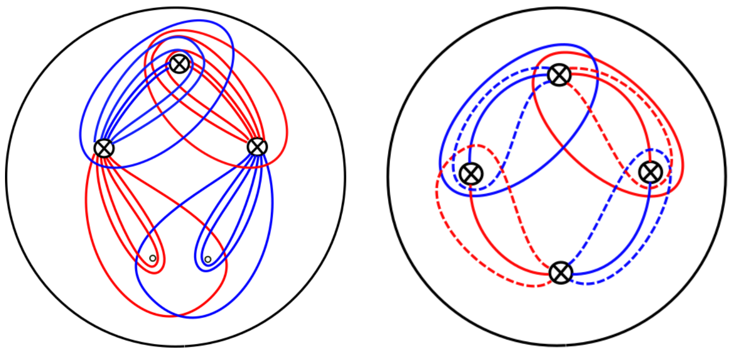

Figure 6.13: A rank 7 collection on $N_{3,2}$ and a rank 5 collection on N_4.

Case IV ($g=4$)**.** *On the right, Figure 6.13 shows a filling collection of curves on N_4 with intersection matrix*

$$i(C,C)=\begin{pmatrix}0&2&2&2&2\\2&0&2&2&2\\2&2&0&2&2\\2&2&2&0&4\\2&2&2&4&0\end{pmatrix}, \tag{6.3}$$

which has rank 5. By dropping both or one of the dashed curves, the remaining three or four curves still fill. The intersection matrices are the upper left 3×3 and 4×4 submatrices of (6.3), which have rank 3 and rank 4, respectively.

Finally, Figure 6.14 shows a filling collection of curves on $N_{4.1}$ with intersection matrix

$$i(C,C)=\begin{pmatrix}0&2&2&2&2&2&4&0\\2&0&2&2&2&2&4&0\\2&2&0&4&4&4&8&0\\2&2&4&0&0&0&0&0\\2&2&4&0&0&2&2&2\\2&2&4&0&2&0&2&2\\4&4&8&0&2&2&0&4\\0&0&0&0&2&2&4&0\end{pmatrix}.$$

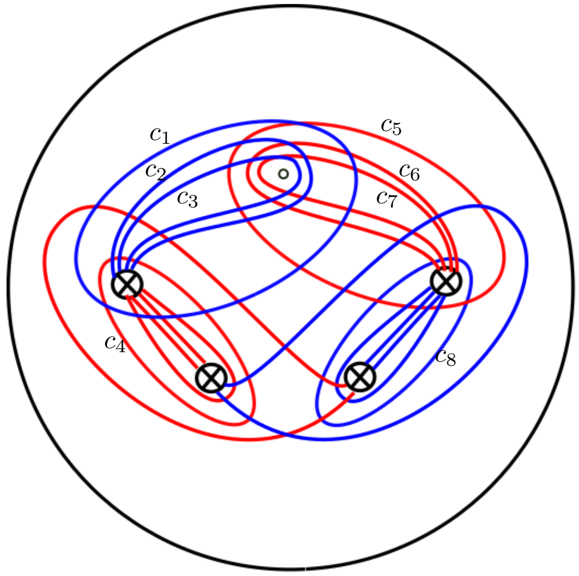

Figure 6.14: A rank 8 collection on $N_{4,1}$.

It is easy to zero out the upper right and lower left 4×4 *block using row and column operations. The remaining* 4×4 *matrices are invertible, hence* $i(C, C)$ *has rank 8.*

This completes the proof of Proposition 6.2. $\square$

7 Two pseudo-Anosov examples for the sporadic cases

We have tried but were not able to construct inconsistently marked collections of curves with rank 6 intersection matrix on N_4 and rank 5 intersection matrix on $N_{3,1}$. Since we have made many attempts using different perspectives and were not successful, we conjecture that such collections of curves do not exist. However, we will show that degree 6 and 5 stretch factors still exist on the surfaces N_4 and $N_{3,1}$.

