Geometry & Topology

Volume 29 (2025)

Issue 4 (pages 1693–2250)

msp

GEOMETRY & TOPOLOGY

msp.org/gt

See inside back cover or msp.org/gt for submission instructions.

The subscription price for 2025 is US $865/year for the electronic version, and $1210/year (+$75, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Geometry & Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Geometry & Topology (ISSN 1465-3060 printed, 1364-0380 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

GT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Geometry & Topology **29**:4 (2025) 1693–1817
DOI: 10.2140/gt.2025.29.1693
Published: 27 June 2025

The conjugacy problem for UPG elements of Out(F_n)

MARK FEIGHN
MICHAEL HANDEL

An element ϕ of the outer automorphism group $\mathsf{Out}(F_n)$ of the rank n free group F_n is *polynomially growing* if the word lengths of conjugacy classes in F_n grow at most polynomially under iteration by ϕ. It is *unipotent* if, additionally, its action on the first homology of F_n with integer coefficients is unipotent. In particular, if ϕ is polynomially growing and acts trivially on first homology with coefficients the integers mod 3, then ϕ is unipotent and also every polynomially growing element has a positive power that is unipotent. We solve the conjugacy problem in $\mathsf{Out}(F_n)$ for the subset of unipotent elements. Specifically, there is an algorithm that decides if two such are conjugate in $\mathsf{Out}(F_n)$.

20F65, 57M07

1 Introduction

In this paper, we consider the conjugacy problem for $\mathsf{Out}(F_n)$, the group of outer automorphisms of the free group of rank n. Namely, given $\phi, \psi \in \mathsf{Out}(F_n)$, find an algorithm that decides if ϕ and ψ are conjugate in $\mathsf{Out}(F_n)$.

The case in which ϕ is fully irreducible, also known as iwip, was first solved by Sela [1995] using his solution to the isomorphism problem for torsion-free word hyperbolic groups. This was recently generalized, using a similar approach, by Dahmani [2016] to the case that ϕ is hyperbolic, or equivalently, that every nontrivial element of F_n has exponential growth under iteration by ϕ. See also [Dahmani 2017]. An alternate approach to the fully irreducible case takes advantage of the fact that the finite set of (unmarked) train track maps that represent a fully irreducible ϕ is a complete invariant for the conjugacy class of ϕ. Los [1996] and Lustig [2007] (see also [Handel and Mosher 2011]) solved the conjugacy problem for fully irreducible ϕ by algorithmically constructing the set of (unmarked) train track maps for ϕ.

On the other end of the growth spectrum, the conjugacy problem for Dehn twists (equivalently rotationless, linearly growing ϕ) was solved by Cohen and Lustig [1999] using, among other things, Whitehead's algorithm (see below). Krstić, Lustig and Vogtmann [Krstić et al. 2001] proved an equivariant Whitehead algorithm and used that to solve the conjugacy problem for all elements with linear growth.

Building on the approach of Sela mentioned above, Dahmani and Touikan [2021] reduce the conjugacy problem for $\mathsf{Out}(F_n)$ to a list of problems about mapping tori of polynomial growing elements. This is applied in their solution to the conjugacy problem for outer automorphisms of free groups whose polynomially growing part is unipotent linear [Dahmani and Touikan 2023].

Dahmani, Francaviglia, Martino and Touikan [Dahmani et al. 2025] solve the conjugacy problem for $\mathsf{Out}(F_3)$.

Lustig [2000; 2001] posted preprints addressing the general case of the conjugacy problem but these have never been published.

Our main theorem addresses the case that ϕ is polynomially growing and rotationless, equivalently ϕ is polynomially growing and induces a unipotent action on $H_1(F_n, \mathbb{Z})$; we write $\phi \in \mathsf{UPG}(F_n)$. Being an element of $\mathsf{UPG}(F_n)$ is a conjugacy invariant and can be checked algorithmically.

It is often the case, when studying $\mathsf{Out}(F_n)$, that the techniques required to treat the $\mathsf{UPG}(F_n)$ case are very different from those needed for the cases in which there is exponential growth. For example, the polynomially growing and exponentially growing cases of the Tits alternative for $\mathsf{Out}(F_n)$ are proved in separate papers; see Bestvina, Feighn and Handel [2005; 2000].

Theorem 1.1 *There is an algorithm that takes as input $\phi, \psi \in \mathsf{UPG}(F_n)$ and outputs* YES *or* NO *depending on whether or not there exists $\theta \in \mathsf{Out}(F_n)$ such that $\phi = \psi^\theta := \theta\psi\theta^{-1}$. Further, if* YES *then the algorithm also outputs such a θ.*

Remark 1.2 If one knows that ϕ and ψ are conjugate, then a conjugator θ can be produced by searching a list of the elements of $\mathsf{Out}(F_n)$. This is not what we do. Rather, the construction of a conjugator, when one exists, is an integral part of the proof of the main statement of Theorem 1.1.

Remark 1.3 Theorem 1.1 is not an abstract existence theorem. It is proved by constructing an explicit algorithm satisfying the conclusions of the theorem. The same is true for other results in this paper that begin with, "There is an algorithm".

A detailed description of the algorithm is given in Section 2 so we restrict ourselves here to four results/observations that underlie our proof.

- Each $\phi \in \mathsf{UPG}$ is rotationless (Lemma 3.18) and so can be represented by a particularly nice relative train track map $f : G \to G$ call a CT; see Section 3.6. There is an algorithm (Theorem 3.20) to construct one such $f : G \to G$ and from this we can compute all of the invariants used in this paper.

- A set equipped with an action by a group G is a G*-set*. A G-set X *satisfies property* W (for Whitehead) if it comes equipped with an algorithm that takes as input $x, y \in X$ and outputs `YES` or `NO` depending on whether or not there exists $\theta \in G$ such that $\theta(x) = y$ together with such a θ if `YES`. We call such an algorithm a W-*algorithm*. The Whitehead/Gersten algorithm is a W-algorithm for the $\mathsf{Out}(F_n)$-set of finite lists of conjugacy classes of finitely generated subgroups of F_n; see [Gersten 1984, Theorems W&M], and also [Kalajdžievski 1992] and [Bestvina et al. 2023].

This can be applied directly to our problem by finding subgroups associated to elements of UPG. For example, there is a free factor system $\mathcal{F}_0(\phi)$ characterized by the fact that a conjugacy class in F_n is carried by $\mathcal{F}_0(\phi)$ if and only if it grows linearly under iteration by ϕ. Since a free factor system is an unordered list of conjugacy classes of free factors, we can check if there exists $\theta \in \mathsf{Out}(F_n)$ such that $\mathcal{F}_0(\psi) = \theta(\mathcal{F}_0(\phi))$. If no such θ exists then ϕ and ψ are not conjugate. If there is such a θ then after replacing ψ by $\psi^{\theta^{-1}}$, we may assume, as far as the conjugacy problem is concerned, that $\mathcal{F}_0(\phi) = \mathcal{F}_0(\psi)$. Moreover, any conjugator will preserve $\mathcal{F}_0(\phi) = \mathcal{F}_0(\psi)$.

