Research Experience for Undergraduates Program
Research Reports

Indiana University, Bloomington

Summer 2016

Contents

Preface

During the summer of 2016 five students participated in the Undergraduate Research Experience program in Mathematics at Indiana University. This program was sponsored by Indiana University and the Department of Mathematics. The program ran for eight weeks, from June 6 through July 31, 2016. Five faculty served as research advisers to the students from Indiana University:

- Chris Connell worked with Emma Brissett.
- Elizabeth Housworth worked with Kees McGahan.
- Chris Judge worked with Yvonne Chazal.
- Jeffrey Meier worked with James Dix.
- Dylan Thurston worked with Baptiste Dejean, Christian Gorski and Emily Tumber.
- Matthias Weber worked with Daniel Freese.

Following the introductory pizza party, students began meeting with their faculty mentors and continued to do so throughout the next eight weeks. The students also participated in a number of social events and educational opportunities and field trips.

Individual faculty gave talks throughout the program on their research, about two a week. Students also received LaTeX training in a series of workshops. Other opportunities included the option to participate in a GRE and subject test preparation seminar. Additional educational activities included tours of the library, the Slocum puzzle collection at the Lilly Library and the IU cyclotron facility, and self guided tours of the art museum. Students presented their work to faculty mentors and their peers at various times. This culminated in their presentations both in poster form and in talks at the statewide Indiana Summer Undergraduate Research conference which we hosted at the Bloomington campus of IU.

On the lighter side, students were treated to a reception by the graduate school as well as the opportunity to a fun filled trip to a local amusement park. They were also given the opportunity to enjoy a night of "laser tag" courtesy of the Department of Mathematics.

The summer REU program required the help and support of many different groups and individuals to make it a success. We foremost thank the Indiana University and the Department of Mathematics for major financial support for this bridge year between two National Science Foundation grants. We especially thank our staff member Mandie McCarty for coordinating the complex logistical arrangments (housing, paychecks, information packets, meal plans, frequent shopping for snacks). Additional logistical support was provided by the Department of Mathematics and our chair, Elizabeth Housworth. We are in particular thankful to Jeff Taylor for the computer support he provided. Thanks to those faculty who served as mentors and those who gave lectures. Thanks to David Baxter of the Center for Exploration of Energy and Matter (nee IU cyclotron facility) for his personal tour of the LENS facility and lecture. Thanks to Andrew Rhoda for his tour of the Slocum Puzzle Collection.

Chris Connell
September, 2016

Figure 1: REU Participants, from left to right: Kees McGahan, Emma Brissett, Daniel Freese, Yvonne Chazal, Bat DeJean, James Dix, Christian Gorski.

Geometry of Hyperbolic Percolation Clusters

Emma Brissett
Advisor: Chris Connell

September 14, 2016

Abstract

Given an infinite connected graph G, we can perform a Bernoulli bond percolation on G in the following way: for each edge in G, with probability p, the edge remains, and with probability $1-p$, the edge is removed. We are left with a random subgraph of G containing a combination of finite and infinite components called clusters. Every infinite connected graph has constants p_c and p_u in $[0,1]$ which are threshold probability values with respect to the number of infinite clusters in the percolation subgraph. Of special interest are percolations on hyperbolic graphs. There are many results concerning the number of infinite clusters in percolation subgraphs of planar hyperbolic graphs, as well as p_c and p_u values for such graphs, and we extend these to include nonplanar hyperbolic graphs. From there, we study the asymptotic behavior of infinite clusters in percolations on Gromov hyperbolic graphs. Finally, we calculate p_c bounds for radially homogeneous graphs.

1 Introduction

An *invariant percolation* on a graph G is a probability measure on the space of subgraphs of G, which is invariant under the automorphisms of G. The connected components of an invariant percolation are called *clusters*. This report will focus on Bernoulli(p) bond percolation, which is the random subgraph of G where all vertices $V(G)$ are included and each edge is in the subgraph with probability p, independently. Let $p_c = p_c(G)$ be the infimum of the set of $p \in [0,1]$ such that the Bernoulli(p) percolation on G has an infinite cluster a.s. We call p_c the critical parameter. Likewise, let $p_u = p_u(G)$ be the infimum of the set of $p \in [0,1]$ such that the Bernoulli(p) percolation on G has a unique infinite cluster a.s.

Let $\mathrm{Aut}(G)$ be the group of automorphisms of G. G is *transitive* if $\mathrm{Aut}(G)$ acts transitively on the vertices $V(G)$. G is *quasi-transitive* if the action of $\mathrm{Aut}(G)$ on $V(G)$ has finitely many orbits. G is *unimodular* if $\mathrm{Aut}(G)$ is unimodular, meaning the left-invariant Haar measure is also right-invariant. G has *one end* if for every finite set of vertices $V_0 \in V(G)$, there is exactly one infinite connected

component of G/V_0. G is *discrete* if $\mathrm{Aut}(G)$ is discrete. Note that discrete graphs are unimodular. G is *amenable* if there exists a sequence of sets $F_n \subset G$ such that

$$\lim_{n\to\infty} \frac{|\partial F_n|}{|F_n|} = 0.$$

G is a *nonelementary* hyperbolic graph if there is an infinite number of points in the ideal boundary $\partial_\infty G$. Note that transitive, nonelementary, discrete, hyperbolic graphs are nonamenable. Without transitivity, a nonelementary, discrete, hyperbolic graph could be made amenable by attaching an amenable graph to a vertex. Of particular interest are nonelementary, discrete, Gromov hyperbolic graphs. G is *Gromov hyperbolic* if all triangles in G are δ-thin. We call a triangle δ-thin if there exists a $\delta > 0$ such that each side lies in the δ-thickening of the other two sides. For a Gromov hyperbolic graph containing points x, y and z, we define the *Gromov product* of y and z as follows:

$$(y,z)_x = \frac{1}{2}(d(x,y) + d(x,z) - d(y,z)).$$

As an introduction to our project, we will examine Bernoulli percolation in the tree case and demonstrate our methods for finding p_c and p_u for T_4.

2 Percolation on T_4

T_4 is the 4-regular tree and the Cayley graph of the free group on two generators. Like the graphs we will study, T_4 is transitive, unimodular, nonamenable, and Gromov-hyperbolic. One key difference is T_4 has infinite ends, while the graphs we will study have one end.

Let ω be the Bernoulli bond percolation on T_4 and let $S(x,n)$ be the sphere with center x and radius n. We know the probability of not reaching $S(x,1)$ is the probability that all four edges are removed: $(1-p)^4$. Therefore, the probability of reaching $S(x,1)$ is $1-(1-p)^4$. The expected number of vertices reached in $S(x,1)$ is $4p$. For example, if $p = \frac{2}{3}$, we expect that $4\left(\frac{2}{3}\right) = \frac{8}{3}$ vertices are reached.

Given that we reached $4p$ vertices in $S(x,1)$, the probability of reaching $S(x,2)$ is $4p(1-(1-p)^3)$, since each vertex in $S(x,1)$ has three outward edges. Then, the expected number of vertices reached in $S(x,2)$ is $4p(3p)$. Continuing on, each vertex will always have three outward edges, so, given that we reached $S(x,n-1)$, the expected number of vertices reached in $S(x,n)$ is $4p(3p)^{n-1}$. Let Z_n be the expected number of vertices reached in $S(x,n)$. Then, $p_c = \frac{1}{3}$, since for $p < \frac{1}{3}$, as $n \to \infty$, $Z_n \to 0$, and for $p > \frac{1}{3}$, as $n \to \infty$, $Z_n \to \infty$.

Lemma 1. *If G is a transitive graph with more than one end, then $p_u(G) = 1$.*

Proof. Suppose for $p_u < p < 1$ there is a unique infinite cluster $\mathcal{C} \in G$. By transitivity, a sequence in $\mathcal{C}$ would be equally likely to occur in any end, so

since $\mathcal{C}$ is unique, it must have multiple ends. However, by the definition of ends, there exists a finite subgraph $T \in G$ such that G/T has disconnected ends. Since $p < 1$, there is a positive probability that T will be removed, which contradicts the assumption that $\mathcal{C}$ has multiple ends, so $p_u = 1$.

□

Since T_4 is transitive and has an infinite number of ends, $p_u(T_4) = 1$. That is, T_4 only has a unique infinite cluster when no edges are removed.

Note that calculating conditional expectation in the tree case is simple, since there are single inward edges and a constant number of outward edges. Some of the graphs we will study include multiple inward and lateral edges, which make expectation calculations more difficult. First, however, we will extend several theorems from [3] to include nonplanar graphs.

3 The Number of Infinite Clusters

The following theorem and proof follow results from Theorem 3.3 in [5].

Theorem 2. *Each rotation system P on a locally finite graph G determines a cellular embedding of G into some oriented surface S such that: i) the 2-cells are in a bijective correspondence with cycles of $P^{-1}r$ and ii) the rotation system of this embedding is equal to P.*

In other words, every graph G of finite degree embeds in a surface such that its cycles surround a disk. Let $\mathcal{K}(G)$ be the 2-skeleton of G for such an embedding. Then, every edge bounds exactly two faces, and we can define the dual $G^\dagger$ as follows: let every vertex $v^\dagger$ of $G^\dagger$ lie in the corresponding face of $\mathcal{K}(G)$ and every edge $e \in E(G)$ intersect only the dual edge $e^\dagger \in E(G^\dagger)$, and only in one point. Such a dual graph will maintain the properties necessary to extend several theorems from [3] to nonplanar graphs.

The following theorem and proof follow results from Theorem 1 in [6].

Theorem 3. *For the random variable N_0 denoting the number of infinite components in a percolation, a.s. $N_0 = 0, 1,$ or ∞.*

Proof. Let Ω be a configuration space $\{0,1\}^{\mathbb{G}}$, P be a translation ergodic probability measure on Ω, and ω be a configuration in Ω. The random variable N_0 and the events $\{N_0 = k\}$ are invariant under any shift T_j; therefore by ergdocity there is some $k_0 = 0, 1, 2, \ldots,$ or ∞ such that $P(N_0 = k_0) = 1$. Suppose $k_0 \neq 0$, 1, or ∞; we wish to obtain a contradiction. Let W_n be the event that $N_0 = k_0$ and each one of the k_0 infinite clusters has nonzero intersection with the ball $B(1, n)$; then $P(W_n) \to P(N_0 = k_0) = 1$ and so for some m, $P(W_m) > 0$.

Let $V = B(1,m)$ and define $\Psi : \Omega \to \Omega$ by

$$(\Psi\omega)_v = \begin{cases} 1, & v \in B(1,m) \\ w_v, & v \notin B(1,m) \end{cases}$$

So, Ψ is the transformation which makes all sites in $B(1,m)$ occupied and leaves all other sites unchanged. By Proposition 9, $P(\Psi(W_m)) > 0$, but clearly $\Psi(W_m) \subset \{N_0 = 1\}$, thus $P(N_0 = 1) > 0$, which contradicts the supposition that $k_0 \neq 0$, 1, or ∞.

□

Let ω be the Bernoulli percolation on G. Then $\omega^\dagger$ will denote the set

$$\omega^\dagger := \{e^\dagger : e \notin \omega\}.$$

Let k be the number of components in ω and $k^\dagger$ be the number of components in $\omega^\dagger$. Since they are both percolations, we know by Theorem 3 that

$$(k, k^\dagger) \in \{(0,0), (0,1), (0,\infty), (1,0), (1,1), (1,\infty), (\infty,0), (\infty,1), (\infty,\infty)\}.$$

Once we extend several theorems from [3] to include nonplanar graphs, we will be able to use these results to narrow down the possibilities for $(k, k^\dagger)$.

The following lemma and proof follow results from Lemma 3.3 in [3]. Note that our definition of dual implies $(G^\dagger)^\dagger = G$, a property used in the following proof. Let K be a finite component in ω and let $\partial_E K$ be the set of edges not in K with one or two vertices in K. Let $K' \subset G^\dagger$ be the set of edges dual to the edges in $\partial_E K$, and let $K'' \subset \{(G^\dagger)^\dagger = G\}$ be a component enclosing $K' \subset G^\dagger$. The following lemma also uses a result from Theorem 2.4 in [1] that says when the expected degree $\mathrm{E}[\deg_\omega v]$ of a vertex v in an invariant percolation on a unimodular nonamenable graph G is sufficiently close to $\deg_G v$, there are infinite clusters in ω with positive probability.

Lemma 4. *Let G be a transitive, nonamenable graph with one end. Let ω be an invariant bond percolation on G. If ω has only finite components a.s., then $\omega^\dagger$ has infinite components a.s.*

Proof. Suppose that both ω and $\omega^\dagger$ have only finite components a.s. Then a.s. given a component K of ω, there is a unique component K' of $\omega^\dagger$ that surrounds it. Similarly, for every component K of $\omega^\dagger$, there is a unique component K' of ω that surrounds it. Let K_0 denote the set of all components of ω. Inductively, set

$$K_{j+1} := \{K'' : K \in K_j\}.$$

For $K \in K_0$ let $r(K) := \sup\{j : K \in K_j\}$ be the *rank* of K, and define $r(v) := r(K)$ if K is the component of v in ω. For each r let ω^r be the set of

edges in $E(G)$ incident with vertices $v \in V(G)$ with $r(v) \leq r$. Then ω^r is an invariant bond percolation and

$$\lim_{r\to\infty} \mathrm{E}[\deg_{\omega^r} v] = \deg_G v.$$

Consequently, by the above result from [1], we find that ω^r has with positive probability infinite components for all sufficiently large r. This contradicts the assumption that ω and $\omega^\dagger$ have only finite components a.s.

□

The following corollary and proof follow results from Corollary 3.6 in [3].

Corollary 5. *Let G be a transitive, nonamenable graph with one end. Let ω be an invariant percolation on G. Suppose that both ω and $\omega^\dagger$ have infinite components a.s. Then, a.s. at least one among ω and $\omega^\dagger$ has infinitely many infinite components.*

Proof. Draw G and $G^\dagger$ in the plane in such a way that every edge e intersects $e^\dagger$ in one point, v_e, and there are no other intersections of G and $G^\dagger$. This defines a new graph $\hat{G}$, whose vertices are $V(G) \cup V(G^\dagger) \cup \{v_e : e \in E(G)\}$. Note that $\hat{G}$ is quasi-transitive. Set

$$\hat{\omega} := \{[v, v_e] \in E(\hat{G}) : v \in V(G), e \in \omega\} \cup \{[v^\dagger, v_e] \in E(\hat{G}) : v^\dagger \in V(G^\dagger), e \notin \omega\}.$$

Then $\hat{\omega}$ is an invariant percolation on $\hat{G}$. Note that the number of infinite components of $\hat{\omega}$ is the number of infinite components of ω plus the number of infinite components of $\omega^\dagger$. By Theorem 3 applied to $\hat{\omega}$, we find that $\hat{\omega}$ has infinitely many infinite components.

□

The following theorem and proof follow results from Theorem 3.1 in [3].

Theorem 6. *Let G be a transitive, nonamenable graph with one end. Let ω be an invariant percolation on G. Let k be the number of infinite components of ω, and let $k^\dagger$ be the number of infinite components of $\omega^\dagger$. Then a.s.*

$$(k, k)^\dagger \in \{(1, 0), (0, 1), (1, \infty), (\infty, 1), (\infty, \infty)\}.$$

Proof. Each of k, $k^\dagger$, is in $\{0, 1, \infty\}$ by Theorem 1. The case $(k, k^\dagger) = (0, 0)$ is ruled out by Lemma 4. The case $(k, k)^\dagger = (1, 1)$ is ruled out by Corollary 5. Since every two infinite components of ω must be separated by some component of $\omega^\dagger$, the situation $(k, k)^\dagger = (\infty, 0)$ is impossible. The same reasoning shows that $(k, k)^\dagger = (0, \infty)$ cannot happen.

□

The following theorem and proof follow results from Theorem 3.7 in [3].

Theorem 7. *Let G be a transitive, nonamenable graph with one end, and let ω be the Bernoulli bond percolation on G. Let k be the number of infinite components of ω, and let $k^\dagger$ be the number of infinite components of $\omega^\dagger$. Then a.s.*

$$(k, k^\dagger) \in \{(1,0), (0,1), (\infty, \infty)\}.$$

Proof. By Theorem 6, it is enough to rule out the cases $(1, \infty)$ and $(\infty, 1)$. Let K be a finite connected subgraph of G. If K intersects two distinct infinite components of ω, then $\omega^\dagger - \{e^\dagger : e \in E(K)\}$ has more than one infinite component. If $k > 1$ with positive probability, then there is some finite subgraph K such that K intersects two infinite components of ω with positive probability. Therefore, we find that $k^\dagger > 1$ with positive probability (since the distribution of $\omega^\dagger - \{e^\dagger : e \in E(K)\}$ is absolutely continuous to the distribution of $\omega^\dagger$). By ergodicity, this gives $k^\dagger > 1$ a.s. An entirely dual argument shows that $k > 1$ a.s. when $k^\dagger > 1$ with positive probability.

□

The following theorem and proof follow results from Theorem 3.8 in [3].

Theorem 8. *Let G be a transitive, nonamenable graph with one end. Then $p_c(G^\dagger) + p_u(G) = 1$ for Bernoulli bond percolation.*

Proof. Let ω_p be Bernoulli(p) bond percolation on G. Then $\omega_p^\dagger$ is Bernoulli($1-p$) bond percolation on $G^\dagger$. It follows from Theorem 7 that the number of infinite components $k^\dagger$ of $\omega^\dagger$ is 1 when $p < p_c(G)$, ∞ when $p \in (p_c(G), p_u(G))$, and 0 when $p > p_u(G)$.

□

The following theorem is Theorem 1.3 from [1].

Theorem 9. *Let G be a transitive, unimodular, nonamenable graph. Then a.s. critical Bernoulli bond percolation on G has no infinite components.*

The following theorem and proof follow results from Theorem 1.1 in [3].

Theorem 10. *Let G be a transitive, unimodular, nonamenable graph with one end. Then $0 < p_c(G) < p_u(G) < 1$, for Bernoulli bond percolation on G.*

Proof. Set $p_c = p_c(G)$. By Theorem 9, w_{p_c} has only finite components a.s. By Theorem 7, $(w_{p_c})^\dagger$ has a unique infinite component a.s. Consequently, by Theorem 9 again, $(w_{p_c})^\dagger$ is supercritical; that is, $p_c(G^\dagger) < 1 - p_c(G)$. An appeal to Theorem 8 now establishes the inequality $p_c(G) < p_u(G)$. Using the well-known inequality $p_c(G) \geq 1/(d-1)$, where d is the maximal degree of the vertices in G, we see that $p_u(G) = 1 - p_c(G^\dagger) \leq 1 - 1/(d^\dagger - 1)$, where $d^\dagger$ is the maximal degree of the vertices in $G^\dagger$, and so we get $p_u(G) < 1$, which completes the proof.

□

4 Geometric Consequences

For an infinite graph G and an infinite set of vertices $V_0 \in V(G)$, we define the Cheeger constant of G as follows:

$$i_E(G) = \inf\{\frac{|\partial_E V_0|}{|V_0|} : \emptyset \neq V_0 \subset V(G), |V_0| < \infty\},$$

where $\partial_E V_0$ is the set of edges emanating from V_0 that do not terminate in V_0.

Note that $i_E = 0$ for amenable graphs and $i_E > 0$ for transitive, nonamenable graphs. The following theorem is Theorem 4.4 from [2].

Theorem 11. *Let G be a graph with a transitive closed unimodular automorphism group $\Gamma \subset \mathrm{Aut}(G)$, and suppose that $i_E > 0$. Let ω be a Γ-invariant percolation on G. Then, for Bernoulli percolation with $p > p_c$, a simple random walk on some infinite cluster of ω has positive speed with positive probability.*

Assumption 12. *Let G be a Gromov hyperbolic graph. Let ω be the Bernoulli bond percolation on G. Then, any infinite cluster $\mathcal{C} \in \omega$ is also Gromov hyperbolic.*

Theorem 13. *Let G be a transitive, unimodular, nonamenable Gromov hyperbolic graph. Let ω be the Bernoulli bond percolation on G. Then, for $p > p_c$, a.s. every simple random walk on an infinite cluster $\mathcal{C} \in \omega$ tends to a unique point in $\partial_\infty G$.*

Proof. Let $X(t)$ be a simple random walk on $\mathcal{C}$. Then, by Theorem 11, there exists a constant $\Lambda > 0$ such that

$$\Lambda|j - i| \leq d(X(i), X(j)) \leq \ell(X(i) \xrightarrow{X} X(j)) \leq |j - i|. \tag{1}$$

Suppose the simple random walk tends to more than one point in $\partial_\infty G$. Then, there exist sequences i_k and j_k such that $X(i_k)$ and $X(j_k)$ tend to different limit points in $\partial_\infty G$. By Assumption 12, $\mathcal{C}$ is Gromov hyperbolic, so as i_k and j_k tend to infinity, the Gromov product is bounded by some constant b. By (1), we know the path from $X(i_k)$ to $X(j_k)$ must remain outside the sphere of radius Λi_k. However, for large i_k and j_k, the length l of the path outside the sphere is exponentially large, so, for constants $b_1, b_2 > 0$, which depend on δ and b, we have

$$\ell(X(i_k) \xrightarrow{X} X(j_k)) \geq b_1 e^{b_2 \Lambda \frac{i_k + j_k}{2}},$$

which, for sufficiently large k, is much greater than $|j_k - i_k|$ and contradicts (1). Therefore, the Gromov product could not have remained bounded and i_k and j_k must tend to the same limit point in $\partial_\infty G$.

□

We call G *radially homogeneous* if G is regular and there exist center-dependent values d_o, d_i, and d_ℓ such that each vertex in G (except the center) has d_o expected outward edges, d_i expected inward edges, and d_ℓ expected lateral edges. We call G *equi-traversed* if G is radially homogeneous and d_o, d_i, and d_ℓ hold for simple random walks on G which start at the center. Then, the speed Λ of a simple random walk on G can be calculated as follows:

$$\Lambda = \frac{d_o - d_i}{d},$$

where $d = d_o + d_i + d_\ell$ is the degree of each vertex.

For an infinite cluster $\mathcal{C} \in \omega$, we have expected outward, inward, and lateral degrees e_o, e_i, and e_ℓ. Given a single $v \in \mathcal{C}$, let d_o^v, d_i^v, and d_ℓ^v be the expected outward, inward, and lateral degrees from v to other vertices in $\mathcal{C}$. Then, the expected progress Λ_v of a simple random walk at a given $v \in \mathcal{C}$ can be calculated as follows:

$$\Lambda_v = \frac{d_o^v - d_i^v}{d_o^v + d_i^v + d_\ell^v}.$$

By Theorem 11, every simple random walk on $\mathcal{C}$ has positive speed, so for the expected net progress $\lambda_\mathcal{C}$ of a simple random walk, we have:

$$\Lambda_\mathcal{C} = E_v\left[\frac{d_o^v - d_i^v}{d_o^v + d_i^v + d_\ell^v}\right] \geq \frac{1}{d}E_v[d_o^v - d_i^v] = \frac{e_o - e_i}{d} \geq 0.$$

Using this and the fact that infinite graphs must have $e_i \geq 1$, we have

$$e_o > e_i \geq 1. \tag{1}$$

Given a path $\sigma(0) \to \sigma(n)$ along an infinite cluster $\mathcal{C} \in \omega$, a vertex $v \in \mathcal{C}$ is *dead* if any path from v to infinity must pass through σ. Otherwise, v is *undead* and there exists a non-retracing path from v to infinity.

Since G is hyperbolic, it has a compactification $\bar{G} = G \cup \partial_\infty G$ and $\mathcal{C} \subset G$ has a compactification $\bar{\mathcal{C}} \subset \bar{G}$. Note that we can re-metrize $\bar{\mathcal{C}}$ to be a compact metric space with the same geodesics using a method such as the Floyd method. A generalized version of the Arzela-Ascoli theorem, Corollary 3.11 from [4], states that any sequence of geodesic segments $\alpha_n \subset \mathcal{C}$ has a convergent subsequence to a geodesic $\alpha \subset \bar{\mathcal{C}}$. In our case, α will be a ray since our starting point is $x_0 = \alpha_n(0)$ and the length of $\alpha(n)$ tends to infinity.

Then, we have the following algorithm to create two distinct $\mathcal{C}$-geodesic rays, η and γ, that tend to different points in $\partial_\infty G$.

1. Let $\sigma(n)$ and $\tau(n)$ be walks on $\mathcal{C}$. Start at the point $x_0 = \sigma(0) = \tau(0)$.

2. Assuming we have chosen $\sigma(0), \sigma(1), ..., \sigma(n-1)$ and $\tau(0), \tau(1), ..., \tau(n-1)$, we choose $\sigma(n)$ and $\tau(n)$ among the undead vertices connected to $\sigma(n-1)$ and $\tau(n-1)$, respectively, such that $d(\sigma(n), \tau(n))$ is maximal. If multiple edge choices establish maximal distance, we choose outward edges over inward and lateral, and we choose lateral edges over inward.

3. Let σ_n and τ_n be $\mathcal{C}$-geodesic segments from x_0 to $\sigma(n)$ and $\tau(n)$, respectively. If multiple $\mathcal{C}$-geodesic segments exist, we choose ones with the minimum Hausdorff distance from σ_{n-1} and τ_{n-1}, respectively.

4. Then, by the generalized Arzela-Ascoli theorem from above, we know σ_n and τ_n have convergent subsequents to geodesics η and γ in $\bar{\mathcal{C}}$.

Theorem 14. *Let G be an equi-traversed, transitive, unimodular, nonamenable, Gromov hyperbolic graph. Let ω be the Bernoulli bond percolation on G. Then, for $p > p_c$, a.s. every infinite cluster $\mathcal{C} \in \omega$ has more than one point in $\partial_\infty G$.*

Proof. At each stage of the second step of our algorithm, we maximize the distance between $\sigma(n)$ and $\tau(n)$. By (1), there is an average branching of outward edges, so we can always increase the distance by some $\epsilon > 0$ on average, so we have

$$d(\sigma(n), \tau(n)) > \epsilon n \to \infty.$$

However, by Assumption 12, $\mathcal{C}$ is Gromov hyperbolic, so if η and γ were $\mathcal{C}$-geodesics tending to the same point in $\partial_\infty G$, they would remain a bounded distance apart. Therefore, η and γ tend to different points in $\partial_\infty G$.

□

5 p_c Bounds for Equi-Traversed Graphs

We can adopt a lower bound for p_c from the uniform tree case, since uniform trees are the least connected equi-traversed graphs. For uniform trees, $p_c = \frac{1}{d-1}$. We can derive an upper bound for p_c from a recursive equation for the expected number of vertices in $S(x, n) \cap \mathcal{C}$.

Let Z_n,Z_{n-1}, and Z_{n+1} be the number of vertices in $A_n = S(x, n) \cap \mathcal{C}$, $A_{n-1} = S(x, n-1) \cap \mathcal{C}$, and $A_{n+1} = S(x, n+1) \cap \mathcal{C}$, respectively. Let $e^k_{o,i,\ell}$ be the expected number of outward, inward, and lateral edges in G from vertices in A_k. Note that $e^k_{o,i,\ell} = d^k_{o,i,\ell}$ for graphs with a constant number of outward, inward, and lateral edges. Let r^k_ℓ be the expected number of vertices in A_k that can only be reached from A_k and r^{k+1}_i be the expected number of vertices in A_k that can only be reached from A_{k+1}. We know Z_n will be at least the expected number of vertices reached from the previous layer: that is, the expected number of outward edges from the previous layer, minus the overcount of the expected number of inward edges for vertices in Z_n. We also take into account the vertices

that can only be reached laterally from A_n, as well as the vertices that can only be reached from A_{n+1}. Then we have:

$$\begin{aligned} Z_n &= \sum_{v \in S(x,n)} \mathbb{P}[v \in \mathcal{C}] \\ &\geq p[e_o^{n-1} Z_{n-1} - (e_i^n - 1) Z_n + r_\ell^n Z_n + r_i^{n+1} Z_{n+1}]. \end{aligned}$$

Since r_i^n and r_ℓ^{n+1} are hard to estimate and likely to be small, we use an underestimate to find a lower bound for the exponential growth rate.

$$\begin{aligned} Z_n &\geq p e_o^{n-1} Z_{n-1} - p(e_i^n - 1) Z_n \\ (1 + p(e_i - 1)) Z_n &\geq p e_o Z_{n-1} \\ Z_n &\geq \left(\frac{p e_o}{1 - p + p e_i} \right)^n \end{aligned}$$

Provided $p > \frac{1}{1+e_o-e_i}$, the exponential growth rate will be greater than one, so $Z_n \to \infty$ as $n \to \infty$. So, we have an upper bound for p_c:

$$p_c \leq \frac{1}{1 + e_o - e_i}.$$

Note that $p_c < 1$ when $e_o > e_i$. In conclusion, for equi-traversed graphs, we have:

$$\frac{1}{d-1} \leq p_c \leq \frac{1}{1 + e_o - e_i}.$$

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

References

[1] I. Benjamini, R. Lyons, Y. Peres, and O. Schramm, *Group-invariant percolation on graphs*, Geom. Funct. Anal. **9** (1999), no. 1, 29-66.

[2] I. Benjamini, R. Lyons, and O. Schramm, *Percolation perturbations in potential theory and random walks*, in M. Picardello and W. Woess, editors, Random Walks and Discrete Potential Theory, Sympos. Math., Cambridge University Press, Cambridge, 1999, pp. 56-84. Papers from the workshop held in Cortona, 1997.

[3] I. Benjamini and O. Schramm, *Percolation in the hyperbolic plane*, J. Amer. Math. Soc. **14** (2000), no. 2, 487-507.

[4] M. Bridson and A. Haefliger, *Metric Spaces of Non-positive Curvature*, Springer, Berlin (1999).

[5] B. Mohar, *Embeddings of infinite graphs*, J. Combin. Theory Ser. **B 44** (1988), no. 1, 29-43.

[6] C. Newman and L. Schulman, *Infinite clusters in percolation models*, J. Statist. Phys. **26** (1981), no. 3, 613-628.

Searching for Solitary Pseudo-Anosovs

Yvonne Chazal*

Summer 2016

Abstract

Pseudo-Anosov homeomorphisms of a surface can be represented by their action on certain polygonal structures, called translation representations. In many cases, pseudo-Anosovs share the same translation representation [3]. If a pseudo-Anosov does not share its translation representation with any pseudo-Anosov other than its powers, we will call it solitary. It is not yet known whether or not solitary pseudo-Anosovs exist. In this report, we consider the action of pseudo-Anosovs on homology and develop tests for pseudo-Anosovs to have the same translation representation in terms of this action. Using Mathematica, we apply these tests to genus 3 surfaces.[1]

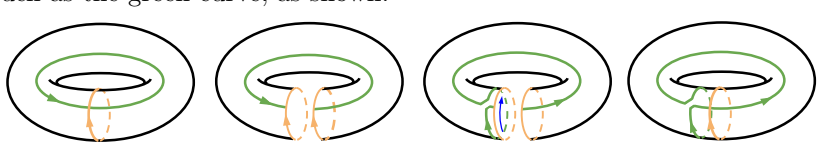
1 Introduction

For this paper, we will assume that when we have a surface, which we will always denote X, it is both closed and orientable with genus g. When we use the term curve, we will assume it is both simple and closed with orientation.

