The Minimum Dilatation of Pseudo-Anosov 5-Braids
Ji-Young Ham and Won Taek Song

CONTENTS
1. Introduction
2. Folding Automata
3. Search for the Minimum Dilatation
4. Implementation
References

The minimum dilatation of pseudo-Anosov 5-braids is shown to be the largest zero \( \lambda_5 \approx 1.72208 \) of \( x^4 - x^3 - x^2 - x + 1 \), which is attained by \( \sigma_1\sigma_2\sigma_3\sigma_4\sigma_1\sigma_2 \).

1. INTRODUCTION

Let \( f: D^2 \to D^2 \) be an orientation-preserving disk homeomorphism that is the identity map on the boundary \( \partial D^2 \), and let \( \{p_i\} \subset \text{int}(D^2) \) be a periodic orbit of \( f \) (or more generally a finite set invariant under \( f \)). The points \( p_i \) move under an isotopy from the identity map on \( D^2 \) to \( f \). Their trajectory forms a geometric braid \( \beta \), a collection of strands in \( D^2 \times [0,1] \) connecting \( p_i \times 1 \) to \( f(p_i) \times 0 \) (see Figure 1). The isotopy class of \( \beta \) determines the homotopy class of \( f \) relative to \( \{p_i\} \cup \partial D^2 \) and vice versa. An \( n \)-braid refers to the isotopy class of a geometric braid with \( n \) strands. The set of \( n \)-braids forms the braid group \( B_n \).

From now on we consider \( f \) as a homeomorphism on a punctured sphere \( f: \text{int}(D^2) - \{p_i\} \to \text{int}(D^2) - \{p_i\} \). In particular, by forgetting the boundary \( \partial D^2 \), we ignore Dehn twists along \( \partial D^2 \) that do not affect the dynamics of the braid \( \beta \). In other words, we consider an \( n \)-braid \( \beta \) as a mapping class on an \((n+1)\)-times punctured sphere with the (so-called) boundary puncture fixed.

1.1 Topological Entropy

The topological entropy \( h_T(\beta) \) of the braid \( \beta \) is defined to be the infimum topological entropy of the disk homeomorphisms representing \( \beta \):

\[
h_T(\beta) = \inf_{g \simeq f} h_T(g).
\]

The topological entropy of a braid is a conjugacy invariant measuring the dynamical complexity of the braid. It is equal to the logarithm of the growth rate of the free group automorphism induced on \( \pi_1(D^2 - \{p_i\}) \). When \( \beta \) is represented by a pseudo-Anosov homeomorphism \( f \) with dilatation factor \( \lambda_f = \lambda(f) \), we have \( h_T(\beta) = \).
log $\lambda_f$. In this case the dilatation $\lambda(\beta)$ of the braid is also given by $\lambda(\beta) = \lambda_f$.

If $f$ is homotopic to a periodic homeomorphism, the braid $\beta$ is called periodic. If there is a collection of disjoint subsurfaces of $\text{int} (D^2) - \{p_i\}$ with negative Euler characteristics that is homotopically invariant under $f$, the braid $\beta$ is called reducible. Since we consider the dynamics of a periodic braid to be trivial, studying the dynamics of braids reduces to the maps on aperiodic irreducible components.

### 1.2 Pseudo-Anosov Homeomorphisms

By the Nielsen–Thurston classification of surface homeomorphisms [Thurston 88, Bers 78, Bestvina and Handel 95, Daskalopoulos and Wentworth 03], an aperiodic irreducible braid is represented by a pseudo-Anosov homeomorphism. A pseudo-Anosov homeomorphism has several nice extremal properties: It realizes the minimum topological entropy and the minimum quasi-conformality constant in its homotopy class. It also has the minimum number of periodic orbits for each period [Birman and Kidwell 82].

A surface homeomorphism $f: F \to F$ is called a pseudo-Anosov homeomorphism relative to a puncture set $\{p_i\} \subset F$ when the following conditions hold: First we need a singular flat metric on $F$ with a finite singularity set $\{q_j\}$ such that $\{p_i\} \subset \{q_j\}$. Each singularity $q_j$ has its cone angle in $\{k\pi \mid k \in \mathbb{Z}_{>0}\}$. If a singularity has cone angle $\pi$, it must be one of the puncture points $p_i$. The homeomorphism $f$ is required to preserve $\{q_j\}$ and to be locally affine (hyperbolic) on $F - \{q_j\}$ with a constant dilatation factor $\lambda_f > 1$. In particular, at a fixed point in $F - \{q_j\}$, the map $f$ is locally written as $\begin{bmatrix} \lambda_f & 0 \\ 0 & \lambda_f^{-1} \end{bmatrix}$.

Thus roughly speaking, if a surface homeomorphism $f$ represents an aperiodic irreducible mapping class, then one can simplify $f$ by pulling it tight everywhere until it becomes linear almost everywhere in an appropriate sense.

The horizontal directions to which $f$ expands by the factor $\lambda_f$ can be integrated to form one invariant measured foliation $\mathcal{F}^s$. The vertical directions perpendicular to $\mathcal{F}^s$ form the other invariant measured foliation $\mathcal{F}^u$. From a singularity $q_j$ with cone angle $k\pi$, $k$ singular leaves of $\mathcal{F}^s$ emanate. In this case, $q_j$ is called a $k$-prong singularity.

Note that in the above definition of pseudo-Anosov homeomorphism we can remove or add punctures while keeping the same map $f: F \to F$. When $\{f^j(x)\}$ is a periodic orbit of unpunctured points, puncturing at $\{f^j(x)\}$ refers to adding them to the puncture set $\{p_i\}$. Conversely, when $\{f^j(p_i)\}$ is a periodic orbit of $k$-prong punctured singularities for $k > 1$, capping them off refers to removing them from the puncture set. For pseudo-Anosov braids, puncturing or capping off corresponds to adding or removing some strands.