In Sections 10–15 we show that the Whitehead/Gersten algorithm can be used as the platform on which to build other useful $\mathsf{Out}(F_n)$-sets that satisfy property W. The $\mathsf{Out}(F_n)$-set of finite lists of finitely generated subgroups of F_n also satisfies property M (for McCool). Namely, it is equipped with an algorithm that takes as input $x \in X$ and outputs a finite presentation for $G_x := \{\theta \in G \mid \theta(x) = x\}$. Although it is not strictly necessary for solving the conjugacy problem, property M is important in its own right and we show that all of the $\mathsf{Out}(F_n)$-sets constructed in Sections 10–15 satisfy property M.

- Lemma 4.21, an adaptation of the recognition theorem [Feighn and Handel 2011, Theorem 5.3], gives necessary and sufficient conditions for $\theta \in \mathsf{Out}(F_n)$ to conjugate $\phi \in \mathsf{UPG}$ to $\psi \in \mathsf{UPG}$. The nonnumerical condition is that $\theta(\mathcal{L}(\phi)) = \mathcal{L}(\psi)$, where $\mathcal{L}(\phi)$ is a certain set of lines associated to ϕ and similarly for $\mathcal{L}(\psi)$. If $f : G \to G$ is a CT representing ϕ then $\mathcal{L}(\phi)$ is the set of lines carried by a finite type Stallings graph $\Gamma(f)$ called the *eigengraph* for f. $\Gamma(f)$ depends on f but the set of lines carried by $\Gamma(f)$ depends only on ϕ. The numerical condition of Lemma 4.21 concerns the "twist coordinates" associated to the linear parts of ϕ and ψ and is relatively easy to handle; see Lemmas 17.1 and 17.8. Almost all of the paper is concerned with the existence or not of θ satisfying $\theta(\mathcal{L}(\phi)) = \mathcal{L}(\psi)$.

- A CT $f: G \to G$ comes equipped with a filtration $G_{i_0} \subset G_{i_1} \subset \cdots \subset G_{i_t}$, where for $j > 0$, each G_{i_j} is an f-invariant core subgraph which is obtained from $G_{i_{j-1}}$ by adding a single topological arc, possibly divided into two edges. Edges of $G_{i_j} \setminus G_{i_{j-1}}$ are said to have height j. $\Gamma(f)$ has a compact core to which finitely many rays $\{R_E\}$ are added, one for each nonfixed nonlinear edge E of G. Understanding the structure of rays is an important step in understanding $\mathcal{L}(\phi)$. Each R_E has initial edge E and $R_E \setminus E$ is a ray that crosses only edges with height strictly less than that of E. (This is most definitely a UPG phenomenon. If E belongs to an exponentially growing stratum then E occurs infinitely often in R_E.) Thus R_E can be studied inductively, working up through the filtration. This is carried out in Section 5 and Sections 15–17.

Example 3.1 gives an illustrative element of $\mathsf{UPG}(F_n)$ and is further developed as we progress through the text.

Acknowledgements We are indebted to a referee for an exceptionally thorough and useful report.

Feighn was supported by the National Science Foundation under grant DMS-1406167 and also under grant DMS-14401040 while in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2016 semester. Handel was supported by National Science Foundation grant DMS-1308710 and by PSC-CUNY grants in Program Years 47 and 49.

2 The algorithm

The logical structure of our proof of Theorem 1.1 is a series of reductions

$$\text{Theorem 1.1} \Longleftarrow \text{Proposition 14.7} \Longleftarrow \text{Proposition 16.4} \Longleftarrow \text{Proposition 17.2},$$

and a proof of Proposition 17.2. The above theorem and propositions produce algorithms that we denote by $\mathrm{ALG}_{1.1}$, $\mathrm{ALG}_{14.7}$, $\mathrm{ALG}_{16.4}$ and $\mathrm{ALG}_{17.2}$, respectively. The proof of Theorem 1.1 $\Longleftarrow$ Proposition 14.7 shows how to use $\mathrm{ALG}_{14.7}$ to construct $\mathrm{ALG}_{1.1}$, and similarly for the other implications. Thus $\mathrm{ALG}_{1.1}$ calls $\mathrm{ALG}_{14.7}$, which calls $\mathrm{ALG}_{16.4}$, which calls $\mathrm{ALG}_{17.2}$.

2.1 Theorem 1.1 $\Longleftarrow$ Proposition 14.7

One way to make progress on the conjugacy problem for UPG is to find *W-invariants* for UPG; ie find $\mathsf{Out}(F_n)$-equivariant maps $J: \mathsf{UPG} \to X$, where X is an $\mathsf{Out}(F_n)$-set with a W-algorithm $W_X(\cdot,\cdot)$. If $W_X(J(\phi), J(\psi)) = \mathtt{NO}$, then ϕ is not conjugate to ψ in $\mathsf{Out}(F_n)$. If $W_X(J(\phi), J(\psi)) = (\mathtt{YES}, \xi)$, then $J(\psi^{\xi^{-1}}) = J(\phi)$. Replacing ψ by $\psi^{\xi^{-1}}$, we may assume that $J(\phi) = J(\psi)$. In this case, any θ conjugating ϕ to ψ is contained in the subgroup of $\mathsf{Out}(F_n)$ that fixes $J(\phi)$.

In Sections 5–14 we construct seven such W-invariants and bundle them into a single invariant $\mathsf{I}_{\mathfrak{c}}(\phi)$. Once this is done, it is easy to use an algorithm satisfying the conclusions of Proposition 14.7 to produce an algorithm satisfying the conclusions of Theorem 1.1. The details are given in the proof of Lemma 14.8.

Items (1)–(4) below outline how our ultimate W-invariant $\mathsf{I}_{\mathfrak{c}}\colon \mathsf{UPG} \to \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ is chosen. Item (5) refers to shrinking the set of potential conjugators from the stabilizer of $\mathsf{I}_{\mathfrak{c}}(\phi)$ to one of its finite-index subgroups $\mathcal{X}_{\mathfrak{c}}(\phi)$.