Definition. *A self-homeomorphism is a bijective map from a surface to itself that is continuous and has a continuous inverse.*

For the remainder of this report, we will use the term *homeomorphism* to denote self-homeomorphism, as all homeomorphisms will be from a surface to itself.

An example of such a function is a left Dehn twist defined geometrically as follows. Let our surface X be a torus. With curve γ as shown below in orange on the left. If we cut along γ as shown, twist the portion to the left along γ, then re-identify along γ, we have a homeomorphism. This affects curves intersecting γ, such as the green curve, as shown.



*North Carolina State University, Indiana University - Bloomington

[1]This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

The Nielson-Thurston Classification [3] gives a classification for the homeomorphisms of a surface. A homeomorphism ϕ is called *periodic* if there exists some integer $k > 0$ such that $\phi^k = id$. ϕ is called *reducible* if some power of ϕ maps a finite union of disjoint simple, closed curves on X back to themselves.

Nielson-Thurston Classification. *If $\phi : X \to X$ is a homeomorphism, one of the following holds:*

1. *ϕ is periodic*
2. *ϕ is reducible*
3. *ϕ is pseudo-Anosov*

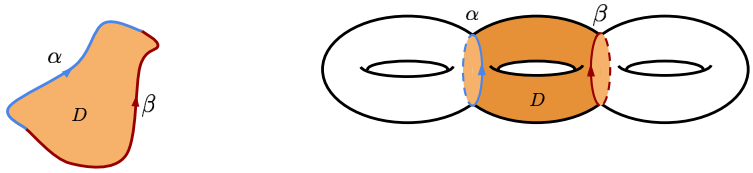
The last type of homeomorphism is of particular interest to us, and is the subject of this project.

Definition. *A homeomorphism $\phi : X \to X$ is called pseudo-Anosov if and only if $\phi^k(\gamma) \neq \gamma$ for all γ and $k > 1$.*

A self-homeomorphism of a surface can be understood by its action on curves on the surface. If two curves are "parallel," then their image under the homeomorphism will also be "parallel." Therefore, we can identify "parallel" curves using a notion of "parallel" that is made precise by the homology equivalence relation.

2 Homology

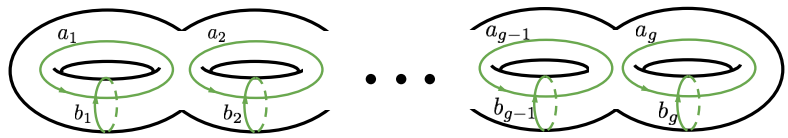
Definition. *Two curves (not necessarily closed) are homologous if their difference bounds a surface.*



In the two examples of homologous curves above, we have α (blue) is homologous to β (red), and $\partial D = \alpha - \beta = \alpha \cup -\beta$, surfaces D colored orange.

Homology is an equivalence relation, meaning we can define equivalency classes of homologous curves which we will call *homology classes*. The homology class with representative γ will be denoted $[\gamma]$. If we take the set of formal linear combinations of curves $\{\Sigma a_i \gamma_i : a_i \in \mathbb{Z}\}$ and mod out by the equivalence relation, these homology classes form a free abelian group together under addition which we will call the *first homology group of* X, denoted $H_1(X, \mathbb{Z})$ [5]. If we extend this group to consider real coefficients, $H_1(X, \mathbb{R})$, it forms a vector space.

On this vector space, we will need to utilize a basis. The basis we will utilize, which we will discuss further in the next section, will look as follows.



$\{a_i, b_i\}_{i=1}^g$ is just one set of representations of our basis elements $\{[a_i], [b_i]\}_{i=1}^g$. Note that for all $\{[a_i], [b_i]\}_{i=1}^g$, $a_j \cap b_j$ is only one point and there are no other intersections. Our first homology group will be all linear combinations of these basis elements.

$$H_1(X, \mathbb{R}) = \left\{ \sum_{i=1}^{g} m_i[a_i] + n_i[b_i] : m_i, n_i \in \mathbb{R} \right\}.$$

For a homeomorphism ϕ, we can induce a linear homeomorphism ϕ_* on the first homology group. We have that if two curves a_1 and a_2 are homologous, denoted $a_1 \sim a_2$, then $\phi(a_1) \sim \phi(a_2)$, thus such a function would be well-defined.[2] For $[\gamma] = a_1[\gamma_1] + a_2[\gamma_2] + \cdots + a_k[\gamma_k]$, define $\phi_*([\gamma]) = a_1\phi_*([\gamma_1]) + a_2\phi_*([\gamma_2]) + \cdots + a_k\phi_*([\gamma_k])$. Thus $\phi_* : H_1(X, \mathbb{R}) \to H_1(X, \mathbb{R})$ is a linear transformation on vector space $H_1(X, \mathbb{R})$. We will use ϕ_* and its matrix representation, which we will denote A_ϕ, interchangeably.

Definition. *Define $i : H_1(X, \mathbb{R}) \times H_1(X, \mathbb{R}) \to \mathbb{R}$. The intersection form of two differentiable curves α and β, denoted $i(\alpha, \beta)$, is the signed count of times α and β intersect, where the sign of an intersection is positive if the order of the vectors aligns with the orientation of the surface and is negative otherwise.*[3]

Note that $i(\alpha, \beta) = -i(\beta, \alpha)$ and that if $i(\alpha, \gamma) = 0$ for all $\gamma \in X$, then α does not intersect any curves on the surface, so it must be 0. Also note that for our basis, we have $i(a_i, b_i) = 1$ and all other combinations are 0.

Definition. *A bilinear form $(\cdot, \cdot)$ on vector space V is symplectic if and only if the following hold:*

- *If $(v, w) = 0$ for all $w \in V$, then $v = 0$.*
- *$(v, w) = -(w, v)$ for all $v, w \in V$.*

Definition. *In general, a collection of curves $\{a_i, b_i\}_{i=1}^g$ representing basis $\{[a_i], [b_i]\}_{i=1}^g$ will be called a geometric set of representatives if $i(a_j, b_j) = 1$*

[2] If $\alpha \sim \beta$, then for a homeomorphism ϕ, we have $\phi(\alpha \cup -\beta) = \phi(\partial D)$, and that $\phi(\alpha) \cup -\phi(\beta) = \partial\phi(D) = \partial D$, so $\phi(\alpha) \sim \phi(\beta)$.

[3] If α and β do not intersect transversely, then we can find homologous curves that do, and the intersection number does not depend on the homology representative.

and there are no other intersections. If $\{[a_i],[b_i]\}$ has a geometric set of representatives, we will call it a geometric symplectic basis.

Thus, the intersection form is symplectic, and the homology basis that we have chosen is a geometric symplectic basis.

Definition. *A linear transformation L is called symplectic with respect to $(\cdot,\cdot)$, if and only if $(Lx, Ly) = (x, y)$ for all $x, y \in V$.*

For any symplectic linear transformation L, we have that $det(L) = 1$, and characteristic polynomial P_L is palindromic, which implies that if λ is an eigenvalue for L, then λ^{-1} is also an eigenvalue for L. It is also true that for J defined below, $L^T JL = J$ for all L.

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & 0 & \dots & 0 & 0 \\ & & \dots & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 0 \end{pmatrix}$$

We have that for all ϕ_*, $i(\phi_*(\alpha), \phi_*(\beta)) = i(\alpha, \beta)$. By definition, this means that ϕ_* is a symplectic linear transformation on vector space $H_1(X, \mathbb{R})$, and thus we have all of the properties listed above for ϕ_* (and its matrix representation A_ϕ).

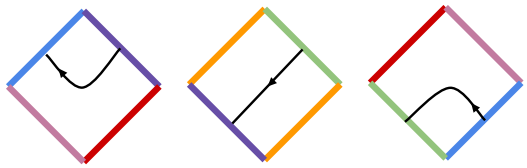
3 Translation Representations and Affine Group

To further examine homeomorphisms on a surface and their effects on curves, one can consider a geometric realization of these topological functions in the plane called the translation representation.

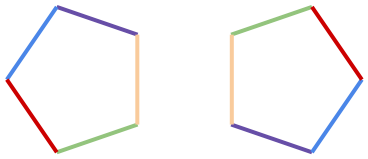
Definition. *A translation representation Σ of surface X is a finite union of polygons with set of edges E coupled with pairing function $p : E \to E$ such that*

1. *$p \circ p = id$*
2. *p has no fixed points*
3. *$p(e)$ has the same length as edge e for all $e \in E$*
4. *The outward normal vector to $p(e)$ is opposite to the outward normal vector of e*
5. *X is homeomorphic to the space obtained by identifying each edge e with $p(e)$ via translation of the plane.*

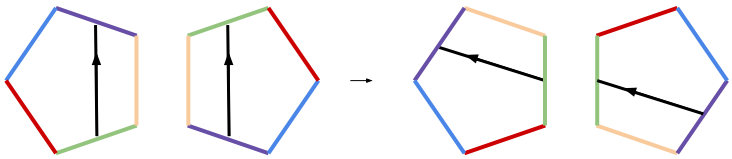
One should note that the number of edges within a translation surface will be even, and that paired edges will be parallel. Curves on such structures will traverse the structure as shown, leaving an edge at the same point that it entered its pair.



The double pentagon is a translation representation for a torus of genus 2. It is comprised of two adjoined regular pentagons with parallel sides identified. Below we have pictured the double pentagon with solid like-colors representing identified sides. The length of each side is 1.

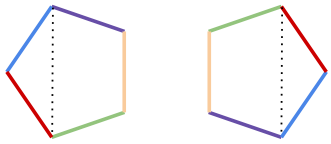


We can use this translation representation to represent homeomorphisms on the genus 2 torus. These homeomorphisms will need to preserve the identifications on paired edges. One such homeomorphism is a rotation about the center of each of the two pentagons in a counter-clockwise direction of $2\pi/5$, as shown below.



As one can see, after being rotated the paired edges remain parallel, meaning the identifications are preserved and we have a valid homeomorphism.

For the second homeomorphism we will consider, we will want to decompose the double pentagon into cylinders. Each cylinder will represent one of the two tori in our genus two torus, and will correspond to two parallelograms within our translation surface. We will use the vertical cylinder decomposition shown below. The black, dashed line represents the separating curve for each of the cylinders.



We are now able to look at homeomorphisms along each of these two cylinders, representing homeomorphisms acting on each of the two tori in the surface. We will call the cylinder bounded by the purple, green, and tan edges and the separating curve the "inner" cylinder. The "outer" cylinder is bounded by the red and blue edges and the separating curve.

Definition. *A homeomorphism is called affine with respect to a translation representation Σ if and only if ϕ is differentiable and the Jacobian $d\phi$ does not depend on the point at which it is taken.*

Proposition 1. *For each affine homeomorphism ϕ on surface X, we have $d\phi \in SL_2(\mathbb{R})$.*

Proof. By the Change of Variables formula, we have

$$\int_X dA = \int_{\phi(X)} det(d\phi)dA$$

Because ϕ is a bijective self-homeomorphism, $\phi(X) = X$, so

$$\int_{\phi(X)} det(d\phi)dA = \int_X det(d\phi)dA = det(d\phi)\int_X dA$$

Thus, by dividing out by the integral, we have that $det(d\phi) = 1$. □

If ϕ is an affine homeomorphism and $\gamma : [0,1] \to \mathbb{R}^2$ is a path, then $(\phi \circ \gamma)$ is also a path. From the Chain Rule, we have

$$(\phi \circ \gamma)' = d(\phi \circ \gamma) = (d\phi) \circ \gamma'$$

It follows by the Fundamental Theorem of Calculus that

$$\begin{aligned}(\phi \circ \gamma)(t) - (\phi \circ \gamma)(0) &= \textstyle\int_0^t (\phi \circ \gamma)'(s)ds \\ &= \textstyle\int_0^t d\phi \circ \gamma'(s)ds \\ &= d\phi \textstyle\int_0^t \gamma'(s)ds \\ &= d\phi(\gamma(t) - \gamma(0))\end{aligned}$$

Thus we have

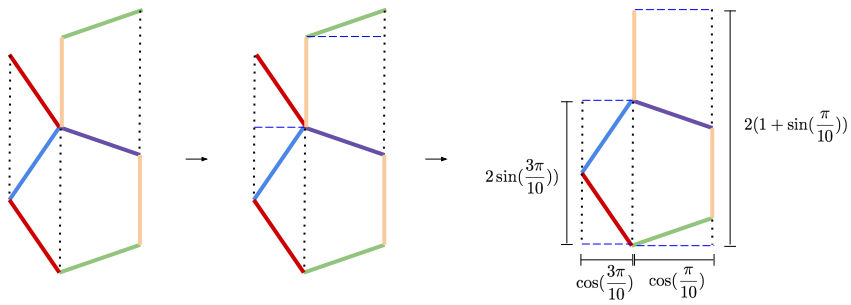
$$\phi \circ \gamma(t) = d\phi(\gamma(t)) - d\phi(\gamma(0)) + \phi \circ \gamma(0)$$

We can combine the vectors that do not depend on t and call it $c \in \mathbb{R}^2$, making our equation

$$\phi(\gamma) = d\phi \cdot \gamma + c$$

Note that if γ is a line segment, then $\phi(\gamma)$ is also a line segment.

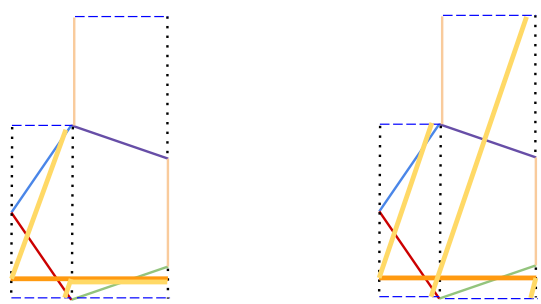
In order to decompose our representation into cylinders, we can cut along the separating curve, and identify the red and purple edges via translation. Then we can make two more cuts (along the pictured blue dashed lines), and identify the remaining blue and green edges.



Using basic geometry, we can find the heights and widths of each cylinder as we have thus far preserved the geometric structure of our translation representation. The moduli of the two cylinders (height/width) are

$$\frac{2\sin(\frac{3\pi}{10})}{\cos(\frac{3\pi}{10})} = 2.7528\ldots \qquad \frac{2(1+\sin(\frac{\pi}{10}))}{\cos(\frac{\pi}{10})} = 2.7528\ldots$$

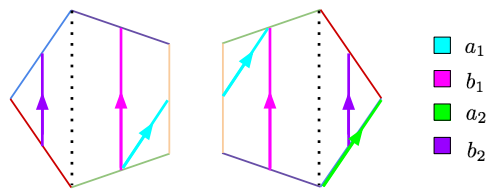
Suppose we have the orange, straight line shown below on each copy of the translation representation, with points along the black dotted line (our separating curve) identified. For a homeomorphism, the points will need to remain identified. The two yellow curves below represent the image of the orange curve under two valid homeomorphisms.



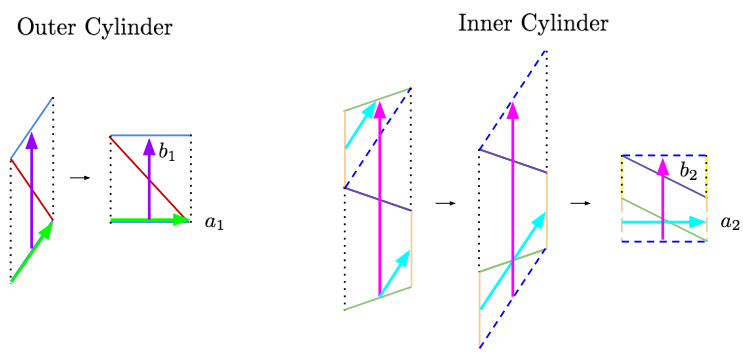
On the left, the two points remain identified, but the yellow curve is not differentiable, and thus it is not affine. On the right, the yellow curve is a straight line, with slope equivalent to the modulus of each of the cylinders. It is because the moduli of these cylinders are rational multiples of each other, namely equivalent, that we are able to produce such a straight line.

The affine homeomorphism shown to the right above is the second homeomorphism we will consider–the left Dehn multitwist–on each of the two tori represented by the two cylinders above. This will be a left Dehn twist about both b_1 and b_2.

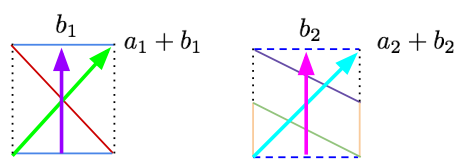
In order to construct the matrix representations on $Sp(4, \mathbb{Z})$ of these two affine homeomorphisms, we need to compute their actions on a homology basis. For this we will use the following basis vectors:



In order to show that these are indeed a valid homology basis, we will again decompose the double pentagon into cylinders, this time, homotoping our cylinders into the square representations of the two tori.



We know that the left Dehn twist about b_1 and b_2 will look as follows on the two tori.



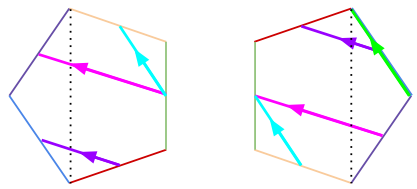
Thus we have the following, and we can construct its matrix representation.

$$\begin{array}{rc} T(a_1) = & a_1 \\ T(b_1) = & a_1 + b_1 \\ T(a_2) = & a_2 \\ T(b_2) = & a_2 + b_2 \end{array} \implies A_T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

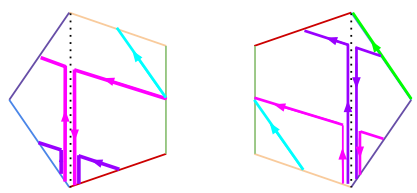
Note that we can separately compute the left Dehn twists about b_1 and b_2 and compose them to retrieve the same matrix representation.

$$A_{T_{b_1}} \circ A_{T_{b_2}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = A_{T_b}$$

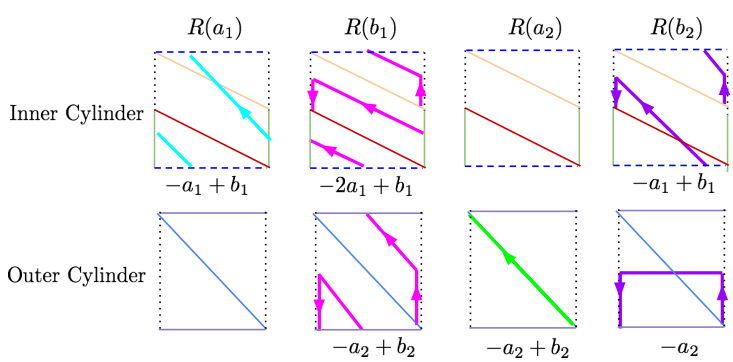
As for the rotation, when applied to the homology basis, we get the following.



In order to determine the homology class of these curves, we must again split the structure into tori, but we must be careful with the curves that traverse the separating curve. For these, we must divide the curve into two curves that cancel along the separating curve. The sum of these two curves will be homologically equivalent to our original curve. We also want the curves to turn in the same direction each time they run into the separating curve. We have chosen for the curves to turn in the counter-clockwise direction, consistent with the standard orientation. This is shown below.



We can homotope each cylinder into their square representations, homotoping $R(a_1), R(b_1), R(a_2)$, and $R(b_2)$ along with them. Below we have each of the four curves split into the two tori and homotoped to the square representations. They are each on their own square representation to reduce clutter. From here, it is fairly easy to see the decomposition of these curves into the basis elements.



Thus, we have

$$\begin{array}{rc} R(a_1) = & -a_1 + b_1 \\ R(b_1) = & -2a_1 + b_1 - a_2 + b_2 \\ R(a_2) = & -a_2 + b_2 \\ R(b_2) = & -a_1 + b_1 - a_2 \end{array} \implies A_R = \begin{bmatrix} -1 & -2 & 0 & -1 \\ 1 & 1 & 0 & 1 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The Dehn multitwist is classified as reducible, as each time the function is applied, b_1 and b_2 remain fixed. The rotation is of finite order, as every power of 5 will be the identity when the pentagons are rotated a full 2π. Thus neither of these homeomorphisms constitute pseudo-Anosovs. However, we can find pseudo-Anosovs in combinations of T and R.

The affine self-homeomorphisms of a translation representation Σ constitute a group together with composition called the *affine group of* Σ, denoted $\mathrm{Aff}(\Sigma)$.

Definition. *Two pseudo-Anosov homeomorphisms $\phi : X \to X$ and $\psi : X \to X$ are called friends if there exists Σ such that $\phi, \psi \in$ Aff(Σ).*

Note that ϕ^k is friends with ϕ for all integers $k \neq 0$.

Definition. *ϕ is called solitary if and only if Aff$(\Sigma) = \langle \phi \rangle$.*

By this definition, ϕ and its powers are the only affine homeomorphism associated with the translation representation Σ. Such homeomorphisms are exactly what we are searching for.

4 Function h

Now that we have a geometric structure of homeomorphisms on a given surface, it is natural to construct a geometric measure on this structure. We will define a linear functional $h : H_1(X, \mathbb{R}) \to \mathbb{R}^2$ with respect to translation surface Σ as follows.

$$h(\gamma) = \begin{bmatrix} h_x(\gamma) \\ h_y(\gamma) \end{bmatrix} = \begin{bmatrix} \int_\gamma dx \\ \int_\gamma dy \end{bmatrix}$$

This function measures the amount γ travels in the direction of e_x and e_y respectively and provides us with a geometric measure for γ.

h is a linear transformation , as both h_x and h_y are integral functions, and are thus linear. To construct the matrix representation of this linear transformation, we only need to see what this function does to the basis elements on the translation structure Σ of surface X. Thus we have the following:

$$B_h = \left[\begin{array}{c|c|c|c|c} h(a_1) & h(b_1) & \dots & h(a_{2g}) & h(b_{2g}) \end{array} \right]$$
$$= \begin{bmatrix} h_x(a_1) & h_x(b_1) & \dots & h_x(a_{2g}) & h_y(b_{2g}) \\ h_y(a_1) & h_y(b_1) & \dots & h_y(a_{2g}) & h_y(b_{2g}) \end{bmatrix}$$

These values are easily computed from the geometry of the translation structure of a surface, by calculating the length of each a_i and b_i in directions x and y.

Proposition 2. *For translation representation Σ of surface X, $h : H_1(X, \mathbb{R}) \to \mathbb{R}^2$ is surjective.*

Proof. Proceed by contradiction. Assume $h_x(\gamma) = 0$ for all $\gamma \in H_1(X, \mathbb{Z})$. Fix a point p_0 in Σ, and let p be another point on Σ. Note that p could be located on a different polygon from p_0.

Let α and β be paths in Σ, joining points p_0 and p. These paths may be a collection of paths inside multiple polygons such that when the edges of these polygons are identified, the paths form one continuous path from p_0 to p. $\alpha \cup -\beta$ is a closed curve, and thus represents an element of $H_1(X, \mathbb{Z})$. By our initial assumption, we have $h_x(\alpha \cup -\beta) = 0$, and thus:

$$\int_\alpha dx - \int_\beta dx = \int_{\alpha \cup -\beta} dx = h_x(\alpha \cup -\beta) = 0$$

so $\int_{\alpha \cup -\beta} dx$ is independent of path.

Let us define $f : X \to \mathbb{R}$ by $p \mapsto \int_\alpha dx$. It is clear that f is differentiable.

The union of finitely many polygons is a closed and bounded set. Thus, by the Heine-Borel theorem, Σ is compact, so because f is continuous, f must reach a maximum value at some $p_{max} \in \Sigma$, i.e. $f(p_{max})$ is the largest value for f.

Consider the following three cases:

1. *Suppose (by contradiction) that p_{max} belongs to the interior of a polygon in Σ.* Then there exists $\epsilon > 0$ such that for $t \in [0, \epsilon]$, point $p_{max} + (t, 0)$ belongs to the polygon containing p_{max}. If we integrate dx over the concatenation of the path from p_0 to p with the path $t \mapsto p_{max} + (t, 0)$, we have

   $$f(p_{max} + (\epsilon, 0)) = f(p_{max}) + \epsilon > f(p_{max})$$

   This is a contradiction as $f(p_{max})$ is the maximum value for f.

2. *Suppose (by contradiction) that p_{max} belongs to the interior of an edge e of a polygon in Σ.* Then the maximum value of f is achieved at every point along e. Thus e is parallel to the y-axis and the outward normal vector of e is $(1, 0)$. Let T be the translation identifying e with $p(e)$. Then the outward normal vector of $p(e)$ is $(-1, 0)$. p_{max} is on the interior of e, so $T(p_max)$ is on the interior of $p(e)$=$T(e)$. Then there exists $\epsilon > 0$ such that for $t \in [0, \epsilon]$, point $T(p_{max} + (t, 0))$ belongs to the polygon containing edge $p(e)$. If we integrate dx over the concatenation of the path from p_0 to p with the path $t \mapsto T(p_{max} + (t, 0))$, we have

   $$f(T(p_{max}) + (\epsilon, 0)) = f(p_{max}) + \epsilon > f(p_{max})$$

   This is a contradiction as $f(p_{max})$ is the maximum value for f.

3. *Suppose (by contradiction) that p_{max} is a vertex of a polygon in Σ.* Then p_{max} may be identified with multiple vertices. Let $v_1, v_2, \ldots, v_k$ be vertices identified with p_{max}, with polygons $P_1, P_2, \ldots, P_k$ containing those vertices and translations $T_1, T_2, \ldots, T_k$ identifying those polygons. Let P be the polygon containing point p_{max}. The union of P and all of the images of $T_i(P_i)$ contains an open disk centered at p_{max}. Thus $T_i(p_{max})$ is a point on polygon P_i. Then there exists $\epsilon > 0$ such that for $t \in [0, \epsilon]$, point $T_i(p_{max} + (t, 0))$ belongs to polygon P_i. If we integrate dx over the concatenation of the path from p_0 to p with the path $t \mapsto T_i(p_{max} + (t, 0))$, we have

$$f(T_i(p_{max}) + (\epsilon, 0)) = f(p_{max}) + \epsilon > f(p_{max})$$

This is a contradiction as $f(p_{max})$ is the maximum value for f.

Therefore, our initial assumption that $h_x(\gamma) = 0$ for all $\gamma \in H_1(X, \mathbb{Z})$ must be false. Then there exists $\gamma_x \in H_1(X, \mathbb{Z})$ such that $h_x(\gamma_x) \neq 0$.

Let $a = h_x(\gamma_x)$ and $b = h_x(\gamma_x)$. Then we have

$$\int_\gamma b \cdot dx - a \cdot dy = b \int_{\gamma_x} dx - a \int_{\gamma_x} dx = b \cdot h_x(\gamma_x) - a \cdot h_x(\gamma_x) = ba - ab = 0$$

Let $\omega = b \cdot dx - a \cdot dy$. Apply the argument above to ω, making the assumption that $\int_\alpha \omega = 0$ for all paths α. Creating a similar function f gives a gradient vector field in the direction $(b, -a)$ at the maxima of f. By the argument above, there exists $\gamma_\omega \in H_1(X, \mathbb{Z})$ such that $\int_{\gamma_\omega} \omega \neq 0$. Because $\int_{\gamma_x} \omega = 0$, $h(\gamma_\omega)$ is independent of $h(\gamma_x) = (a, b)$. Thus, the dimension of $\mathrm{Im}(h)$ is at least 2, so because $\mathbb{R}^2$ is two dimensional, h is surjective. $\square$

In addition to our original definition of h_x and h_y, we have its relation to the intersection number.

$$h_x(\gamma) = \int_\gamma dx = i(\gamma, h_x^*) \tag{1}$$

So what is this h_x^*? Let's look at $\gamma = a_i$ and $\gamma = b_i$. Recall, $h_x^* \in H_1(X, \mathbb{R})$, so $h_x^* = m_1 a_1 + n_1 b_1 + \cdots + m_g a_g + n_g b_g$.

$$\begin{gathered}\gamma = a_i \implies h_x(a_i) = i(a_i, h_x^*) = i(a_i, m_1 a_1 + n_1 b_1 + \cdots + m_g a_g + n_g b_g) \\ = m_1 i(a_i, a_1) + n_1 i(a_i, b_1) + \cdots + m_g i(a_i, a_g) + n_g i(a_i, b_g) = n_i i(a_i, b_i) = n_i \\ \implies n_i = h_x(a_i)\end{gathered}$$

$$\begin{gathered}\gamma = b_i \implies h_x(b_i) = i(b_i, h_x^*) = i(b_i, m_1 a_1 + n_1 b_1 + \cdots + m_g a_g + n_g b_g) \\ = m_1 i(b_i, a_1) + n_1 i(b_i, b_1) + \cdots + m_g i(b_i, a_g) + n_g i(b_i, b_g) = m_i i(b_i, a_i) = -m_i \\ \implies m_i = -h_x(b_i)\end{gathered}$$

and similarly for h_y^*. Thus, we have

$$
\begin{aligned}
h_x^* &= -h_x(b_1)a_1 + h_x(a_1)b_1 + \cdots - h_x(b_g)a_g + h_x(a_g)b_g \\
&= \sum_{i=1}^{g} -h_x(b_i)a_i + h_x(a_i)b_i \\
h_y^* &= \sum_{i=1}^{g} -h_y(b_i)a_i + h_y(a_i)b_i.
\end{aligned}
$$

Combining our definition of h and h_x in terms of h_x^* (as well as h_y), we have

$$
h(\gamma) = \begin{bmatrix} i(\gamma, h_x^*) \\ i(\gamma, h_y^*) \end{bmatrix}.
$$

Proposition 3. *h_x^*, h_y^* span $ker(h)^\perp$.*

Proof. For all curves γ in $ker(h)$, $h(\gamma) = (0,0)$. By definition of h, this means $h_x(\gamma) = 0$ and $h_y(\gamma) = 0$. Thus $i(\gamma, h_x^*) = 0$ and $i(\gamma, h_y^*) = 0$ for all $\gamma \in ker(h)$. By definition of $\perp$, this means that $h_x^*, h_y^* \in ker(h)^\perp$.

Suppose (by contradiction) that h_x^* and h_y^* are not linearly independent. Then there exists some constant c such that $h_y^* = c \cdot h_x^*$. If this is the case, then we have

$$
h(\gamma) = \begin{bmatrix} i(\gamma, h_x^*) \\ i(\gamma, h_y^*) \end{bmatrix} = \begin{bmatrix} i(\gamma, h_x^*) \\ c \cdot i(\gamma, h_x^*) \end{bmatrix} = \begin{bmatrix} 1 \cdot i(\gamma, h_x^*) \\ c \cdot i(\gamma, h_x^*) \end{bmatrix} = i(\gamma, h_x^*) \begin{bmatrix} 1 \\ c \end{bmatrix}
$$

Therefore, for all γ, $h(\gamma)$ is a multiple of $\begin{bmatrix} 1 \\ c \end{bmatrix}$. Thus, the image of h lies in the line spanned by $\begin{bmatrix} 1 \\ c \end{bmatrix}$. Thus $Im(h)$ has dimension 1. But this contradiction because from Proposition 2, we have that $Im(h)$ has dimension 2. Thus, h_x^* and h_y^* are linearly independent.