Due to the lack of suitable collections of curves, these constructions will not use Penner's construction. Instead, we will describe the mapping classes explicitly and compute the stretch factors using transition matrices of train track maps.

Proposition 7.1. *We have* $6 \in D(N_4)$ *and* $5 \in D(N_{3,1})$.

Proof. We will now represent the surface $N_{3,1}$ as a polygon with its sides identified. On Figure 7.1, the open circles on the boundary of each rectangle divide the boundary to six segments. The two parallel pairs of side which are not marked by an arrow are identified by a translation. The pair marked by an arrow is identified by a flip. The six open circles on the boundary identify to a single point, representing the puncture.

The first picture of Figure 7.1 shows a train track τ on this surface. The train track has two switches, seven branches and four complementary regions. Three of these regions is a trigon, one is a monogon.

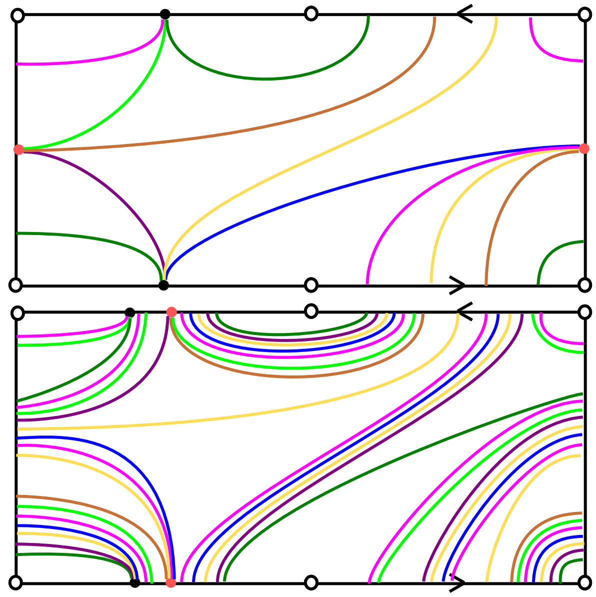

Figure 7.1: A train track embedded into $N_{3,1}$ in two different ways.

The second picture shows the same train track, embedded in $N_{3,1}$ in a different way. The map from τ to this second train track extends to a homeomorphism of the surface which well-defined up to homotopy. In other words, Figure 7.1 describes some $f \in \mathrm{Mod}(N_{3,1})$ and the second train track is $f(\tau)$.

Note that $f(\tau)$ is carried on τ and the matrix describing the f-action on the branches of τ is

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

This matrix is Perron–Frobenius and the minimal polynomial of the Perron–Frobenius eigenvalue is $x^5 - 3x^4 - x^3 + x^2 - x - 1$. Hence f is a pseudo-Anosov mapping class with degree 5 stretch factor.

For our second example, consider the pictures on Figure 7.2 which describe an element of $\mathrm{Mod}(N_{4,1})$ in the same way as in the previous example. The matrix describing