Let $\tilde{f}: \tilde{F} \to \tilde{F}$ be a lift of $f$ on a finite-fold cover $\tilde{F}$ of $F$ branched at some finite set of points invariant under $f$. Then by pulling back the flat metric on $F$ to $\tilde{F}$, the lift $\tilde{f}$ is also a pseudo-Anosov homeomorphism with the same dilatation factor $\tilde{\lambda}_f = \lambda_f$.

### 1.3 Train-Track Representative

Using a Markov partition (or its associated train-track representative), the flat metric and the pseudo-Anosov homeomorphism can be described quite concretely (see [Fathi et al. 79, Exposé 9] for the definition and see Figure 2 for an example). Let $\{R_i\}$ be a Markov partition for a pseudo-Anosov homeomorphism $f$. The transition matrix $M_f = (m_{ij})$ is defined by setting the $i,j$ entry $m_{ij}$ to be the number of times that $f(R_i)$ crosses over $R_j$.

The transition matrix $M_f$ is Perron–Frobenius: For some $k > 0$, each entry of $M_f^k$ is strictly positive. In particular, the largest eigenvalue of $M_f$ is real and has an eigenvector with strictly positive coordinates [Seneta 73, Theorem 1.1].

The widths $v_i$ and the heights $w_i$ of $R_i$ satisfy the equations

$$\lambda_f v_i = \sum_j m_{ij} v_j, \quad w_j = \frac{1}{\lambda_f} \sum_i m_{ij} w_i.$$
FIGURE 2. The boundary of the L-shaped region reads, counterclockwise from the puncture on the top, \( abb^{-1} cdd^{-1} e f f^{-1} g \) with \( ac = g^{-1} e^{-1} \). After side pairing, the L-shaped region becomes a four-times punctured sphere with a flat metric. The pseudo-Anosov homeomorphism maps each rectangle to a longer and thinner horizontal strip running over other rectangles. In the train-track representative, each edge is assigned tangential and transverse measures coming from the width and the height of the corresponding rectangle.

In particular, the dilatation factor \( \lambda_f \) appears as the eigenvalue of \( M_f \) whose eigenvector has strictly positive coordinates.

We use train-track representatives as a notational simplification for Markov partitions. As in Figure 2, each expanding edge of the invariant train track corresponds to a rectangle in the Markov partition. Once we know the transition matrix of the graph map, it is easy to recover the heights and widths of the rectangles.

1.4 Main Question
Let us consider the set \( \Lambda_{g,n} \) of the dilatation factors for pseudo-Anosov homeomorphisms on an \( n \)-times punctured genus-\( g \) surface \( F_{g,n} \):

\[
\Lambda_{g,n} = \{ \lambda_f \mid f : F_{g,n} \to F_{g,n} \text{ pseudo-Anosov homeomorphisms} \}
\]

Since we can bound the number of rectangles in Markov partitions using the Euler characteristic of the punctured surface, \( \Lambda_{g,n} \) consists of eigenvalues of Perron–Frobenius matrices with bounded dimension. In particular, the set \( \Lambda_{g,n} \) is discrete and has a minimum. Our current work is motivated by the following question:

Question 1.1. What is \( \min \Lambda_{g,n} \)?

The question asks for the simplest pseudo-Anosov homeomorphism on the surface.

A pseudo-Anosov homeomorphism \( f \) induces an isometry on the Teichmüller space equipped with the Teichmüller metric. The pair of invariant measured foliations \( (\mathcal{F}^s, \mathcal{F}^u) \) determines a geodesic axis in the Teichmüller space on which \( f \) acts as a translation by \( \log \lambda_f \). The axis projects down to a closed geodesic in the moduli space, which is the quotient of the Teichmüller space by the action of the mapping class group.

Conversely, any closed geodesic in the moduli space represents the conjugacy class of some pseudo-Anosov mapping class. Therefore, Question 1.1 can be rephrased as asking for the shortest closed geodesic in the moduli space.

The hyperbolic volume of the mapping torus is another natural complexity measure for a pseudo-Anosov homeomorphism. We notice that a pseudo-Anosov homeomorphism with small dilatation tends to have a mapping torus with small hyperbolic volume and vice versa.

The question for the minimum volume of the hyperbolic mapping tori on a given surface seems to be much more difficult than Question 1.1. In [Cao and Meyerhoff 01] the minimum volume for orientable cusped hyperbolic 3-manifolds is computed. An extensive use of computer programs is involved in its proof. In this paper we also use a computer program for the proof of the main theorem, but the algorithm and the actual code are much simpler than those of [Cao and Meyerhoff 01].

1.5 Related Results
The question of the minimum dilatation of pseudo-Anosov homeomorphisms on a given surface still remains largely unanswered since the Nielsen–Thurston classification of surface homeomorphisms. The existence of a Markov partition [Fathi et al. 79, Exposé 10] for a pseudo-Anosov homeomorphism implies that a dilatation should appear as the largest eigenvalue of a Perron–Frobenius
matrix of bounded dimension, hence in particular should be an algebraic integer. However, it is not clear how the restriction that the symbolic dynamical system dictated by a Perron–Frobenius matrix be from a homeomorphism on a given surface actually affects the possible values of entropy (the logarithm of the dilatation).