Because h is surjective, $ker(h)$ has dimension $2g - 2$, by the Rank-Nullity Theorem. A general fact in symplectic linear algebra gives us that $dim(ker(h)) + dim(ker(h)^\perp) = 2g$. Therefore, $ker(h)^\perp$ has dimension 2. Because h_x^* and h_y^* are in $ker(h)^\perp$ and are linearly independent, they span $ker(h)^\perp$. $\square$

Theorem 1. *Let $\phi : \Sigma \to \Sigma$ be an affine homeomorphism, inducing a linear function $\phi_* : H_1(X, \mathbb{R}) \to H_1(X, \mathbb{R})$, and let $\alpha : [0,1] \to \Sigma$ be a simple, closed curve with homology class $[\alpha] \in H_1(X, \mathbb{R})$. Then $h(\phi_* \circ \alpha) = d\phi \cdot h(\alpha)$.*

Proof. $[0,1] \subset \mathbb{R}$ is spanned by a standard basis vector $\{e_t\}$, which we can think of as 1. $\Sigma \subset \mathbb{R}^2$ is spanned by standard basis vectors $\{e_x, e_y\}$.

$$
h_x \circ \phi_*([\alpha]) = h_x([\phi \circ \alpha])
$$

As stated in Fulton's *Algebraic Topology*, for all $\alpha \sim \beta$, $\int_\alpha dx = \int_\beta dx$ [1], so when considering a homology class, we need only look at a single representative. Here, we will examine $\phi \in [\phi]$ and $\alpha \in [\alpha]$. By definition, we have

$$h_x(\phi \circ \alpha) = \int_{\phi\circ\alpha} dx$$

We can then use the following theorem from Guillemin-Pollack's *Differential Topology* [2]:

Change of Variables in $\mathbb{R}^2$. *If U and V are open subsets of $\mathbb{R}^k$, $f : U \to V$ is an orientation-preserving diffeomorphism, and ω is an integrable k-form on U, then $\int_U \omega = \int_V f^*\omega$.*

Here, f^* is the transpose map of f, and $f^*\omega$ is called the pullback of ω by f. In our case, this gives us

$$\int_{\phi\circ\alpha} dx = \int_0^1 (\phi \circ \alpha)^* dx \tag{2}$$

Let us consider the integrand here pointwise, where $v \in T_{t_0}([0,1])$ for $t_0 \in [0,1]$. By the definition of pullback of dx by $\phi \circ \alpha$ [2],

$$(\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)}(v) = v \cdot (\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)}(e_t) \tag{3}$$

We will drop v for now and retrieve it later. It follows from the definition of transpose map [2]

$$\begin{aligned}(\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)}(e_t) &= (dx)_{(\phi\circ\alpha)(t_0)}\left[\frac{\partial((\phi\circ\alpha)_x)_{t_0}}{dt}(dt) + \frac{\partial((\phi\circ\alpha)_y)_{t_0}}{dt}(dt)\right] \\ &= \frac{\partial((\phi\circ\alpha)_x)_{t_0}}{dt}\end{aligned}$$

Retrieving v from (3),

$$\begin{aligned}v \cdot (\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)}(e_t) &= v \cdot \frac{\partial((\phi\circ\alpha)_x)_{t_0}}{dt} \\ &= (dt)_{t_0}(v) \cdot \frac{\partial((\phi\circ\alpha)_x)_{t_0}}{dt} \\ &= \frac{\partial((\phi\circ\alpha)_x)_{t_0}}{dt}(dt)_{t_0}(v)\end{aligned}$$

From here, we can distribute $_x$, as it will "trickle down" through the partials, then we can apply the chain rule.

$$= \frac{\partial(\phi_x \circ \alpha_x)_{t_0}}{dt}(dt)_{t_0}(v)$$
$$= \frac{(\partial\phi_x)_{(\phi\circ\alpha)(t_0)} \cdot (\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}(v)$$

Thus, we have

$$(\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)}(v) = \frac{(\partial\phi_x)_{(\phi\circ\alpha)(t_0)} \cdot (\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}(v) \quad \forall v \in T_{t_0}([0,1])$$

And because we have equality pointwise on the whole interval, we have that the functions are equivalent.

$$(\phi \circ \alpha)^*(dx)_{(\phi\circ\alpha)(t_0)} = \frac{(\partial\phi_x)_{(\phi\circ\alpha)(t_0)} \cdot (\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}$$

Returning to (2) with our integral, we now have

$$\int_0^1 (\phi \circ \alpha)^* dx = \int_0^1 \frac{(\partial\phi_x)_{(\phi\circ\alpha)(t_0)} \cdot (\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}(v)$$

However, because ϕ is an affine homeomorphism, $\partial(\phi_x)$ does not depend on t. Thus, we have

$$\int_0^1 \frac{(\partial\phi_x)_{(\phi\circ\alpha)(t_0)} \cdot (\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}(v) = \partial\phi_x \int_0^1 \frac{(\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0}$$

However, we have that

$$\int_0^1 \frac{(\partial\alpha_x)_{t_0}}{dt}(dt)_{t_0} = \int_0^1 \alpha^*(dx)$$

By retrieving $\partial\phi_x$, applying the Change of Variables Theorem, and using the definition of h, respectively

$$\partial\phi_x \cdot \int_0^1 \alpha^*(dx) = \partial\phi_x \cdot \int_\alpha dx = \partial\phi_x \cdot h_x(\alpha)$$

Thus, we have

$$h_x(\phi \circ \alpha) = \partial\phi_x \cdot h_x(\alpha)$$

Using the same procedure with h_y,

$$h_y(\phi \circ \alpha) = \partial\phi_y \cdot h_y(\alpha)$$

then we have the following:

$$h(\phi_* \circ \alpha) = \begin{bmatrix} h_x(\phi_* \circ \alpha) \\ h_y(\phi_* \circ \alpha) \end{bmatrix} = \begin{bmatrix} \partial\phi_x \cdot h_x(\alpha) \\ \partial\phi_y \cdot h_y(\alpha) \end{bmatrix} = d\phi \cdot \begin{bmatrix} h_x(\alpha) \\ h_y(\alpha) \end{bmatrix} = d\phi \cdot h(\alpha)$$

□

To further demonstrate the relationship between h, ϕ_*, and $d\phi$, consider the following commutative diagram. In this case, we say h *intertwines* ϕ_* and $d\phi$.

$$\begin{tikzcd} H_1(X,\mathbb{R}) \arrow[r, "\phi_*"] \arrow[d, "h"] & H_1(X,\mathbb{R}) \arrow[d, "h"] \\ \mathbb{R}^2 \arrow[r, "d\phi"] & \mathbb{R}^2 \end{tikzcd}$$

Let us now switch to thinking of ϕ_* as its matrix representation A_ϕ. Thus we have

$$h(A_\phi \cdot v) = d\phi \cdot h(v) \tag{4}$$

Suppose A_ϕ has eigenvalue λ with corresponding eigenvector $v*$. We then have the following:

$$d\phi \cdot h(v^*) = h(A_\phi \cdot v^*) = h(\lambda v^*) = \lambda \cdot h(v^*) \tag{5}$$

Of course v^* could be in the kernel of h, making $h(v^*) = 0$, and thus making the equation trivial. However, if this is not the case, then λ is an eigenvalue for both A_ϕ and $d\phi$, and $h(v*)$ is an eigenvector of $d\phi$. And vectors that satisfy these conditions are of particular interest to us.

Theorem 2 (Thurston [3])**.** *A homeomorphism $\phi \in Aff(\Sigma)$ is pseudo-Anosov if and only if $|tr(d\phi)| > 2$.*

Proposition 4. *If $\phi \in Aff(\Sigma)$ is pseudo-Anosov, then A_ϕ has two distinct eigenpairs (λ_+, v_+) and (λ_-, v_-) such that the following hold true:*

1. $v_+, v_- \in ker(h)^\perp$
2. *$h(v_+)$ and $h(v_-)$ are eigenvectors of $d\phi$*
3. $|\lambda_+ + \lambda_-| > 2$
4. $\lambda_+ \cdot \lambda_- = 1$.

Proof. We will first consider the pullback of dx and dy by ϕ. Note that there is a difference between ϕ^* and ϕ_*.

$$\phi^*(dx) = d\phi_x = \frac{\partial\phi_x}{dx}dx + \frac{\partial\phi_x}{dy}dy$$

$$\phi^*(dy) = d\phi_y = \frac{\partial\phi_y}{dx}dx + \frac{\partial\phi_y}{dy}dy$$

Let F denote the vectors space of 1-forms of the form $a \cdot dx + b \cdot dy$ where $a, b \in \mathbb{R}$. Then ϕ^* acts on F in the following way.

$$\begin{aligned}
\phi^*(a \cdot dx + b \cdot dy) &= a\tfrac{\partial \phi_x}{\partial x}dx + b\tfrac{\partial \phi_x}{\partial y}dy + a\tfrac{\partial \phi_y}{\partial x}dx + b\tfrac{\partial \phi_y}{\partial y}dy \\
&= \begin{bmatrix} a\frac{\partial \phi_x}{\partial x}dx + b\frac{\partial \phi_x}{\partial y}dy \\ a\frac{\partial \phi_y}{\partial x}dx + b\frac{\partial \phi_y}{\partial y}dy \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \phi_x}{\partial x} & \frac{\partial \phi_x}{\partial y} \\ \frac{\partial \phi_y}{\partial x} & \frac{\partial \phi_y}{\partial y} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = (d\phi)^T \begin{bmatrix} a \\ b \end{bmatrix}
\end{aligned}$$

If the reader will recall, $d\phi$ is the Jacobian of ϕ. Because ϕ is an affine, pseudo-Anosov homeomorphism, we have that $d\phi$ exists and has $|tr(d\phi)| > 2$. Because $det(d\phi) = 1$, $d\phi$ is conjugate to a diagonal matrix with its two eigenvalues along the diagonal. Thus the determinant, 1, is the product of the two eigenvalues, so the eigenvalues are inverses, say λ and λ^{-1}. Trace is also preserved by conjugation, so $|\lambda + \lambda^{-1}| = |tr(d\phi)| > 2$.

Because of this, $d\phi^T$ will have eigenvectors, say ω_+ and ω_- with the such eigenvalues. Eigenpairs for $d\phi^T$ are also eigenpairs for ϕ^*, so (λ, ω_+) and $(\lambda^{-1}\omega_-)$ are eigenpairs for ϕ^*.

From here onward, we will mainly discuss ω_+, but the same arguments can be carried out on ω_-. Let $\omega_+ = a_+dx + b_+dy$. Define a function $h_{\omega_+} : H_1(X, \mathbb{R}) \to H_1(X, \mathbb{R})$ such that

$$h_{\omega_+}(\gamma) = a_+ \int_\gamma dx + b_+ \int_\gamma dy = a_+ h_x(\gamma) + b_+ h_y(\gamma).$$

Thus, we have that the following. The matrix C is defined below in-context.

$$\begin{bmatrix} h_{\omega_+} \\ h_{\omega_-} \end{bmatrix} = \begin{bmatrix} a_+ & b_+ \\ a_- & b_- \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = C \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Let $h^*_{\omega_+} \in H_1(X, \mathbb{R})$ be defined as follows for all curves $\gamma \in H_1(X, \mathbb{R})$.

$$h_{\omega_+}(\gamma) = i(\gamma, h^*_{\omega_+}) = \int_\gamma \omega_+$$

Note that $h^*_{\omega_+} \in ker(h)^\perp$ because $h^*_{\omega_+} \in \text{span}\{h^*_x, h^*_y\}$.

It follows that

$$\lambda i(\gamma, h^*_{\omega_+}) = \lambda \int_\gamma \omega_+ = \int_\gamma \phi^*(\omega_+)$$

By the definition of pullback [2],

$$\int_\gamma \phi^*(\omega_+) = \int_{\phi_*(\gamma)} \omega_+$$

By the definition of h_{ω_+}, we have

$$\int_{\phi_*(\gamma)} \omega_+ = h_{\omega_+}(\phi_*(\gamma)) = i(\phi_*(\gamma), h^*_{\omega_+})$$

Because $i(\cdot,\cdot)$ and ϕ_* are symplectic, we have that

$$i(\phi_*(\gamma), h^*_{\omega_+}) = i(\gamma, \phi_*^{-1}(h^*_{\omega_+}))$$

and therefore, $\lambda i(\gamma, h^*_{\omega_+}) = i(\gamma, \phi_*^{-1}(h^*_{\gamma_+}))$. By basic algebra and the non-degeneracy of $i(\cdot,\cdot)$, we have the following for all $\gamma \in H_1(X, \mathbb{R})$.

$$\begin{aligned}
\lambda i(\gamma, h^*_{\omega_+}) = i(\gamma, \lambda h^*_{\omega_+}) = i(\gamma, \phi_*^{-1}(h^*_{\gamma_+})) & \\
\implies \quad & i(\gamma, \lambda h^*_{\omega_+}) - i(\gamma, \phi_*^{-1}(h^*_{\gamma_+})) = 0 \\
\implies \quad & i(\gamma, \lambda h^*_{\omega_+} - \phi_*^{-1}(h^*_{\gamma_+})) = 0 \\
\implies \quad & \lambda h^*_{\omega_+} - \phi_*^{-1}(h^*_{\gamma_+}) = 0 \\
\implies \quad & \lambda h^*_{\omega_+} = \phi_*^{-1}(h^*_{\gamma_+})
\end{aligned}$$

Thus, we have $\phi_*(h^*_{\omega_+}) = \lambda^{-1} h^*_{\omega_+}$. By a similar argument, we have $\phi_*(h^*_{\omega_-}) = \lambda h^*_{\omega_-}$. Thus, ϕ_* has eigenpairs $(\lambda^{-1}, h^*_{\omega_+})$ and $(\lambda, h^*_{\omega_-})$.

Let $v_+ = h^*_{\omega_+}$, $\lambda_+ = \lambda^{-1}$, $v_- = h^*_{\omega_-}$, and $\lambda_- = \lambda$. □

Definition. *The eigenvectors and eigenvalues as found in the theorem above are called affine.*

Definition. *Eigenvalues λ_1 and λ_2 are called hyperbolic if $\lambda_1 \cdot \lambda_2 = 1$ and $|\lambda_1 + \lambda_2| > 2$.*

Proposition 5. *If pseudo-Anosovs ϕ and ψ are friends, then the affine eigenvectors of ϕ and ψ span the same plane.*

Proof. If ϕ and ψ are friends, then they both belong to $\mathrm{Aff}(\Sigma)$ for some translation representation Σ with associated h. Because $\phi, \psi \in \mathrm{Aff}(\Sigma)$, we have that their affine eigenvectors are in $ker(h)^\perp$. These eigenvectors are linearly independent, and thus span the same plane. □

Thus, if the affine eigenvectors of A_ϕ do not span the same symplectic plane as the affine eigenvectors of any other pseudo-Anosov homeomorphism, then ϕ may be solitary. However, this is almost impossible to check as there are infinitely many homeomorphisms on any given surface.

Let (λ, v_+) and (λ^{-1}, v_-) be affine eigenpairs for A_ϕ. By definition of eigenvalue, we have $A_\phi v_+ = \lambda v_+$. However, by dividing by both λ and A_ϕ, we get $\lambda^{-1} v_+ = A_\phi^{-1} v_+$. Now let us consider $A_\phi^{-1} + A_\phi$.

$$(A_\phi^{-1} + A_\phi) v_+ = A_\phi^{-1} v_+ + A_\phi v_+ = \lambda^{-1} v_+ + \lambda v_+ = (\lambda^{-1} + \lambda) v_+$$

Similarly, for v_-, we have

$$(A_\phi^{-1} + A_\phi)v_- = A_\phi^{-1}v_- + A_\phi v_- = \lambda v_- + \lambda^{-1}v_- = (\lambda + \lambda^{-1})v_-$$

Thus, we have that $(\lambda^{-1} + \lambda)$ is an eigenvalue of $A_\phi^{-1} + A_\phi$ with eigenspace $\langle v_+, v_- \rangle$. Thus we have that if ϕ and ψ are friends, and $\{\lambda, \lambda^{-1}\}$ and $\{\mu, \mu^{-1}\}$ are affine eigenvalues of A_ϕ and A_ψ respectively, then eigenvalue $(\lambda^{-1} + \lambda)$ of $(A_\phi^{-1} + A_\phi)$ and $(\mu^{-1} + \mu)$ of $(A_\psi^{-1} + A_\psi)$ share common eigenspaces.

Thus, if the sum of the affine eigenvalues of A_ϕ constitute an eigenspace of $(A_\phi^{-1} + A_\phi)$ that is not shared by that of any other homeomorphism, then A_ϕ does not have any friends, and thus may be solitary. However, as before, this is almost impossible to check. But we are able to use this fact to check for homeomorphisms that are potentially solitary, by checking them against large numbers of pseudo-Anosovs.

5 Computing

We used the principles outlined above to explore sets of homeomorphisms that we could generate. For all computation, we used Mathematica.

The Burkhardt Generators below span $Sp(2g, \mathbb{Z})$ [4].

Transvection (T):

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ & & \dots & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Factor Rotation (R):

$$\begin{bmatrix} 0 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ & & \dots & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Factor Mix (M):

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & 1 & \dots & 0 & 0 \\ & & \dots & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Factor Swap (S_1) ($1 \leftrightarrow 2$):

$$\begin{bmatrix} 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ & & \dots & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

When considering a surface with genus g, there are $g - 1$ factor swap generators swapping adjacent factors $\{a_i, b_i\} \leftrightarrow \{a_i + 1, b_i + 1\}$ for all $1 \leq i \leq g$.

In order to generate many symplectic matrices of genus 2 and 3, we took combinations of these generators and their inverses in tuples of a certain word length. We only considered genus 2 and genus 3.

Farb-Margalit assert that we can further dissect these generators into Dehn twists as follows, where T_γ denotes a left Dehn twist about curve γ [4].

Transvection	T_{b_1}
Factor Rotation	$T_{b_1} \cdot T_{a_1} \cdot T_{b_1}$
Factor Mix	$T_{b_1}^{-1} \cdot T_{b_2}^{-1} \cdot T_{b_2 - b_1}$
Factor Swap ($i \leftrightarrow i+1$)	$(T_{a_{i+1}} \cdot T_{b_{i+1}} \cdot T_{a_{i+1}-b_i} \cdot T_{a_i} \cdot T_{b_i})^3$

After generating the matrices and eliminating duplicates that arose, we checked for pseudo-Anosovs by checking for both hyperbolic eigenvalues and satisfaction of the Casson-Bleiler Criteria below, adopted from Farb-Margalit's *Primer on Mapping Class Groups* [4].

Casson-Bleiler Criteria for Pseudo-Anosovs. *Suppose ϕ is a homeomorphism with symplectic matrix representation A_ϕ. Let $P_A(x)$ be the characteristic polynomial for A_ϕ. If each of the following conditions hold true, then ϕ is pseudo-Anosov.*

1. *$P_A(x)$ cannot be written as a product of two polynomials that are the characteristic polynomials of symplectic matrices.*
2. *$P_A(x)$ is not a cyclotomic polynomial.*
3. *$P_A(x)$ cannot be expressed as a polynomial in x^k for any $k > 1$.*

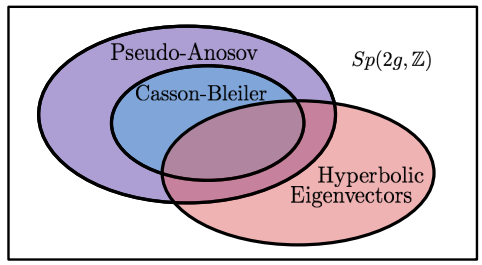
We began by checking all unique characteristic polynomials of matrices generated by combinations of the generators and their inverses. This gave us the following percentage of characteristic polynomials of the specified word length satisfying the Casson-Bleiler Criteria out of the total unique characteristic polynomials of that specific word length.

Word Length	3	4	5	6
$g = 2$	16.667%	25.000%	26.000%	58.475%
$g = 3$	0%	0%	17.549%	29.310%

This same set of characteristic polynomials gave us the following percentage of characteristic polynomials that had hyperbolic eigenvectors out of the total unique characteristic polynomials of that specific word length.

Word Length	3	4	5	6
$g = 2$	16.667%	45.833%	62.000%	71.186%
$g = 3$	0%	25.926%	50.877%	56.0356%

Note the discrepancy in these values. The set of Casson-Bleiler pseudo-Anosovs and the matrices with hyperbolic eigenvectors are not the same, neither is one a subset of the other. Their relationship is described by the following Venn diagram.



If the reader will recall, in order for a homeomorphism ϕ to be pseudo-Anosov if and only if ϕ has hyperbolic eigenvalues, we must know that ϕ is affine with respect to some Σ. If this is not the case, we can define a 2:1 double covering of the translation representation such that ϕ will be affine [6]. However, this increases the genus of the representation, so we will not consider these matrices further.

Among the matrices representing homeomorphisms on a genus 3 surface of word length 6, we examined those that satisfied the Casson-Bleiler Criteria but that have no hyperbolic eigenvalues, of which there were 11, by examining their characteristic polynomials and their generators.

Our first thought was to consider their decomposition into Dehn Twists. This was when we realized that for any pseudo-Anosov in genus 3, there will be at least two factor swaps. If there is not, then it will be reducible. This means that there are at least 30 Dehn twists comprising each pseudo-Anosov, which would be very difficult to visualize.

In an effort to simplify these lengthy combinations of Dehn twists, I considered which Dehn twists would commute with each other, and composed the following commutativity table. A ✓ denotes that two Dehn twists commute with each other.

	T_{a_i}	T_{b_i}	$T_{a_i}^{-1}$	$T_{b_i}^{-1}$	T_{a_j}	T_{b_j}	$T_{a_j}^{-1}$	$T_{b_j}^{-1}$
T_{a_i}	✓		✓	✓	✓		✓	✓
T_{b_i}		✓	✓	✓		✓	✓	✓

Unfortunately, this did not contribute to any significant decrease in length of the combinations of Dehn twists. However it did spark my curiosity in whether or not characteristics of matrices (Casson-Bleiler, hyperbolic eigenvectors, etc.) could be determined by the generating combination of Burkhardt matrices. This will be left to future work.

Once we had these sets of Casson-Bleiler pseudo-Anosov matrices, we tested for friends among the matrices of the same word length in genus 3. It has been proven that there are no solitary matrices among genus 2 surfaces. To do this, we checked for friends using the following criteria discussed in the previous section.

1. If ϕ and ψ are friends, then the affine eigenvectors of ϕ and ψ span the same symplectic plane.

2. If ϕ and ψ are friends, and $\{\lambda, \lambda^{-1}\}$ and $\{\mu, \mu^{-1}\}$ are affine eigenvalues of A_ϕ and A_ψ respectively, then eigenvalue $(\lambda^{-1} + \lambda)$ of $(A_\phi^{-1} + A_\phi)$ and $(\mu^{-1} + \mu)$ of $(A_\psi^{-1} + A_\psi)$ share common eigenspaces.

At first, we found that all of the matrices had friends. However, after a bit of deliberation, we realized that this was due to the fact that for every matrix we considered, we were also considering its inverse, which would always constitute a friendship. Thus, we took out all matrix inverses, both cutting down significantly on computation time and giving us more interesting results.

As there were no Casson-Bleiler pseudo-Anosovs for word lengths 3 or 4 in genus 3, we did not test these word lengths for friends. For word length 5, testing for friends with the first criteria given above, we found that the affine eigenvectors for 54/222 matrices did not share a symplectic plane with the affine eigenvectors of any of the other matrices in word length 5. We found the same number of matrices with no friends by the second criteria.

In word length 6, we found that 378 matrices were not friends with any of the other 1,928 unique matrices by both the first and second criteria, with 518 of these 1,928 having no affine eigenvectors. When the lists of friendless matrices among word length 5 and word length 6 were concatenated and run through the code, no additional friendships were found.

It is important to note these matrices may have friends in Casson-Bleiler matrices with longer word lengths, or the may have friends that are pseudo-Anosov, but do not satisfy the Casson-Bleiler Criteria. Thus, we cannot state that these are solitary pseudo-Anosovs; however, it is possible that there are solitary pseudo-Anosovs among these matrices.

The next step in our work would be to produce an algorithm for constructing the translation representations from affine eigenvectors. This would help us to investigate why these matrices have no friends among their word length, and perhaps we could construct friends from these translation surfaces in order to rule out matrices that are not solitary.

One could also work to make the Mathematica code more efficient. Computational time and memory was a limiting factor on what we could do, as the number of matrices generated the five generators and their inverses for genus 3 is already 1,000,000 matrices for word length 6, increasing tenfold for each added word. By eliminating calculations for matrix representations, we could expand the number of matrices that we would be able to consider.

References

[1] Fulton, William. *Algebraic Topology: A First Course. New York: Springer-Verlag 1995. Print.*

[2] Guillemin, Victor, and Alan Pollack. *Differential Topology. Englewood Cliffs: Prentice Hall 1974. Print.*

[3] Thurston, William P. *On the geometry and dynamics of diffeomorphisms of surfaces. Bull. Amer. Math. Soc. (N.S.) 19 (1988), no. 2, 417–431.*

[4] Farb, Benson, and Dan Margalit. *A Primer on Mapping Class Groups. 5.0th ed. Princeton: Princeton UP, 2012. Print.*

[5] Hatcher, Allen. *Algebraic Topology. Cambridge: Cambridge UP, 2002. Print.*

[6] Gutkin, Eugene, and Judge, Chris. *Affine mappings of translation surfaces: geometry and arithmetic. Duke Math. J. 103 (2000), no. 2, 191–213.*

Klein Bottle Constructions

James Dix
UT Austin

Advisor: Jeffrey Meier
Indiana University

September 14, 2016

Abstract

In this report we consider knotted Klein bottles in S^4 generated from an inversion of a knot. We use this construction to prove the existence of a knotted Klein bottle which does not decompose as a connected sum of an unkotted projective plane and a knotted projective plane.

1 Background

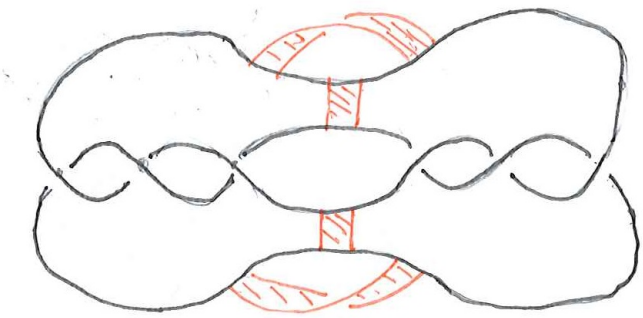
The theory of knotted surfaces is analogous to classical knot theory; it studies the smooth embeddings of a surface in S^4 up to smooth isotopy. A main reason this topic can be difficult to analyze is that the surfaces and isotopies of the surfaces in S^4 are much harder to visualize. One way to describe a knotted surface Σ is to use a banded link diagram.

A banded link diagram is an unlink in $\mathbb{R}^3$ with bands attached so that surgering along these bands gives another unlink. To construct a surface from this, view the banded link in $\mathbb{R}^3$ as a 3-dimensional hyperplane cross-section of a surface in $\mathbb{R}^4$. This cross section splits $\mathbb{R}^4$ into $\mathbb{R}^4_+$ and $\mathbb{R}^4_-$. Disks can be attached in $\mathbb{R}^4_-$ to the pre-surgery unlink in the banded diagram and similarly disks can be attached in $\mathbb{R}^4_+$ to the post-surgery unlink. The result of this gluing is a surface in $\mathbb{R}^4$. Using Morse theory, one can show every knotted surface can be specified by a banded link. The critical points of index of the Morse function 0 and 2 correspond to the disks attached in $\mathbb{R}^4_-$ and $\mathbb{R}^4_+$, while the critical points of index 1 correspond to the bands in the banded link diagram. Thus attaching the correct bands is a matter of finding the saddle points of the surface. The knot group $\pi_1(S^4 \backslash \Sigma)$ can be computed from this diagram using a form of the Wirtinger presentation.

The following are some interesting conjectures about knotted surfaces which were relevant to our project:

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061

Figure 1: A banded link diagram



Conjecture 1 (Unkotting Conjecture)**.** Any orientable surface with knot group $\mathbb{Z}$ or any non-orientable surface with knot group $\mathbb{Z}_2$ is unkotted.

Conjecture 2. Any knotted projective plane can be decomposed into the connected sum of an unkotted projective plane and a knotted sphere.

2 Construction

A half-spun Klein bottle can be constructed using a knot $K \subset B^3$ and an inversion of K (an orientation-reversing isotopy). A surface Σ' in $B^3 \times I$ is created applying the inversion as one proceeds along the interval. Associating $B^3 \times \{0\}$ with $B^3 \times \{1\}$ gives a Klein bottle embedded in $B^3 \times S^1$. $B^3 \times S^1$ can then be embedded in the a standard way into S^4 to create a knotted Klein bottle.

The banded link diagram created by this mirrors the diagram created by creating a knotted torus or a knotted sphere via spinning.

The Klein bottle construction can be thought of as removing the top half of the spun torus and applying the inversion before re-gluing. For example, the banded link for the unkotted Klein bottle can be thought of this way. (Insert diagrams)

When the knot inversion is a simple rotation, the banded link diagram can be used to show that an unkotted projective plane can be removed as a connected summand of the spun Klein bottle. The result is a decomposition of the spun Klein bottle into a knotted projective plane and an unkotted projective plane. In fact, this knotted projective matches a construction given by Price and Roseman, up to a change in Euler number arising from a choice of cross-cap.

Price and Roseman prove that two different inversions of the trefoil can produce different knotted projective planes. These two inversions also produce different knotted Klein bottles. Interestingly, the inversions are the same when

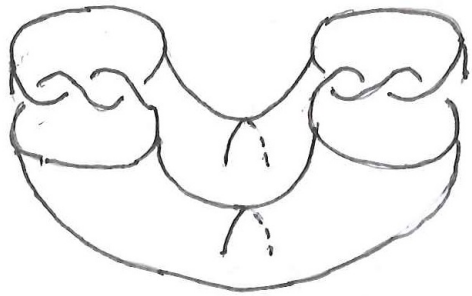


Figure 2: Bottom half of a spun trefoil torus

considering the trefoil inside S^3. A different construction of the same surface provides some insight into the difference created the two similar inversions.

Start this time with a knot $K \subset S^3$ and an inversion of K, and repeat the same construction as before to create a surface embedded in $S^3 \times S^1$. The two inversions of the trefoil will now produce the same knotted surface Σ inside of $S^3 \times S^1$, though they do not when embedded in $B^3 \times S^1$.

To get from a surface Σ embedded in $S^3 \times S^1$ to a surface in S^4, one can surger along a $B^3 \times S^1$ in $S^3 \times S^1$ which wraps once along $S^3 \times S^1$. Surgery involves the removal of $B^3 \times S^1$ and replacing it with with $S^2 \times D^2$ in a standard way. So long as the $B^3 \times S^1$ we surger along does not intersect the surface at all, the result of surgery is a Klein bottle embedded in S^4. However, the knotting of the Klein bottle in S^4 depends on the choice of $B^3 \times S^1$ in S^4.