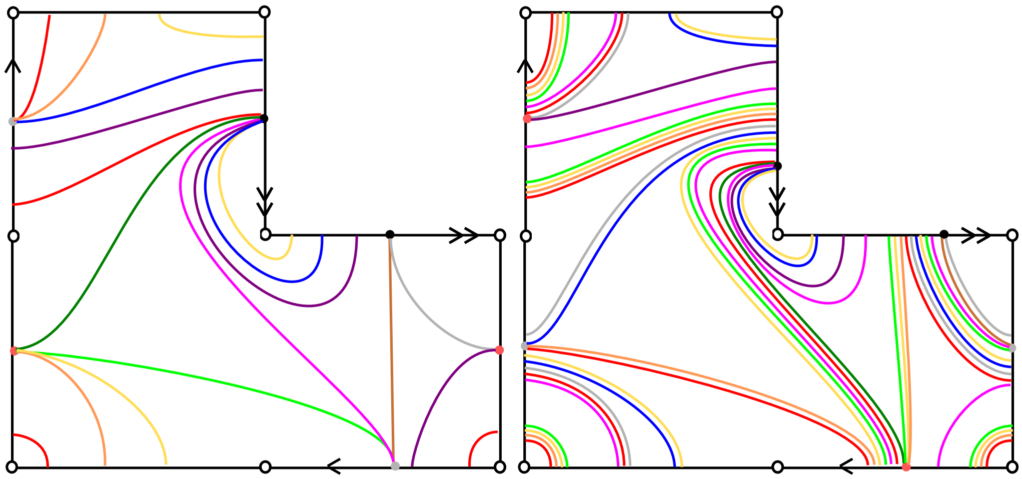

Figure 7.2: A train track embedded into $N_{4,1}$ in two different ways. Segments of the boundaries of the polygons are identified by translations except for the sides marked by arrows. The open circles on the boundary identify to a single point, representing the puncture.

the action on the train track is

$$\begin{pmatrix}1&0&1&1&0&0&0&0&1&1\\0&0&1&0&1&0&0&0&0&1\\0&1&1&1&1&0&0&0&1&0\\0&1&0&1&0&1&0&0&1&0\\0&0&1&1&0&0&0&0&1&1\\0&0&0&0&0&0&0&1&0&0\\0&0&0&1&0&0&0&0&0&0\\1&0&0&0&0&0&1&0&1&0\\0&1&0&0&0&0&1&0&2&0\\0&0&0&0&0&0&0&0&1&0\end{pmatrix}$$

This is again Perron–Frobenius and the minimal polynomial of the largest eigenvalue is $x^6 - 3x^5 - x^4 - x^2 + x + 1$. Hence the mapping class is pseudo-Anosov with degree 6 stretch factor.

We are not done, because we need a degree 6 example on N_4, not on $N_{4,1}$. But note that the complementary region of the train track containing the puncture is a bigon, hence our pseudo-Anosov map has a 2-pronged singularity at the puncture. So we can fill in the puncture to obtain a pseudo-Anosov map on N_4 with the same degree 6 stretch factor. □

We remark that the two candidate mapping classes, each as a product of three Dehn twists, were found with the help of the computer software `flipper` by Mark Bell [Bel16]. The train track maps and the transition matrices were then computed entirely by hand. We have mentioned this in the proof already but we would like to reiterate that verifying the correctness of the examples is a much easier process than finding them, and this process can also be done—and have been done—without using computers: one only needs to check that in each case the two train tracks that we claim to be isomorphic are indeed isomorphic and then use the pictures to verify that the transition matrices are also as shown above.

8 Proofs of the main theorems

In this section, we give the proofs of the main theorems.

8.1 Bounds on the algebraic degree

Determining the sets $D(S)$ and $D^+(S)$ has two parts: ruling out the integers that do not arise as degrees and constructing examples realizing integers that have not been ruled out. The first part has almost entirely been done before, this section is for summarizing the relevant results.

General bounds. The first upper bound on the degree is due to Thurston [Thu88]. Recall that $\mathrm{Teich}(S)$ stands for Teichmüller space of the surface S.

Proposition 8.1. *For any finite type surface S, we have*

$$\max D(S) \leq \dim(\mathrm{Teich}(S)).$$

Next, we give a bound on the largest degree in $D^+(S)$. This was certainly also known before, but since we have not found a proof a literature, we include one here.

First we need a fact about degrees of powers of algebraic numbers. The algebraic degree of some algebraic numbers—for example, $\sqrt{2}$—decreases under taking powers. We will use the fact that this does not happen for pseudo-Anosov stretch factors.

Lemma 8.2. *If λ is a pseudo-Anosov stretch factor, then $\deg(\lambda) = \deg(\lambda^k)$ for every positive integer k.*

Proof. Let $p(x) = (x - \lambda)(x - \lambda_2) \cdots (x - \lambda_n) \in \mathbb{Q}[x]$ be a minimal polynomial of λ, factored over the complex numbers. Note that $n = \deg(\lambda)$.