(1) **Dynamical invariants of $\phi \in \mathsf{UPG}$**

- The finite multiset $\mathsf{Fix}(\phi)$ of conjugacy classes of *fixed subgroups of* ϕ; see Definition 3.14.
- The *linear free factor system* $\mathcal{F}_0(\phi)$; see Definition 6.5.
- The finite set $\{\mathfrak{c}\}$ of *special ϕ-chains*; see Section 6.1 and in particular Notation 6.8.
- The finite set $\mathsf{A}_{\mathrm{or}}(\phi)$ of *axes* for ϕ; see Section 4.2.
- The finite set $\mathsf{SA}(\phi)$ of *strong axes* for ϕ; see Section 4.2.
- The finite set $\Omega_{\mathsf{NP}}(\phi)$ of all nonperiodic *limit lines* for all eigenrays of ϕ; see Section 5.
- For each one-edge extension $\mathfrak{e}$ of each $\mathfrak{c}$, the set $\mathsf{L}_{\mathfrak{e}}(\phi)$ of *added lines with respect to* $\mathfrak{e}$; see Definition 6.14.

The invariants in the first four items are *algebraic* in that they take values in $\mathsf{Out}(F_n)$-sets that can be expressed in terms of conjugacy classes of finitely generated subgroups of F_n or more generally are *iterated sets* (Section 10.1). In particular, they take values in $\mathsf{Out}(F_n)$-sets with W-algorithms and so can be used as they are. The others must be modified.

(2) **Algebraic versions of dynamical invariants** For the last three dynamical invariants, define corresponding (but weaker) algebraic invariants. The last two depend on a choice of special chain $\mathfrak{c}$; see Section 13.1.

- The finite set of *algebraic strong axes*; see Section 13.6.
- The finite set $\{\mathsf{H}_{\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}$ of *algebraic limit lines*; see Section 13.8.
- For each one-edge extension $\mathfrak{e}$ in $\mathfrak{c}$, the finite set $\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)$ of *algebraic added lines with respect to* $\mathfrak{e}$; see Section 13.7.

Remark 2.1 If the seven dynamical invariants in (1) take the same values on ϕ and ψ then, using Lemma 4.21, it is easy to check if ϕ and ψ are conjugate. The same is not true for the seven algebraic invariants in (1) and (2). Too much information was lost in translation.

(3) **W-invariants** Iterated sets, and in particular $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$, are defined in Sections 10 and 11. By construction, all of our algebraic invariants take values in the iterated set $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. We construct a W-algorithm for $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ (and all other iterated sets).

(4) **The total invariant $\mathsf{I}_{\mathfrak{c}}(\phi)$** This is defined by combining the algebraic invariants in (1) and (2) into a single algebraic invariant that takes values in $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$; see Definition 13.13.

(5) **Reduce potential conjugators** Elements of $\mathcal{X}_{\mathfrak{c}}(\phi) < \mathsf{Out}(F_n)$ not only stabilize the algebraic invariants in (1) and (2) making up $\mathsf{I}_{\mathfrak{c}}(\phi)$, they also induce trivial permutations on those invariants that are finite sets; see Definition 14.1.

As mentioned above Lemma 14.8 is proved by constructing $\mathrm{ALG}_{1.1}$ using $\mathrm{ALG}_{14.7}$ and properties of $\mathsf{I}_{\mathfrak{c}}(\phi)$. Hence to prove Theorem 1.1, we are reduced to proving:

Proposition 14.7 *There is an algorithm that takes as input $\phi, \psi \in \mathsf{UPG}(F_n)$ and a chain $\mathfrak{c}$ such that*

- *$\mathfrak{c}$ is special for both ϕ and ψ, and*
- *$\mathsf{I}_{\mathfrak{c}}(\phi) = \mathsf{I}_{\mathfrak{c}}(\psi)$,*

and that outputs `YES` *or* `NO` *depending whether or not there is $\theta \in \mathcal{X}_{\mathfrak{c}}(\phi)$ conjugating ϕ to ψ. Further, if* `YES` *then such a θ is produced.*

2.2 Proposition 14.7 $\Longleftarrow$ Proposition 16.4

$\mathrm{ALG}_{14.7}$ and $\mathrm{ALG}_{16.4}$ differ only in the subgroup of potential conjugators that must be considered. In Proposition 14.7 it is $\mathcal{X}_{\mathfrak{c}}(\phi)$ and in Proposition 16.4 it is an infinite-index subgroup $\mathsf{Ker}(\bar{Q}^{\phi}) < \mathcal{X}_{\mathfrak{c}}(\phi)$ defined in Definition 16.3. See statement of Proposition 14.7 below.

The set of (eigen)rays $\mathcal{R}(\phi)$ (Definition 3.14) is a fundamental dynamical invariant of ϕ. Each $r \in \mathcal{R}(\phi)$ is the conjugacy class $[\tilde{r}]$ of a point $\tilde{r} \in \partial F_n$. There is no W-algorithm for ∂F_n so we work with a weaker algebraic invariant, the conjugacy class $F_{\mathfrak{c}}(r)$ of a free factor determined by r and a special chain $\mathfrak{c}$; see Section 13.4. We do not list this in (2) because it is built into the set of algebraic lines and the set of algebraic added lines. The great advantage of $\mathsf{Ker}(\bar{Q}^{\phi})$ over $\mathcal{X}_{\mathfrak{c}}(\phi)$ is that in the proof of Proposition 17.2 we need only consider conjugating elements that preserve r. (See Lemmas 15.45 and 17.9.) Instead of having to check if two rays are conjugate, we need only check if they are equal.

The definition of $\bar{Q}^{\phi}(\xi)$ for $\xi \in \mathcal{X}_{\mathfrak{c}}$ is given in Definition 16.3. The key result, from the algorithmic point of view, is:

Proposition 16.6 *There is an algorithm that produces a finite set $\{\eta_i\} \subset \mathcal{X}$ so that the union of the cosets of $\mathsf{Ker}(\bar{Q}^{\phi})$ determined by the η_i contains each $\theta \in \mathcal{X}$ that conjugates ϕ to ψ.*

The proof of Proposition 16.6 requires a detailed understanding of the structure of eigenrays and is the most technical part of the paper. The proof of Lemma 16.5 shows how to quickly construct $\mathrm{ALG}_{14.7}$ using $\mathrm{ALG}_{16.4}$ and the coset representatives produced by the algorithm of Proposition 16.6. In other words, to prove Proposition 14.7, we are reduced to proving:

Proposition 16.4 *There is an algorithm that takes as input $\phi, \psi \in \mathsf{UPG}(F_n)$ and a chain $\mathfrak{c}$ such that*

- *$\mathfrak{c}$ is a special chain for ϕ and ψ, and*
- *$\mathsf{I}_{\mathfrak{c}}(\phi) = \mathsf{I}_{\mathfrak{c}}(\psi)$,*

and that outputs `YES` *or* `NO` *depending on whether or not there is $\theta \in \mathsf{Ker}(\bar{Q}^{\phi})$ conjugating ϕ to ψ. Further, if* `YES` *then such a θ is produced.*

2.3 Proposition 16.4 ⟸ Proposition 17.2

This is an easy step. The details are given in the proof of Proposition 16.4 (assuming Lemma 17.1 and Proposition 17.2) following the statement of Proposition 17.2. After this step, we may assume that the restrictions of ϕ and ψ to the linear free factor system $\mathcal{F}_0(\phi) = \mathcal{F}_0(\psi)$ are equal. This provides the basis for an inductive argument completed in the next step.