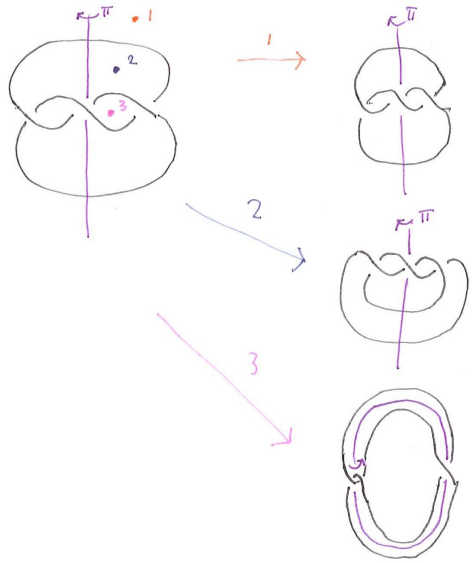
One way to choose a $B^3 \times S^1$ to surger along is to look at the immersed surface in S^3 cut out by the inversion of K. Choosing a $B \cong B^3$ in S^3 that does not intersect this immersed surface means that $B \times S^1$ will not intersect Σ, thus one can surger along it. The different choices of B correspond to the different surfaces generated by inverting the trefoil. To construct the banded link diagram, draw a knot diagram for K in S^3 with B at the point at infinity while keeping track of the inversion.

With the surgery point at infinity, the entire surface Σ can be placed in a $B^3 \times S^1$ that is embedded in the standard way into S^4. Thus the banded link diagram is constructed in the exact same way as described before. Another way to think of this is that the previous banded link construction implicitly had a surgery at infinity.

3 Non-decomposable Surfaces

The third surgery point in the figure above gives a more interesting surface. When constructing the banded link diagram, moving the surgery point out to infinity causes the inversion to no longer act like a rotation in B^3. As a result,

Figure 3: Surgery points for the trefoil



the banded link diagram no longer easily decomposes to pull out an unkotted projective plane as a connected summand. This specific surface can be shown to be trivial and as such can be decomposed into the connected sum of two unkotted projective planes

However for more complicated knots it's possible that a non-rotation inversion produces a knotted Klein bottle without a trivial projective plane summand. In fact, this is the case with the surface generated by the knot in figure 4 and inversion has this property.

Example 3.1. Using the Wirtinger presentation, the knot group of this Klein bottle is given by

$$\begin{aligned} G = \langle a, b, c | a^4, a^2 = b^2 = c^2, aba = b, bab = a, aca = cac, bcb = cbc, \\ a^{-1}ca = c^{-1}ac, b^{-1}cb = c^{-1}bc \rangle \end{aligned} \tag{1}$$

If an unkotted projective plane could be pulled out of the Klein, then all meridians would square to 1 since all merdians are conjugate and the meridian of the unkotted projective plane squares to 1. Letting $H = G/\langle a^2 \rangle$, then $G \cong H$ if a trivial projective plane can be pulled out of the Klein bottle. However, entering the groups into the Magma computer program lets us know that $|G| = 48$ and $|H| = 24$. Thus an unkotted projective plane cannot be pulled out. Other decompositions such as splitting into two knotted projective planes can be tested using $H_2(G)$ but as of now we do not know this group.

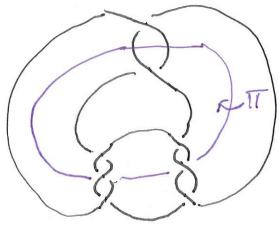


Figure 4: Inversion that generates a non-decomposable Klein bottle

More surfaces with the same indecomposability can be generated by simply adding twists to the bottom part of the knot. These types of inversions can also be used to generate surfaces with knot group $\mathbb{Z}_2$ which may not decompose. This would be a counterexample to the unkotting conjecture, however we have no method to prove indecomposability without using the knot group.

On Surfaces that are Intrinsically Surfaces of Revolution

Daniel Freese
Advisor: Matthias Weber

September 14, 2016

Abstract

We consider surfaces in Euclidean space parametrized on an annular domain such that the first fundamental form and the principal curvatures are rotationally invariant, and the principal curvature directions only depend on the angle of rotation (but not the radius). Such surfaces generalize the Enneper surface. We show that they are necessarily of constant mean curvature, and that the rotational speed of the principal curvature directions is constant. We classify the minimal case. The (non-zero) constant mean curvature case has been classified by Smyth.

[1]

[1]This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

1 Introduction

We begin this paper with some background on surface theory and briefly describe some of the concepts we cover.

Definition 1.1. Let $\mathcal{U} \subset \mathbb{R}^2$ be an open set. A parametrized surface is an immersion $f : \mathcal{U} \to \mathbb{R}^3$.

We consider the parameters $(u, v) \in \mathcal{U}$.

Recall that an immersion is a mapping whose Jacobian matrix Jf has maximal rank, that is, whose columns are linearly independent. When considering a surface parametrization f, this reduces to the vectors $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$ being linearly independent at every point.

Definition 1.2. The tangent space of $\mathbb{R}^3$ at a point $P \in \mathbb{R}^3$ is defined to be the vector space $\{P\} \times \mathbb{R}^3$ with vector addition and scalar multiplication defined as follows:

$$(P, X) + (P, Y) = (P, X + Y), \qquad c(P, X) = (P, cX) \;\; \forall\, X, Y \in \mathbb{R}^3,\; c \in \mathbb{R}\ .$$

Definition 1.3. Let $f : \mathcal{U} \to \mathbb{R}^3$ be a parametrized surface and $p \in \mathcal{U}$. The vectors $\frac{\partial f}{\partial u}|_p$ and $\frac{\partial f}{\partial v}|_p$ form a basis for a 2-dimensional subspace of $T_{f(p)}\mathbb{R}^3$, which we call the tangent plane $T_p f$ of f at p.

Note that for a point $p \in \mathcal{U}$, $T_p f$ is isomorphic to $T_p\mathcal{U}$.

Definition 1.4. Given an immersion $f : \mathcal{U} \subset \mathbb{R}^2 \to \mathbb{R}^3$, the isomorphism $df|_p : T_p\mathcal{U} \to T_p f$ defined by

$$df|_p(X) = J_p f X$$

is called the differential or push-forward of f.

The differential relates vectors in the tangent plane of f to the tangent space of $\mathcal{U}$. We call $d_f X$ the directional derivative of f in the X-direction, and also denote it $d_X f$.

Definition 1.5. The symmetric bilinear form $I_p : T_p\mathcal{U} \to \mathbb{R}$ defined by

$$I_p(X, Y) = \langle df|_p X, df|_p Y \rangle \qquad \forall X, Y \in T_p\mathcal{U}$$

is called the first fundamental form of f at p.

The first fundamental form is represented as the 2×2 matrix

$$I = \begin{pmatrix} \langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \rangle & \langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \rangle \\ \langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial u} \rangle & \langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \rangle \end{pmatrix} .$$

The first fundamental form allows us to calculate angles between tangent vectors, as well as their norms. It also allows us to calculate arc length and surface area. The intrinsic geometry of a surface is the geometry that depends only on the first fundamental form. Thus, if two surfaces that have the same first fundamental form, they have intrinsic geometry. For example, they have the same concept of distance, even if they are different extrinsically. The next notion reveals extrinsic information of the surface.

The unit-normal vector of a surface defined by $N = \frac{\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial u}}{\| \frac{\partial f}{\partial u} \times \frac{\partial f}{\partial u} \|}$ at each point can be viewed as a mapping into $\mathbb{R}^3$, which we call the Gauss map ν.

Definition 1.6. The endomophism $S_p : T_p\mathcal{U} \to T_p f$ defined by

$$S_p X = d_\nu|_p \circ (d_f|_p)^{-1} X \qquad \forall X \in T_p f$$

is called the shape operator of f at p. It is represented by its standard matrix.

The shape operator gives us a notion of the curvature of the surface. When studying curves in space we can calculate the curvature of a given curve, which tells us how much the curve bends. At a point p on a surface, the shape operator tells us the curvature of a curve going in each direction in the tangent plane.

The extremal curvatures are called the principal curvatures, and the directions where the curvatures are extremal are called the principal curvature directions. These correspond respectively to the eigenvalues and eigenvectors of the shape operator.

For more detailed background on these concepts, we refer the reader to [4]. As we proceed, we will omit the point-dependence from our notation.

Now we approach the subject of our study.

The Enneper surface (see Figure 1) is given in conformal polar coordinates as

$$f(u,v) = \frac{1}{6}e^u \begin{pmatrix} 3\cos(v) - e^{2u}\cos(3v) \\ -3\sin(v) - e^{2u}\sin(3v) \\ 3e^u\cos(2v) \ . \end{pmatrix}$$

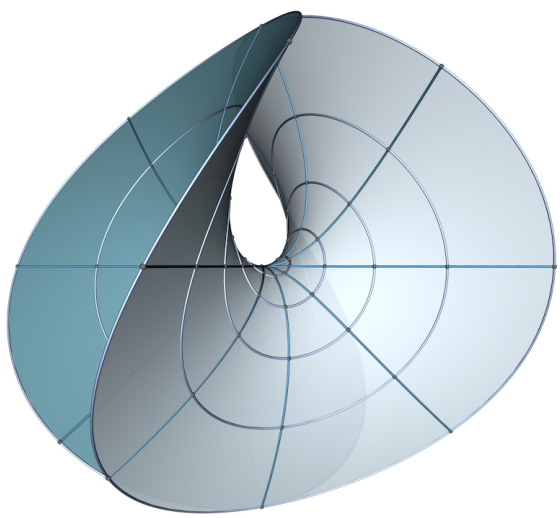


Figure 1: The Enneper Surface

It was discovered in 1871 by Alfred Enneper [2]. Its first fundamental form is given by

$$I = \frac{1}{4}e^{2u}\left(1+e^{2u}\right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ .$$

This means that the Enneper surface is *intrinsically* a surface of revolution (but obviously not extrinsically).

Definition 1.7. An intrinsic surface of revolution is a surface with first fundamental form of the shape

$$I = I_\rho = \rho(u)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ ,$$

where $\rho(u)$ is a positive function.

Of course any surface of revolution is also intrinsically a surface of revolution.

The shape operator of the Enneper surface is also rather special:

$$S = \frac{4}{(1+e^{2u})^2}\begin{pmatrix} \cos(2v) & -\sin(2v) \\ -\sin(2v) & -\cos(2v) \end{pmatrix} = R^{-v}\begin{pmatrix} \frac{4}{(1+e^{2u})^2} & 0 \\ 0 & -\frac{4}{(1+e^{2u})^2} \end{pmatrix} R^{v}$$

where

$$R^v = \begin{pmatrix} \cos(v) & -\sin(v) \\ \sin(v) & \cos(v) \end{pmatrix}$$

is the counterclockwise rotation by v. This is in contrast to the shape operator of a surface of revolution which always takes diagonal form in polar coordinates. It is, however, rather special, because the principal curvature directions rotate with constant speed independent of u and the principal curvatures are independent of v.

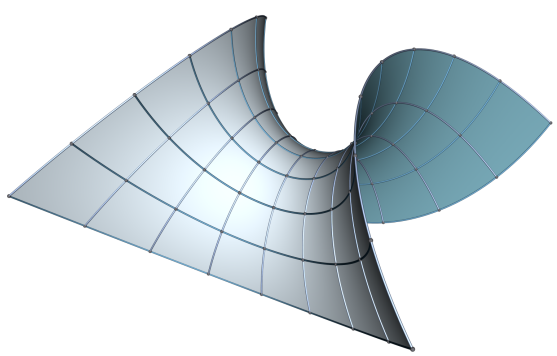


Figure 2: The Enneper Surface with curvature lines

We are generalizing this property of the Enneper surface by introducing the following concept:

Definition 1.8. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be a C^1-function. We say a surface has twist α if its shape operator is of the form

$$S = R^{-\alpha(v)}\begin{pmatrix} \lambda_1(u) & 0 \\ 0 & \lambda_2(u) \end{pmatrix} R^{\alpha(v)} \ . \tag{1}$$

Note that this precisely means that the principal curvature directions are independent of u, and the principal curvatures are independent of v.

In summary, the Enneper surface is an example of an intrinsic surface of revolution with twist $\alpha(v) \equiv v$. A standard surface of revolution, on the other hand, has twist $\alpha(v) \equiv 0$.

Note that while this is a generalization of the minimal Enneper surface, we make no assumption on the principal curvatures, besides their independence of v.

In this paper, we consider these general intrinsic surfaces of revolution with twist α. Our goal is to find conditions on the conformal factor $\rho(u)$ and the twist function $\alpha(v)$ necessary for these surfaces to exist and to construct and classify these surfaces.

Now we can formulate our main theorem, which is a consequence of the Codazzi equations.

Theorem 1.9. *Let Σ be an intrinsic surface of revolution with twist function α. Assume that α is not identically equal to 0 or any other integral multiple of $\pi/2$, on any open interval. Assume furthermore that the surface has no open set of umbilic points. Then Σ has constant mean curvature, and the twist function is linear $\alpha(v) = av$.*

Constant mean curvature surfaces that are intrinsic surfaces of revolution have been studied by Smyth, see [5]. Thus our result complements Smyth's result by replacing his assumption about constant mean curvature with a geometric assumption.

I order to begin a complete classification, we will invoke Bonnet's theorem to prove:

Theorem 1.10. *Let Σ be an intrinsic surface of revolution of constant mean curvature $H = \lambda_1 + \lambda_2$, first fundamental form I_ρ for $\rho : (u_1, u_2) \to \mathbb{R}^{>0}$, and linear twist $\alpha(v) = av$. Then ρ satisfies the differential equation*

$$\rho'(u)^2 - \rho(u)\rho''(u) = \frac{1}{4}H^2\rho(u)^4 - b^2e^{4au} \tag{2}$$

for a constant b.

Vice versa, given H, and $\alpha(v) = av$, a constant b and ρ satisfying Equation (2), define

$$\lambda_{1,2}(u) = \frac{1}{2}H \pm b\frac{e^{2au}}{\rho(u)^2} \ .$$

Then the first fundamental form I_ρ and the shape operator S given by Equation (1) satisfy the Gauss- and Codazzi equations and thus define an intrinsic surface of revolution with constant twist $\alpha(v) = av$ and constant mean curvature H.

In the special case of minimal surfaces, we can achieve a complete classification.

Theorem 1.11. *Let Σ be an intrinsic surface of revolution that is also minimal with constant twist $\alpha(v) = av$ with $a > 0$. Then Σ belongs to an explicit 2-parameter family of minimal surfaces with Weierstrass data given by*

$$G(w) = -\frac{1}{A}w^B \qquad \text{and} \qquad dh = \frac{1}{2B}w^{2a-1}\,dw \ ,$$

with parameters A and B.

In the case that the twist function is $\alpha(v) \equiv a = 0$, we prove:

Theorem 1.12. *Given a conformal factor $\rho(u)$ on an interval $u_1 < u < u_2$ and a constant c such that $c^2\rho(u) > |\rho'(u)|$ for all $u_1 < u < u_2$. Then there is an intrinsic surface of revolution defined on the domain $(u_1, u_2) \times \mathbb{R}$ with first fundamental form I_ρ and twist $\alpha(v) \equiv 0$. Moreover, this surface can be realized as an actual surface of revolution in $\mathbb{R}^3$ of the form*

$$f(u, v) = (g(u)\cos(cv), g(u)\sin(cv), h(u))$$

with suitable functions $g, h : (u_1, u_2) \times \mathbb{R}$.

The paper is organized as follows:

- In section 2, we compute the Gauss- and Codazzi equations for intrinsic surfaces of revolution, reduce them to a single ODE for ρ, and prove Theorems 1.9 and 1.10.
- In section 3, we specialize this equation to the minimal case, integrate the surface equations, find the Weierstrass representation of the surfaces, prove Theorem 1.11, and give examples.

- In section 4, we briefly discuss the constant mean curvature case by connecting our approach to Smyth's. While we are not able to find explicit solutions for the Smyth surfaces, we can find numerical solutions and make images.

- In section 5, we consider the case of twist 0, prove Theorem 1.12, and show that sectors of the Enneper surface are isometric to sectors (with different angle) of surfaces of revolution.

2 Gauss- and Codazzi equations for intrinsic surfaces of revolution with twist $a > 0$

In order to determine when the first fundamental form and shape operator of an intrinsic surface of revolution with twist function α are induced by an actual surface in $\mathbb{R}^3$, we make use of an important theorem of surface theory, Bonnet's theorem. But first, we'll introduce some of the machinery behind this theorem.

We first introduce the covariant derivative. For a surface f and tangent vector fields X, Y on f, the covariant derivative, or Levi-Civita connection, $D_X Y$ measures how Y varies when it goes along a curve in the X-direction. However, it also acts on general tensor fields on the surface. We present some of its properties, which are useful in the derivations that follow.

Lemma 2.1. *Given a surface f, for all tangent vector fields X, Y, Z and endomorphisms F, G of the tangent plane and for every scalar function ϕ along f , we have*

$$
\begin{aligned}
&(1)\ D_{\phi X} Y = \phi D_X Y,\\
&(2)\ D_X(\phi Y) = \phi D_X Y + (d_X \phi) Y,\\
&(3)\ D_X I(Y, Z) = I(D_X Y, Z) + I(Y, D_X Z)\\
&(4)\ D_X Y - D_Y X = [X, Y] = d_X Y - d_Y X,\\
&(5)\ D_X(F(Y)) = (D_X F) Y + F(D_X Y)\\
&(6)\ (D_X F G) = (D_X F) G + F(D_X G).
\end{aligned}
$$

where d_X is the directional derivative in the X-direction.

We refer to these properties by their numbers in what follows.

The Gauss- and Codazzi Equations are two essential equations of surfaces in Euclidean space. These can be written in terms of the covariant derivative, as we will show when we derive their simplified forms.

Bonnet's theorem states that, given any positive definite first fundamental form and shape operator satisfying the Gauss- and Codazzi equations, there exists a unique surface in $\mathbb{R}^3$ with these first fundamental form and shape operator, up to rigid motion. Thus, we need only derive the Gauss- and Codazzi equations for intrinsic surfaces of revolution with twist α and solve them to determine the necessary conditions on the first fundamental form and shape operator.

In order to derive the Gauss- and Codazzi equations we first determine the relevant covariant derivatives. Much of this preparation is standard.

Introduce

$$U = \frac{1}{\rho(u)}\frac{\partial}{\partial u} \qquad V = \frac{1}{\rho(u)}\frac{\partial}{\partial v} \tag{3}$$

as the normalized coordinate vector fields. Their orthonormality with respect to I_ρ follows from the orthonormality of $\frac{\partial}{\partial u}$ and $\frac{\partial}{\partial v}$ with respect to the standard inner product and the definition of I_ρ. Since they form a basis of $\mathbb{R}^2$, our calculations only involve these two vectors. First, we have

Lemma 2.2. *The Levi-Civita connection of the first fundamental form I_ρ is given by*

$$\begin{aligned} D_U U &= 0 & D_U V &= 0 \\ D_V U &= \frac{\rho'(u)}{\rho(u)^2} V & D_V V &= -\frac{\rho'(u)}{\rho(u)^2} U \ . \end{aligned}$$

Proof. By the v-invariance of the first fundamental form, the curves $s \mapsto (s, v)$ are geodesics, and U is the unit-length tangent vector to these curves. This implies $D_U U = 0$. Next $V = R^{\pi/2} U$ and intrinsic rotations are parallel, so that $D_U V = 0$ as well.

Using that D is torsion-free and metric, we compute

$$\begin{aligned}
D_V U &= D_U V + [V, U] \qquad \textit{by } (4) \\
&= 0 + d_V U - d_U V \\
&= \frac{\partial}{\rho(u)\partial v}\frac{\partial}{\rho(u)\partial u} - \frac{\partial}{\rho(u)\partial u}\frac{\partial}{\rho(u)\partial v} \\
&= -\frac{1}{\rho(u)}\frac{\partial}{\partial u}\frac{\partial}{\rho(u)\partial v} \\
&= \frac{1}{\rho(u)}\frac{\rho'(u)}{\rho(u)^2}\frac{\partial}{\partial v} \\
&= \frac{\rho'(u)}{\rho(u)^2} V
\end{aligned}$$

and

$$\begin{aligned}
D_V V &= I(D_V V, U)U + I(D_V V, V)V \\
&= d_V I(V, U)U - I(V, D_V U)U + 0 \qquad \textit{by } (3) \\
&= d_V(0)U - I\left(V, \frac{\rho'(u)}{\rho(u)^2}V\right)U = -\frac{\rho'(u)}{\rho(u)^2}U,
\end{aligned}$$

where $I(D_V V, V)V = 0$ because

$$\begin{aligned}
I(D_V V, V) &= d_V I(V, V)V - I(V, D_V V)V \qquad \textit{by } (3) \\
&= d_V I(V, V)V - I(V, D_V V)V \\
&= d_V(1)V - I(V, D_V V)V = -I(D_V V, V)V.
\end{aligned}$$

□

Lemma 2.3. *The Gauss equation is equivalent to*

$$\lambda_1(u)\lambda_2(u) = \frac{\rho'(u)^2 - \rho(u)\rho''(u)}{\rho(u)^4} \ .$$

Proof. The Gauss equation gives us:

$$I(\mathcal{R}(U,V)V,U) = det(S) = \lambda_1(u)\lambda_2(u)$$

where $\mathcal{R}$ is the Riemann curvature tensor given by

$$\mathcal{R}(X,Y) = D_X D_Y - D_Y D_X - D_{[X,Y]}$$

for all tangent vector fields X, Y.

Let us calculate $\mathcal{R}(U,V)V$.

$$\begin{aligned}
\mathcal{R}(U,V)V &= D_U D_V V - D_V D_U V - D_{[U,V]}V \\
&= D_U\left(-\frac{\rho'(u)}{\rho(u)^2}U\right) - D_V(0) - D_{-\frac{\rho'(u)}{\rho(u)^2}V}V \\
&= -\frac{\rho'(u)}{\rho(u)^2}D_U U - d_U\left(\frac{\rho'(u)}{\rho(u)^2}\right)U + \frac{\rho'(u)}{\rho(u)^2}D_V V \qquad \textit{by } (2) \ \& \ (1) \\
&= 0 - \frac{1}{\rho(u)}\frac{\partial}{\partial u}\left(\frac{\rho'(u)}{\rho(u)^2}\right)U + \frac{\rho'(u)}{\rho(u)^2}\left(-\frac{\rho'(u)}{\rho(u)^2}U\right) \\
&= -\frac{1}{\rho(u)}\left(\frac{\rho''(u)\rho(u)^2 - \rho'(u)(2\rho(u)\rho'(u))}{\rho(u)^4}\right)U - \frac{\rho'(u)^2}{\rho(u)^4}U \\
&= \frac{-\rho''(u)\rho(u) + 2\rho'(u)^2 - \rho'(u)^2}{\rho(u)^4}U \\
&= \frac{\rho'(u)^2 - \rho(u)\rho''(u)}{\rho(u)^4}U \ ,
\end{aligned}$$

and the claim follows. □

In order to derive the Codazzi equations, we need to compute a few more covariant derivatives regarding some endomorphisms of the tangent plane.

Lemma 2.4. *The covariant derivatives of the twist rotation are given by*

$$D_U R^{\alpha(v)} = 0 \qquad \textit{and} \qquad D_V R^{\alpha(v)} = \frac{\alpha'(v)}{\rho(u)}R^{\alpha(v)+\frac{\pi}{2}} \ .$$

Proof. The first equations follows because intrinsic rotations by a constant

angle are parallel and α is independent of u. For the second, we have,

$$\begin{aligned}
(D_V R^{\alpha(v)})U &= D_V(R^{\alpha(v)}U) - R^{\alpha(v)}D_V U \qquad \textit{by } (5) \\
&= D_V(\cos(\alpha(v))U + \sin(\alpha(v))V) - \frac{\rho'(u)}{\rho(u)^2}R^{\alpha(v)}V \\
&= d_V(\cos(\alpha(v))U + \cos(\alpha(v)D_V U + d_V(\sin(\alpha(v)))V + \sin(\alpha(v))D_V V \\
&\quad - \frac{\rho'(u)}{\rho(u)^2}(-\sin(\alpha(v))U + \cos(\alpha(v))V) \qquad \textit{by } (2) \\
&= \frac{1}{\rho(u)}\left(\frac{\partial}{\partial v}\cos(\alpha(v)U + \frac{\partial}{\partial v}\sin(\alpha(v)V\right) + \cos(\alpha(v)D_V U + \sin(\alpha(v))D_V V \\
&\quad - (\sin(\alpha(v))D_V V + \cos(\alpha(v))D_V U) \\
&= \frac{1}{\rho(u)}(-\alpha'(v)\sin(\alpha(v)U + \alpha'(v)\cos(\alpha(v))V) \\
&= \frac{\alpha'(v)}{\rho(u)}\left(\cos\left(\alpha(v) + \frac{\pi}{2}\right)U + \sin\left(\alpha(v) + \frac{\pi}{2}\right)V\right) \\
&= \frac{\alpha'(v)}{\rho(u)}R^{\alpha(v)+\frac{\pi}{2}}U \ ,
\end{aligned}$$

and likewise $D_V R^{\alpha(v)}V = \frac{\alpha'(v)}{\rho(u)}R^{\alpha(v)+\frac{\pi}{2}}V$. □

Lemma 2.5. *The covariant derivatives of the eigenvalue endomorphism Λ are given by*

$$\begin{aligned}
D_U\Lambda(u) &= \frac{1}{\rho(u)}\Lambda'(u) \\
(D_V\Lambda(u))U &= \frac{\rho'(u)}{\rho^2(u)}(\lambda_1(u) - \lambda_2(u))V \\
(D_V\Lambda(u))V &= \frac{\rho'(u)}{\rho^2(u)}(\lambda_1(u) - \lambda_2(u))U \ .
\end{aligned}$$

Proof. First note that $\Lambda U = \lambda_1 U$ and $\Lambda V = \lambda_2 V$. We derive the first

equation as follows:

$$\begin{aligned}
(D_U\Lambda)U &= D_U(\Lambda U) - \Lambda D_U U \\
&= D_U(\lambda_1 U) - \Lambda 0 \\
&= (d_U\lambda_1)U + \lambda_1 D_U U \qquad \textit{by } (2) \\
&= \frac{1}{\rho}\frac{\partial\lambda_1}{\partial u}U + 0 \\
&= \frac{1}{\rho}\lambda_1 U = \frac{1}{\rho(u)}\Lambda'(u)U \ ,
\end{aligned}$$

and likewise $(D_U\Lambda)V = \frac{1}{\rho(u)}\Lambda'(u)V$.

For the second, we compute

$$\begin{aligned}
(D_V\Lambda)U &= D_V(\lambda_1 U) - \Lambda D_V U \\
&= \lambda_1 D_V U - \lambda_2 D_V U \\
&= \frac{\rho'(u)}{\rho^2(u)}(\lambda_1(u) - \lambda_2(u))V \ .
\end{aligned}$$

The third equation is proven the same way. □

Lemma 2.6. *The covariant derivatives of the shape operator are given by*

$$\begin{aligned}
(D_U S)V &= \frac{1}{\rho(u)}R^{-\alpha(v)}\Lambda'(u)R^{\alpha(v)}V \\
(D_V S)U &= (\lambda_1(u) - \lambda_2(u))\frac{\rho'(u) - \rho(u)\alpha'(v)}{\rho^2(u)}R^{-2\alpha(v)}V \ .
\end{aligned}$$

Proof. In the statement, we have indicated the dependence of each function by their respective variables. To improve legibility, we will drop the variables in the computations below. Observe, however, that derivatives like α' and ρ' are always taken with respect the proper variables.

For both equations we differentiate the definition of S according to the

product rule, and then use the lemmas above. For the first equation, we have

$$\begin{aligned}
(D_U S)V &= D_U(R^{-\alpha}\Lambda R^{\alpha})V \\
&= (D_U(R^{-\alpha})\Lambda R^{\alpha} + V + R^{-\alpha}D_U(\Lambda)R^{\alpha} + R^{-\alpha}\Lambda(D_U R^{\alpha}))V \qquad \textit{by } (6) \\
&= (0 + R^{-\alpha}D_U(\Lambda)R^{\alpha} + 0)V \\
&= \frac{1}{\rho(u)}R^{-\alpha(v)}\Lambda'(u)R^{\alpha(v)}V \ .
\end{aligned}$$

For the second, we obtain

$$\begin{aligned}
(D_V S)U &= D_V(R^{-\alpha}\Lambda R^{\alpha})(U) \\
&= \left((D_V R^{-\alpha})\Lambda R^{\alpha} + R^{-\alpha}(D_V\Lambda)R^{\alpha} + R^{-\alpha}\Lambda(D_V R^{\alpha})\right)(U) \\
&= R^{-\alpha}\left(-\frac{\alpha'}{\rho}R^{\frac{\pi}{2}}\Lambda R^{\alpha} + (D_V\Lambda)R^{\alpha} + \frac{\alpha'}{\rho}\Lambda R^{\alpha+\frac{\pi}{2}}\right)(U) \\
&= R^{-\alpha}(-\frac{\alpha'}{\rho}R^{\frac{\pi}{2}}\Lambda(\cos(\alpha)U + \sin(\alpha)V) + \\
&\qquad (D_V\Lambda)(\cos(\alpha)U + \sin(\alpha)V) + \frac{\alpha'}{\rho}\Lambda(-\sin(\alpha)U + \cos(\alpha)V)) \\
&= R^{-\alpha}(-\frac{\alpha'}{\rho}R^{\frac{\pi}{2}}(\lambda_1\cos(\alpha)U + \lambda_2\sin(\alpha)V) + \\
&\qquad \frac{\rho'}{\rho^2}(\lambda_1 - \lambda_2)(\cos(\alpha)V + \sin(\alpha)U) + \frac{\alpha'}{\rho}(-\lambda_1\sin(\alpha)U + \lambda_2\cos(\alpha)V)) \\
&= (\lambda_2 - \lambda_1)(-\frac{\alpha'}{\rho} + \frac{\rho'}{\rho^2})R^{-\alpha}(-\sin(\alpha)U - \cos(\alpha)V) \\
&= (\lambda_1 - \lambda_2)(-\frac{\alpha'}{\rho} + \frac{\rho'}{\rho^2})R^{-2\alpha}V
\end{aligned}$$

□

Corollary 2.7. *The Codazzi equations are equivalent to*

$$\begin{aligned}
&(\lambda_1' + \lambda_2')\sin(\alpha)\cos(\alpha) = 0 \\
&\frac{1}{\rho}(\lambda_1 - \lambda_2)(\rho' - \rho\alpha') = -\lambda_1'\sin^2(\alpha) + \lambda_2'\cos^2(\alpha)
\end{aligned}$$

Proof. The Codazzi equations state that $(D_X S)Y = (D_Y S)X$ for any pair of tangent vectors X and Y. As we are in dimension 2 and the equation is symmetric, it suffices to verify this for $X = U$ and $Y = V$. By the previous theorem, this is equivalent to

$$\rho\Lambda' R^{\alpha} V = (\lambda_1 - \lambda_2)(\rho' - \rho\alpha') R^{-\alpha} V$$

Pairing both sides with $I(\cdot, R^{-\alpha}U)$ and $I(\cdot, R^{-\alpha}V)$ respectively gives

$$\begin{aligned}
&\rho I(\Lambda' R^{\alpha} V, R^{-\alpha} U) = 0 \\
&\rho I(\Lambda' R^{\alpha} V, R^{-\alpha} V) = (\lambda_1 - \lambda_2)(\rho' - \rho\alpha')
\end{aligned}$$

as rotations preserve I, which gives $I(R^{-\alpha}V, R^{-\alpha}U) = I(V, U) = 0$ and $I(R^{-\alpha}V, R^{-\alpha}V) = 1$.