Consider the polynomial $p_k(x) = (x - \lambda^k)(x - \lambda_2^k) \cdots (c - \lambda_n^k)$. The group $G = Aut(\mathbb{Q}(\lambda)/\mathbb{Q})$ of automorphisms of the number field $\mathbb{Q}(\lambda)$ permutes the roots of $p_k(x)$, hence acts trivially on the coefficients. It follows that $p_k(x) \in \mathbb{Q}[x]$. We obtain that $\deg(\lambda^k) \leq \deg(\lambda)$ for every positive integer k. This holds in general, we have not yet used the assumption that λ is a pseudo-Anosov stretch factor.

For the reverse inequality, it suffices to show that $p_k(x)$ is irreducible over $\mathbb{Q}$ for every positive integer k if λ is a pseudo-Anosov stretch factor. If this was not the case, the length of the G-orbit of λ^k would be less than n, so there would be a nontrivial element of G fixing λ^k. As a consequence, $|\lambda_i| = |\lambda|$ would hold for some $2 \leq i \leq n$. This is impossible, since λ is the unique largest root of $p(x)$ in absolute value. $\square$

The second lemma we will use relates degrees on punctured surfaces to degrees on closed surfaces. Recall that $\overline{S}$ denotes the closed surface obtained from S by filling in the punctures.

Lemma 8.3. *For any finite type surface S, we have $D^+(S) = D^+(\overline{S})$.*

Proof. If ψ is a pseudo-Anosov map on S with an orientable invariant foliation, then it has no 1-pronged singularities. So the pseudo-Anosov map on $\overline{S}$ obtained by extending ψ to the punctures is still pseudo-Anosov and has the same stretch factor as ψ. This shows that $D^+(S) \subset D^+(\overline{S})$.

If ψ is a pseudo-Anosov map on $\overline{S}$ with stretch factor λ, then some power of it has at least as many fixed points as the number of punctures of S. Hence there is a pseudo-Anosov map on S whose stretch factor is some power of λ. By Lemma 8.2, the algebraic degree of λ is preserved under powers, hence $D^+(\overline{S}) \subset D^+(S)$. ☐

Proposition 8.4. *For any finite type surface S, we have*

$$\max D^+(S) \leq \dim(H^1(\overline{S}, \mathbb{R})).$$

Proof. By Lemma 8.3 it suffices to show that $\max D^+(S) \leq \dim(H^1(S, \mathbb{R}))$ if S is closed. This follows from the fact that the orientable invariant foliation corresponds to a 1-form on S, so it is an eigenvector of the action of the $H^1(S, \mathbb{R})$. ☐

Bounds on odd degrees. The next group of results is on restrictions on odd degrees.

Proposition 8.5. *If S is orientable, $d \in D(S)$ and d is odd, then $d \leq \frac{1}{2}\dim(\mathrm{Teich}(S))$.*

For closed surfaces, this result is due to Long [Lon85, Theorem 3.3]. McMullen later gave an different proof [Shi16, Theorem 10] which works for punctured surfaces as well. An analogous argument (where the action on the space of projective measured foliations is replaced by the action on homology) also yields the following.

Proposition 8.6. *If S is orientable, $d \in D^+(S)$ and d is odd, then $d \leq \frac{1}{2}\dim(H^1(\overline{S}, \mathbb{R}))$.*

Propositions 8.5 and 8.6 have no analogs for nonorientable surfaces as demonstrated by Proposition 6.2.

Two facts about nonorientable surfaces. Degree two stretch factors occur on the torus and the four times punctured sphere, so they arise on all higher complexity orientable surfaces as well via puncturing and branched coverings. In sharp contrast, degree two stretch factors do not occur on nonorientable surfaces at all. This requires only a small observation, but as far as we know, this fact has not been noticed before.