2.4 Proof of Proposition 17.2

Proposition 17.2 is the inductive step of an argument up the filtration induced by $\mathfrak{c}$. There are six items labeled (1)–(5), (7) that are sequentially checked. If any of these is false then return `NO`. Otherwise, construct the desired conjugator following pages 1809–1811.

3 Background

3.1 Standard notation

The free group on n generators is denoted by F_n. For $a \in F_n$, conjugation by a is denoted by i_a, ie $i_a(x) = axa^{-1}$ for $x \in F_n$. The group of automorphisms of F_n, the group of inner automorphisms of F_n and the group of outer automorphisms of F_n are denoted by $\mathsf{Aut}(F_n)$, $\mathsf{Inn}(F_n) := \{i_a \mid a \in F_n\}$ and $\mathsf{Out}(F_n) = \mathsf{Aut}(F_n)/\mathsf{Inn}(F_n)$, respectively.

For subgroups $H < F_n$, $[H]$ denotes the conjugacy class of H and, for elements $a \in F_n$, $[a]$ denotes the conjugacy class of a.

An outer automorphism $\phi \in \mathsf{Out}(F_n)$ has *polynomial growth*, written $\phi \in \mathsf{PG}$, if for each $a \in F_n$ there is a polynomial P such that reduced word length of $\phi^k([a])$ with respect to a fixed set of generators of F_n is bounded above by $P(k)$. Equivalently, the set of attracting laminations for ϕ [Bestvina et al. 2000, Section 3] is empty. The set $\mathsf{UPG}(F_n)$ of *unipotent outer automorphisms* is the subset of $\mathsf{Out}(F_n)$ consisting of polynomially growing ϕ whose induced action on $H_1(F_n, \mathbb{Z})$ is unipotent. We sometimes write $\phi \in \mathsf{UPG}$ instead of $\phi \in \mathsf{UPG}(F_n)$. In Section 3.5 we show that $\phi \in \mathsf{UPG}$ if and only if $\phi \in \mathsf{PG}$ and ϕ is rotationless in the sense of [Feighn and Handel 2011, Definition 3.13] (where it is called forward rotationless). There is $K_n > 0$ such that if $\phi \in \mathsf{PG}$ then $\phi^{K_n} \in \mathsf{UPG}$ [Feighn and Handel 2018, Corollary 3.14].

The graph with one vertex $*$ and with n edges is the *rose* R_n. Making use of the standard identification of $\pi_1(R_n, *)$ with F_n, there are bijections between $\mathsf{Aut}(F_n)$ and the group of pointed homotopy classes of homotopy equivalences $f\colon (R_n, *) \to (R_n, *)$ and between $\mathsf{Out}(F_n)$ and the group of free homotopy classes of homotopy equivalences $f\colon R_n \to R_n$.

If G is a graph without valence one vertices then a homotopy equivalence $\mu\colon R_n \to G$ is called a *marking* and G, equipped with a marking, is called a *marked graph*. A marking μ induces an identification,

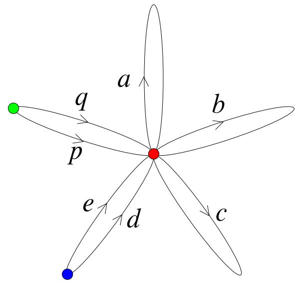


Figure 1: The graph G of Example 3.1.

well-defined up to inner automorphism, of the fundamental group of G with the fundamental group of R_n and hence with F_n. This in turn induces an identification of the group of homotopy classes of homotopy equivalences $f: G \to G$ with $\mathsf{Out}(F_n)$. If $\phi \in \mathsf{Out}(F_n)$ corresponds to the homotopy class of $f: G \to G$ then we say that $f: G \to G$ *represents* ϕ. In Section 3.6 we recall the existence of very well-behaved homotopy equivalences $f: G \to G$ representing an element of $\mathsf{UPG}(F_n)$.

Example 3.1 Here is an example of a homotopy equivalence $f: G \to G$ of a marked graph that represents an element ϕ of $\mathsf{UPG}(F_5)$. Let F_5 be represented as the fundamental group of the rose R_5, let G be the subdivision of R_5 pictured in Figure 1, and let $f: G \to G$ be given by $a \mapsto a$, $b \mapsto ba$, $c \mapsto cb$, $d \mapsto db^2$, $e \mapsto eb^3$, $p \mapsto pa^2$, $q \mapsto qc$. To see that ϕ has polynomial growth, note that the edge q has cubic growth in that

$$|f^k(q)| = |q \cdot c \cdot f(c) \cdot f^2(c) \cdot f^3(c) \cdot \ldots \cdot f^{k-1}(c)| = |q \cdot c \cdot cb \cdot cbba \cdot \ldots \cdot cbba \ldots ba^{k-1}| = \frac{k^3 + 5k + 6}{6}$$

and that no edge grows at a higher rate. In particular conjugacy classes of F_n have at most cubic growth. As we progress through this paper, we will expand upon this example.

3.2 Paths, circuits and lines

A *path* in a marked graph G is a proper immersion of a closed interval into G. In this paper, we will assume that the endpoints of a path, if any, are at vertices. If the interval is degenerate then the path is *trivial*; if the interval is infinite or bi-infinite then the path is a *ray* or a *line*, respectively. We do not distinguish between paths that differ only by a reparametrization of the domain interval. Thus, every nontrivial path has a description as a concatenation of oriented edges and we will use this *edge path* formulation without further mention. Reversing the orientation on a path σ produces a path denoted by either $\overline{\sigma}$ or σ^{-1}. A *circuit* is an immersion of S^1 into G. Unless otherwise stated, a circuit is assumed to have an orientation. Circuits have cyclic edge decompositions. Each conjugacy class in F_n is represented by a unique circuit in G. The conjugacy class in F_n represented by the circuit σ is denoted by $[\sigma]$.

Notation 3.2 Each $\Phi \in \mathsf{Aut}(F_n)$ induces an equivariant homeomorphism of ∂F_n. To simplify notation somewhat, we refer to this extension as Φ rather than, say, $\partial\Phi$. In situations where this might cause confusion, we write $\Phi|\partial F_n$ for the induced homeomorphism of ∂F_n. For example, $\mathsf{Fix}(\Phi)$ is the subgroup of F_n fixed by Φ, and $\mathsf{Fix}(\Phi|\partial F_n)$ is the set of points in ∂F_n fixed by the induced homeomorphism.