There are several known results relevant to this question of the minimum dilatation of pseudo-Anosov homeomorphisms. Penner [Penner 91] gives a lower bound $2^{1/(12g-12+4n)}$ for the dilatations on $F_{g,n}$ a genus-$g$ surface with $n$ punctures. In [Penner 91, Bauer 92, Brinkmann 04], pseudo-Anosov homeomorphisms on $F_{g,0}$ with small dilatations are constructed showing that the minimum dilatation on $F_{g,0}$ converges to 1 as the genus $g$ increases. Fehrenbach and Los [Fehrenbach and Los 99] compute a lower bound $(1+\sqrt{2})^{1/n}$ for the dilatations of pseudo-Anosov disk homeomorphisms (braids) that permute the punctures in one cycle. In [Song 05b], a lower bound $2+\sqrt{5}$ for the dilatations of pseudo-Anosov pure braids is given. A pseudo-Anosov disk homeomorphism is represented by a transitive Markov-tree map preserving the endpoint set of the tree with the same topological entropy. Baldwin [Baldwin 01] gives a lower bound log 3 for the topological entropy of transitive Markov-tree maps fixing each endpoint.

The exact values of the minimum dilatations are known only for a few simple cases. Zhirov [Zhirov 95] shows that if a pseudo-Anosov homeomorphism on $F_{2,0}$ has an orientable invariant foliation, then its dilatation is not less than the largest zero $\lambda_5$ of $x^4-x^3-x^2-x+1$, and he gives an example of a pseudo-Anosov homeomorphism realizing the dilatation $\lambda_5$.

The pseudo-Anosov 3-braid $\sigma_3\sigma_1^{-1}$ is shown to be the minimum in the forcing partial order among pseudo-Anosov 3-braids by Matsuoka [Matsuoka 85] and Handel [Handel 97]; hence it attains the minimum dilatation. The pseudo-Anosov 4-braid $\sigma_3\sigma_2\sigma_1^{-1}$ is claimed in [Song et al. 02] to have the minimum dilatation, but the proof given there unfortunately contains an error.

1.6 Outline

In this paper we prove the following theorem, giving at the same time a corrected proof of the minimality of the dilatation of $\sigma_3\sigma_2\sigma_1^{-1} \in B_4$.

**Theorem 1.2.** The 5-braid $\sigma_1\sigma_2\sigma_3\sigma_4\sigma_1\sigma_2$ attains the minimum dilatation of pseudo-Anosov 5-braids.

The dilatation of a pseudo-Anosov braid is invariant under several operations such as conjugation, compositing with a full twist, taking the inverse, and taking the reverse. It turns out that for braid indices 3 to 5, the pseudo-Anosov braids realizing the minimum dilatations are essentially unique, modulo the aforementioned operations. This could be just a coincidence. It would be a nice surprise if some uniqueness property were to be proved for the minimum-dilatation pseudo-Anosov braids.

The two main ingredients of the proof of Theorem 1.2 are the construction of folding automata for generating candidate pseudo-Anosov braids for the minimum dilatation, and the following lemma for bounding the word lengths of the candidate braids.

**Lemma 1.3.** If $M$ is a Perron–Frobenius matrix of dimension $n$ with $\lambda > 1$ its largest eigenvalue, then

$$\lambda^n \geq |M| - n + 1,$$

where $|M|$ denotes the sum of entries of $M$.

This lemma improves on [Papadopoulos and Penner 90, Theorem 6] and replaces the erroneous Lemmas 3 and 4 of [Song et al. 02].

Given a pseudo-Anosov homeomorphism $f: (F, \{p_i\}) \to (F, \{p_i\})$ on a surface $F$ with punctures $p_i$ with negative Euler characteristic $\chi(F - \{p_i\}) < 0$, there exists a train-track representative of $f$. There exists an invariant train track $\tau \subset F - \{p_i\}$ that carries $f(\tau)$. In particular, there is a splitting sequence

$$\tau = \tau_0 \succ \tau_1 \succ \cdots \succ \tau_k = f(\tau)$$

from $\tau$ to $f(\tau)$, where $\tau_j \succ \tau_{j+1}$ is an elementary splitting move.

By observing that there are only finitely many diffeomorphism types of the pair $(F - \{p_i\}, \tau_j)$, one can effectively construct a splitting automaton, which is a finite graph with train tracks as its vertices and with splitting moves as its arrows.

The existence of the train-track representative, in particular of the splitting sequence, implies that every pseudo-Anosov homeomorphism appears, up to conjugacy, as a closed path in some splitting automaton (see [Papadopoulos and Penner 87]). Papadopoulos and Penner [Papadopoulos and Penner 90, Theorem 6] also give a lower bound for the dilatation in terms of word length in automata.

In this paper we use folding automata as in [Song et al. 02], which are finite graphs with embedded train tracks.
as vertices and with elementary folding maps as arrows. An elementary folding map is an inverse of a splitting move.

If we are given an upper bound for the word length in terms of the dilatation, then on a fixed folding automaton, the search for the minimum dilatation in the automaton reduces to checking for finitely many closed paths.

Lemma 1.3, which is an improvement of [Papadopoulos and Penner 90, Theorem 6], not only gives an upper bound of the word lengths of mapping classes with dilatation bounded by a fixed number, but also trims out many branches that appear in the course of searching a big tree, namely the set of paths with bounded length. In fact, Lemma 1.3 implies that it suffices to consider only those paths with the property that every subpath has a transition matrix with bounded norm.

For the minimum dilatation of 5-braids, the previously mentioned restriction on paths by transition matrix norm and another restriction by Lemma 3.3 significantly reduce the number of candidate braids, making the computation feasible.

We think that the same method for computing the minimum dilatation would still work for a few simpler cases such as that of a genus-2 closed surface, although it would involve a more complicated computer-aided search.

2. FOLDING AUTOMATA

Given a pseudo-Anosov homeomorphism \( f : (F, \{ p_i \}) \rightarrow (F, \{ p_i \}) \) on a closed surface \( F \) with punctures \( \{ p_i \} \), there exists an invariant train track \( \tau \subset F - \{ p_i \} \), and \( f \) is represented by a train-track map \( f_\tau : \tau \rightarrow \tau \) [Papadopoulos and Penner 87].