Proposition 8.7. *If S is nonorientable, then $2 \notin D(S)$.*

Proof. If $\deg(\lambda) = 2$, then its minimal polynomial has the form $x^2 \pm kx \pm 1$ for some $k \in \mathbb{Z}$, since λ in algebraic unit. But this is impossible, since $\pm 1/\lambda$ is never a Galois conjugate of λ when the pseudo-Anosov map is supported on a nonorientable surface. (This is stated for $1/\lambda$ in [Str17, Proposition 2.3], but the same proof works for $-1/\lambda$ as well.) ☐

Finally, we include the following well-known fact for completeness.

Proposition 8.8. *The closed nonorientable surface of genus 3 does not admit pseudo-Anosov maps.*

Proof. There is a unique one-sided curve on the surface whose complement is a one-holed torus [Sch82, Lemma 2.1]. This curve is fixed by every mapping class. □

8.2 Proof of the degree theorem

We are now ready to state and prove a more general version of Theorem 1.1. We use the notations $[a, b]_{\text{even}}$ and $[a, b]_{\text{odd}}$ from Theorem 1.1, and abuse the interval notation $[a, b]$ to mean the set of integers in the corresponding real interval.

Theorem 8.9. *Let S be a finite type orientable surface.*

(i) We have

$$D^+(S) = \big[2, \dim(H_1(\overline{S}, \mathbb{R}))\big]_{\text{even}} \cup \left[3, \frac{1}{2}\dim(H_1(\overline{S}, \mathbb{R}))\right]_{\text{odd}}. \tag{8.1}$$

(ii) If S has an even number of punctures or $\frac{1}{2}\dim(\mathrm{Teich}(S))$ is even, then

$$D(S) = \big[2, \dim(\mathrm{Teich}(S))\big]_{\text{even}} \cup \left[3, \frac{1}{2}\dim(\mathrm{Teich}(S))\right]_{\text{odd}}. \tag{8.2}$$

(iii) If S has an odd number of punctures and $\frac{1}{2}\dim(\mathrm{Teich}(S))$ is odd, then either (8.2) holds or

$$D(S) = \big[2, \dim(\mathrm{Teich}(S))\big]_{\text{even}} \cup \left[3, \frac{1}{2}\dim(\mathrm{Teich}(S)) - 1\right]_{\text{odd}}. \tag{8.3}$$

Let N be a finite type nonorientable surface.

(iv) We have

$$D^+(N) = \big[3, \dim(H_1(\overline{N}, \mathbb{R}))\big].$$

(v) If N is not the closed nonorientable surface N_3 of genus 3, then

$$D(N) = \big[3, \dim(\mathrm{Teich}(N))\big].$$

The reason why the cases (ii) and (iii) are separated is that we cannot realize the odd degree $\frac{1}{2}\dim(\mathrm{Teich}(S))$ when the surface is not a double cover of a nonorientable surface. It seems possible that realizing this degree is not possible using Penner's construction. We do not know if it is possible using other constructions. Randomized computer experiments yield almost exclusively even degree stretch factors, so searching for these degrees with computers also seems a nontrivial task.

The exclusion of the surface N_3 is necessary, since $\dim(\mathrm{Teich}(N_3)) = 3$, so the interval $[3, \dim(\mathrm{Teich}(N_3))]$ is nonempty. However, this surface does not admit pseudo-Anosov maps (Proposition 8.8).

Proof. The fact that the sets cannot be larger than stated follows from the results in Section 8.1. It remains to prove that all the degrees claimed in the theorem can be realized.

The statements (iv) and (v) follow from Theorem 5.4 and Proposition 6.2 with the exception of the two sporadic cases: $6 \in D(N_4)$ and $5 \in D(N_{3,1})$. (Compare Table 8.1 and Table 6.1.) Examples for these two cases were given in Section 7.

n\g	1	2	3	4	5
0	$\emptyset$	$\emptyset$	$\emptyset$	A	A
1	$\emptyset$	$\emptyset$	A	A	A
2	$\emptyset$	A	A	A	A
3	A	A	A	A	A

Table 8.1: Nonorientable surfaces admitting (A) and not admitting ($\emptyset$) pseudo-Anosov maps.