The action of F_n on ∂F_n is by conjugation, ie by $a \cdot P = i_a(P)$ for $P \in \partial F_n$. For each nontrivial $a \in F_n$, i_a fixes two points in ∂F_n: a repeller a^- and an attractor a^+.

A marking μ induces an identification, well-defined up to inner automorphism, of the set of ends of $\widetilde{G}$ with ∂F_n, and likewise the group of covering translations of $\widetilde{G}$ with $\mathsf{Inn}(F_n)$. We choose such an identification once and for all. The covering translation corresponding to i_a is denoted by T_a as is the extension of T_a to a homeomorphism of ∂F_n. We have $T_a|\partial F_n = i_a|F_n$. If $f : G \to G$ represents ϕ then each lift $\tilde{f} : \widetilde{G} \to \widetilde{G}$ induces an equivariant homeomorphism, still called $\tilde{f}$, of ∂F_n; see, for example, Section 2.3 of [Feighn and Handel 2011]. There is a bijection between the set of lifts $\tilde{f}$ of $f : G \to G$ and the set of automorphisms Φ representing ϕ defined by $\tilde{f} \leftrightarrow \Phi$ if $\tilde{f}|\partial F_n = \Phi|\partial F_n$.

A line $\widetilde{L}$ in the universal cover $\widetilde{G}$ of a marked graph G is a bi-infinite edge path. The ends of $\widetilde{L}$ determine ends of $\widetilde{G}$ and hence points in ∂F_n. In this way, the *space of oriented lines in the tree* $\widetilde{G}$ can be identified with the space $\widetilde{\mathcal{B}}$ of ordered pairs of distinct elements of ∂F_n. The *space of oriented lines in* G is then identified with the space $\mathcal{B}$ of F_n-orbits of elements of $\widetilde{\mathcal{B}}$. The topology on ∂F_n induces a topology on $\widetilde{\mathcal{B}}$ and hence a topology on $\mathcal{B}$ called the *weak topology*.

3.3 Free factor systems

The subgroup system $\mathcal{F} = \{[A_1], \dots, [A_m]\}$ is a *free factor system* if $A_1, \dots, A_m$ are nontrivial free factors of F_n and either $F_n = A_1 * \cdots * A_m$ or $F_n = A_1 * \cdots * A_m * B$ for some nontrivial free factor B. The $[A_i]$ are the *components* of $\mathcal{F}$. If G is a marked graph and K is a subgraph whose noncontractible components are $K_1, \dots, K_m$ then $\mathcal{F}(K, G) := \{[\pi_1(K_1)], \dots, [\pi_1(K_m)]\}$ is a free factor system that is *realized by* $K \subset G$. Every free factor system $\mathcal{A}$ can be realized by $K \subset G$ for some marked graph G and some core subgraph $K \subset G$. Recall that a graph is *core* if through every edge there is an immersed circuit and that the *core of a graph* is the union of the images of its immersed circuits.

We write $\mathcal{F}_1 \sqsubset \mathcal{F}_2$ and say that $\mathcal{F}_1$ *is contained in* $\mathcal{F}_2$ if for each component $[A_i]$ of $\mathcal{F}_1$ there is a component $[B_j]$ of $\mathcal{F}_2$ so that A_i is conjugate to a subgroup of B_j. Equivalently, there is a marked graph G with core subgraphs $K_1 \subset K_2$ such that $\mathcal{F}_1 = \mathcal{F}(K_1, G)$ and $\mathcal{F}_2 = \mathcal{F}(K_2, G)$. If one can choose K_1 and K_2 so that $K_2 \setminus K_1$ is a single edge then we say that $\mathcal{F}_1 \sqsubset \mathcal{F}_2$ is a *one-edge extension*. For example, $\{[A]\} \sqsubset \{[B]\}$ is a one-edge extension if and only if $\mathrm{rank}(B) = \mathrm{rank}(A) + 1$ and $\{[A_1], [A_2]\} \sqsubset \{[B]\}$ is a one-edge extension if and only if $\mathrm{rank}(B) = \mathrm{rank}(A_1) + \mathrm{rank}(A_2)$.

Example 3.3 Suppose that H_1 is a subgraph of a marked graph G, that $H_2 = H_1 \cup E_2 \subset G$, where E_2 is an edge that forms a loop that is disjoint from H_1, and that $H_3 = H_2 \cup E_3$, where $E_3 \subset G$ is an edge with one endpoint in H_1 and the other at the unique endpoint of E_2. Then $\mathcal{F}(H_1, G) \sqsubset \mathcal{F}(H_2, G)$ and $\mathcal{F}(H_2, G) \sqsubset \mathcal{F}(H_3, G)$ are proper inclusions and $\mathcal{F}(H_1, G) \sqsubset \mathcal{F}(H_3, G)$ is a one-edge extension.

This is essentially the only way in which a one-edge extension can be "reducible". We record a specific consequence of this in the following lemma.

Lemma 3.4 *Suppose that $\mathcal{F}(H_1, G) \sqsubset \mathcal{F}(H_2, G)$ and $\mathcal{F}(H_2, G) \sqsubset \mathcal{F}(H_3, G)$ are proper inclusions and that $\mathcal{F}(H_1, G)$ and $\mathcal{F}(H_2, G)$ have the same number of components. Then $\mathcal{F}(H_1, G) \sqsubset \mathcal{F}(H_3, G)$ is not a one-edge extension.*

Proof This follows from [Handel and Mosher 2020, Part 2, Definition 2.4 and Lemma 2.5]. □

If $\mathcal{F} = \{[A_1], \dots, [A_m]\}$ and $a \in F_n$ is conjugate into some A_i then $[a]$ is *carried* by $\mathcal{F}$. A line $L \in \mathcal{B}$ is *carried* by $\mathcal{F}$ if it is a limit of periodic lines corresponding to conjugacy classes that are carried by $\mathcal{F}$. Equivalently, $[a]$ or L is *carried* by $\mathcal{F}$ if for some, and hence every, $K \subset G$ realizing $\mathcal{F}$, the realization of $[a]$ or L in G is contained in K. For every collection of conjugacy classes and lines there is a unique minimal (with respect to $\sqsubset$) free factor system that carries each element of the collection [Bestvina et al. 2000, Corollary 2.6.5].