A train track \( \tau \) is a smooth branched 1-manifold embedded in the surface \( F - \{ p_i \} \) such that each component of the complement \( F - \{ p_i \} - \tau \) is either a once-punctured \( k \)-gon for \( k \geq 1 \) or an unpunctured \( k \)-gon for \( k \geq 3 \). The train track \( \tau \) is called invariant under \( f \) if \( f(\tau) \) smoothly collapses onto \( \tau \) in \( F - \{ p_i \} \), inducing a smooth map \( f_\tau : \tau \rightarrow \tau \) that maps branch points to branch points. In this case one may repeatedly fold (or zip) \( f(\tau) \) near cusps to obtain a train track isotopic to \( \tau \) in \( F - \{ p_i \} \) (see Figure 3 and [Kleinberg and Menasco 98, Figures 4, 5]).

Let \( f_\tau : \tau \rightarrow \tau \) be a train-track representative of a pseudo-Anosov homeomorphism \( f \). An edge \( e \) of \( \tau \) is called infinitesimal if it is eventually periodic under \( f_\tau \), that is, \( f_\tau^{N+k}(e) = f_\tau^{N}(e) \) for some \( N, k > 0 \). An edge of \( \tau \) is called expanding if it is not infinitesimal.

An expanding edge \( e \) actually has a positive length in the sense that \( \lim_{N \rightarrow \infty} |f_\tau^N(e)|/\lambda_f^N \) is positive, where \( |\cdot| \) denotes the word length of a path, and \( \lambda_f = \lambda(f) \) denotes the dilatation factor for \( f \).

A graph map is called Markov if it maps vertices to vertices, and is locally injective at points that do not map into vertices. Given a Markov map \( g : \tau \rightarrow \tau' \), the transition matrix \( M_g = (m_{ij}) \) is defined by the condition that the \( j \)th edge \( (e_j')_{\pm 1} \) of \( \tau' \) occurs \( m_{ij} \) times in the path \( g(e_i) \), the image of the \( i \)th edge of \( \tau \). When \( \tau' = \tau \), the transition matrix is square, and considering its spectral radius makes sense.

The spectral radius of \( M_{f_\tau} \) equals the dilatation factor \( \lambda(f) \) for the pseudo-Anosov homeomorphism \( f \). Coordinates of the corresponding eigenvectors of \( M_{f_\tau} \) and its transpose \( M_{f_\tau}^T \) are tangential and transverse measures of edges of \( \tau \), which are projectively invariant under \( f \).

An elementary folding map \( \pi : \tau \rightarrow \tau' \) is a smooth Markov map between two train tracks \( \tau \) and \( \tau' \) such that for only one edge \( e \) of \( \tau \) does the image \( \pi(e) \) have word length 2, the other edges mapping to paths of length 1. In other words, the transition matrix \( M_{\pi} \) is of the form \( P + B \) for some permutation matrix \( P \) and for some elementary matrix \( B \).

When the train tracks are embedded in a surface, as in our case of concern, the pairs of edges that are folded should be adjacent in the surface: The two segments of \( \tau \) that are identified by the elementary folding map are two sides of an open triangle in \( F - \{ p_i \} - \tau \) (see Figure 3).

**Proposition 2.1.** A train-track representative \( f_\tau : \tau \rightarrow \tau \) of a surface homeomorphism \( f : (F, \{ p_i \}) \rightarrow (F, \{ p_i \}) \) admits a folding decomposition as follows:

\[
f_\tau = \rho \circ \pi_k \circ \cdots \circ \pi_1,
\]
where \( \pi_j : \tau_j \to \tau_{j+1} \) are elementary folding maps, \( \tau_1 = \tau_{k+1} = \tau \), and \( \rho : \tau \to \tau \) is an isomorphism induced by a periodic surface homeomorphism

\[
(F - \{p_i\}, \tau) \to (F - \{p_i\}, \tau)
\]

preserving \( \tau \).

Proof: The result follows from [Stallings 83]. See [Papadopoulos and Penner 87, Song et al. 02] for more details.

By observing that there are only finitely many possible diffeomorphism types for the pairs \( (F - \{p_i\}, \tau_j) \) appearing in the folding decomposition, we can construct folding automata. A folding automaton is a connected directed graph with diffeomorphism types of train tracks as vertices, with elementary folding maps and isomorphisms as arrows. See Figure 7 for a simplified version of a folding automaton. The train tracks in Figure 7 admit no nontrivial isomorphisms: that is, if \( h : (D_5, \tau) \to (D_5, \tau) \) is an orientation-preserving diffeomorphism fixing \( \tau \) in the automaton, then \( h \) is isotopic to the identity map. So in Figure 7 there are no arrows corresponding to isomorphisms.

Corollary 2.2. All train-track representatives of pseudo-Anosov homeomorphisms are represented by closed oriented paths in folding automata.

Each closed path based at a train track \( \tau \) in a folding automaton has an associated train-track representative \( f_\tau : \tau \to \tau \) of some homeomorphism \( f : (F, \{p_i\}) \to (F, \{p_i\}) \). If the disk homeomorphism \( f \) is pseudo-Anosov, it admits a train-track representative \( f_\tau \) whose transition matrix \( M_{f_\tau} \) is Perron–Frobenius (also called primitive) modulo infinitesimal edges: For some \( N > 0 \), the power \( M_{f_\tau}^N \) is strictly positive in the block of expanding edges. To find out whether \( M \) is Perron–Frobenius, it suffices by [Holladay and Varga 58, Wielandt 50], [Seneta 73, Theorem 2.8] to check whether \( M^{n^2 - 2n + 2} \) has all nonzero entries, where \( n \) is the dimension of the matrix \( M \).