The first equation simplifies to

$$\begin{aligned}
0 &= I(-\Lambda'(\sin(\alpha)U + \cos(\alpha)V), \cos(\alpha)U - \sin(\alpha)V) \\
&= I(-\lambda_1' \sin(\alpha)U + \lambda_2' \cos(\alpha)V), \cos(\alpha)U - \sin(\alpha)V) \\
&= -(\lambda_1' + \lambda_2') \sin(\alpha)\cos(\alpha),
\end{aligned}$$

giving us the first result, and the second to

$$\begin{aligned}
\frac{1}{\rho}(\lambda_1 - \lambda_2)(\rho' - \rho\alpha') &= I(\Lambda' R^{\alpha} V, R^{-\alpha} V) \\
&= I(-\lambda_1' \sin(\alpha)U + \lambda_2' \cos(\alpha)V), \sin(\alpha)U + \cos(\alpha)V) \\
&= -\lambda_1' \sin^2(\alpha) + \lambda_2' \cos^2(\alpha)
\end{aligned}$$

as claimed. □

We are now ready to prove Theorem 1.9

Proof. By assumption, the twist function α is not identically equal to an integral multiple of $\pi/2$ on any open interval. By the first Codazzi equation, the mean curvature $H(u) = \lambda_1(u) + \lambda_2(u)$ is constant except possibly at

isolated points. As we assume that H is at least C^1, this implies that H is constant.

Using this simplifies to the second Codazzi equation to

$$\frac{1}{\rho(u)}(2\lambda_1(u) - H)(\rho'(u) - \rho(u)\alpha'(v)) = -\lambda_1'(u)$$

As the right hand side is independent of v, so is the left hand side. This can only be the case if $\alpha'(v)$ is a constant as claimed, or that $H = 2\lambda_1(u)$ on an open interval. In the latter case we have on the same interval that $\lambda_1(u) = \lambda_2(u) = \lambda$ for a constant λ. But this means that this portion of the surface is umbilic, which we have excluded. □

Observe that we have not used the Gauss equations in the above proof. We will now use the second Codazzi equation to eliminate λ_1 and λ_2 from the Gauss equations.

Lemma 2.8. *For $\alpha(v) = av$ and $H = \lambda_1(u) + \lambda_2(u)$ a constant, the second Codazzi equation has the general solution*

$$\lambda_1(u) = \frac{1}{2}H + b\frac{e^{2au}}{\rho(u)^2} \ ,$$

where b is any real number.

Proof. Define

$$\mu(u) = \rho^2(u)\left(\lambda_1(u) - \frac{1}{2}H\right) \ .$$

The second Codazzi equation is then equivalent to

$$\mu'(u) = 2a\mu(u) \ .$$

Integrating and substituting back gives the claim. □

The following corollary proves Theorem 1.10.

Corollary 2.9. *A first fundamental form I_ρ with $rho = \rho(u)$ and shape operator S as in Equation (1) such that $H = \lambda_1(u) + \lambda_2(u)$ is constant and $\alpha(v) = av$ satisfy the Gauss and Codazzi equations if and only if*

$$\rho'(u)^2 - \rho(u)\rho''(u) = \frac{1}{4}H^2\rho(u)^4 - b^2e^{4au} \tag{4}$$

In particular, by Bonnet's theorem, these data determine and intrinsic surface of revolution, and every such surface arises this way.

Proof. This follows by using the explicit solutions for λ_1 and λ_2 from Lemma 2.8 in the Gauss equation from Lemma 2.3, and simplifying. $\square$

To classify all intrinsic surfaces of revolution, we would need to find all solutions to the differential equation (4), and then to integrate the surface equation to obtain a parametrization. We will discuss the solutions of 4 for $H = 0$ in Section 3.

We end this section by carrying out the first integration step of the surface equation, which is quite explicit and shows that special coordinate curves are planar.

Assume that $\rho(u)$ is a solution of 4. To determine the surface parametrization, we will first determine a differential equation for the curve $\tilde{c} = f \circ c$ with $c(s) = (s, 0)$.

Recall from Equations 3 and 2.2 that

$$X(s) = U(s, 0) \qquad \text{and} \qquad Y(s) = V(s, 0)$$

are a parallel frame field along $c(s)$ with respect to the first fundamental form.

Following the proof of Bonnet's theorem, we derive a Frenet-type differential equation for the orthonormal frame $\tilde{X}(s) = dfX(s)$, $\tilde{Y}(s) = dfY(s)$, and $\tilde{N}(s) = \tilde{X}(s) \times \tilde{Y}(s)$.

$$\begin{aligned}
\tilde{X}'(s) &= df\frac{D}{ds}X(s) + \langle \tilde{X}'(s), \tilde{N}(s)\rangle \tilde{N}(s) \\
&= -\langle \tilde{X}(s), \tilde{N}'(s)\rangle \tilde{N}(s) \\
&= -\langle dfX(s), dfS\frac{\partial}{\partial u}\rangle \tilde{N}(s) \\
&= -\rho(s) I(X(s), SX(s))\tilde{N}(s)
\end{aligned}$$

where the second equality is due to the fact that $X(s)$ is parallel and that $\langle \tilde{X}'(s), \tilde{N}(s)\rangle + \langle \tilde{X}(s), \tilde{N}'(s)\rangle = \frac{d}{ds}\langle \tilde{X}(s), \tilde{N}(s)\rangle = 0$. Similarly,

$$\tilde{Y}'(s) = -\rho(s) I(Y(s), SX(s))\tilde{N}(s)$$

Finally,

$$\begin{aligned}\tilde{N}'(s) &= \langle \tilde{N}'(s), \tilde{X}(s)\rangle \tilde{X}(s) + \langle \tilde{N}'(s), \tilde{Y}(s)\rangle \tilde{Y}(s) \\ &= \langle dfS\frac{\partial}{\partial u}, dfX(s)\rangle \tilde{X}(s) + \langle dfS\frac{\partial}{\partial u}, dfY(s)\rangle \tilde{Y}(s) \\ &= I(S\frac{\partial}{\partial u}, X(s))\tilde{X}(s) + I(dfS\frac{\partial}{\partial u}, Y(s))\tilde{Y}(s)\end{aligned}$$

In our case, using the explicit formula for the shape operator and the principal curvatures in terms of ρ and a, b, this simplifies to give the following lemma:

Lemma 2.10.

$$\begin{aligned}\tilde{X}'(s) &= -\left(\frac{e^{2as}b}{\rho(s)} + \frac{1}{2}H\rho(s)\right)\tilde{N}(s) \\ Y'(s) &= 0 \\ \tilde{N}'(s) &= \left(\frac{e^{2as}b}{\rho(s)} + \frac{1}{2}H\rho(s)\right)\tilde{X}(s)\end{aligned}$$

Corollary 2.11. *The space curve $f(s, 0)$ is planar.*

Proof. This is immediate because $\tilde{Y}(s)$ is constant. Note that this only works because $v = 0$. □

This is as far as we can get in the general case. For the minimal case, we will solve the Equation (4) explicitly and be able to integrate the surface equations further.

3 The minimal case

In the minimal case $H = 0$ the differential equation for ρ simplifies to

$$\rho'(u)^2 - \rho(u)\rho''(u) = -b^2 e^{4au} \tag{5}$$

Without much loss of generality, we can assume $b = 1$ by scaling ρ by a positive constant. There is one exception, namely when $b = 0$. In this case, $\lambda_1 = \lambda_2 = 0$, so that the surface is a plane, which we disregard.

Lemma 3.1. *All positive solutions of*

$$\rho'(u)^2 - \rho(u)\rho''(u) = -e^{4au} \tag{6}$$

defined in any open interval are given by

$$\rho(u) = \frac{e^{2au}}{2B}\left(Ae^{Bu} + \frac{e^{-Bu}}{A}\right)$$

for arbitrary $A, B > 0$.

Proof. It is easy to check that ρ satisfies Equation (6). To show that every local solution σ is of this form, it suffices to show that for any fixed real u, the initial values $\sigma(u) > 0$ and $\sigma'(u)$ are equal to the initial data $\rho(u)$ and $\rho'(u)$ for a suitable choice of $A > 0$ and $B > 0$. Then the local uniqueness theorem for ordinary differential equations implies that $\rho = \sigma$ near u and hence everywhere.

To this end, we have to solve

$$\begin{aligned}
\sigma(u) &= \frac{1}{2B}e^{2au}\left(e^{Bu}A + \frac{e^{-Bu}}{A}\right) \\
\sigma'(u) &= \frac{1}{2B}e^{2au}\left(ABe^{Bu} - \frac{Be^{-Bu}}{A}\right) + \frac{1}{B}ae^{2au}\left(e^{Bu}A + \frac{e^{-Bu}}{A}\right)
\end{aligned}$$

for A and B. Surprisingly, this is possible explicitly. The strategy is to solve the first equation for A, choosing the larger solution

$$A = e^{-Bu}\left(Be^{-2au}\sigma(u) + \sqrt{B^2e^{-4au}\sigma(u)^2 - 1}\right) \ .$$

Inserting this into the second equation and simplifying gives

$$\sigma'(u) - 2a\sigma(u) = \sqrt{B^2\sigma(u)^2 - e^{4au}}$$

which can be solved for B. Again choosing the positive solution gives

$$B = \frac{1}{\sigma(u)}\sqrt{e^{4au} + (\sigma'(u) - 2a\sigma(u))^2} \ .$$

Note that $\sigma(u) > 0$ as we are only interested in positive conformal factors. This in turn makes the radicand in the preliminary expression for A, and hence A itself, positive. Explicity:

$$A = e^{-\frac{u\left(2a\sigma(u)+\sqrt{(\sigma'(u)-2a\sigma(u))^2+e^{4au}}\right)}{\sigma(u)}} \left(-2a\sigma(u) + \sigma'(u) + \sqrt{(\sigma'(u) - 2a\sigma(u))^2 + e^{4au}}\right) .$$

□

Remark 3.2. The Enneper solution ρ_{Enn} corresponds to $a = 1$, $A = B = 1$.

Using the solutions for ρ from Lemma 3.1 in Lemma 2.10 (and remembering that we normalized $b = 1$), straightforward computations gives

$$\begin{aligned}
\tilde{X}'(s) &= -\frac{2ABe^{Bs}}{A^2e^{2Bs}+1}\tilde{N}(s) \\
\tilde{Y}'(s) &= 0 \\
\tilde{N}'(s) &= \frac{2ABe^{Bs}}{A^2e^{2Bs}+1}\tilde{X}(s)
\end{aligned}$$

Integrating gives the following lemma:

Lemma 3.3. *Up to a motion in space, the solution to this equation is given by*

$$\tilde{X}(s) = \frac{1}{1+e^{2Bs}A^2}\begin{pmatrix} 1-A^2e^{2Bs} \\ 0 \\ -2Ae^{Bs} \end{pmatrix}, \ \tilde{Y}(s) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \tilde{N}(s) = \frac{1}{1+e^{2Bs}A^2}\begin{pmatrix} 2Ae^{Bs} \\ 0 \\ 1-A^2e^{2Bs} \end{pmatrix}$$

We have normalized the frame to that for $s = -\infty$, $\tilde{X} = (1, 0, 0)$ and $\tilde{N} = (0, 0, 1)$ in agreement with our parametrization of the Enneper surface.

Corollary 3.4. *The space curve $\tilde{c}(s) = f(s, 0)$ is given by*

$$\tilde{c}(s) = -\frac{e^{2as}}{2B}\left(\frac{e^{Bs}A}{B+2a} + \frac{e^{-Bs}}{A(B-2a)}, 0, \frac{1}{a}\right)$$

if $B \neq \pm 2a$. If $B = 2a$ (say, the other case being similar), we have

$$\tilde{c}(s) = -\frac{e^{2as}}{4a^2}\left(\frac{1}{4}Ae^{2as}, 0, 1\right) + \frac{s}{4aA}(1, 0, 0) .$$

Proof. This follows by integrating

$$\begin{aligned}\tilde{c}'(s) &= \frac{d}{ds} f(s,0) \\ &= I(\frac{\partial}{\partial s}, X(s))\tilde{X}(s) + I(\frac{\partial}{\partial s}, Y(s))\tilde{Y}(s) \\ &= -\frac{e^{2as}}{2B}\left(Ae^{Bs} - \frac{e^{-Bs}}{A}, 0, 2\right)\end{aligned}$$

using the previous lemma, and simplifying. □

Instead of now integrating the surface equations likewise along the curves $s \mapsto (u,s)$ for fixed u, we will use the Björling formula [1] to obtain the parametrization more easily.

Recall that given a real analytic curve $\tilde{c} : (u_1, u_2) \to \mathbb{R}^3$ and a real analytic unit normal field $\tilde{N} : (u_1, u_2) \to \mathbb{R}^3$ satisfying $\langle \tilde{c}'(u), \tilde{N}(u)\rangle = 0$, there exists a unique minimal surface containing $\tilde{c}$ and having surface normal $\tilde{N}$ along $\tilde{c}$, and this surface can be given in a neighborhood of $(u_1, u_2) \subset \mathbb{C}$ by

$$f(z) = \mathrm{Re}\left(\tilde{c}(z) - i\int^z \tilde{N}(w) \times \tilde{c}'(w)\, dw\right)$$

where we write $z = u + iv$ and have extended $\tilde{c}$ and $\tilde{N}$ to holomorphic maps into $\mathbb{C}^3$.

In our case, we obtain the explicit formula for f:

$$f(u,v) = \frac{e^{2au}}{2B}\begin{pmatrix} \frac{e^{-Bu}\cos((2a-B)v)}{2aA-AB} - \frac{Ae^{Bu}\cos((2a+B)v)}{2a+B} \\ \frac{e^{-Bu}\sin((2a-B)v)}{2aA-AB} + \frac{Ae^{Bu}\sin((2a+B)v)}{2a+B} \\ -\frac{\cos(2av)}{a} \end{pmatrix}$$

for $B \neq 2a$, and

$$f(u,v) = \frac{1}{4a^2}\begin{pmatrix} \frac{au}{A} - \frac{1}{4}Ae^{4au}\cos(4av) \\ \frac{av}{A} + \frac{1}{4}Ae^{4au}\sin(4av) \\ -e^{2au}\cos(2av) \end{pmatrix}$$

for $B = 2a$. Note that in the last case scaling a by a constant and (u,v) by the reciprocal only scales the surface, so we can as well assume that $a = 1$ in this case.

The Weierstrass data [1] of these surfaces are particularly simple. Using $z = u + iv$, let (also for $B = 2a$)

$$G(z) = \frac{1}{A} e^{-Bz} \qquad \text{and} \qquad dh = -\frac{1}{B} e^{2az}\, dz \ .$$

be the Gauss map and height differential of the Weierstrass representation formula

$$f(z) = \operatorname{Re} \int^z \begin{pmatrix} \frac{1}{2}(1/G - G) \\ \frac{i}{2}(1/G + G) \\ 1 \end{pmatrix} dh \ .$$

This gives the surfaces $f(u, v)$ above. This can be verified either by evaluating the integral or by solving the Björling integrand $\tilde{c}'(z) - i\tilde{N}(z) \times \tilde{c}'(z)$ for G and dh.

Of particular interest are the cases when B and $2a$ are integers. Then the substitution $z = -\log(w)$ changes the Weierstrass data into

$$G(z) = \frac{1}{A} w^B \qquad \text{and} \qquad dh = \frac{1}{B} w^{-2a-1}\, dw \ ,$$

defined on the punctured plane $\mathbb{C}^*$ and being minimal surfaces of finite total curvature.

A substitution in the domain of the form $w \mapsto \lambda w$ will scale G and dh by powers of λ, so we can assume without loss of generality that $A = 1$.

Some of the minimal surfaces we have obtained are described in [3]. We will now discuss examples.

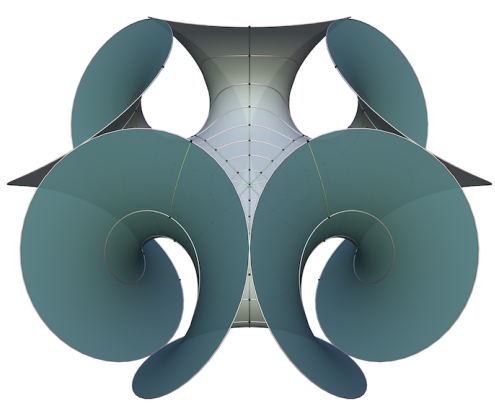
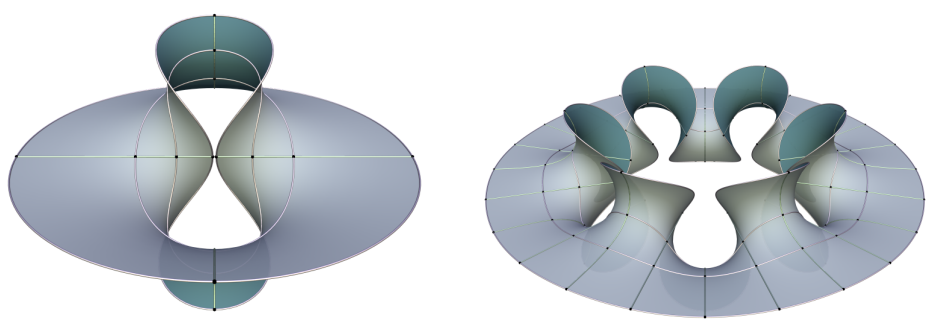


Figure 3: The Enneper Surface of order 5

In case that $B = 2a - 1 \in \mathbb{N}$, we obtain the Enneper surfaces of cyclic symmetry of order $B + 1$, see Figure 3. For $B = 2a - 1 = 1$, we obtain the original Enneper surface.

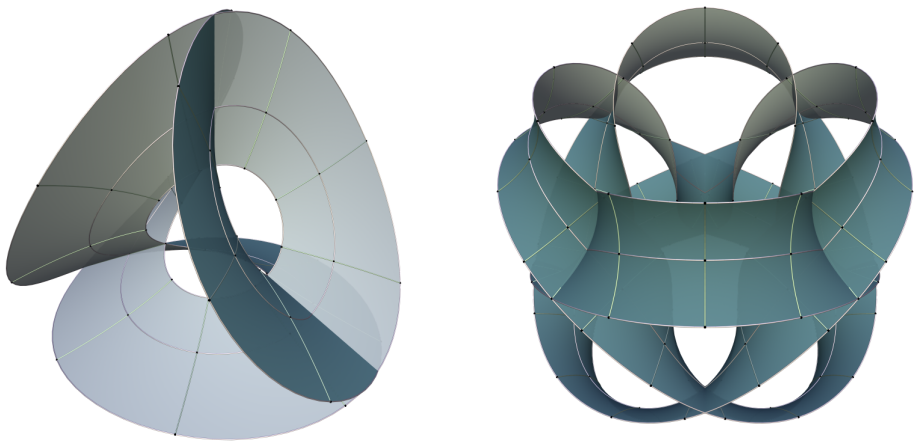
The *planar Enneper surfaces* of order n are given by choosing $B = n + 1$ and $2a = n$. See Figure 4 for the cases $n = 1$ and $n = 6$. These surfaces feature an Enneper type end and a planar end. Remarkably, in the non-zero CMC case, there are only one-ended solutions [5].



(a) order 1 (b) order 6

Figure 4: Planar Enneper surfaces

Other choices of a and B lead to more wildly immersed examples. In Figure 5 we show images of thin annuli $u_1 < u < u_1$.



(a) $B = 1$, $a = 3/2$ (b) $B = 7$, $a = 2$

Figure 5: Generalized Enneper surfaces

There is one case that deserves attention: If $B = 2a$, the Weierstrass 1-forms have residues, and hence the surface can become periodic. The prototype case here is $B = 1$ and $a = 1/2$ (see Figure 6) which leads to a translation invariant surface that hasn't made it into the literature to our knowledge. It deserves attention because it is in the potentially classifiable list of minimal surfaces in the space form $\mathbb{R}^3/\mathbb{Z}$ (where $\mathbb{Z}$ acts through a cyclic group of translations) of finite total curvature -4π. Other surfaces in this list include the helicoid and the singly periodic Scherk surfaces.

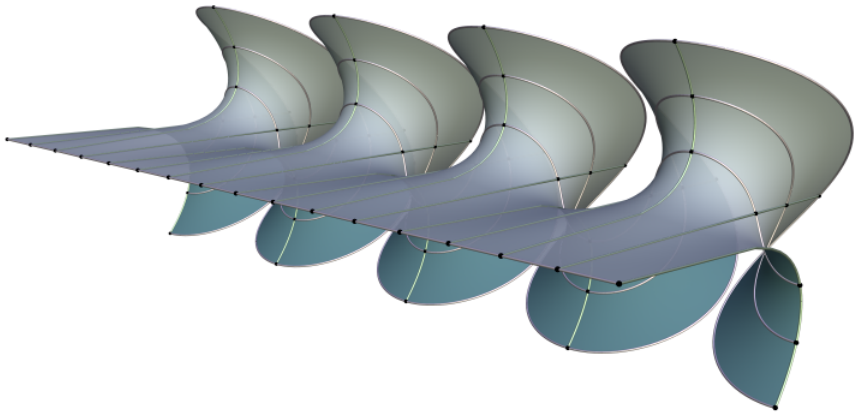


Figure 6: The Translation Invariant Enneper Surface

4 Constant Mean Curvature

In [5], Smyth considers intrinsical surfaces of revolution under a different viewpoint: He assumes from the beginning that his surfaces have constant mean curvature, but does not make further assumptions about the shape operator. Nevertheless, we both end up with the same class of surfaces. Therefore we would like to connect our approach with Smyth's in the CMC case.

First we can compute the Hopf differential using the coordinate $z = u+iv$: Using the Definitions 1.7 of I and 1.8 of S, and the formulas for α, λ_1 and λ_2 from Theorem 1.10, a straightforward computation shows that

$$
\begin{aligned}
\Omega &= I\left(S \cdot \frac{d}{dz}, \frac{d}{dz}\right) \\
&= I\left(S \cdot \frac{1}{2}\begin{pmatrix}1\\-i\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1\\-i\end{pmatrix}\right) \\
&= \frac{1}{2} b e^{2au}\left(\cos(2av) + i\sin(2av)\right) \\
&= \frac{1}{2} b e^{2az}
\end{aligned}
$$

is indeed holomorphic and agrees with Smyth's computation. Secondly, to show that our equation for ρ is equivalent to Smyth's equation, we substitute

$$
\begin{aligned}
\rho(u) &= e^{\phi(u)/2} \\
\phi(u) &= F(u) - 2au + \log(b)
\end{aligned}
$$

and obtain

$$
F''(u) = -4be^{-2au}\sinh(F(u))
$$

in the case that $H = 2$ (which is Smyth's case $H = 1$). This again agrees with Smyth's equation, up to a normalization of constants.

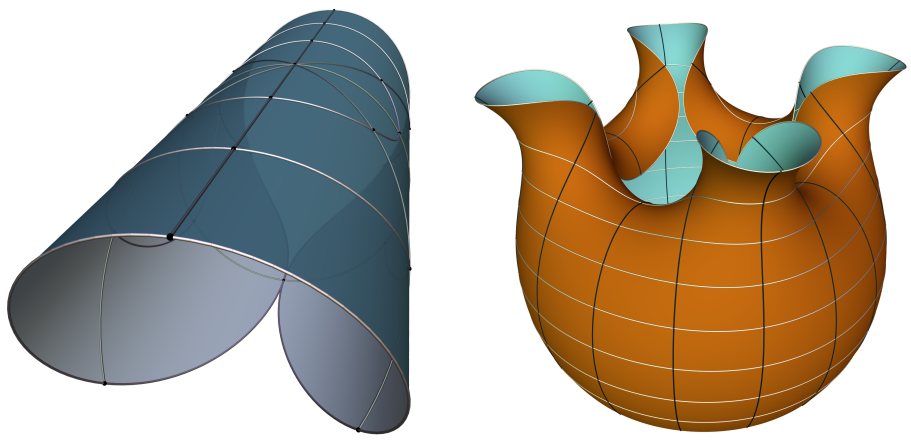
In general, there are apparently no explicit solutions to Equation (4) for $H \neq 0$ in the literature. There is, however, one explicit solution given by

$$
\rho(u) = \frac{\sqrt{2}\sqrt{b}e^{au}}{\sqrt{H}} \ .
$$

By Lemma 2.8, the principal curvatures become simply $\lambda_1 = H$ and $\lambda_2 = 0$. This implies that the surface under consideration is in fact a cylinder. This is somewhat surprising, as the standard parametrization of a cylinder over a circle of radius $1/H$ as an extrinsic surface of revolution has twist 0. In our case, however, the cylinder is parametrized using geodesic polar coordinates (see the left image in Figure 7) as

$$
f(u,v) = \frac{1}{H}\begin{pmatrix}\cos\left(\frac{1}{a}\sqrt{2bH}e^{au}\cos(av)\right)\\ \sin\left(\frac{1}{a}\sqrt{2bH}e^{au}\cos(av)\right)\\ \frac{1}{a}\sqrt{2bH}e^{au}\sin(av)\end{pmatrix}
$$

For other initial data of the Equation (4), only numerical solutions are available. These can be obtained easily by integrating the surface equations. The right image in Figure 7 was obtained using $a = 1$, $b = 4.2625$, $H = 1/2$, and $\rho(0) = \rho'(0) = 2$.



(a) Cylinder in polar coordinates

(b) Intrinsic CMC surface of revolution (numerical solution)

Figure 7: Two CMC surfaces

5 The untwisted case

In this section, we will consider the exceptional case of Theorem 1.9 where $\alpha(v) = a$ with a being an integral multiple of $\pi/2$, and prove Theorem 1.12.

Thus we are given a first fundamental form I_ρ and shape operator

$$S = \pm \begin{pmatrix} \lambda_1(u) & 0 \\ 0 & \lambda_2(u) \end{pmatrix} \qquad \text{or} \qquad S = \pm \begin{pmatrix} \lambda_2(u) & 0 \\ 0 & \lambda_1(u) \end{pmatrix} ,$$

depending on the congruence class of a modulo 2π. Without loss of generality, we will assume $a = 0$ and therefore

$$S = \begin{pmatrix} \lambda_1(u) & 0 \\ 0 & \lambda_2(u) \end{pmatrix} .$$

The Gauss- and Codazzi equations become

$$\lambda_1(u)\lambda_2(u) = \frac{\rho'(u)^2 - \rho(u)\rho''(u)}{\rho(u)^4}$$

and

$$\frac{\rho'}{\rho}(\lambda_1 - \lambda_2) = \lambda_2'$$

Eliminating λ_1 from the first equation using the second equation leads to the differential equation

$$\frac{\rho'(u)^2 - \rho(u)\rho''(u)}{\rho(u)^4} = \lambda_2(u)\left(\lambda_2(u) + \frac{\rho(u)\lambda_2'(u)}{\rho'(u)}\right)$$

for λ_2. Surprisingly, this equation can be solved explicitly by

$$\begin{aligned}\lambda_1(u) &= \frac{\rho(u)\rho''(u) - \rho'(u)^2}{\rho(u)^2\sqrt{c^2\rho(u)^2 - \rho'(u)^2}}\\ \lambda_2(u) &= -\frac{\sqrt{c^2\rho(u)^2 - \rho'(u)^2}}{\rho(u)^2}\end{aligned}$$

for any choice of c that makes the radicand positive.

We now show that any untwisted surface is a general surface of revolution. Recall that typically a surface of revolution is being parametrized as:

$$f(u,v) = (g(u)\cos(v), g(u)\sin(v), h(u))$$

However, by changing the speed of rotation, a surface of revolution can also be given by

$$f(u,v) = (g(u)\cos(cv), g(u)\sin(cv), h(u))$$

where c is a positive constant.

We now show that we can find g and h defined on the interval (u_1, u_2) having the first fundamental form and shape operator of the untwisted intrinsic surface of revolution above, with the rotational speed-up c being the constant c in Theorem 1.12 introduced above as an integration constant.

The first fundamental form of f is given by:

$$I = \begin{pmatrix} g'(u)^2 + h'(u)^2 & 0 \\ 0 & c^2 g(u)^2 \end{pmatrix} .$$

Comparing this to the definition of I_ρ gives the following equations:

$$\begin{aligned} g'(u)^2 + h'(u)^2 &= \rho(u)^2 \\ c^2 g(u)^2 &= \rho(u)^2 \end{aligned}$$

This determines $g(u) = \frac{\rho(u)}{c}$ and $h(u)$ by $h'(u) = \frac{1}{c}\sqrt{c^2\rho(u)^2 - \rho'(u)^2}$. Note that the radicand is positive by our assumption about c.

Straightforward computation shows that the shape operator of $f(u, v)$ with g and h as above coincides with the shape operator S of the intrinsic surface of revolution.

This completes the proof of Theorem 1.12.

Example 5.1. Knowing this, we can find surfaces of revolution with speed-up $c \geq 3$ that are locally isometric to the Enneper surface.

For the Enneper surface, we have

$$\rho(u) = \frac{1}{4} e^{2u} \left(e^{-u} + e^{u}\right)$$

so that the radicand $c^2\rho(u)^2 - \rho'(u)^2$ becomes

$$\frac{1}{16} e^{2u} u \left(\left(c^2 - 9\right) e^{4u} + \left(2c^2 - 6\right) e^{2u} + c^2 - 1\right) .$$

Thus for $c \geq 3$, we can find g and h as needed. The integral for h is generally not explicit, but for $c = 3$ we can obtain

$$\begin{aligned} g(u) &= \frac{1}{12} e^{2u} \left(e^{-u} + e^{u}\right) \\ h(u) &= \frac{1}{36} \left(2\sqrt{3} \sinh^{-1}\left(\sqrt{\frac{3}{2}} e^{u}\right) + 3e^{u}\sqrt{2 + 3e^{2u}}\right) . \end{aligned}$$

This means that the surface of revolution in Figure 8 is isometric to one third of the Enneper surface, punctured at the "center".