Notation 3.5 $\mathsf{Out}(F_n)$ acts on the set of conjugacy classes $[F]$ of free factors F. If $\phi \in \mathsf{Out}(F_n)$ fixes $[F]$ then we say that $[F]$ *is ϕ-invariant* and write $\phi|[F]$ for the *restriction of ϕ to* $[F]$ (which is well-defined because F is its own normalizer in F_n). We often say that F is ϕ-invariant and write $\phi|F$ just to simplify notation. [Bestvina et al. 2005, Proposition 4.44] implies that if ϕ is UPG then $\phi|F$ is UPG. If $\mathcal{F} = \{[A_1], \dots, [A_m]\}$ is a free factor system and each $[A_i]$ is ϕ-invariant then we say that $\mathcal{F}$ is ϕ-invariant and denote $\{\phi|A_1, \dots, \phi|A_m\}$ by $\phi|\mathcal{F}$.

3.4 $\mathsf{Fix_N}(\phi)$, principal lifts and $\mathcal{R}(\phi)$

We continue with Notation 3.2. If $P \in \mathsf{Fix}(\Phi|\partial F_n)$ and if there is a neighborhood U of P in ∂F_n such that $\Phi(U) \subset U$ and such that $\bigcap_{i=1}^{\infty} \Phi^i(U) = P$ then P is *attracting*. If P is an attracting fixed point for $\Phi^{-1}|\partial F_n$ then it is a *repelling* fixed point for $\Phi|\partial F_n$. By $\mathsf{Fix}_+(\Phi)$, $\mathsf{Fix}_-(\Phi)$ and $\mathsf{Fix_N}(\Phi)$ we denote the *set of attracting fixed points* for $\Phi|\partial F_n$, the *set of repelling fixed points* for $\Phi|\partial F_n$ and the *set of nonrepelling fixed points* for $\Phi|\partial F_n$, respectively; thus $\mathsf{Fix_N}(\Phi) = \mathsf{Fix}(\Phi|\partial F_n) \setminus \mathsf{Fix}_-(\Phi)$. Note that all of these sets are contained in ∂F_n.

If $A < F_n$ is a finitely generated subgroup then the inclusion of A into F_n is a quasi-isometric embedding and so extends to an inclusion of ∂A into ∂F_n with the property that $\{a^{\pm} \mid \text{nontrivial } a \in A\}$ is dense in ∂A. In particular, since the subgroup $\mathsf{Fix}(\Phi)$ consisting of elements in F_n that are fixed by Φ is finitely generated [Gersten 1987] (see also [Bestvina and Handel 1992]), we have $\partial\,\mathsf{Fix}(\Phi) \subset \partial F_n$. The following lemma implies that $\partial\,\mathsf{Fix}(\Phi) \subset \mathsf{Fix}(\Phi|\partial F_n)$, and that $\mathsf{Fix}_+(\Phi)$, $\mathsf{Fix}_-(\Phi)$ and $\mathsf{Fix_N}(\Phi)$ are $\mathsf{Fix}(\Phi)$-invariant.

Lemma 3.6 *Let $\Phi \in \mathsf{Aut}(F_n)$ and $0 \neq a \in F_n$. The following are equivalent*:

- $a \in \mathsf{Fix}(\Phi)$.
- *Either a^- or a^+ is contained in $\partial\,\mathsf{Fix}(\Phi)$.*
- *Both a^- and a^+ are contained in $\partial\,\mathsf{Fix}(\Phi)$.*

- *The automorphism* i_a *commutes with* Φ.
- *The automorphism* $i_a|\partial F_n$ *commutes with* $\Phi|\partial F_n$.

Proof This is well known; see eg Lemmas 2.3 and 2.4 of [Bestvina et al. 2004] and Proposition I.1 of [Gaboriau et al. 1998]. □

Lemma 3.7 *If* $P \in \partial F_n$ *is fixed by automorphisms* $\Phi \neq \Phi'$ *representing* $\phi \in \mathrm{Out}(F_n)$ *then* $P = a^{\pm}$ *for some nontrivial* $a \in F_n$.

Proof There exists a nontrivial $a \in F_n$ such that $i_a = \Phi^{-1}\Phi'$ fixes P. □

Definition 3.8 An automorphism Φ representing $\phi \in \mathrm{UPG}(F_n)$ is *principal* if $\mathrm{Fix}_{\mathrm{N}}(\Phi)$ contains at least two points and if $\mathrm{Fix}_{\mathrm{N}}(\Phi) \neq \{a^-, a^+\}$ for any nontrivial $a \in F_n$. The *set of principal automorphisms representing* ϕ is denoted by $\mathcal{P}(\phi)$. See Section 3.2 of [Feighn and Handel 2011] for complete details.

Lemma 3.9 *If* Φ *is principal then* $\mathrm{Fix}_{\mathrm{N}}(\Phi)$ *is the disjoint union of* $\partial\,\mathrm{Fix}(\Phi)$ *and* $\mathrm{Fix}_+(\Phi)$. *Moreover,* $\mathrm{Fix}_+(\Phi)$ *is a union of finitely many* $\mathrm{Fix}(\Phi)$ *orbits.*

Proof The first assertion follows from Proposition I.1 of [Gaboriau et al. 1998]. The second is obvious if $\mathrm{Fix}_+(\Phi)$ is finite and follows from Lemma 2.5 of [Bestvina et al. 2004] if $\mathrm{Fix}_+(\Phi)$ is infinite. □

Remark 3.10 We sometimes say that $P \in \partial F_n$ is periodic if it is fixed by i_a for some nontrivial $a \in \mathcal{F}_n$. Nonperiodic points are dense in $\mathrm{Fix}_{\mathrm{N}}(\Phi)$ for each $\Phi \in \mathcal{P}(\phi)$. Lemmas 3.6 and 3.9 imply that no element of $\mathrm{Fix}_+(\Phi)$ is periodic. If $\mathrm{Fix}_+(\Phi) \neq \varnothing$ then $\mathrm{Fix}_+(\phi)$ is dense in $\mathrm{Fix}_{\mathrm{N}}(\phi)$ and we are done. Otherwise, $\mathrm{Fix}(\Phi)$ has rank at least two and $\mathrm{Fix}_{\mathrm{N}}(\Phi) = \partial\,\mathrm{Fix}(\Phi)$.

Definition 3.11 Two automorphisms Φ_1 and Φ_2 are in the same *isogredience class* if there exists $a \in F_n$ such that $\Phi_2 = i_a \Phi_1 i_a^{-1}$, in which case $\mathrm{Fix}_{\mathrm{N}}(\Phi_2) = i_a\,\mathrm{Fix}_{\mathrm{N}}(\Phi_1)$ and similarly for $\mathrm{Fix}_-(\Phi_2)$, $\mathrm{Fix}_+(\Phi_2)$ and $\mathrm{Fix}(\Phi_2)$. It follows that if Φ_1 and Φ_2 are isogredient then $[\mathrm{Fix}(\Phi_1)] = [\mathrm{Fix}(\Phi_2)]$ and $[\mathrm{Fix}_{\mathrm{N}}(\Phi_1)] = [\mathrm{Fix}_{\mathrm{N}}(\Phi_2)]$, where $[\ ]$ denotes the orbit under the action of F_n on sets of points in ∂F_n. It is easy to see that isogredience defines an equivalence relation on $\mathcal{P}(\phi)$. The *set of isogredience classes* of $\mathcal{P}(\phi)$ is denoted by $[\mathcal{P}(\phi)]$.