Now we discuss simplifying the train-track maps so that we can restrict to simplified folding automata. If the pseudo-Anosov homeomorphism \( f \) fixes a distinguished puncture \( p_0 \), that is, \( f(p_0) = p_0 \) (for instance when \( f \) is from a disk homeomorphism and \( p_0 \) is the boundary puncture), then we can give a restriction to the train-track map \( f_\tau : \tau \to \tau \), thereby reducing the size of the folding automata needed in our computation.

We first assume that only the component of \( F - \tau \) containing \( p_0 \) has expanding edges on its sides: The other components of \( F - \tau \) not containing \( p_0 \) are bounded only by infinitesimal edges. If one is given a train-track representative \( f_\tau : \tau \to \tau \) not satisfying this assumption, one may apply a splitting operation [Bestvina and Handel 95, Section 5] near \( p_0 \) (when \( p_0 \) is enclosed only by infinitesimal edges), then apply a sequence of folding operations [Bestvina and Handel 92, p. 15], [Los 96, Section 2.2] near other punctures \( p_i \), \( i \neq 0 \), until all the components of the train-track complement not containing \( p_0 \) shrink to become infinitesimal, to obtain a new train-track representative satisfying the assumption [Bestvina and Handel 92, Proposition 3.3].

Applying some more folding operations (see Figure 4), we can also remove any cusp occurring between an expanding edge and an infinitesimal edge. We assume that cusps occur only at corners of infinitesimal multigons. If one is given a train-track representative with a cusp incident only to expanding edges not satisfying this assumption, one may apply a splitting operation at the cusp until the cusp hits an infinitesimal multigon (see Figure 5). Therefore a pseudo-Anosov braid has an invariant train track that is locally modeled by infinitesimal \( k \)-gons to which expanding edges are joined (possibly) forming cusps only between expanding edges (see Figure 6).

In this paper we use simplified versions of folding automata whose train tracks satisfy the previously given conditions, and each arrow is either an isomorphism or a composite of two elementary folding maps whereby one expanding edge and one infinitesimal edge are absorbed into another expanding edge. It is not hard to see that simplified folding automata also generate all the conjugacy classes of pseudo-Anosov homeomorphisms.
FIGURE 5. A splitting operation that shifts a cusp incident only to expanding edges.

FIGURE 6. Local models for train tracks in simplified folding automata.

FIGURE 7. A folding automaton for pseudo-Anosov 5-braids with two unpunctured 3-prong singularities.
In this paper our subject of interest is pseudo-Anosov homeomorphisms on a 5-times punctured disk \( D_5 \), or equivalently on a 6-times punctured sphere \( F_{0,6} \) with a distinguished boundary puncture.

We explain how to read Figure 7, which depicts a simplified version of a folding automaton. Each train track is embedded in a 5-times punctured disk, with each puncture enclosed by an infinitesimal monogon. Each embedding is chosen arbitrarily, and only the orientation-preserving diffeomorphism types of embedded train tracks count.

An arrow is a composite of two elementary folding maps, one involving an infinitesimal edge and another involving only expanding edges. We ignore the infinitesimal edges in computing the transition matrix because the occurrences of infinitesimal edges do not affect the resulting dilatation factor.

An arrow is drawn dashed if it induces a homeomorphism isotopic to the identity, and it is drawn solid otherwise. Note that a folding map \( \pi: \tau \rightarrow \tau' \) determines a disk homeomorphism

\[ f: D_5 \rightarrow D_5 \]

up to isotopy when the embeddings \( \tau \rightarrow D_5 \) and \( \tau' \rightarrow D_5 \) of the two train tracks are fixed. In particular, \( \tau' \circ f(\tau) \); that is, \( f(\tau) \) folds to be \( \tau' \), inducing the folding map \( \pi \).

To each solid arrow a braid word is assigned representing the associated disk homeomorphism.

Edges of a train track are numbered \( \{1, 2, \ldots, 6\} \) in such a way that in the peripheral word running clockwise from a cusp, new edges appear in increasing order. This naming of edges amounts to fixing a groupoid homomorphism from paths in the automaton to transition matrices; that is, for two paths \( \gamma \) and \( \delta \), \( M(\gamma \cdot \delta) = M(\gamma)M(\delta) \) if \( \gamma \) ends at the starting vertex of \( \delta \), where \( M(\gamma) \) denotes the transition matrix for \( \gamma \).

Each arrow is associated with a permutation \( i_1i_2i_3i_4i_5i_6 \) and a rule \( m \rightarrow n \), meaning that under the elementary folding map, the edge \( j \) maps to \( i_j \) for \( j \neq m \), and \( m \) maps to \( i_m \cdot n \). (Here we are concerned with only the transition matrix, so that the direction of edges and the order of concatenation are irrelevant.)

Given an adjacent pair \((e_1, e_2)\) of edges with a cusp between them, there are two possible folding maps: one such that the image of \( e_1 \) passes over that of \( e_2 \), and the other vice versa. Therefore, from each train track in Figure 7, two arrows of elementary folding maps emanate. Likewise, two arrows point toward each train track, because at each cusp two different elementary splittings are possible.

3. **Search for the Minimum Dilatation**

In this section we prove that the largest zero \( \lambda_5 \) of \( x^4 - x^3 - x^2 - x + 1 \) is indeed the minimum dilatation for pseudo-Anosov 5-braids.

The problem for the minimum dilatation reduces to a search in a finite set of closed paths in folding automata because by [Papadopoulos and Penner 90, Theorem 6] or by Lemma 1.3 the dilatation grows as the norm of the transition matrix, and there are only finitely many closed paths whose transition matrices have norm bounded by a given number. For instance, if a closed path in folding automata has length \( N \), then its associated transition matrix has norm at least \( N \).

We first restate and prove Lemma 1.3.