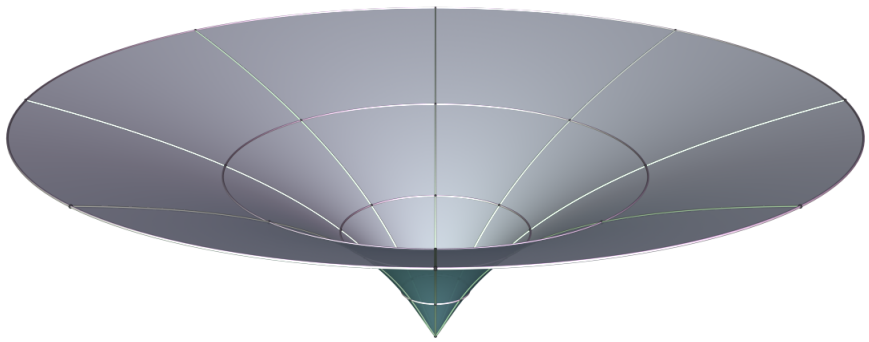


Figure 8: Surface of revolution isometric to one third of the Enneper Surface

In contrast, if $c = 1$, the radicand is negative for all u, which implies that no piece of the Enneper surface can be isometrically realized as a standard surface of revolution (with no speed-up).

References

[1] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab. *Minimal surfaces. I*, volume 295 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1992. Boundary value problems.

[2] A. Enneper. Weitere bemerkungen über asymptotische linien. *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts- Universität zu Göttingen*, pages 2–23, 1871.

[3] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry*, pages 1–96, 1989. University of Tokyo, 1989, and Lecture Notes No. 12, SFB256, Bonn, 1989.

[4] W. Kühnel. *Differential Geometry: Curves - Surfaces - Manifolds, Second Edition*, volume 16 of *Student Mathematical Library*. American Mathematical Society, 2006.

[5] Brian Smyth. *A Generalization of a Theorem of Delaunay on Constant Mean Curvature Surfaces*, pages 123–130. Springer New York, New York, NY, 1993.

Some Results in the Theory of Elastic Networks

Bat Dejean, Christian Gorski
Advisor: Dylan Thurston

September 14, 2016

Abstract

Elastic networks can be thought of as networks of rubber bands which can be stretched in different configurations. Stretching an elastic network in Euclidean space has a direct analogy to the study of resistor networks. However, stretching elastic networks along more general graphs is more complicated. A fundamental question is how to determine when one elastic network is "looser" than another. One way to attack this problem is to find the energy of the network as a piecewise-quadratic function of the lengths of the graph in which it is stretched. This is done by examining so-called "train track structures" on the network. We also use geometric methods to prove that an elastic tripod is "looser" than any electrically equivalent graph and that an elastic delta is "tighter" than any electrically equivalent graph. We give some generalizations of these results. We give an example of a class of elastically equivalent graphs and assorted proofs that a tripod is "looser" than the electrically equivalent delta. [1]

Introduction

An elastic network, informally speaking, can be thought of as a network of rubber bands, which has different energies when stretched into different configurations. Mathematically, the act of "stretching" an elastic network is simply a map (up to some notion of homotopy) from the network to some target space. When one stretches an elastic network in a Euclidean space, there is a direct analogy with the well-understood theory of resistor networks (see 2). However, we want to stretch networks in more general spaces; we focus on the case mapping to trees, but one result generalizes to maps to Hadamard spaces.

[1] This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

This paper is largely self-contained, with the exception of a few, more laborious proofs.

In 1, we give basic definitions and some criteria for harmonicity of maps. In 2 we make explicit the analogy to resistor networks. In 3 we introduce train track structures and show how they are used to relate the energy of a map to the dimensions of the target tree.

We then use some geometry and the electrical analogy to prove some results about "looseness." In 4 we give a proof that tripods are in some sense "minimal-energy" elastic networks for maps to trees and that deltas are "maximal-energy" elastic networks for maps to Hadamard spaces.

As a sort of appendix, we include several other proofs that a tripod is "looser" than its electrically equivalent delta, in the hope that these different perspectives may be generalized further.

1 Definitions

Throughout this paper, *graphs* will denote finite 1-dimensional CW complexes, with a linear structure on each open edge (1-cell). Maps between graphs are continuous maps, which are not required to map vertices to vertices. However, we will be focusing on *marked graphs*, which are graphs with a distinguished finite subset whose points are designated as "marked." Maps between marked graphs are continuous maps which map marked points to marked points. Whenever we refer to homotopy, we mean homotopy relative to marked points; that is, all maps in a given homotopy class are equal when restricted to the marked points.

With this in mind, we can make precise the notion of an elastic network.

Definition 1.1. An **elastic structure** α on a graph Γ is an assignment of a positive real number to each edge of Γ. We call the real number assigned to an edge e the **elasticity** of e. The pair (Γ, α) is called an **elastic graph** or an **elastic network**. Usually we will use a single symbol, say G, for (Γ, α).

The greater the elasticity of an edge, the "looser" it is. If we treat the elasticities of edges as lengths, the elastic structure induces a metric on the graph in an obvious way, and this, in turn, induces a measure on the graph in an obvious way.

For most of this paper, we will consider the case of mapping our network to a graph (usually a tree) with specified edge lengths, which we think of as stretching a network of rubber bands within a network of thin pipes.

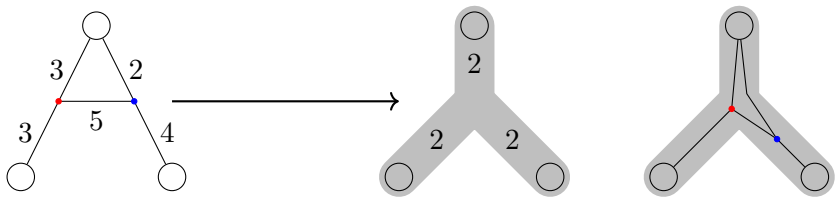


Figure 1: We visualize a map from an elastic graph to a length graph as stretching out a network of rubber bands along a pipe network. The map shown above is quite "unnatural" (that is, it is not harmonic), since the two loosest edges are stretched the least.

Definition 1.2. A **length structure** L on a graph Γ is an assignment of a nonnegative real number to each edge of Γ. We call the real number assigned to each edge e of Γ the **length** of e. The pair (Γ, L) is called a **length graph** or a **pipe graph**. Usually we will use a single symbol, say K, for (Γ, L).

If all edges have nonzero length, the length structure induces a metric structure on the graph in an obvious way. Otherwise, the length structure on a graph induces a pseudometric on the graph in an obvious way, and there is an associated metric graph given by collapsing to a single point all points which have distance zero from one another. This new graph is called the "collapsed graph." If the length graph is denoted K, we denote the collapsed graph by K^*.

When we refer to, for example, a map $f : G \to K$, where G is a marked elastic graph and K a marked length graph, we simply mean maps of the underlying marked graphs.

Now, we rigorously define what we mean by the "energy" of a stretched network. In fact, there are several notions of energy for maps between graphs with different structures (see [1, Appendix A]), but we will be focusing on the so-called *Dirichlet energy*. If the target space is Euclidean space or a graph, then the definition below is fairly straightforward. It requires some delicacy to flesh out when the target is simply a metric space.

Definition 1.3. Let $G = (\Gamma, \alpha)$ be an elastic graph, M be a metric space. Let $f : G \to M$ be a Lipschitz map. Then the **Dirichlet energy** of f is the quantity

$$\operatorname{Dir} f = \int_{x \in \Gamma} |f'(x)|^2 dx.$$

Moreover, we define the Dirichlet energy of the marked homotopy class $[f]$

to be

$$\mathrm{Dir}[f] = \inf_{g \in [f]} \mathrm{Dir}\, g.$$

If M is Euclidean space or a graph, then we can define $|f'(x)|$ to be

$$\lim_{y \to x} \frac{d(f(x), f(y))}{d(x, y)},$$

and this will be defined almost everywhere if f is Lipschitz. More generally, we define $|f'(x)|$ to be the infimum over all local Lipschitz constants at x, where a local Lipschitz constant at x is just a Lipschitz constant for a restriction of f to a sufficiently small neighborhood of x.

If f minimizes Dirichlet energy within its homotopy class, then $|f'|$ is constant on each edge, and for each edge e, we have

$$|f'(x)| = \frac{\ell(f(e))}{\alpha(e)}$$

for each $x \in e$, where $\ell(f(e))$ is the length of the image of e, in a sense appropriate to M. Moreoever, the Dirichlet energy is then

$$\mathrm{Dir}\, f = \sum_{e \in \mathrm{Edge}(\Gamma)} \frac{\ell(f(e))^2}{\alpha(e)}.$$

In the case that M is a tree or Euclidean space, this just turns out to be the distance between the images of the endpoints of e. Also note that, if the target is a geodesic space, the images of edges under a harmonic map are geodesics. We call maps which minimize energy within their homotopy class *harmonic maps*. The existence of harmonic maps is not obvious, and is not true in general. A proof for existence in the case of maps into length graphs can be found in Theorem 4 of [1].

We will be dealing primarily with maps into trees in this paper, and there is a simple criterion (precisely stated in Proposition 5.2 of [1]) which determines whether a map from a marked elastic graph to a length graph is harmonic. We can think of $|f'|$ as the tension at a point, and then harmonic maps correspond to a sort of physical equilibrium. Specifically, tensions are constant at each edge, and "balance" at each vertex. That is, the total tension pulling in any one direction cannot exceed the sums of the (magnitudes of the) tensions pulling in all other directions. This condition is referred to as satisfying the "triangle inequalities." (See 2.) Notice that if there are only two directions going out from a point, the inequalities force an equality,

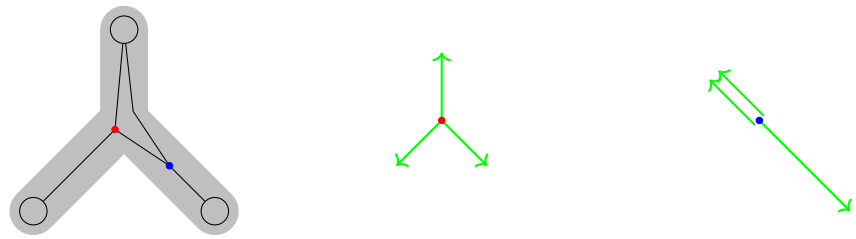


Figure 2: A harmonic map must be such that the tensions satisfy the "triangle inequalities" at each vertex. Since there are only two directions going out from the blue vertex, the total tension in one direction equals the total tension in the other.

and in fact the sum of the tensions in one direction is equal to the sum of the tensions in the other.

One of the problems we were particularly interested in was determining if one elastic network is "looser" than another in the following sense:

Definition 1.4 (Looseness)**.** Given marked elastic graphs G_1 and G_2 with a bijection between their marked vertices, G_1 is looser than G_2, or $G_1 \preceq G_2$, if for any choice of target points in any pipe tree L for the marked vertices, the corresponding harmonic maps $f_1 : G_1 \to L$ and $f_2 : G_2 \to L$ satisfy $\operatorname{Dir} f_1 \leq \operatorname{Dir} f_2$.

Notice that as the target is a pipe tree, specifying the images of the marked points of G_1 and G_2 specifies unique marked homotopy classes, so there's no ambiguity in which homotopy classes are under consideration.

We obtain some results in 4, and the methods of 3 can be used to obtain a decision procedure for this relationship.

2 The electrical case

2.1 Analogies between resistor networks and elastic networks

Consider the following dictionary for converting between resistor and elastic networks:

$$\begin{aligned}
\text{Rubber Band} &\Leftrightarrow \text{Resistor} \\
\text{Tension} &\Leftrightarrow \text{Current} \\
\text{Position} &\Leftrightarrow \text{Voltage} \\
\text{Elasticity} &\Leftrightarrow \text{Resistance} \\
\text{Hooke's Law} &\Leftrightarrow \text{Ohm's Law} \\
\text{Dirichlet energy} &\Leftrightarrow \text{Power dissipated}
\end{aligned}$$

This dictionary translates between Hooke's law $T = \frac{L}{\alpha}$ and Ohm's law $I = \frac{V}{R}$, and the last correspondence is a corollary of the others. This also allows us to consider elastic graphs as resistor networks: simply let the resistances be the elasticities.

This dictionary is useful for the following reason. If we identify position with voltage, $f : G \to \mathbb{R}^1$ is harmonic iff it satisfies Kirchhoff's junction rule at unmarked nodes: both are equivalent to

$$\sum_{e \text{ incident to } u} \frac{f\left(\text{other node incident to e}\right) - f\left(u\right)}{\alpha\left(e\right)} = 0$$

for every unmarked node u. (Kirchhoff's loop rule comes for free, as shown in 3.) Thus, under this dictionary, the theories of harmonic maps to $\mathbb{R}^1$ and solutions to Kirchhoff's laws are the same.

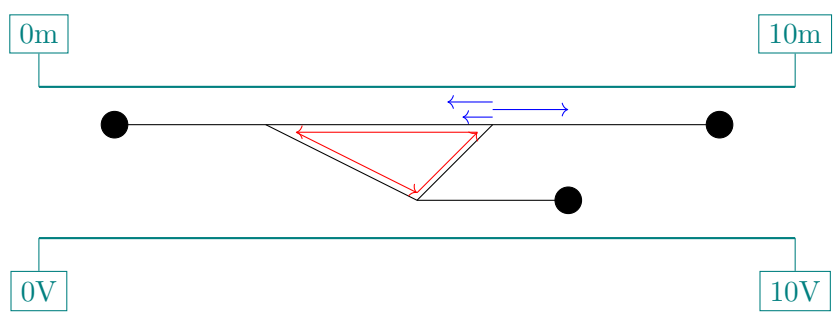


Figure 3: That the sum of the red displacements is 0 is Kirchhoff's loop rule, and that the sum of the blue tensions is 0 is Kirchhoff's junction rule.

This allows us to "import" facts about resistor networks, which are better-understood, into the context of harmonic maps to $\mathbb{R}^1$.

We can talk about resistor networks being equivalent in the following sense:

Definition 2.1 (Response matrix)**.** For a resistor network with n marked points, the vector of currents flowing into the resistor network at each exterior node is a linear function of the voltages at those nodes. The *response matrix* of the resistor network is the corresponding $n \times n$ matrix.

Remark 2.2. The response matrix is symmetric. Additionally, interpreted as a bilinear form on the space of boundary voltages, the norm-squared of a given set of voltages is the power dissipated in the network when given those voltages.

Definition 2.3 (Electrical equivalence)**.** Two resistor networks with a bijection between their exterior nodes are *electrically equivalent* if they have the same response matrix. Two elastic graphs are electrically equivalent if they're electrically equivalent as resistor networks.

Remark 2.4. By 2.2, electrically equivalent resistor networks dissipate the same power when given the same voltages. Therefore electrically equivalent elastic networks have the same Dirichlet energy when "given the same voltages;" that is, when the two networks map harmonically to $\mathbb{R}^1$ and corresponding marked vertices have the same image in $\mathbb{R}^1$.

This does *not* mean that electrically equivalent graphs have the same Dirichlet energy when mapping harmonically to a pipe network; voltage is $\mathbb{R}^1$-valued, so the analogy fails in this situation.

Example 2.5 (Resistors in series and in parallel)**.** Resistors in series and parallel follow the familiar laws: two resistors α_1 and α_2 in series are electrically equivalent to a single resistor $\alpha_1 + \alpha_2$, and two resistors in parallel are electrically equivalent to a single resistor $\frac{1}{1/\alpha_1 + 1/\alpha_2}$.

Example 2.6 ($Y - \Delta$ equivalence)**.** If Y and Δ are a tripod and delta, respectively, with the resistances below, then Y and Δ are electrically equivalent if and only if $\beta_i = \frac{\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3}{\alpha_i}$ for $i \in \{1, 2, 3\}$.

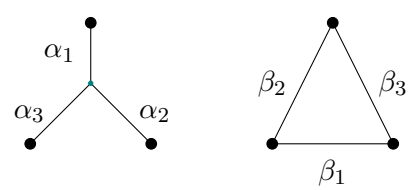
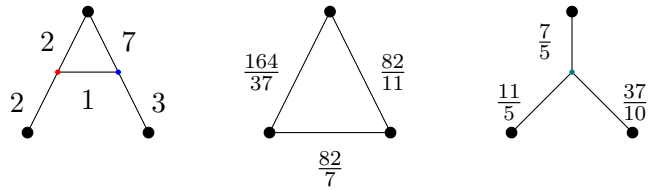


Figure 4: Y-Δ equivalence

Example 2.7. Using the above examples, the following are electrically equivalent:



2.2 Classification of networks with three marked vertices

The above "A"-shaped graph was electrically equivalent to a tripod and a delta; is this always true? The answer is no: for example, a network with no edges isn't equivalent to any tripod or delta. If we include some limiting cases corresponding to tripods/deltas with "zero elasticity" or "infinite elasticity," the statement is true. More specifically, define the following:

Definition 2.8. A *generalized delta* is a delta or one of the graphs in 5 and a *generalized tripod* is a tripod or one of the graphs in 5.

Figure 5: The generalized tripods/deltas

Then we have

Proposition 2.9 (Classification of networks with three marked vertices)**.** *Every elastic network with three marked vertices is electrically equivalent to a unique elastic generalized tripod and a unique elastic generalized delta.*

Proof. We first show it's electrically equivalent to a unique generalized Δ. The response matrix of the delta in 4 can be calculated to be

$$\begin{bmatrix} \beta_2^{-1}+\beta_3^{-1} & -\beta_3^{-1} & -\beta_2^{-1} \\ -\beta_3^{-1} & \beta_1^{-1}+\beta_3^{-1} & -\beta_1^{-1} \\ -\beta_2^{-1} & -\beta_1^{-1} & \beta_1^{-1}+\beta_2^{-1} \end{bmatrix}.$$

With the convention $\frac{1}{\infty}=0$, the response matrices of the other generalized deltas are this matrix in the cases that one or more of the $\beta_i$is is ∞, and all such cases have a generalized delta realizing them. That is, the response matrix of a generalized delta is a generic matrix of the above form with $\beta_i \in (0,\infty]$, which is a generic symmetric matrix with nonpositive off-diagonal terms and rows summing to 0. It therefore suffices to show every network has such a response matrix.

Symmetry is taken care of by 2.2. That the off-diagonal terms are non-positive is simply the fact that when two exterior nodes are at 0V and the other at 1V there's a net flow out from the 0V nodes. That the rows sum to 0 is the fact that no current flows when all exterior nodes are brought to 1V, as then all interior nodes will be at 1V also.

All the conditions are satisfied, so every network is electrically equivalent to a generalized delta.

This implies the network is also electrically equivalent to a generalized tripod: if the generalized delta is a delta, it's $Y - \Delta$ equivalent to a tripod. Otherwise, it's already a generalized tripod.

Uniqueness is easy to check. □

2.3 Harmonic maps to $\mathbb{R}^n$

In this section we generalize 2.4 into 2.12.

The following definition is analogous to the case of mapping to a tree, and agrees with our previous definition in the case $n = 1$:

Definition 2.10 (Harmonic map)**.** A continuous map $f : \Gamma \to \mathbb{R}^n$ is harmonic if

1. Every edge of Γ is sent linearly to the segment connecting its endpoints (and continuously, so no edges are "flipped")
2. $\sum\limits_{e \text{ incident to u}} \frac{f(\text{other node incident to e}) - f(u)}{\alpha(e)} = 0$ for every unmarked vertex u of Γ.

Condition 1 states that each edge is "taut," and condition 2 states that the tensions on each edge cancel.

These maps behave exactly as in the electrical (i.e. $\to \mathbb{R}^1$) case:

Proposition 2.11. *Let $f : \Gamma \to \mathbb{R}^n$ be PL from an elastic graph Γ to $\mathbb{R}^n$, and let $f_i = \pi_i \circ f : \Gamma \to \mathbb{R}^1$ be the ith projection of f. Then* $\operatorname{Dir} f = \sum\limits_i \operatorname{Dir} f_i$. *Further, TFAE:*

(1) f is harmonic

(2) Every f_i is harmonic

(3) Every f_i minimizes Dirichlet energy within its marked homotopy class

(4) f minimizes Dirichlet energy within its marked homotopy class

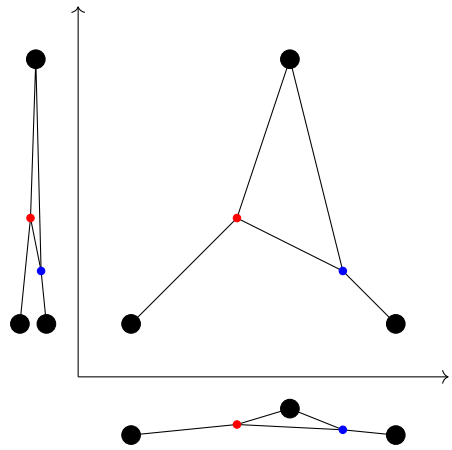


Figure 6: Illustration of 2.11

Proof. The first equation is the Pythagorean Theorem: $\operatorname{Dir} f = \int_\Gamma \|f'\|^2 \, dx = \int_\Gamma \sum_i |f_i'|^2 \, dx = \sum_i \int_\Gamma |f_i'|^2 \, dx = \sum_i \operatorname{Dir} f_i$.

(1) $\Leftrightarrow$ (2): The only nontriviality is that the tensions of f on a node cancel iff the tensions of every f_i on the node cancel. However, this is also simple, as a tension of f_i is the ith component of the corresponding tension of f.

(2) $\Leftrightarrow$ (3): Easy to verify.

(3) $\Leftrightarrow$ (4): By $\operatorname{Dir} f = \sum_i \operatorname{Dir} f_i$. □

As a corollary, 2.4 also holds for maps to $\mathbb{R}^n$:

Corollary 2.12. *For electrically equivalent Γ_1 and Γ_2 and harmonic maps $f : \Gamma_1 \to \mathbb{R}^n$ and $g : \Gamma_2 \to \mathbb{R}^n$ sending corresponding marked vertices to the same point,* $\operatorname{Dir} f = \operatorname{Dir} g$.

Proof. $\operatorname{Dir} f = \sum_i \operatorname{Dir} f_i = \sum_i \operatorname{Dir} g_i = \operatorname{Dir} g$. □

3 Train tracks

3.1 Calculating Energy

One way to try to determine whether an elastic network is "looser" than another is to try to get a sort of "energy function" for each, and compare them directly. For resistor networks, the power dissipated in the network is

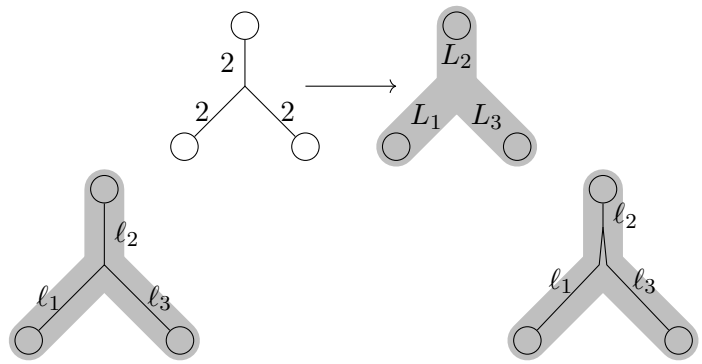


Figure 7: We must make combinatorial assumptions about the harmonic map in order to predict how the stretched lengths relate to the target lengths. On the left, we assume the central elastic vertex is mapped to the central pipe vertex. On the right, the central elastic vertex has been "pulled up into an edge."

just a quadratic function of the boundary voltages. Analogously, we would like to determine the Dirichlet energy of a map from an elastic graph to pipe graph as a function of the pipe lengths. More precisely, given a marked elastic graph G, a marked graph Γ, and a marked homotopy class $[f]$ of maps from G to Γ, there is a harmonic map f_L in $[f]$ for each length structure L on Γ. A length structure on Γ can be viewed as a point in $\mathbb{R}^n_{\geq 0} := \{(L_1, ..., L_n) \in \mathbb{R}^n : L_i \geq 0\}$, where n is the number of edges in Γ. What we want to determine is the energy function $E_{[f]} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$, $E_{[f]}(L) = \operatorname{Dir} f_L$.

If Γ is topologically a line segment, then the analogy to resistor networks tells us that $E_{[f]}$ is simply a quadratic function of the length(s) of Γ. Moreover, the electrical analogy tells us that we can find the stretched lengths of G on Γ as a linear function of the length(s) of Γ.

More generally, the stretched lengths of G are a *piecewise*-linear function of the target lengths, and $E_{[f]}$ is *piecewise*-quadratic. The reason that our case is more complicated than the electrical case is that positions are not parametrized by real numbers, and so we have to consider cases where different combinations of tensions work with or against each other separately.

Example 3.1. Consider the elastic tripod Y pictured in Figure 7. We map it to another marked tripod K, and attempt to calculate the stretched lengths (ℓ_1, ℓ_2, ℓ_3) of Y as a function of the lengths (L_1, L_2, L_3) of the edges of K. There are (up to symmetry) two cases to consider. If the central

vertex of Y is mapped to the central vertex of K, we clearly have

$$\begin{aligned}\ell_1 &= L_1\\ \ell_2 &= L_2\\ \ell_3 &= L_3.\end{aligned}$$

If the central vertex of Y is "pulled up" into an edge of K, then we have two concerns: the forces at the central vertex must balance, and lengths must "match." That is, the sum of the lengths of the stretched edges in any path which goes "straight" from one marked vertex to another must equal the length of the unique geodesic between the marked vertices given by the length structure on K. So we have

$$\frac{\ell_1}{2} + \frac{\ell_3}{2} - \frac{\ell_2}{2} = 0$$

and

$$\begin{aligned}\ell_1 + \ell_2 &= L_1 + L_2\\ \ell_2 + \ell_3 &= L_2 + L_3.\end{aligned}$$

Row reducing then gives the relationship

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}.$$

□

From the above example, we see that we need to make some assumptions about the combinatorics of the harmonic map to calculate what the relationship between stretched lengths and target lengths is. It turns out the relevant combinatorial information is recorded by "train track structures" on the source graph. These are designed to keep track of which "directions" on the source graph are mapped "in the same direction" on the target.

Intuitively, the "directions" at a point in a graph are simply the directions one can take on a path from the point. Formally, we have the following definition.

Definition 3.2. If Γ is a graph, and $x \in \Gamma$, then a **direction** at x is the germ of a piecewise-linear map $(\mathbb{R}_{\geq 0}, 0) \to (\Gamma, x)$, considered up to piecewise-linear reparametrization.

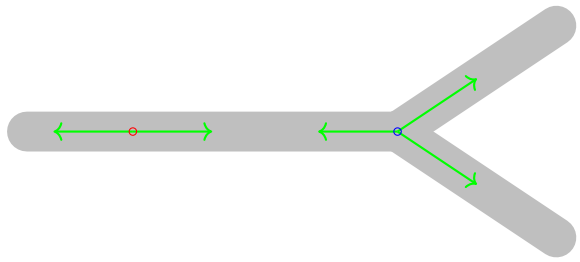


Figure 8: There are two nonzero directions at the red point, and three nonzero directions at the blue point.

The germ of the constant map is called the *zero* direction. There are two nonzero directions at any point in the interior of an edge, but there may be more at a vertex (see 8). Directions push forward by piecewise-linear maps in an obvious way; if $f : \Gamma \to \Gamma'$ is piecewise-linear, and $\gamma : (\mathbb{R}_{\geq 0}, 0) \to (\Gamma, x)$ is a representative of a direction d at $x \in \Gamma$, then the germ of the composition $f \circ \gamma$ is a direction at $f(x)$, which we denote by $f(d)$.

Definition 3.3. A **train track structure** τ on marked graph Γ is a partition of the nonzero directions into equivalence classes, called **gates**, with at least two gates at each unmarked point.

Directions and train tracks are introduced (in the context of maps between graphs) in [1].

Given a piecewise-linear map $f : \Gamma \to \Gamma'$, the *train track of* f is the train track structure τ on Γ such that

$$(d_1 \sim_\tau d_2) \Leftrightarrow (f(d_1) = f(d_2)).$$

That is, τ identifies precisely the directions which are mapped to the same direction under f.

Since at each point in the interior of an edge, there are only two directions, and these must belong to two different gates, we are really only concerned about the gates at each vertex of our source graph. We can record the train track diagrammatically by connecting edges which are identified with arcs (see 9). In this way, we see that we can also think of the train track structure as a partition at each vertex of the edges adjacent to that vertex. (If some edge only one boundary vertex, i.e. is a loop, than that edge will be counted twice, once for each direction out of its boundary vertex). We will also call these equivalence classes of edges gates.

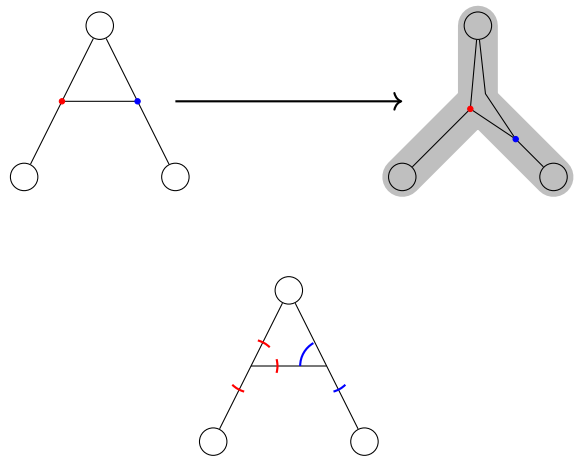


Figure 9: A map and its associated train track.

The key observation is that a train track structure on G tells us how to "balance tensions" and "match path lengths." This gives a linear system, which in turn gives a linear relationship between the stretched lengths and the target lengths which holds for all harmonic maps which have a given train track.

Example 3.4. Consider mapping the A graph in 10 to a tripod with varying lengths. Suppose we assume the harmonic map has the train track structure shown. Then we get the system of equations:

$$\frac{\ell_1}{3} + \frac{\ell_3}{5} - \frac{\ell_2}{3} = 0$$
$$\frac{\ell_3}{5} + \frac{\ell_4}{2} - \frac{\ell_5}{4} = 0$$

by balancing forces and

$$\ell_1 + \ell_2 = L_1 + L_2$$
$$\ell_4 + \ell_5 = L_2 + L_3$$
$$\ell_2 + \ell_3 + \ell_5 = L_2 + L_3$$

by matching lengths. We can row reduce to get the relationship

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \\ \ell_5 \end{bmatrix} = \frac{1}{47} \begin{bmatrix} 28 & 25 & -3 \\ 19 & 22 & 3 \\ -15 & -5 & 10 \\ 4 & 17 & 13 \\ -4 & 30 & 34 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}.$$

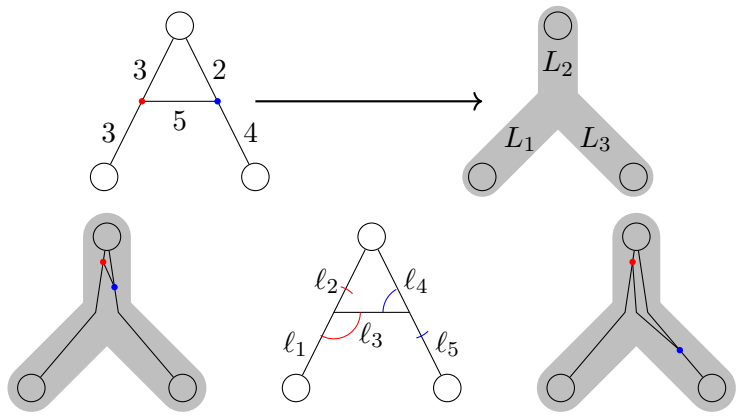


Figure 10: Calculating the relationship between stretched and target lengths for a specific train track. Notice that although the blue vertex is mapped to a different edge on the left and on the right, both maps have the same train track.