We recall the following result from Remark 3.9 of [Feighn and Handel 2011]; see also Lemma 3.25 of this paper.

Lemma 3.12 $\mathcal{P}(\phi)$ *is a finite union of isogredience classes.*

Our next lemma states that $[\mathrm{Fix}_{\mathrm{N}}(\Phi)]$ determines the isogredience class of $\Phi \in \mathcal{P}(\phi)$.

Lemma 3.13 *Suppose that* $\Phi_1, \Phi_2 \in \mathcal{P}(\phi)$. *Then* Φ_1 *and* Φ_2 *are isogredient if and only if* $[\mathrm{Fix}_{\mathrm{N}}(\Phi_1)] = [\mathrm{Fix}_{\mathrm{N}}(\Phi_2)]$. *More precisely,* $\Phi_2 = i_a \Phi_1 i_a^{-1}$ *if and only if* $\mathrm{Fix}_{\mathrm{N}}(\Phi_2) = i_a\,\mathrm{Fix}_{\mathrm{N}}(\Phi_1)$.

Proof It is obvious that if $\Phi_2 = i_a \Phi_1 i_a^{-1}$ then $\mathsf{Fix}_{\mathsf{N}}(\Phi_2) = i_a \mathsf{Fix}_{\mathsf{N}}(\Phi_1)$. For the converse note that if $\mathsf{Fix}_{\mathsf{N}}(\Phi_2) = i_a \mathsf{Fix}_{\mathsf{N}}(\Phi_1) = \mathsf{Fix}_{\mathsf{N}}(i_a \Phi_1 i_a^{-1})$ then $\Phi_2^{-1} i_a \Phi_1 i_a^{-1}$ is an inner automorphism whose induced action on ∂F_n fixes $\mathsf{Fix}_{\mathsf{N}}(\Phi_2)$ and so is not equal to $\{a^-, a^+\}$ for any nontrivial a. This proves that $\Phi_2^{-1} i_a \Phi_1 i_a^{-1}$ is trivial and so $\Phi_2 = i_a \Phi_1 i_a^{-1}$. □

Definition 3.14 Define sets

$$\mathsf{Fix}_{\mathsf{N}}(\phi) := \{[\mathsf{Fix}_{\mathsf{N}}(\Phi_1)], \dots, [\mathsf{Fix}_{\mathsf{N}}(\Phi_m)]\} \quad \text{and} \quad \mathcal{R}(\phi) := \left\{ [P] \;\middle|\; P \in \bigcup_{i=1}^{m} \mathsf{Fix}_+(\Phi_i) \right\} \subset \partial F_n / F_n ,$$

and a multiset (repeated elements allowed)

$$\mathsf{Fix}(\phi) := \{[\mathsf{Fix}(\Phi_1)], \dots, [\mathsf{Fix}(\Phi_m)]\},$$

where the Φ_i are representatives of the isogredience classes in $\mathcal{P}(\phi)$. Thus $\mathsf{Fix}_{\mathsf{N}}(\phi)$ is a finite set of F_n-orbits of subsets of ∂F_n, and $\mathcal{R}(\phi)$ is a finite set of F_n-orbits of points in ∂F_n.

Definition 3.15 For us a *natural invariant of a group G* is a map $I : G \to X$, where X is a G-set and, for all $\phi, \theta \in G$, we have $I(\phi^\theta) = \theta(I(\phi))$.

The following lemma says that $[\mathcal{P}(\phi)]$, $\mathsf{Fix}(\phi)$, $\mathsf{Fix}_{\mathsf{N}}(\phi)$ and $\mathcal{R}(\phi)$ are natural invariants of $\mathsf{Out}(F_n)$.

Lemma 3.16 *Suppose that $\Theta \in \mathsf{Aut}(F_n)$ represents $\theta \in \mathsf{Out}(F_n)$ and that $\psi = \theta\phi\theta^{-1}$. Then:*

(1) *The map $\Phi \mapsto \Psi := \Theta\Phi\Theta^{-1}$ defines a bijection between $\mathcal{P}(\phi)$ and $\mathcal{P}(\psi)$, and induces a bijection $[\mathcal{P}(\phi)] \leftrightarrow [\mathcal{P}(\psi)]$.*

(2) $\mathsf{Fix}(\Psi) = \Theta(\mathsf{Fix}(\Phi))$, $\mathsf{Fix}_{\mathsf{N}}(\Psi) = \Theta(\mathsf{Fix}_{\mathsf{N}}(\Phi))$ *and* $\mathsf{Fix}_+(\Psi) = \Theta(\mathsf{Fix}_+(\Phi))$.

(3) $\mathsf{Fix}(\psi) = \theta(\mathsf{Fix}(\phi))$, $\mathsf{Fix}_{\mathsf{N}}(\psi) = \theta(\mathsf{Fix}_{\mathsf{N}}(\phi))$ *and* $\mathcal{R}(\psi) = \theta(\mathcal{R}(\phi))$.

Proof The automorphism Ψ represents $\theta\phi\theta^{-1} = \psi \in \mathsf{Out}(F_n)$. If $\Phi' = i_c \Phi i_c^{-1}$ then $\Psi' := \Theta\Phi'\Theta^{-1} = i_{\Theta(c)}(\Theta\Phi\Theta^{-1}) i_{\Theta(c)}^{-1} = i_{\Theta(c)} \Psi i_{\Theta(c)}^{-1}$ so conjugation by Θ maps isogredience classes of ϕ to isogredience classes of ψ. The items in (2) are easy standard facts about conjugation. Since $\Theta(a^\pm) = (\Theta(a))^\pm$, it follows that Ψ is principal if Φ is principal. The induced map $\mathcal{P}(\phi) \to \mathcal{P}(\psi)$ is obviously invertible and is hence a bijection. This completes the proof of (1). If Θ is replaced by $i_a\Theta$ then Ψ is replaced by $i_a \Psi i_a^{-1}$ and $\mathsf{Fix}(\Psi)$, $\mathsf{Fix}_{\mathsf{N}}(\Psi)$ and $\mathsf{Fix}_+(\Psi)$ are replaced by $i_a(\mathsf{Fix}(\Psi))$, $i_a(\mathsf{Fix}_{\mathsf{N}}(\Psi))$ and $i_a(\mathsf{Fix}_+(\Psi))$ respectively. Thus $\theta([\mathsf{Fix}(\Phi)]) = [\mathsf{Fix}(\Psi)]$, $\theta([\mathsf{Fix}_{\mathsf{N}}(\Phi)]) = [\mathsf{Fix}_{\mathsf{N}}(\Psi)]$ and $\theta([\mathsf{Fix}_+(\Phi)]) = [\mathsf{Fix}_+(\Psi)]$. This verifies (3). □

The following lemma is used implicitly throughout the paper.