The construction of even degrees for (i), (ii) and (iii) follow from Theorem 5.4 and Proposition 6.1.

It remains to construct odd degrees on orientable surfaces. We obtain these by lifting pseudo-Anosov maps from nonorientable surfaces. Lifting preserves the stretch factor hence also the degree.

Suppose S is an orientable surface with an even number of punctures, and let $S \to N$ be a covering where N is a nonorientable surface. We have $\dim(\mathrm{Teich}(S)) = 2\dim(\mathrm{Teich}(N))$ and $\dim(H_1(\overline{S})) = 2\dim(H_1(\overline{N}))$. If $S \neq S_2$, then $N \neq N_3$, and it follows that $[3, \frac{1}{2}\dim(\mathrm{Teich}(S))] \subset D(S)$ and $[3, \frac{1}{2}\dim(H_1(\overline{S}))] \subset D^+(S)$. This proves (8.1) and (8.2) when S has an even number of punctures and $S \neq S_2$. In the case $S = S_2$, we only need to show that $3 \in D(S)$. This is shown by an example in Section 7 of [Shi16]. Note that Lemma 8.3 implies that (8.1) holds also when S has an odd number of punctures. This completes the proof of (i).

Finally, suppose S is an orientable surface that has an odd number of punctures. Let S' be the surface obtained from S by filling in one puncture. We have already shown that (8.2) holds for S', and we will use this to show that it also holds (or almost holds) for S.

To prove that (8.2) holds for S when $\frac{1}{2}\dim(\mathrm{Teich}(S))$ is even, it suffices to show that the left hand side contains the right hand side. Note that since $\frac{1}{2}\dim(\mathrm{Teich}(S))$ is even, there is no difference between the right hand sides of (8.2) for S and S'. On the other hand, the argument in the proof of Lemma 8.3 shows that $D(S') \subset D(S)$. So since $D(S')$ is already proven to contain the right hand side, the same is true for $D(S)$. This completes the proof of statement (ii).

Now suppose $\frac{1}{2}\dim(\mathrm{Teich}(S))$ is odd. By an analogous argument, the fact that $D(S)$ contains the right hand side of (8.3) follows from the fact that $D(S') \subset D(S)$ and the fact that $D(S')$ contains the right hand side. This completes the proof of (iii) and hence the proof of the theorem. □

8.3 Proof of the trace field theorem

In this section, we give the proof of Theorem 1.3. First we need the following lemma which will be used to show that the stretch factor λ and its reciprocal λ^{-1} tend to be Galois conjugates when they arise from Penner's construction on an orientable surface.

Lemma 8.10. *Suppose the surface S is orientable and let Ω be the intersection matrix of a collection of curves satisfying the hypotheses of Penner's construction. If M is a product of the matrices Q_i with Perron–Frobenius eigenvalue λ, then λ^{-1} is also an eigenvalue of M.*

Proof. Recall from the proof of Proposition 5.1 the linear map $\ell_M : \widehat{V} \to \widehat{V}$ induced by the left action of M on $\widehat{V} = \mathbb{R}^n / \mathrm{Nul}(\Omega)$. We will show that ℓ_M is a symplectic transformation. Since λ is an eigenvalue of ℓ_M and eigenvalues of symplectic transformations come in reciprocal pairs [MHO09, Chapter 2], this will complete the proof.

Since S is orientable, our collection of curves is a union of two multicurves A and B. Therefore Ω has the block form

$$\Omega = \begin{pmatrix} 0 & X \\ X^T & 0 \end{pmatrix}$$

where X is an $a \times b$ matrix where a and b are the number of curves in A and B, respectively.