□

In general, we generate the equations thus. For each vertex of G which has exactly two gates g_1 and g_2,

$$\sum_{e \in g_1} \frac{\ell(e)}{\alpha(e)} - \sum_{e \in g_2} \frac{\ell(e)}{\alpha(e)} = 0$$

since the tensions must balance. Notice that the left side is a linear combination of the stretched edge lengths. "Matching path lengths" is a little more involved.

We say that a (piecewise-linear) path $\gamma : I \to G$ is *straight* (or *makes valid turns*) if it does not "backtrack" with respect to the train track structure τ on G; that is, at no point $x \in I$ do two distinct nonzero directions from x map under γ to directions identified by τ. Put another way, if $g : G \to K$ is a map inducing the train track structure on G, then the train track structure induced on I by the composition $g \circ \gamma : I \to K$ never identifies two different nonzero directions. We say a path $\gamma : I \to K$ is straight if it simply does not "backtrack;" for all $x \in I$, no two nonzero directions from x map under γ to the same direction in K. Put another way, the train track structure induced by γ never identifies two different nonzero directions.

We can define a similar notion of "straight" for *chains*. A *chain* is an alternating sequence of vertices and edges, where each edge connects the

vertices on either side of it. A chain $(v_0, e_1, v_1, ..., e_m, v_m)$ is called straight if it does not "backtrack." That is, if the chain lies in a graph with a train track structure, then we require that e_i and e_{i+1} lie in different gates at v_i for each $1 \leq i \leq m-1$. If the chain lies in a graph with no train track structure, then we must have $e_i \neq e_{i+1}$ for all $1 \leq i \leq m-1$, unless e_i is a loop whose two directions out from its boundary vertex are not identified.

Given a straight path between vertices, we say a chain *parametrizes* the path if some orientation on the edges is such that the path given by concatenating all the edges in order is homotopic to the original path. Thence, given our elastic graph G, our length graph K, our homotopy class $[f] : G \to K$, and a train track structure on τ, we have a way to associate straight chains with marked endpoints on G to straight chains with marked endpoints on K. Specifically, we associate our chain in G to a (marked) straight path γ in G which it parametrizes; the homotopy class $[f \circ \gamma]$ contains (up to PL reparametrization) a unique representative which is straight. (In fact, if $g \in [f]$ induces the train track structure τ on G, then $g \circ \gamma$ is a straight path between two marked points in K). We then parametrize the straight path in $[f \circ \gamma]$ by a straight chain $(v'_0, e'_1, v'_1, ..., e'_q, v'_q)$. You can think of this chain as the "image" of our first chain. (We can do a similar process with straight cycles and straight loops).

Clearly, the total lengths of associated straight chains must match; the first chain is a chain of stretched lengths with no backtracking, and the second is the path of target lengths along which they are stretched. Thus, having fixed our train track structure, we get

$$\sum_{i=0}^{m} \ell(e_i) = \sum_{j=0}^{q} L(e'_j)$$

for each straight chain $(v_0, e_1, v_1, ..., e_m, v_m)$ (either with marked endpoints or with $v_0 = v_m$) in G and each associated straight chain $(v'_0, e'_1, v'_1, ..., e'_q, v'_q)$. Notice that the left side a linear combination of stretched lengths and the right is a linear combination of target lengths.

Note that if K is a tree, then any straight chain in K is determined by its endpoints. This makes computations a lot easier; one simply must find straight chains with marked endpoints in G and keep track of where the endpoints map in K. (Since K has no cycles, there will be no straight cycles in G if τ is actually induced by a map to K). This method is fairly straightforward to program; in fact, we used this method to write a program giving the relationship between stretched lengths and target lengths and the energy function for maps from 3-marked elastic graphs to a tripod.

In all of our experiments mapping different 3-marked networks to the tripod, the system produced had a unique solution for each assignment of target lengths (although most of the time there were redundant equations). We do not yet have a proof that these systems of equations have unique solutions, but we suspect that they do in the case that K is a tree.

Assuming the systems are sufficiently determined, then for a specified train track τ, by row-reducing our system, we get a matrix we get a matrix T such that

$$\begin{bmatrix} \ell_1 \\ \vdots \\ \ell_N \end{bmatrix} = T \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix}.$$

If

$$A = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_N \end{bmatrix}$$

is the diagonal matrix of the elasticities of the source graph, then the Dirichlet energy of a map is

$$\begin{aligned} \operatorname{Dir} g = \begin{bmatrix} \ell_1 & \cdots & \ell_N \end{bmatrix} A \begin{bmatrix} \ell_1 \\ \vdots \\ \ell_N \end{bmatrix} &= \left(T \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix} \right)^t A \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix} \\ &= \begin{bmatrix} L_1 & \cdots & L_n \end{bmatrix} M \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix}, \end{aligned}$$

where $M = T^t A T$. Thus, we have explicitly obtained a piece of our energy function.

3.2 Regions of Validity

In order to get our system of equations, we had to assume that the harmonic representative of $[f]$ had a specific train track structure. However, the assignment of target lengths $(L_1, ..., L_n)$ already determines the harmonic representative(s), which in turn determines a train track structure (or structures) which may differ from our assumption. Therefore, we want to determine which assignments of target lengths correspond to which train track structures, and thence how the "pieces" of our piecewise-quadratic energy function are shaped and arranged.

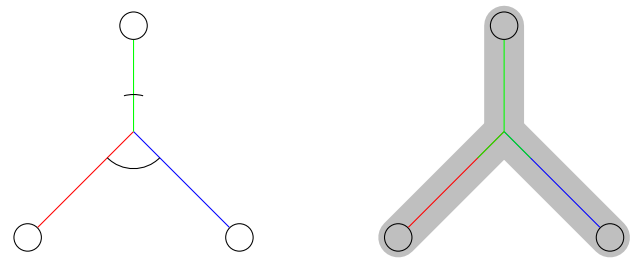


Figure 11: Map an elastic tripod to a tripod. If we assume that our harmonic map has the train track structure on the left, and then try to calculate the stretched lengths for L_2 large compared to L_1 and L_3, we get the nonsensical answer that $\ell_2 > L_2$. Although this solution satisfies our system of equations, it clearly does not correspond to a continuous map.

Definition 3.5. Given a marked elastic graph G, a marked graph Γ, and a homotopy class $[f] : G \to \Gamma$, we say that a train track structure τ is **valid** for a length structure L on Γ if there exists $g \in [f]$ such that $g : G \to (\Gamma, L)$ is harmonic and τ is the train track of g. We call the subset $\{L \in \mathbb{R}^n_{\geq 0} : \tau \text{ is valid for } L\}$ the **region of validity** for τ.

Conjecture 3.6. A train track τ to is valid for a length structure $L = (L_1, ...L_n)$, if and only if following criteria must be satisfied:

1. The stretched lengths must be nonnegative. That is, if T is the matrix obtained from τ as in 3.1, then each component of TL must be greater than or equal to zero.

2. The triangle inequalities must be satisfied at each vertex of G. The cases where only two nonzero directions go out from a vertex are taken care of in the course of the calculation of T, but we must still apply the triangle inequalities for vertices with three or more nonzero directions.

3. The track cannot "split hairs."

This last criterion is a bit difficult to characterize. An example of "splitting hairs" is given in 11.

The explicit inequalities prohibiting "splitting hairs" compare the lengths of straight chains in G and K. That is, the length of a straight chain from an unmarked vertex to a marked vertex in G cannot exceed the length of a corresponding straight chain in K. The trouble is that the straight chain in K no longer has fixed endpoints, and so is not determined by the homotopy

class. Instead, the corresponding chain is the smallest that contains the image of the original under any map which has the given train track.

Clearly, the criteria in 3.6 are necessary for a train track to be valid. The conjecture is that they are sufficient. More work needs to be done to prove this, and perhaps find a better characterization of the last condition.

So long as these conditions are the correct ones, we then know that the regions of validity are polyhedra. That is, they are intersections Euclidean half-spaces. This is because each of the conditions manifests itself as a system of linear inequalities of the stretched lengths, which, through our calculated relationship T from 3.1 are transformed into linear inequalities of the target lengths. (The solution sets of linear inequalities are, of course, polyhedra). Moreover, we know that the regions of validity are cones (that is, invariant under positive scalar multiplication), since scaling up the target lengths uniformly will not change the combinatorics of the harmonic map. Alternatively, one may note that the inequalities correspond to whether some linear transformation maps into the first quadrant.

Since the regions of validity are cones, we can get all the information we need about them from their projection onto the standard $(n-1)$-simplex (the subset on which the sum of all the target lengths is 1). This is especially convenient, since then for maps to a tripod, we have a clear 2-dimensional representation of the regions of validity.

Using code written by the authors, as well as code written by Andrew Henderson and code written by Dylan Thurston, we were able to experimentally determine and plot the regions of validity for different train track structures on some maps into tripods. These are presented in 12 and 13.

Understanding regions of validity allows one to understand the global structure of the energy functions of maps between graphs. For each n, there are only finitely many trees with n boundary vertices, and so we can understand the energies of maps from a fixed elastic graph with n marked points into a tree by examining finitely many homotopy classes of maps and finitely many train track structures. This would give a decision procedure to check whether one network is looser than another directly; we just check–on finitely many sufficiently small regions of $\mathbb{R}^n_{\geq 0}$–whether one quadratic function is greater than another. Geometric approaches to getting results about looseness are given in the next section.

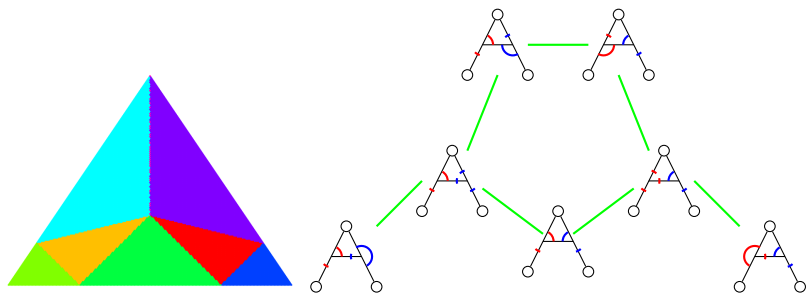


Figure 12: On the left, we have pictured experimentally determined regions of validity for maps from an A graph with uniform elasticities graph into the tripod. The different colors represent different train track structures on A. On the right, we have all the different train track structures that will occur for a map from any elastic A to the tripod. Edges are drawn between train track structures which are in some sense similar. We see that this graph corresponds nicely to the regions on the left.

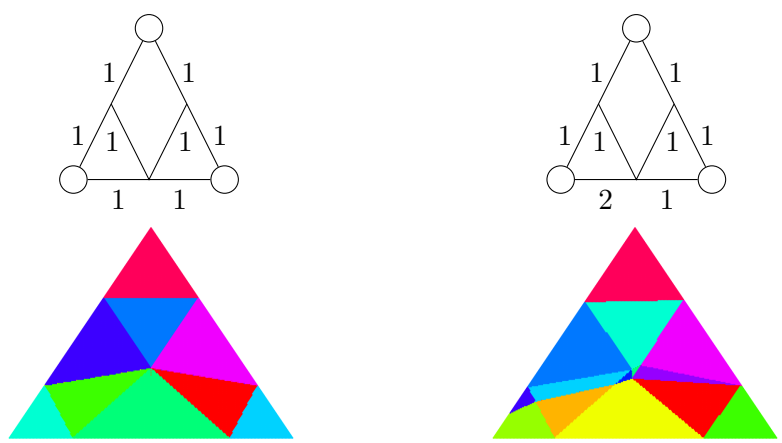


Figure 13: We map the elastic ("M") graphs shown in the top row to the tripod and plot the regions of validity of the train tracks that appear. For the uniformly elastic M, only 9 distinct train tracks are valid in a 3-dimensional region. When we break the symmetry as in the graph on the right, 5 more regions appear. If we try to make a graph of all possible train track structure, as in 12, we get a much larger graph that does not correspond nicely to either of these, and, in fact, does not embed nicely into the plane. Further work must be done to characterize how regions of validity collapse and evolve as we vary elasticities on the source.

4 The tripod and the delta are extremal

We now prove one of our main results, that elastic tripods, in some sense, minimize Dirichlet energy, while elastic deltas maximize it. Our methods do not make use of the train track structures detailed above, but rather some Euclidean geometry. The precise result is as follows:

Proposition 4.1. *Let G be any 3-marked elastic network, and let Y be the electrically equivalent generalized tripod and Δ the electrically equivalent generalized delta. Then $Y \preceq G \preceq \Delta$.*

The proof is simplified because the image of a harmonic map from a 3-marked elastic network to a tree lies is a 3-marked tripod, 2- or 3-marked interval, or a single marked vertex (14 should make this clear, but this also makes an instructive exercise).

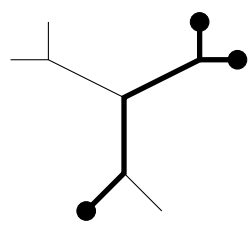


Figure 14: If any of the image lies outside the bold tripod, it would retract onto the tripod, decreasing Dirichlet energy, thus contradicting that harmonic maps minimize Dirichlet energy

The proposition is trivial in the case of the single marked vertex, and it follows trivially from electrical equivalence in the case of the interval. Therefore, we really only need to consider maps to a length tripod K; that is, it suffices to show the following:

Proposition 4.2. *Let G, Δ, and Y be as in 4.1, let L be a length tripod, and let $f_G : G \to L$, $f_\Delta : \Delta \to L$, and $f_Y : Y \to L$ be harmonic and send corresponding marked points of the three graphs to the same outer vertices of L, hitting all three vertices. Then* $\operatorname{Dir} f_Y \leq \operatorname{Dir} f_G \leq \operatorname{Dir} f_\Delta$.

4.1 The tripod is extremal

First we prove $\operatorname{Dir} f_Y \leq \operatorname{Dir} f_G$.

The idea of the proof is that f_Y "looks like" a harmonic map to Euclidean space, while f_G may be "tighter" than a harmonic map to Euclidean space. Specifically, we use the following lemma:

Lemma 4.3. *Let H be a marked elastic graph and K be a marked length tree. Let $f : H \to K$ be harmonic. Suppose there exists a choice of marked points in $\mathbb{R}^n$ and a map $\iota : K \to \mathbb{R}^n$ which is isometric onto its image on each edge of K and such that the composition $\iota \circ f$ is also harmonic. Then, for any elastic graph H' which is electrically equivalent to H, we have*

$$\operatorname{Dir} f \leq \operatorname{Dir} g,$$

where $g : H' \to K$ is harmonic and in the homotopy class corresponding to f.

Proof of 4.3. Since ι is isometric on each edge, clearly $\operatorname{Dir} \iota \circ f = \operatorname{Dir} f$ and $\operatorname{Dir} \iota \circ g = \operatorname{Dir} g$. Moreover, since H and H' are electrically equivalent, their corresponding harmonic maps to Euclidean space have equal energy; that is, $\operatorname{Dir}[\iota \circ f] = \operatorname{Dir}[\iota \circ g]$. Therefore, we have

$$\operatorname{Dir} f = \operatorname{Dir}(\iota \circ f) = \operatorname{Dir}[\iota \circ f] = \operatorname{Dir}[\iota \circ g] \leq \operatorname{Dir}(\iota \circ g) = \operatorname{Dir} g$$

as desired. □

The trick is actually constructing such a map ι, as follows.

Proof of $\operatorname{Dir} f_Y \leq \operatorname{Dir} f_G$. Apply 4.3 with $H = Y$ $H' = G$, and $K = L$. We need only find an appropriate ι. Pick any point in $\mathbb{R}^2$ for the central vertex of L to go to. As ι is isometric on the edges of L, it's determined uniquely once the angles of its edges are specified, so we need only specify angles making $\iota \circ f_Y$ harmonic.

If Y is a generalized tripod which is not a tripod, the correct choice of ι lays 2 edges at a 180° angle to one another; the reader can fill in the details. If Y is a tripod and the central vertex of Y maps to the interior of a leg of L, lay that leg at a 180° angle to the other two.

More interesting, however, is the last case, that Y is a tripod and its central vertex maps to L's central vertex. Here, if we can pick three vectors with the stretched tensions of Y as magnitudes whose sum is 0, just let ι orient each leg parallel to the corresponding tension vector. We now only need to choose these three vectors.

As in this case the tensions satisfy the triangle inequalities, a triangle with those tensions as sides exists, so the vectors corresponding to the sides of the triangle, oriented say counterclockwise, suffice, as shown in 15.

□

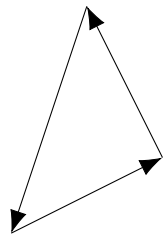


Figure 15: These vectors sum to 0, so take these as the tensions.

Remark. In fact, the proof above can easily be adapted to show that every n-star (configuration of n marked points each connected by one edge to a central vertex) is looser than any elastic graph electrically equivalent to it. Once again, it suffices to consider maps from elastic n-stars to length n-stars (after considering some degenerate cases), and we simply need to construct the map ι from 4.3. The tensions on the central vertex satisfy all the triangle inequalities, so there exists an n-gon in $\mathbb{R}^2$ whose side lengths are the (magnitudes of the) tensions. We once again orient each leg in the direction of the corresponding tension in the n-gon to get the map ι.

4.2 The delta is extremal

Now we prove that $\operatorname{Dir} f_G \leq \operatorname{Dir} f_\Delta$.

Definition 4.4 (Comparison triangles)**.** Given 3 points (p, q, r) in a metric space, a (Euclidean) comparison triangle is a choice of 3 points (p', q', r') in $\mathbb{R}^2$ with matching distances; that is, $d(p, q) = d(p', q')$ and similarly for the other pairs.

Remark. As metric spaces satisfy the triangle inequality, (possibly degenerate) comparison triangles always exist.

Proposition 4.5. *Let Δ be any elastic generalized delta with its 3 vertices marked, and let G be any electrically equivalent elastic network with 3 marked vertices. Then $G \preceq \Delta$.*

We need the following lemma:

Lemma 4.6. *If L is a pipe tripod and (p', q', r') is a comparison triangle for the outer vertices (p, q, r) of L, a short (i.e. distance-decreasing) PL map π from the solid triangle $\triangle p'q'r'$ (i.e. the convex hull of the three points) to L exists sending $p' \mapsto p$, $q' \mapsto q$, and $r' \mapsto r$.*

Proof of 4.6. A map from the boundary of the comparison triangle (that is, the union of the segments connecting two of the comparison points) to L exists which is isometric on each edge: send segment $p'q'$ isometrically to the geodesic pq, and so on. Informally, we can think of this map as adding a hinge to the comparison triangle at each of its vertices and each of the three preimages of the central vertex, and then "collapsing" the triangle, as shown in 16.

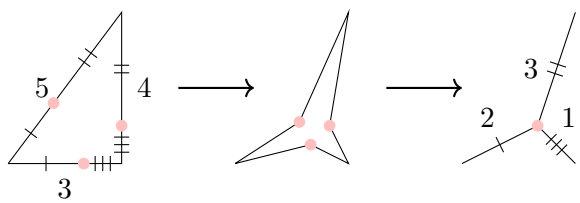


Figure 16: A comparison triangle collapsing onto its tripod

Notice that the pink preimages of the central vertex are the points where the incircle of $\triangle p'q'r'$ meets the triangle; this can be seen, for instance, by the fact that only one choice of three pink points satisfies the congruences in the first triangle of figure 1 and the points where the incircle meets the triangle also satisfy them.

We can therefore extend this collapse to the entire solid triangle: subdivide the triangle into 3 via its angle bisectors and orthogonally project each subtriangle onto its corresponding side as shown in 17 before collapsing. Notice this is well-defined on the angle bisectors: no matter which side we project onto to resolve the ambiguity, the image in the final tripod is the same. Call the extension π.

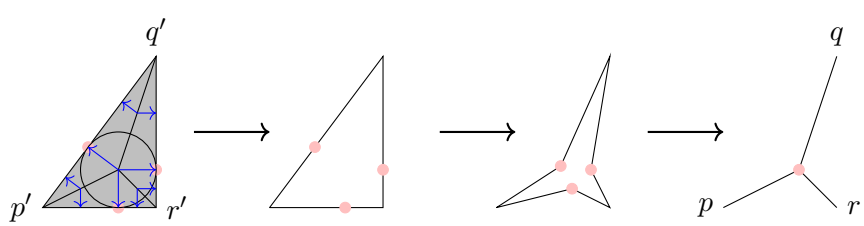


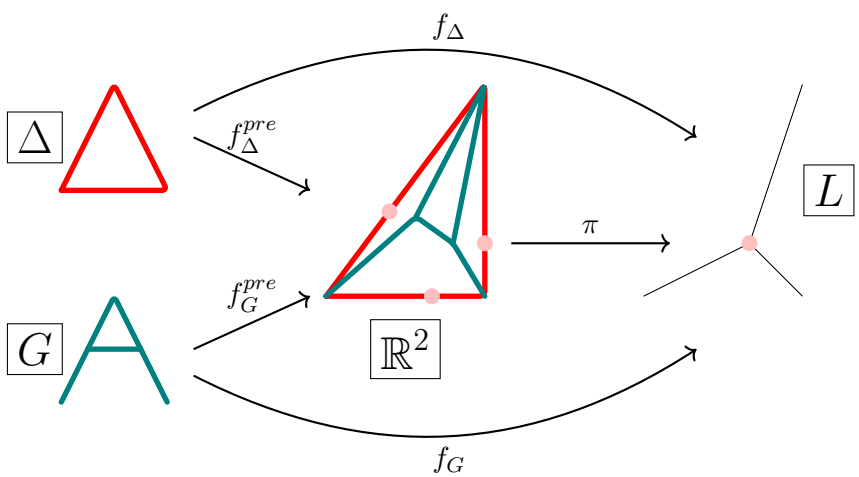
Figure 17: The extended map π

Now π is short, as desired: the pullback metric (in the sense of differential geometry) from the tripod, which is defined a.e., is simply the appropriate orthogonal projection of the Euclidean metric, and so measures distances no

greater than the Euclidean metric. That $p' \mapsto p$, $q' \mapsto q$, $r' \mapsto r$, and that π is PL are clear. $\square$

Proof of 4.5. Let L be a pipe tripod, and let $f_G : G \to L$ and $f_\Delta : \Delta \to L$ be harmonic and send corresponding marked points to the same outer nodes of L. We wish to show $\operatorname{Dir} f_G \leq \operatorname{Dir} f_\Delta$.

Let $f_G^{pre} : G \to \mathbb{R}^2$ and $f_\Delta^{pre} : \Delta \to \mathbb{R}^2$ be harmonic and send the marked points of G and Δ to the corresponding vertices of a comparison triangle for L. Let π be the short map of 4.6. The image of f_G^{pre} lies entirely in the solid comparison triangle, as otherwise we could retract f_G^{pre} onto the comparison triangle, decreasing distances and therefore Dirichlet energy, contadicting harmonicity of f_G^{pre}. Thus $\pi \circ f_G^{pre}$ makes sense, and the following diagram commutes up to marked homotopy:



Now we have

$$\operatorname{Dir} f_G \overset{(1)}{\leq} \operatorname{Dir}\left(\pi \circ f_G^{pre}\right) \overset{(2)}{\leq} \operatorname{Dir} f_G^{pre} \overset{(3)}{=} \operatorname{Dir} f_\Delta^{pre} \overset{(4)}{=} \operatorname{Dir} f_\Delta,$$

where (1) follows as f_G is harmonic and harmonic maps minimize Dirichlet energy in their homotopy class, (2) follows as π is short and therefore decreases energies, (3) follows from 2.12, and (4) follows as both f_Δ and f_Δ^{pre} stretch the sides of Δ to the same lengths. $\square$

Remark 4.7. The above proof almost generalizes to arbitrary Hadamard targets rather than pipe tripods; the only thing that doesn't generalize easily is the construction of the short map π. However, [4, Theorem 5.4] guarantees

its existence (the construction is amusing: fill it in one point at a time!) Thus deltas are also extremal when mapping to Hadamard spaces.

5 Elastically equivalent graphs

Another interesting relationship between graphs is the following:

Definition 5.1 (Elastic equivalence)**.** Define G_1 and G_2 to be *elastically equivalent* if $G_1 \preceq G_2$ and $G_2 \preceq G_1$; that is, whenever the two have the same Dirichlet energies when stretched out in a tree.

Of course, replacing an edge of a graph with two edges in series or in parallel gives another elastically equivalent graph, but are there any examples of elastically equivalent graphs which aren't "trivially" elastically equivalent in this way?

There in fact are.

Proposition 5.2. *Let g_1 and g_2 be any* ***electrically*** *equivalent graphs with 3 marked vertices pasted into triangles as shown below (with their marked vertices placed on the pink vertices), let G_1 and G_2 be the resulting networks, and let $\alpha_1\beta_1\gamma_1 = \alpha_2\beta_2\gamma_2$. Then the graphs of 18 are* ***elastically*** *equivalent.*

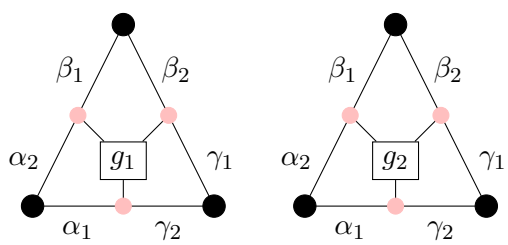


Figure 18: Elastically equivalent networks

Some intuition can be gained as follows: the condition $\alpha_1\beta_1\gamma_1 = \alpha_2\beta_2\gamma_2$ roughly translates to "the amounts the g_is get twisted in the clockwise and counterclockwise directions cancel," forcing them to flatten, which is just the electrical case; thus we only see the electrical properties of the g_is.

Proof. We prove this in the case that g_2 is a generalized delta; 2.9 and transitivity of elastic equivalence take care of the rest.

Again we only have to consider maps to an arbitrary pipe tripod L. It suffices to show that the harmonic map $f : G_2 \to L$ doesn't pull the three

pink vertices into different legs of the tripod, for then they lie along a single geodesic of L, so by electrical equivalence we can "cut" out g_2 from f and "paste" in g_1 without affecting harmonicity or Dirichlet energy.

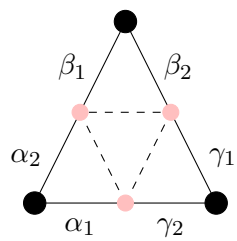


Figure 19: G_2 if g_2 is a generalized delta; some or all of the dashed lines may be missing.

We prove the impossibility of f pulling the three pink vertices into different legs of L by casework. All cases except the counterclockwise one in 20 and its clockwise "mirror" are rather trivial: a node maps to the middle of an edge of L, but the only tensions on it pull it to the center of L, so f can't be harmonic. For brevity, we omit the details of those cases.

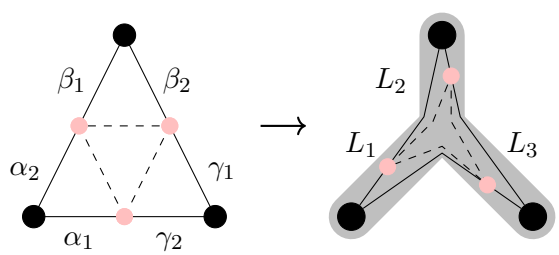


Figure 20: The main case to be eliminated

In this case, the tension of α_2, which is $< \frac{L_1}{\alpha_2}$, must be at least strong enough to cancel β_1's tension, which is $> \frac{L_2}{\beta_1}$; that is,

$$\frac{L_2}{\beta_1} < \frac{L_1}{\alpha_2}.$$

Similarly,

$$\frac{L_3}{\gamma_1} < \frac{L_2}{\beta_2} \text{ and}$$
$$\frac{L_1}{\alpha_1} < \frac{L_3}{\gamma_2}.$$

Therefore

$$L_1 < \frac{\alpha_1 L_3}{\gamma_2} < \frac{\alpha_1 \gamma_1 L_2}{\gamma_2 \beta_2} < \frac{\alpha_1 \gamma_1 \beta_1 L_1}{\gamma_2 \beta_2 \alpha_2} = L_1,$$

contradiction. The impossibility of the "mirror" case is similar. □

6 Open questions and conjectures

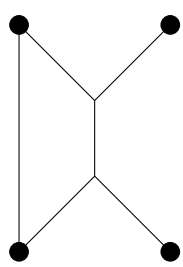
We raise the following questions.

6.1 Generalizing to more marked vertices

Conjecture 6.1. Trees are extremal; that is, if T is any elastic tree with some number of marked vertices and G is any electrically equivalent elasticc graph, $T \preceq G$.

Remark. Directly generalizing 4.1's argument results in problems if two disjoint bits of the elastic graph run through the same pipe node; that is, if the preimage of small neighborhoods of that node are disconnected. Then the disjoint bits might try to angle the adjacent pipes in two different configurations, making ι impossible to construct.

Question 6.2. When we move to 4 or more marked vertices, not everything is equivalent to a tree; for example, with 4 marked vertices we need to add in networks of the following form:



(This can be seen, say, by [2, Theorem 2]: depending on the elasticities, the possible response matrices of this graph have 6 degrees of freedom,

whereas trees can only have up to 5.) What is the correct generalization of the tripod's extremality in this case?

Question 6.3. What analogue does the delta have in cases with more marked vertices?

More generally, we have the following question.

Question 6.4. What other criteria are there for looseness?

6.2 Elastic equivalence

Question 6.5. Is there a better proof of 5.2 than ad hoc casework?

Elastic equivalence seems a rare phenomenon; therefore it may be possible to give a characterization.

Question 6.6. Is there a characterization of elastic equivalence?

6.3 Generalizations to Hadamard space targets

As pointed out in 4.7, the delta is still extremal for Hadamard target spaces.

Conjecture 6.7. The tripod is also still extremal when mapping to Hadamard target spaces.

Conjecture 6.8. The graphs of 5 are still equivalent when mapping to Hadamard targets. (We expect this is false.)

6.4 Exploring Regions of Validity

Question 6.9. Are the conjectured criteria for validity in in 3.6 sufficient?

Question 6.10. Is there a more straightforward description of the third criterion in 3.6?

Question 6.11. How do we characterize the evolution of regions of validity as elasticities are varied on the source graph?

6.5 Assorted unexplained phenomena

6.5.1 "Refraction Patterns"

Question 6.12. The plots in 13 and 21 have an appearance suggestive of a refraction pattern. Is the evolution of regions of validity at all related to refraction?