**Lemma 3.1.** If \( M \) is a Perron–Frobenius matrix of dimension \( n \) with \( \lambda > 1 \) its largest eigenvalue, then

\[ \lambda^n \geq |M| - n + 1, \]

where \( |M| \) denotes the sum of entries of \( M \).

**Proof:** Let \( M = (m_{ij}) \) and let \( (v_i) \) be the eigenvector given by the equation

\[ \lambda v_i = \sum_{j=1}^{n} m_{ij} v_j \]

for \( v_i > 0, 1 \leq i \leq n \).

The matrix \( M \) is the transition matrix of a graph \( G \) with vertex set \( V(G) = \{1, 2, \ldots, n\} \) such that the number of oriented edges from \( i \) to \( j \) is \( m_{ij} \). Let \( M^n = (k_{ij}) \). The number of paths with length \( n \) from \( i \) to \( j \) is \( k_{ij} \). For each pair \((i,j)\) of vertices there exists an oriented path from \( i \) to \( j \), since \( M \) is Perron–Frobenius.

Note that \((v_i)\) is also the eigenvector of \( M^n \) with eigenvalue \( \lambda^n \). Choose \( v_p = \min_i v_i \) the smallest coordinate of \((v_i)\):

\[ \lambda^n v_p = \sum_j k_{pj} v_j \geq \left( \sum_j k_{pj} \right) v_p \quad \text{since} \quad v_j \geq v_p. \]

The inequality \( \lambda^n \geq \sum_j k_{pj} \) reads that \( \lambda^n \) is not less than the number of length-\( n \) paths from the vertex \( p \) of \( G \).

Take a maximal positive tree \( T \subset G \) rooted at \( p \): Each vertex of \( G \) is connected to \( p \) by a unique oriented path in \( T \). Since \( T \) is maximal, \( |V(T)| = |V(G)| = n \), so that the number of edges of \( T \) is \( |E(T)| = n - 1 \).

If an oriented path \( \gamma \) from \( p \) has length \( n \), it must digress from \( T \) at some point. Define \( a(\gamma) = e \in E(G) - E(T) \) to be the first edge not in \( E(T) \) of \( \gamma \). This defines...
Lemma 3.3. Let $\gamma$ be a closed path in a folding automaton. Let $N,k > 0$ be numbers such that the transition matrices $M(\gamma^{N+i+k})$ and $M(\gamma^{N+i})$ have the same pattern and $M(\gamma^{N+i+k}) \geq M(\gamma^{N+i})$ for any $i \geq 0$. Then a closed path of the form $\alpha \cdot \gamma^{N+i+k} \cdot \delta$ in the folding automaton represents a pseudo-Anosov homeomorphism if and only if $\alpha \cdot \gamma^{N+i} \cdot \delta$ does. Furthermore, in this case we have the inequality

$$\lambda(\alpha \cdot \gamma^{N+i} \cdot \delta) \leq \lambda(\alpha \cdot \gamma^{N+i+k} \cdot \delta)$$

between their dilatation factors.

Proof: It suffices to prove the lemma for $\delta \cdot \alpha \cdot \gamma^{N+i+k}$ and $\delta \cdot \alpha \cdot \gamma^{N+i}$, since conjugation affects neither dilatation factor nor being pseudo-Anosov.

Since $M(\gamma^{N+i+k})$ and $M(\gamma^{N+i})$ have the same pattern, $M(\delta \cdot \alpha)M(\gamma^{N+i+k})$ and $M(\delta \cdot \alpha)M(\gamma^{N+i})$ also have the same pattern. In particular, one is Perron–Frobenius if and only if the other is, which proves the first claim of the lemma.

From

$$M(\gamma^{N+i}) \leq M(\gamma^{N+i+k})$$

we have

$$M(\delta \cdot \alpha \cdot \gamma^{N+i}) \leq M(\delta \cdot \alpha \cdot \gamma^{N+i+k}),$$

which by [Seneta 73, Theorem 1.1(e)] implies the inequality of the lemma.

Remark 3.4. Let $\gamma, N, k$ be given as in Lemma 3.3. Then the lemma implies that when we search just for the minimum dilatation factor for pseudo-Anosov homeomorphisms, it suffices to search in the set of paths that do not contain $\gamma^{N+k}$ as a subpath.

In the search in the automaton in Figure 7, we exclude paths containing several closed paths, for example

$$\begin{pmatrix} 123564 & 123456 \end{pmatrix}^6 \begin{pmatrix} 1 \to 4 & 4 \to 1 \end{pmatrix}^6 \begin{pmatrix} 312456 \end{pmatrix}^6 \begin{pmatrix} 4 \to 1 & 4 \to 3 \end{pmatrix}^6,$$

and second iterates of length-1 loops. This reduces the size of the set of candidate braids for minimum dilatation to the extent that the computation in the proof of Theorem 3.5 becomes possible on a personal computer.

Theorem 3.5. If a pseudo-Anosov 5-braid has an invariant foliation with two unpunctured 3-prong singularities, then its dilatation is not less than the largest zero $\approx 2.01536$ of $x^5 - x^4 - 4x^3 - x + 1$.

Proof: It is easy to check that there are only nine different diffeomorphism types of train tracks in $D_5$, locally modeled by infinitesimal multigons with outgoing expanding-edge legs as in Figure 6. By computing the elementary folding maps among them (more precisely composites of two elementary folding maps, one of them involving an infinitesimal edge), we have a folding automaton as depicted in Figure 7.

By a computer-aided search [Song 05a] in the set of paths $\gamma$ with $|M_\gamma| \leq 2.02^6 + 5 < 73$, we conclude that up to conjugacy and multiplication by central elements, $\sigma_3 \sigma_1^{-1} \sigma_2^{-1}$ with dilatation $\approx 2.01536$ is the only such pseudo-Anosov 5-braid with dilatation less than 2.02.
Lemma 3.6. If a pseudo-Anosov 5-braid has an invariant foliation with an unpunctured 4-prong singularity, then its dilatation is not less than the largest zero, \( \approx 2.15372 \), of \( x^4 - 3x^3 + 3x^2 - 3x + 1 \).