Define the alternating bilinear form $\langle \cdot, \cdot \rangle_\Delta$ on $\mathbb{R}^n$ by the formula

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_\Delta = \mathbf{v}_1^T \Delta \mathbf{v}_2,$$

where

$$\Delta = \begin{pmatrix} 0 & X \\ -X^T & 0 \end{pmatrix}.$$

The matrices Ω and Δ are related by the equations $\Delta = U\Omega = -\Omega U$, where

$$U = \begin{pmatrix} I_a & 0 \\ 0 & -I_b \end{pmatrix} \tag{8.4}$$

and I_a and I_b are the $a \times a$ and $b \times b$ identity matrices.

Next we show that $\langle \cdot, \cdot \rangle_\Delta$ descends to a symplectic form on $\widehat{V}$. If $\mathbf{v}_0, \mathbf{v}_0' \in \mathrm{Nul}(\Omega)$, then

$$(\mathbf{v}' + \mathbf{v}_0')^T \Delta (\mathbf{v} + \mathbf{v}_0) = \mathbf{v}'^T \Delta \mathbf{v},$$

because $\Delta \mathbf{v}_0 = U\Omega \mathbf{v}_0 = 0$ and $\mathbf{v}_0'^T \Delta = -\mathbf{v}_0'^T \Omega U = 0$. As a consequence, Δ gives rise to a well-defined alternating form on $\widehat{V}$ that we also denote by $\langle \cdot, \cdot \rangle_\Delta$. This form is nondegenerate, because for any $\mathbf{v} \in \mathbb{R}^n - \mathrm{Nul}(\Omega)$, we have $\Delta \mathbf{v} \neq 0$, therefore there exists $\mathbf{v}' \in \mathbb{R}^n$ with $\mathbf{v}'^T \Delta \mathbf{v} \neq 0$. Hence $\langle \cdot, \cdot \rangle_\Delta$ is a symplectic form on $\widehat{V}$.

Finally, we show that the left actions of the Q_i preserve this symplectic form on $\widehat{V}$. Using the fact that the diagonal matrices U and D_i commute, we have

$$\begin{aligned} Q_i^T \Delta Q_i &= (I + \Omega D_i) U \Omega (I + D_i \Omega) = \\ &= U\Omega + \Omega D_i U \Omega + U \Omega D_i \Omega + \Omega D_i U \Omega D_i \Omega = \\ &= \Delta - U\Omega D_i \Omega + U \Omega D_i \Omega + \Omega U D_i \Omega D_i \Omega = \Delta, \end{aligned}$$

where the last term vanishes because the diagonal entries of Ω are zero, so $D_i \Omega D_i = 0$. □

Proof of Theorem 1.3. All even degrees between 2 and $6g-6$ can be realized on S_g as the algebraic degree of stretch factors by Theorem 8.9. The construction of these examples occurs on the surfaces S_g themselves (as opposed to odd degrees that are constructed by lifting from nonorientable surfaces). By Lemma 8.10, both λ_k and λ_k^{-1} are eigenvalues of the matrices $M_{\gamma,k}$ in Theorem 5.3, so λ_k and λ_k^{-1} are Galois conjugates if k is large enough. It follows that $\mathbb{Q}(\lambda) : \mathbb{Q}(\lambda + \lambda^{-1}) = 2$, and we obtain that the trace field $\mathbb{Q}(\lambda + \lambda^{-1})$ can have any degree from 1 to $3g-3$.

On the other hand, the degree of the trace field cannot be larger than $3g-3$. If it were, then we would have $\mathbb{Q}(\lambda) : \mathbb{Q}(\lambda + \lambda^{-1}) = 1$, otherwise the degree of λ would be bigger than $6g-6$, contradicting Proposition 8.1. In other words, λ and λ^{-1} would not be Galois conjugates. But if they are not Galois conjugates, then McMullen's argument that proves Proposition 8.5 would prove that $\deg(\lambda) \leq 3g-3$. □

References

[AY81] Pierre Arnoux and Jean-Christophe Yoccoz. Construction de difféomorphismes pseudo-Anosov. *C. R. Acad. Sci. Paris Sér. I Math.*, 292(1):75–78, 1981.