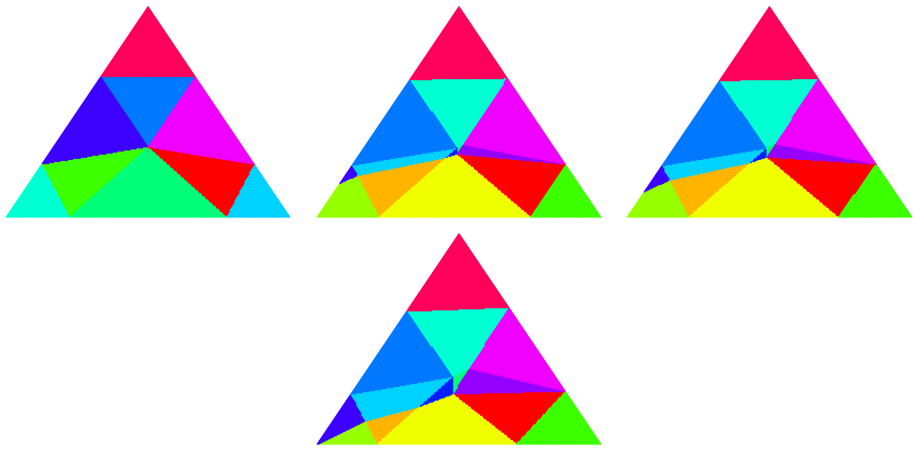


Figure 21: Comparing regions of validity with different source elasticities reveals "refraction patterns."

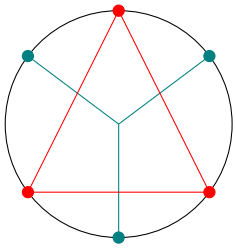


Figure 22: Duality.

6.5.2 Duality

The tripod and the delta can be considered "dual" to one another, as follows:

That is, they're in some sense "duals on a circle;" this easily generalizes to planar graphs embedded in a circle with their marked nodes on the boundary of the circle. The best way of describing the dual of an elastic graph makes the following conversions:

$$\begin{aligned}
\text{Tension} &\Rightarrow \text{Length} \\
\text{Length} &\Rightarrow \text{Tension} \\
\text{Elasticity} &\Rightarrow \frac{1}{\text{Elasticity}} \\
\text{Hooke's Law} &\Rightarrow \text{Hooke's Law} \\
\text{Dirichlet energy} &\Rightarrow \text{Dirichlet energy}
\end{aligned}$$

The most remarkable things, however, are the following:

$$\begin{aligned}
\text{Kirchoff's loop rule} &\Rightarrow \text{Kirchoff's junction rule} \\
\text{Kirchoff's junction rule} &\Rightarrow \text{Kirchoff's loop rule}
\end{aligned}$$

This has been explored for resistor networks; for example, in [3, Section 3] the electrical case with 2 marked vertices is examined.

Due to a number of suggestive facts, we believe this should generalize somehow to harmonic maps from elastic graphs to pipe graphs. Firstly, the triangle inequality condition for harmonicity is exactly dual to the metric triangle inequality between the endpoints of a pipe tripod, generalizing the swap between Kirchoff's laws in the electrical case. In addition, a number of matrices that seemed related to duality (from calculations of how duality should behave) appeared in other calculations. Even the two triangles in the proofs of 4 are dual to one another!

A full description of duality seems elusive, however. We hope future work can shed some light on this.

A Assorted proofs that $Y \preceq \Delta$

Over the course of our research, we found several proofs of the following proposition:

Proposition A.1. *Let Y be an elastic tripod, let Δ be an electrically equivalent delta, and let K be a length tripod. If $f : Y \to K$ is harmonic, and $g : \Delta \to K$ is harmonic in the corresponding homotopy class, then* $\operatorname{Dir} f \leq \operatorname{Dir} g$.

This is, of course, implied by our main results in 4. However, it may be fruitful to examine the different proofs we found and search for generalizations. Therefore, we include several proofs (sometimes under more assumptions) of A.1 below.

Proof (Algebraic, Vertex-to-vertex case only). Let $f : Y \to K$ be a harmonic map from an elastic tripod to a length tripod, both with outer vertices marked. If f maps the central vertex of Γ to the central vertex then we know that the triangle inequalities are satisfied:

$$\frac{L_1}{\alpha_1} \leq \frac{L_2}{\alpha_2} + \frac{L_3}{\alpha_3}, \quad \frac{L_2}{\alpha_2} \leq \frac{L_3}{\alpha_3} + \frac{L_1}{\alpha_1}, \quad \frac{L_3}{\alpha_3} \leq \frac{L_1}{\alpha_1} + \frac{L_2}{\alpha_2},$$

where L_i are the lengths on K and α_i are the corresponding elasticities on Y. We define elasticities on a delta graph Δ by the familiar Y-Δ-transformation:

$$\alpha_a = \frac{\alpha}{\alpha_1}, \quad \alpha_b = \frac{\alpha}{\alpha_2}, \quad \alpha_c = \frac{\alpha}{\alpha_3}$$

where $\alpha := \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1$. Doing the inverse transform gives another set of triangle inequalities:

$$L_1 \frac{\alpha_a + \alpha_b + \alpha_c}{\alpha_b\alpha_c} \leq L_2 \frac{\alpha_a + \alpha_b + \alpha_c}{\alpha_a\alpha_c} + L_3 \frac{\alpha_a + \alpha_b + \alpha_c}{\alpha_a\alpha_b}$$

and so on, which, dividing by $\frac{\alpha_a+\alpha_b+\alpha_c}{\alpha_a\alpha_b\alpha_c}$, in turn yields

$$\alpha_a L_1 \leq \alpha_b L_2 + \alpha_c L_3, \quad \alpha_b L_2 \leq \alpha_c L_3 + \alpha_a L_1, \quad \alpha_c L_3 \leq \alpha_a L_1 + \alpha_b L_2.$$

Combining two triangle inequalities at a time gives

$$|\alpha_a L_1 - \alpha_b L_2| \leq \alpha_c L_3, \quad |\alpha_b L_2 - \alpha_c L_3| \leq \alpha_a L_1, \quad |\alpha_c L_3 - \alpha_a L_1| \leq \alpha_b L_2,$$

which gives the inequality:

$$(\alpha_a L_1 - \alpha_b L_2)^2 + (\alpha_a L_2 - \alpha_b L_3)^2 + (\alpha_a L_3 - \alpha_b L_1)^2 \leq (\alpha_a L_1)^2 + (\alpha_b L_2)^2 + (\alpha_c L_3)^2.$$

Expanding and rearranging terms gives

$$\alpha_a^2 L_1^2 + \alpha_b^2 L_2^2 + \alpha_c^2 L_3^2 \leq 2(\alpha_a\alpha_b L_1 L_2 + \alpha_b\alpha_c L_2 L_3 + \alpha_c\alpha_a L_3 L_1),$$

and dividing both sides by $\alpha_a\alpha_b\alpha_c$ gives

$$\frac{\alpha_a}{\alpha_b\alpha_c} L_1^2 + \frac{\alpha_b}{\alpha_c\alpha_a} L_2^2 + \frac{\alpha_c}{\alpha_a\alpha_b} L_3^2 \leq 2\left(\frac{L_1 L_2}{\alpha_c} + \frac{L_2 L_3}{\alpha_a} + \frac{L_3 L_1}{\alpha_b}\right).$$

Adding $\frac{\alpha_b+\alpha_c}{\alpha_b\alpha_c}L_1^2 + \frac{\alpha_c+\alpha_a}{\alpha_c\alpha_a}L_2^2 + \frac{\alpha_a+\alpha_b}{\alpha_a\alpha_b}L_3^2$ to both sides and factoring gives

$$\frac{\alpha_a+\alpha_b+\alpha_c}{\alpha_b\alpha_c}L_1^2 + \frac{\alpha_a+\alpha_b+\alpha_c}{\alpha_c\alpha_a}L_2^2 + \frac{\alpha_a+\alpha_b+\alpha_c}{\alpha_a\alpha_b}L_3^2 \\ \leq \frac{(L_2+L_3)^2}{\alpha_a} + \frac{(L_3+L_1)^2}{\alpha_b} + \frac{(L_1+L_2)^2}{\alpha_c}.$$

The right side is easily seen to be the energy of the harmonic map from Δ to K. Applying the inverse Y-Δ transform to the left side gives $\frac{L_1^2}{\alpha_1} + \frac{L_2^2}{\alpha_2} + \frac{L_3^2}{\alpha_3}$, which is the energy of the harmonic map $f : Y \to K$. $\square$

Proof (linear algebra bash, vertex-to-vertex case only). We need only consider the case that the central vertex of the tripod maps to the central vertex of the pipe tripod; that is, the case that $\frac{\ell_1}{\alpha_1}, \frac{\ell_2}{\alpha_2}$, and $\frac{\ell_3}{\alpha_3}$ satisfy the triangle inequalities. This is true iff the vector

$$L = \begin{pmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{pmatrix}$$

is a nonnegative linear combination of

$$\begin{pmatrix} 0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ 0 \\ \alpha_3 \end{pmatrix}, \text{ and } \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ 0 \end{pmatrix};$$

that is, iff

$$L = \begin{pmatrix} 0 & \alpha_1 & \alpha_1 \\ \alpha_2 & 0 & \alpha_2 \\ \alpha_3 & \alpha_3 & 0 \end{pmatrix} L'$$

for some L' in the first quadrant of $\mathbb{R}^3$.

The Dirichlet energy of Y is

$$
\begin{aligned}
&L^T \begin{pmatrix} \alpha_1^{-1} & & \\ & \alpha_2^{-1} & \\ & & \alpha_3^{-1} \end{pmatrix} L \\
&\quad = L'^T \begin{pmatrix} 0 & \alpha_2 & \alpha_3 \\ \alpha_1 & 0 & \alpha_3 \\ \alpha_1 & \alpha_2 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1^{-1} & & \\ & \alpha_2^{-1} & \\ & & \alpha_3^{-1} \end{pmatrix} \begin{pmatrix} 0 & \alpha_1 & \alpha_1 \\ \alpha_2 & 0 & \alpha_2 \\ \alpha_3 & \alpha_3 & 0 \end{pmatrix} L' \\
&\qquad = L'^T \begin{pmatrix} \alpha_2+\alpha_3 & \alpha_3 & \alpha_2 \\ \alpha_3 & \alpha_1+\alpha_3 & \alpha_1 \\ \alpha_2 & \alpha_1 & \alpha_1+\alpha_2 \end{pmatrix} L'. \quad \text{(A.1)}
\end{aligned}
$$

Similarly, if we let $\sigma = \alpha_1 \cdot \alpha_2 + \alpha_1 \cdot \alpha_3 + \alpha_2 \cdot \alpha_3$, the Dirichlet energy of Δ is

$$
\begin{aligned}
&L^T \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1^{-1} & & \\ & \beta_2^{-1} & \\ & & \beta_3^{-1} \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} L \\
&\qquad = \frac{1}{\sigma} L^T \begin{pmatrix} \alpha_2+\alpha_3 & \alpha_3 & \alpha_2 \\ \alpha_3 & \alpha_1+\alpha_3 & \alpha_1 \\ \alpha_2 & \alpha_1 & \alpha_1+\alpha_2 \end{pmatrix} L \\
&= \frac{1}{\sigma} L'^T \begin{pmatrix} 0 & \alpha_2 & \alpha_3 \\ \alpha_1 & 0 & \alpha_3 \\ \alpha_1 & \alpha_2 & 0 \end{pmatrix} \begin{pmatrix} \alpha_2+\alpha_3 & \alpha_3 & \alpha_2 \\ \alpha_3 & \alpha_1+\alpha_3 & \alpha_1 \\ \alpha_2 & \alpha_1 & \alpha_1+\alpha_2 \end{pmatrix} \begin{pmatrix} 0 & \alpha_1 & \alpha_1 \\ \alpha_2 & 0 & \alpha_2 \\ \alpha_3 & \alpha_3 & 0 \end{pmatrix} L' \\
&\quad = L'^T \left(\begin{pmatrix} \alpha_2+\alpha_3 & \alpha_3 & \alpha_2 \\ \alpha_3 & \alpha_1+\alpha_3 & \alpha_1 \\ \alpha_2 & \alpha_1 & \alpha_1+\alpha_2 \end{pmatrix} + \frac{2\alpha_1\alpha_2\alpha_3}{\sigma} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \right) L' \\
&\qquad = \text{(A.1)} + \frac{2\alpha_1\alpha_2\alpha_3}{\sigma} L'^T \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} L' \quad \text{(A.2)}
\end{aligned}
$$

Clearly, as L' is in the first quadrant and the α_is are positive, (A.2)$\geq$(A.1), as desired. $\square$

References

[1] Dylan Thurston, *Elastic Graphs*, Preprint, 2016, arXiv:1607.00340v1.

[2] E.B. Curtis, D.Ingerman, and J.A. Morrow, *Circular Planar Graphs and Resistor Networks*. Linear Algebra and its Applications 283 (1998) 115-150.

[3] R. L. Brooks, C. A. B. Smith, A. H. Stone, and W. T. Tutte, *The Dissection of Rectangles into Squares*. Duke Math. J. Volume 7, number 1 (1940), 312-340.

[4] S. Alexander, V. Kapovitch, and A. Petrunin, *Alexandrov meets Kirszbraun*, Preprint, 2010, arXiv:1012.5636

Final Report
GTR+Γ+I Model Identifiability using a Discrete Gamma Distribultion

Kees McGahan
Mentor: Elizabeth Houseworth

July 2016

Abstract

Phylogenetics is the study of trees that describe the evolution of organisms. Mathematics provides models for DNA evolution that enable Biologists to determine trees in a rigorous statistical framework. In order for these models to be useful, the models need to be identifiable. That is, with an infinite amount of data from the true model, we should be able to recover all the parameters in that model uniquely. One commonly used model, especially in phylogenetic software, is the GTR+Γ+I model, whose parameters include: T, the phylogenetic tree, Q, the General Time Reversible (GTR) Markov rate matrix, and R, a random mutation rate taken from a mixture of a Gamma distribution and invariant sites. This model has already been shown to be identifiable when the Gamma distribution is continuous. However, all known software packages use a discrete version of the Gamma distribution instead. In this REU project, we are exploring both numerically and analytically whether the model with only two rate classes and no invariant sites is identifiable.[1]

[1]This material is based upon work supported by the National Science Foundation under Grant No. DMS-1461061.

This summer, I pursued a problem in mathematical biology, in the identifiability of trees. Phylogenetics is the study of trees that describe how species evolve. Biologists find themselves intrigued with the actual trees themselves, while the math involved revolves around the models used to generate said trees. Our work was centered around one such model, the GTR+Γ+I Model and an attempt to prove the identifiability of this model for a specific case.

The main term to be aware of here is identifiabilty which means: if given an infinite amount of data all of the model's parameters can be uniquely determined. In our case the data comes in the form of DNA sequences of our species. We shall assume two important details about our sequences: the first is that we have at least 4 species, as there is only one unrooted tree topology involving 3 species, and the second being that we assume our sequences contain an infinite amount of data. Through using these infinite sequences we hope to show that we can uniquely determine the parameters of the model. For the model we're using, the GTR+Γ+I Model, the parameters are Q, T and R.

Our parameter Q will be some general time reversible rate matrix with at least two non-zero distinct eigenvalues. The T that we are hoping to determine is the phylogenetic tree itself, which we will consider to be unrooted and as having three or more distinct inter-species distances. The final parameter R, represents a rate of evolution. It has already been proved that if R is a a continuous gamma distribution with mean 1 than the model will be identifiable. My work this summer was with using R as a two rate class parameter still with mean 1. We chose to work with a discrete Gamma distribution as this typically is the distribution that software packages use, although they typically use four rate classes. We hoped to be able to show that working with 2 rate classes would give some insight into using a discrete distribution, and possibly have our work extended to more rate classes.

Thus starting with 4 or more DNA sequences our hope is from pairwise sequence comparisons we can uniquely determine the above parameters. Due to having an assumed infinite amount of data for pairwise comparisons we can easily determine the stationary frequencies for each nucleotide and hence probability matrix for any pair of species. We know through Markov chains that we can arrive at an equation that looks like

$$P_{i,j}(t) = Ee^{t_{i,j}QR}$$

where we take the expected value of R, our discrete distribution of two rate classes. It's known that the eigenvectors of P are the eigenvectors of Q. Thus

our eigenvalues of Q are related to those of P through the equations:

$$m_1 = (e^{tr\lambda_1} + e^{(2-r)t\lambda_1})/2,$$

$$m_2 = (e^{tr\lambda_2} + e^{(2-r)t\lambda_2})/2,$$

$$m_3 = (e^{tr\lambda_3} + e^{(2-r)t\lambda_3})/2$$

Where $\lambda_1, \lambda_2, \lambda_3$ are our 3 distinct non-zero eigenvalues of Q and where m_1, m_2, m_3 are the associated eigenvalues from P. We can generalize this to just thinking of the eigenvalues of P of functions of two parameters such that we get something like this:

$$m_1 = \mu(t\lambda_1, r),$$

$$m_2 = \mu(t\lambda_2, r),$$

$$m_3 = \mu(t\lambda_3, r).$$

We can then assume by contradiction that for one of our eigenvalues of P, say m_1 that we have functions of different parameters that generate it:

$$\mu(t\lambda_1, r) = m_1 = \mu(TL_1, s)$$

where we define $0 < r < s < 1$. This would create a problem as it would mean that the same data m_1 is being generated by different sets of parameters. As our functions are already of two parameters essentially, this can happen a few times, but not too many. From this point we shall invert one side of our equality with respect to the TL_1 component giving:

$$\mu^{-1}(\mu(t\lambda_1, r), s) = m_1 = TL_1.$$

We can also do this using the same pair of species (same t/T values), and with a different distinct eigenvalue to get an equation of the form:

$$\mu^{-1}(\mu(t\lambda_2, r), s) = m_2 = TL_2.$$

Dividing one equation by the other gives a ratio of the form:

$$\frac{\mu^{-1}(\mu(t\lambda_1, r), s)}{\mu^{-1}(\mu(t\lambda_2, r), s)} = \frac{TL_1}{TL_2}$$

Which through some simplification of variables can be written as

$$C = \frac{\mu^{-1}(\mu(\lambda_1, r), s)}{\mu^{-1}(\mu(B\lambda_1, r), s)}$$

where C is a constant and B is a constant with value greater than 1. Both C and B are ratios of eigenvalues which is why we are allowed to assume that B is greater than 1 as we can arbitrarily assume 1 eigenvalue greater than the other to get a value greater than 1. To make sure that this model is identifiable we needed to show that for our ratio doesn't equally C too many times.

Previous work, notably a proof of identifiability of the continuous case showed solved the similar problem through calculus and proving that the line C was crossed only once or twice. We attempt a similar approach looking to find a finite small amount of crossings of C. One way to do this is search for the minima/maxima of the function as these will bound how many times our ratio can cross C. Thus we took the derivative of our ratio. One way to prove that our function only crosses C one time is if our derivative can be broken up into two parts; one increasing and the other decreasing. We attempted to do this and ended up with:

$$\frac{d}{dx}[\frac{\mu^{-1}(\mu(x, r), s)}{\mu^{-1}(\mu(Bx, r), s)}] =$$

$$\frac{\mu^{-1}(\mu(Bx, r)s)\frac{1}{\mu'(\mu^{-1}(\mu(x,r),s),s)}\mu'(x, r) - \mu^{-1}(\mu(x, r)s)\frac{1}{\mu'(\mu^{-1}(\mu(Bx,r),s),s)}\mu'(Bx, r)B}{\mu^{-1}(\mu(Bx, r), s)^2}.$$

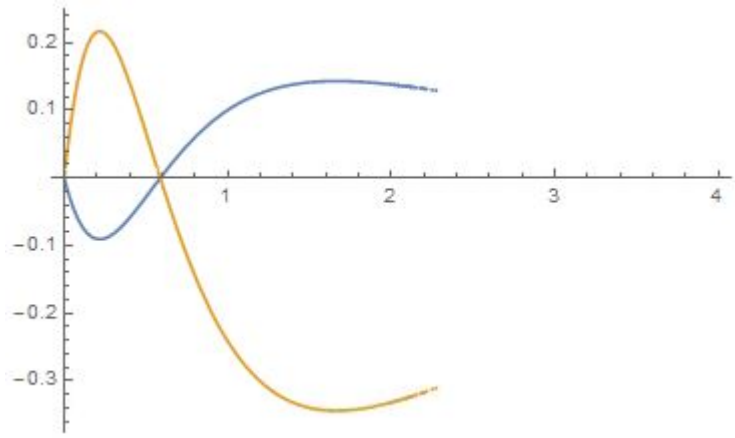
Focusing only on the numerator we can get this too look like the following by utilizing the fact that we have a difference in the numerator:

$$LHS = \frac{1}{b}[\frac{\mu^{-1}(\mu(Bx, r)s)}{\mu^{-1}(\mu(x, r)s)}][\frac{\mu'(x, r)}{\mu'(\mu^{-1}(\mu(x, r), s), s)}][\frac{\mu'(\mu^{-1}(\mu(Bx, r), s), s)}{\mu'(Bx, r)}] - 1$$

and

$$RHS = (\frac{1}{b}-1)[\frac{\mu^{-1}(\mu(Bx, r)s)}{\mu^{-1}(\mu(x, r)s)}][\frac{\mu'(x, r)}{\mu'(\mu^{-1}(\mu(x, r), s), s)}][\frac{\mu'(\mu^{-1}(\mu(Bx, r), s), s)}{\mu'(Bx, r)}]+B-1.$$

These two functions yields a graph of the form:



We realized that this isn't the desired result of one increasing and one decreasing function, however the functions are neatly symmetric and meet at the axis. This changed the goal to showing our LHS and RHS functions have this form, which would give us one critical point in our ratio and hence at most 2 crossings of the line C. To show that our LHS and RHS functions have this form we can only work with one due to the resulting symmetry. Working with the LHS we must show that it is decreasing at 0, at some point crosses the x-axis, and then goes off to some asymptote without crossing the axis again.

Part 1: Increasing at 0

Theorem 1. *The Derivative of the left hand side is negative, hence showing that our LHS function is decreasing and making the RHS function increasing through symmetry*

Proof. The first part we wished to show was that our function LHS decreases at 0. As shown, we break the function up into three parts and individually take their derivatives. The parts

$$[\frac{\mu'(x,r)}{\mu'(\mu^{-1}(\mu(x,r),s),s)}]$$

and

$$[\frac{\mu'(\mu^{-1}(\mu(Bx,r),s),s)}{\mu'(Bx,r)}]$$

derive and evaluated at 0 pose no problems but when we take the derivative of

$$[\frac{\mu^{-1}(\mu(Bx,r)s)}{\mu^{-1}(\mu(x,r)s)}]$$

and evaluate it at 0, we run into some 0/0 terms. As a work around we approximate the μ^{-1} term with a second order Taylor series as shown in the aside.

ASIDE:

We know the second order Taylor series of $\mu(x,r)$ is

$$\mu(x,r) = 1 + x + \frac{(r^2) + (2-r)^2}{4}x^2 + ...$$

To find the inverse of this expansion we shall first find the inverse of the function

$$(\mu(x,r) - 1) = x + \frac{(r^2) + (2-r)^2}{4}x^2 + ...$$

Thus we know that $(\mu - 1)^{-1}(x,s)$ has no constant term and thus has some expansion that looks like

$$\mu - 1)^{-1}(x,s) = b_1 x + b_2 x^2 + ...$$

Since we know that

$$(\mu-1)^{-1}((\mu(x,r)-1),s) = x = b_1(x+\frac{(r^2) + (2-r)^2}{4}x^2+...)+b_2(x+\frac{(r^2) + (2-r)^2}{4}x^2+...)^2+...$$

we can easily solve for the coefficients b_1 and b_2 which we can see are $b_1 = 1$ and

$$b_2 = -\frac{(r^2) + (2-r)^2}{4}$$

This yields the second order Taylor series of

$$(\mu - 1)^{-1}(x,s) = x - \frac{(s^2) + (2-s)^2}{4}x^2 + ...$$

However this can also be used as the Taylor series for

$$\mu^{-1}(\mu(x,r) - 1),s) = (\mu - 1) - \frac{(s^2) + (2-s)^2}{4}(\mu - 1)^2 + ...$$

because we just need to make sure that the Taylor series is still expanded about 0. By evaluating our function at $(\mu(x,r)-1)$ which at $x=0$ is 0, our μ^{-1} is being expanded about 0 as desired. Hence we can use this form, and then substituting in for the above equations for μ yields

$$\mu^{-1}(\mu(x,r)-1,s) = [x+\frac{(r^2)+(2-r)^2}{4}x^2+...]-\frac{(s^2)+(2-s)^2}{4}[x+\frac{(r^2)+(2-r)^2}{4}x^2+...]^2$$

Which simplifying to only the second order terms of our Taylor expansion about zero gives

$$\mu^{-1}(\mu-1(x,r),s) = x+[\frac{(r^2)+(2-r)^2}{4}-\frac{(s^2)+(2-s)^2}{4}]x^2+...$$

BACK TO WORK

Now we can go back to looking at our function LHS and its derivative. We have:

$$LHS = \frac{1}{b}[\frac{\mu^{-1}(\mu(Bx,r)s)}{\mu^{-1}(\mu(x,r)s)}][\frac{\mu'(x,r)}{\mu'(\mu^{-1}(\mu(x,r),s),s)}][\frac{\mu'(\mu^{-1}(\mu(Bx,r),s),s)}{\mu'(Bx,r)}]-1$$

Looking at this in its three parts as

$$LHS = \frac{1}{b}XYZ-1$$

where

$$X = [\frac{\mu^{-1}(\mu(Bx,r)s)}{\mu^{-1}(\mu(x,r)s)}]$$

$$Y = [\frac{\mu'(x,r)}{\mu'(\mu^{-1}(\mu(x,r),s),s)}]$$

$$Z = [\frac{\mu'(\mu^{-1}(\mu(Bx,r),s),s)}{\mu'(Bx,r)}]$$

we see that we're solving for

$$LHS'(0) = X'YZ + XY'Z + XYZ'.$$

Note that we're excluding the $1/B$ term as it wont effect whether our funciton is positive or negative when evaluated at 0. The derivatives of each of the parts are:

$$X' = \frac{B\mu'(Bx,r)\mu^{-1}(\mu(x,r),s)\mu'(\mu^{-1}(\mu(x,r),s),s) - \mu'(x,r)\mu^{-1}(\mu(Bx,r),s)\mu'(\mu^{-1}(\mu(Bx,r),s),s)}{[\mu^{-1}(\mu(x,r),s)]^2\mu'(\mu^{-1}(\mu(Bx,r),s),s)\mu'(\mu^{-1}(\mu(x,r),s),s)}$$

$$Y' = \frac{\mu'(\mu^{-1}(\mu(x,r),s),s)\mu''(x,r) - \mu'(x,r)\mu''(\mu^{-1}(\mu(x,r),s),s)\frac{1}{\mu'(\mu^{-1}(\mu(x,r),s),s)}\mu'(x,r)}{[\mu'(\mu^{-1}(\mu(x,r),s),s)]^2}$$

$$Z' = \frac{\mu'(Bx,r)\mu''(\mu^{-1}(\mu(Bx,r),s),s)\frac{1}{\mu'(\mu^{-1}(\mu(Bx,r),s),s)}\mu'(Bx,r) - \mu'(\mu^{-1}(\mu(Bx,r),s),s)\mu''(Bx,r)}{[\mu'(Bx,r)]^2}.$$

We also know what the components evaluated at 0 are:

$$\mu(0,r) = (e^0 + e^0)/2 = 1$$

$$\mu'(0,r) = (re^0 + (2-r)e^0)/2 = 1$$

$$\mu''(0,r) = (r^2e^0 + (2-r)^2e^0)/2 = (r^2 + (2-r)^2)/2$$

Using the above information it is easy to see that both Y and Z evaluated at 0 are both 1. Evaluating X at 0 would give 0/0 but using a first order Taylor series we can see goes to B. So far that leaves us with $LHS(0) = X' + B*Y' + B*Z'$. Then utilizing the information from the ASIDE for the $\mu^{-1}(\mu(x,r),s)$ terms that appear in X and evaluating at our derivatives we get something that simplifies out to

$$LHS'(0) = B(1-B)(\frac{r^2 + (2-r)^2 - s^2 - (2-s)^2}{4} + \frac{r^2 + (2-r)^2 - s^2 - (2-s)^2}{2}).$$

All that's left is to show that based on $0 < r < s < 1$ and $B > 1$ this is negative. The only part whose sign isn't immediately determined is whether

$$r^2 + (2-r)^2 - s^2 - (2-s)$$

is positive. We can see that this equals

$$(r+s)(r-s) + (s-r)(4-r-s) =$$

$$(r-s)(r+s-4+r+s) =$$
$$2(r-s)[r+s-2].$$

Which based on restrictions is positive. And therefore the equation

$$LHS'(0) = 2B\frac{r^2+(2-r)^2-s^2-(2-s)^2}{4}(1-B)$$

is indeed negative due to the $(1-B)$ term. Hence we've shown that derivative of the LHS at 0 is negative, and so due to symmetry our LHS is decreasing and our RHS is increasing at $x=0$. □

Part 2: Goes to asymptote

I was unable to show that my functions LHS and RHS have asymptotes at some constant as X goes to infinity. However we were able to find the asymptote at infinity of the original ratio. Furthermore as this asymptote is a constant, we could hypothesize that the LHS and RHS would at infinity go to 0 as they are derivatives of our original ratio. Knowing this could provide more unique bounds for Part 3.

Lemma 1. *The original ratio of* $[\frac{\mu^{-1}(\mu(x,r)s)}{\mu^{-1}(\mu(Bx,r)s)}]$ *goes to an asymptote of* $1/B$ *at infinity.*

Proof. Looking at the ratio we have

$$[\frac{\mu^{-1}(\mu(x,r)s)}{\mu^{-1}(\mu(Bx,r)s)}].$$

We know that

$$\mu(x,r) = \frac{e^{rx}+e^{(2-r)x}}{2} = \frac{e^{(2-r)x}}{2}(1+e^{-2(1-r)x}).$$

This will be dominated by the $\frac{e^{(2-r)x}}{2}$ part as x goes to infinity. Thus as x goes to infinity

$$\mu^{-1}(x,r) = \frac{\ln(2x)}{2-r}.$$

Therefore we can see that

$$\mu^{-1}(\mu(x,r)s) = \frac{\ln(2e^{(2-r)x})}{2-r} = \frac{2-r}{2-s}x.$$

Using the same logic we can see that

$$\mu^{-1}(\mu(Bx,r)s) = \frac{\ln\left(2e^{(2-r)Bx}\right)}{2-r} = \frac{2-r}{2-s}Bx.$$

Therefore as x goes to infinity our ratio

$$[\frac{\mu^{-1}(\mu(x,r)s)}{\mu^{-1}(\mu(Bx,r)s)}]$$

goes to $1/B$.

$\square$

Part 3: Crosses a small finite amount of times

We were unable to prove this but we do have a hypothesis that the LHS (and hence through symmetry RHS) both only cross the axis a finitely small amount of times. In fact our hypothesis would by they cross only the axis only once each. We believe it to be a true hypothesis due to extensive work in mathematica, and that we have yet to find a counterexample for any values of our parameters of r, s, B. A lot of our work in this area has been hoping to find bounds that we could use to approximate our functions and work with them instead. As of yet we haven't managed to complete our goal.

References

Chai, Juanjuan, and Elizabeth A. Housworth. "On Rogers' proof of identifiability for the GTR+ Γ+ I model." Systematic biology (2011): syr023.