Proof: The folding automaton for this case is similar to the one in Figure 8. As in the proof of Theorem 3.2, a computer-aided search [Song 05a] in the set of closed paths up to length \( 56 > 2.2^3 + (5 - 1) \) shows that the largest zero of \( x^4 - 3x^3 + 3x^2 - 3x + 1 \) is the minimum dilatation factor for a pseudo-Anosov braids with three prongs at the boundary puncture. We can assume that the punctured 3-prong singularity is the boundary puncture, since it should be fixed by the homeomorphism \( f \). Now we use the folding automaton that generates such pseudo-Anosov braids with three prongs at the boundary puncture.

There are eleven diffeomorphism types of train tracks to consider for this case (see Figure 9). There are fifty arrows in the automaton, which are too many to be drawn in a figure in this paper. See [Song 05a] for details.

By the same kind of computer-aided search as before, in the set of closed paths in the folding automaton up to length \( 12 \geq (\lambda_5)^4 + (4 - 1) \), we conclude that \( \lambda_5 \) is the minimum dilatation factor for pseudo-Anosov braids in this automaton.

The dilatation is achieved by \( \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 \). □

Lemma 3.7. If a pseudo-Anosov 5-braid has an invariant foliation with a punctured 3-prong singularity, then its dilatation is not less than \( \lambda_3 \), the largest zero of \( x^4 - x^3 - x^2 - x + 1 \).

Proof: Let \( f: F_{0.6} \rightarrow F_{0.6} \) be a pseudo-Anosov homeomorphism with an invariant foliation \( \mathcal{F} \). If the invariant measured foliation \( \mathcal{F} \) on a 6-times punctured sphere has a punctured 3-prong singularity, then it has five other punctured 1-prong singularities and no more.

We can assume that the punctured 3-prong singularity is the boundary puncture, since it should be fixed by the homeomorphism \( f \). Now we use the folding automaton that generates such pseudo-Anosov braids with three prongs at the boundary puncture.

There are eleven diffeomorphism types of train tracks to consider for this case (see Figure 9). There are fifty arrows in the automaton, which are too many to be drawn in a figure in this paper. See [Song 05a] for details.

By the same kind of computer-aided search as before, in the set of closed paths in the folding automaton up to length \( 12 \geq (\lambda_5)^4 + (4 - 1) \), we conclude that \( \lambda_5 \) is the minimum dilatation factor for pseudo-Anosov braids in this automaton.

The dilatation is achieved by \( \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 \). □

Remark 3.8. In [Zhirov 95], \( \lambda_5 \) is proved to be the minimum dilatation factor for a pseudo-Anosov homeomorphism with an orientable invariant foliation on a closed genus-2 surface. The proof in [Zhirov 95] seems to have a gap. To complete the proof one needs to show that the golden ratio \((1 + \sqrt{5})/2 \approx 1.61803\), the largest zero of \( x^4 - 3x^2 + 1 \), cannot be a dilatation factor for such a pseudo-Anosov homeomorphism. In [Franks and Rykken 99], it is proved that such a pseudo-Anosov homeomorphism with quadratic dilatation factor is a lift of an Anosov homeomorphism via a branched covering. Lemmas 3.6 and 3.7 follow from [Zhirov 95] by taking double covers branched at odd-prong singularities.

By collecting all the results, we conclude this section by restating and proving the main theorem, Theorem 1.2:

Theorem 3.9. The 5-braid \( \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 \) attains the minimum dilatation of pseudo-Anosov 5-braids.

Proof: Let \( f: F_{0.6} \rightarrow F_{0.6} \) be a pseudo-Anosov homeomorphism on a 6-times punctured sphere with a punctured point fixed by \( f \). Let \( \mathcal{F} \) be its invariant measured foliation. Since \( \mathcal{F} \) has exactly six punctures, the formula \( 2 = \chi(F_{0.6}) = \sum_k (1 - k/2)n_k \), where \( n_k \) denotes the number of \( k \)-prong singularities, says that no singularity of \( \mathcal{F} \) can have more than four prongs.

Here is a list of possible types of \( \mathcal{F} \) according to its singularity type:

1. six punctured 1-prong singularities and one unpunctured 4-prong singularity: \( n_1 = 6, \ n_4 = 1 \);
2. six punctured 1-prong singularities and two unpunctured 3-prong singularities: \( n_1 = 6, \ n_3 = 2 \);
3. five punctured 1-prong singularities and one punctured 3-prong singularity: \( n_1 = 5, \ n_3 = 1 \);
4. five punctured 1-prong singularities, one punctured 2-prong singularity, and one unpunctured 3-prong singularities: \( n_1 = 5, \ n_2 = 1, \ n_3 = 1 \);
5. four punctured 1-prong singularities and two punctured 2-prong singularities: \( n_1 = 4, \ n_2 = 2 \).

Case 5 is, by capping off 2-prong singularity punctures, that of a pseudo-Anosov homeomorphism on a four-times
FIGURE 9. Train tracks for pseudo-Anosov 5-braids with a 3-pronged boundary puncture.

punctured sphere, which lifts to an Anosov homeomorphism on a torus via branched double covering. Therefore in this case, \( \lambda(f) \geq (3 + \sqrt{5})/2 > \lambda_5 \).

Case 4 reduces to case 3 by capping off the punctured 2-prong singularity and puncturing at the 3-prong singularity.

For cases 1 and 3 we have \( \lambda(f) \geq \lambda_5 \) by Lemmas 3.6 and 3.7 or by Remark 3.8.