[Bel16] Mark Bell. flipper (computer software). Version 0.10.3, 2013–2016.

[BF07] Mladen Bestvina and Koji Fujiwara. Quasi-homomorphisms on mapping class groups. *Glas. Mat. Ser. III*, 42(62)(1):213–236, 2007.

[BF17] Mladen Bestvina and Koji Fujiwara. Handlebody subgroups in a mapping class group. In *In the Tradition of Ahlfors–Bers, VII*, volume 696 of *Contemp. Math.*, pages 29–50. Amer. Math. Soc., Providence, RI, 2017.

[BT92] Alberto Borobia and Ujué R. Trías. A geometric proof of the Perron-Frobenius theorem. *Rev. Mat. Univ. Complut. Madrid*, 5(1):57–63, 1992.

[Fat92] Albert Fathi. Démonstration d'un théorème de Penner sur la composition des twists de Dehn. *Bull. Soc. Math. France*, 120(4):467–484, 1992.

[FM12] Benson Farb and Dan Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[FR99] J. Franks and E. Rykken. Pseudo-Anosov homeomorphisms with quadratic expansion. *Proc. Amer. Math. Soc.*, 127(7):2183–2192, 1999.

[GJ00] Eugene Gutkin and Chris Judge. Affine mappings of translation surfaces: geometry and arithmetic. *Duke Math. J.*, 103(2):191–213, 2000.

[HMSZ06] Pascal Hubert, Howard Masur, Thomas Schmidt, and Anton Zorich. Problems on billiards, flat surfaces and translation surfaces. In *Problems on mapping class groups and related topics*, volume 74 of *Proc. Sympos. Pure Math.*, pages 233–243. Amer. Math. Soc., Providence, RI, 2006.

[KM14] Sadayoshi Kojima and Greg McShane. Normalized entropy versus volume. 11 2014.

[KS00] Richard Kenyon and John Smillie. Billiards on rational-angled triangles. *Comment. Math. Helv.*, 75(1):65–108, 2000.

[Lon85] Darren D. Long. Constructing pseudo-Anosov maps. In *Knot theory and manifolds (Vancouver, B.C., 1983)*, volume 1144 of *Lecture Notes in Math.*, pages 108–114. Springer, Berlin, 1985.

[LR17] Christopher J. Leininger and Alan W. Reid. Personal communication, 2017.

[McM03] Curtis T. McMullen. Teichmüller geodesics of infinite complexity. *Acta Math.*, 191(2):191–223, 2003.

[MHO09] Kenneth R. Meyer, Glen R. Hall, and Dan Offin. *Introduction to Hamiltonian dynamical systems and the N-body problem*, volume 90 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2009.

[Pen88] Robert C. Penner. A construction of pseudo-Anosov homeomorphisms. *Trans. Amer. Math. Soc.*, 310(1):179–197, 1988.

[Sch82] Martin Scharlemann. The complex of curves on nonorientable surfaces. *J. London Math. Soc. (2)*, 25(1):171–184, 1982.

[Shi16] Hyunshik Shin. Algebraic degrees of stretch factors in mapping class groups. *Algebr. Geom. Topol.*, 16(3):1567–1584, 2016.

[SS15] Hyunshik Shin and Balázs Strenner. Pseudo-Anosov mapping classes not arising from Penner's construction. *Geom. Topol.*, 19(6):3645–3656, 2015.

[Str17] Balázs Strenner. Lifts of pseudo-Anosov homeomorphisms of nonorientable surfaces have vanishing SAF invariant. *To appear in Math. Res. Lett.*, 2017.

[Thu88] William P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2):417–431, 1988.

[Yaz17] Mehdi Yazdi. Personal communication, 2017.

[Zor06] Anton Zorich. Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pages 437–583. Springer, Berlin, 2006.