Finally, case 2 is covered by Theorem 3.5, so that we have \( \lambda(f) > 2.01 > \lambda_5 \). In fact, this is the only part of the proof that actually requires a computer-aided search if one uses Zhirov’s result [Zhirov 95].

Collecting all of these, we conclude that \( \lambda(f) \geq \lambda_5 \approx 1.72208 \). It is easily checked that \( \beta = \sigma_1\sigma_2\sigma_3\sigma_4\sigma_1\sigma_2 \) realizes this dilatation \( \lambda(\beta) = \lambda_5 \).

4. IMPLEMENTATION

To search for the minimum-dilatation pseudo-Anosov homeomorphism on a given surface, we first need to generate a collection of folding automata. For 5-braids it is possible to build the necessary folding automata manually. On surfaces with more punctures and greater genus, we also need a computer program, genauto, to generate the folding automata. This paper will not cover the details of its implementation. The following is pseudocode for genauto.

Algorithm 4.1. genauto

input: Genus \( g \) and the number of punctures \( n \).
output: Folding automata on \( F_{g,n} \).

step 1. Generate the finite set of diffeomorphism types of embedded train tracks \( \tau_i \subset F_{g,n} \).

step 2. For each \( \tau_i \), compute all the elementary folding maps \( f_{ij} : \tau_i \to \tau'_{ij} \) from \( \tau_i \) if any. Compute isomorphisms \( h_{ij} : \tau'_{ij} \to \tau_k \) from the train track \( \tau'_{ij} \) to one in the set \( \{\tau_i\} \).

step 3. For each \( \tau_i \), compute the isomorphisms \( g_{il} : \tau_i \to \tau_i \) if any.

step 4. The elementary folding maps \( h_{ij} \circ f_{ij} \) and the isomorphisms \( g_{il} \) form the arrows of the folding automata. Compute their transition matrices after labeling each edge of all the train tracks \( \tau_i \).

Note that for step 1 one needs to solve the isomorphism problem for embedded train tracks. Once step 1 is done, implementing the other steps is more straightforward.

By running genauto, we obtain the folding automata as a collection of connected directed graphs with each arrow labeled by a transition matrix. The goal is to enumerate in the folding automata all the closed paths representing pseudo-Anosov mapping classes with an upper bound for the dilatation.

In this paper we deal with 5-braids using simplified folding automata. We ignore infinitesimal edges when computing transition matrices. Therefore a pseudo-Anosov braid is represented by a closed path in a folding automaton whose associated transition matrix is Perron–Frobenius.
The following is pseudocode for our program \texttt{fbrmin}. See [Song 05a] for details.

**Algorithm 4.2. \texttt{fbrmin}**

**input:** A directed graph $\mathcal{G}$ with arrows labeled by transition matrices, an upper bound $\lambda$ for the minimum dilatation, and a set $\mathcal{W}$ of subwords that are to be avoided during the search.

**output:** The list of closed paths in $\mathcal{G}$ representing pseudo-Anosov braids with dilatation less than $\lambda$.

**step 1.** Set $\text{maxnorm} = |\lambda^n + n - 1|$, where $n$ is the dimension of the transition matrices, and set $\text{archive} = \emptyset$, in which closed paths with small dilatation are to be stored.

**step 2.** Set $\text{tmp}_1$ to be the set of length-one paths in $\mathcal{G}$.

**step 3.** For each $i$ from 2 to $\text{maxnorm}$,

a. Compute $\text{childrenpaths}_i$ by appending paths in $\text{tmp}_{i-1}$ to paths in $\text{tmp}_{i-1}$, in all the ways possible in $\mathcal{G}$.

b. Compute $\text{tmp}_i$, the subset of $\text{childrenpaths}_i$ consisting of paths $\beta$ without any subword from the avoided-word set $\mathcal{W}$, with transition matrix $M_\beta$ such that $|M_\beta| \leq \text{maxnorm}$, and $M_\beta$ has at least one row and one column whose row (column) sum is less than 3.

c. Take the subset $\text{selectedcans}_i$ of $\text{tmp}_i$ consisting of closed paths representing pseudo-Anosov braids with dilatation less than $\lambda$, and append it to $\text{archive}$.

**step 4.** Return $\text{archive} = \cup_i \text{selectedcans}_i$.

In step 3b, we use Lemmas 1.3 and 3.3 to trim out much of unnecessary computation (see Remark 3.4).

When the row sums of a transition matrix $M_\beta$ all exceed 3, then the spectral radius of $M_\beta$ is greater than 3. In this case the same holds for every transition matrix of the form $M_{\beta \cdot \gamma} = M_\beta M_\gamma$, since $M_\gamma \geq P$ for some permutation matrix $P$. Therefore, since we are looking for transition matrices with spectral radius less than 3, we can safely remove such paths $\beta \cdot \gamma$ from consideration, as done in step 3b.

For computational aspects, the proof of Theorem 3.5 using the automaton in Figure 7 is the main part, which consumes most of the time and memory. On a 2.40-GHz machine, it took 1000 seconds of time and 150 megabytes of memory. During the search, each of roughly 85,000 matrices was actually tested for its largest eigenvalue.

We do not know how far the same kind of computation would work for more-complicated surfaces. We expect that at least the case for 6-braids, hence for genus-2 closed surfaces, can be done on a personal computer without too much difficulty.

**REFERENCES**


Ji-Young Ham, Department of Mathematics, University of California Santa Barbara, CA 93106 (ham@mail.math.ucsb.edu)

Won Taek Song, School of Engineering, Information and Communications University, 119 Munjiro, Yuseong-gu, Daejeon 305-732, Korea (cape@icu.ac.kr)

Received August 8, 2005; accepted in revised form July 21, 2006.