Lemma 3.17 *If A is a ϕ-invariant free factor then the inclusion of ∂A into ∂F_n induces an inclusion of $\mathcal{R}(\phi|A)$ into $\mathcal{R}(\phi)$.*

Proof An automorphism $\Phi'\colon A \to A$ representing $\phi|A$ extends to an automorphism $\Phi\colon F_n \to F_n$ representing ϕ. We claim that if $P \in \partial A$ then $P \in \mathsf{Fix}_+(\Phi')$ if and only if $P \in \mathsf{Fix}_+(\Phi)$. Symmetrically, $P \in \mathsf{Fix}_-(\Phi')$ if and only if $P \in \mathsf{Fix}_-(\Phi)$. It follows that $\mathsf{Fix}_{\mathsf{N}}(\Phi') \subset \mathsf{Fix}_{\mathsf{N}}(\Phi)$ and hence that Φ is principal if Φ' is principal. This will complete the proof of the lemma.

To prove the claim, extend a basis $\mathcal{A}$ for A to a basis $\mathcal{B}$ for F_n. Following [Gaboriau et al. 1998], we view $P \in \partial F_n$ as an infinite word $P = x_1x_2x_3\cdots$ with each $x_i \in \mathcal{B}$. For each $i \in \mathbb{N}$, let $x_1,\dots,x_{k(i)}$ be the common initial segment of $\Phi(x_1\dots x_i)$ and P. Then $P \in \mathsf{Fix}_+(\Phi)$ if and only if $k(i)-i \to \infty$ [Gaboriau et al. 1998, Proposition I.1]. If $P \in \partial A$ then each $x_i \in \mathcal{A}$ and each $\Phi(x_1\dots x_i) = \Phi'(x_1\dots x_i) \in A$ so $k(i)$ is the same whether we compute using Φ or Φ'. □

3.5 UPG is rotationless

Relative train track theory is most effective when applied to elements of $\mathsf{Out}(F_n)$ that are rotationless as defined in [Feighn and Handel 2011, Definition 3.13 and Remark 3.14]. In this section, we show that for PG elements, ϕ is rotationless if and only if ϕ is UPG. The exact definition of rotationless plays no role in this paper so is not repeated here.

Lemma 3.18 *Each* $\phi \in \mathsf{UPG}$ *is rotationless.*

Proof By [Bestvina et al. 2000, Proposition 5.7.5], there is a sequence $\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_K$ of ϕ-invariant one-edge extensions where $\mathcal{F}_0$ is trivial and $\mathcal{F}_K = \{[F_n]\}$. We may assume without loss of generality that $\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_K$ is a maximal such chain.

[Feighn and Handel 2011, Theorem 2.19], which makes no assumptions on ϕ, proves the existence of a relative train track map $f\colon G \to G$ and filtration $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_N = G$ representing ϕ and satisfying a certain list of five properties, two of which are denoted by (P) and (NEG). Additionally, the filtration realizes $\mathcal{F}_1 \sqsubset \mathcal{F}_2 \sqsubset \cdots \sqsubset \mathcal{F}_K$ in the sense that each $\mathcal{F}_k$ is represented by a core filtration element G_{i_k}. Since $\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_K$ is maximal, G_{i_k} is obtained from $G_{i_{k-1}}$ by adding either a topological circle that is disjoint from $G_{i_{k-1}}$ or a topological arc with both endpoints in $G_{i_{k-1}}$ [Handel and Mosher 2020, Part II, Lemma 2.5]. We denote the closure of $G_{i_k} \setminus G_{i_{k-1}}$, equipped with the simplicial structure inherited from G_{i_k}, by $\widehat{H}_k$. Since ϕ is PG, there are no EG strata.

We use the following consequences of properties (P) and (NEG).

(1) The terminal endpoint of a nonperiodic edge in $\widehat{H}_k$ is contained in $G_{i_{k-1}}$.

(2) If $\widehat{H}_k$ contains a periodic edge then it is a single periodic stratum [Feighn and Handel 2011, Lemma 2.20(1)].

If $\widehat{H}_k$ is a circle that is disjoint from $G_{i_{k-1}}$, then its conjugacy class is fixed by some iterate of ϕ and so is fixed by ϕ [Bestvina et al. 2005, Proposition 3.16]. There are two possibilities; $\widehat{H}_k$ is a single fixed edge; or $\widehat{H}_k$ has more than one edge and $f|\widehat{H}_k$ is a nontrivial rotation with one orbit of edges. In the latter case, we say that $\widehat{H}_k$ is a *rotating circle*.

If $\widehat{H}_k$ intersects $G_{i_{k-1}}$ then it is a topological arc E_k whose ends may or may not be identified. Either $f(E_k) = v_k E_k u_k$ or $f(E_k) = v_k \bar{E}_k u_k$ for some paths $u_k, v_k \subset G_{i_{k-1}}$ [Bestvina et al. 2000, Corollary 3.2.2]. Since ϕ is UPG, the latter is ruled out by [Bestvina et al. 2005, Proposition 5.7.5(2); see the second paragraph on page 595]. If both u_k and v_k are trivial then E_k is a single fixed edge. If exactly one of u_k and v_k is trivial then E_k is a single nonperiodic NEG edge. If neither u_k and v_k are trivial then E_k consists of two nonperiodic NEG edges with a common fixed initial endpoint. In all three cases, the directions determined by E_k and $\bar{E}_k$ are either nonperiodic or fixed.

An easy induction argument on k shows that:

(a) If a vertex v is not contained in a rotating circle then v is fixed by f and each periodic direction based at v is fixed by f.

(b) Each rotating circle is a component of $\mathrm{Per}(f)$, the set of periodic points for f, and each point in a rotating circle has exactly two periodic directions.

These are exactly the conditions needed to verify that $f : G \to G$ is rotationless in the sense of [Feighn and Handel 2011, Definition 3.18]. Proposition 3.19 of [loc. cit.] states that the existence of a rotationless $f : G \to G$ satisfying the conclusions of [Feighn and Handel 2011, Theorem 2.19] is equivalent to ϕ being rotationless. □

Remark 3.19 The converse of Lemma 3.18, that every rotationless PG ϕ is UPG, is also true. We make no use of this fact but include a proof for completeness. See Section 3.6 for a review of CTs. Since ϕ is rotationless and PG, it is represented by a CT $f : G \to G$ without EG or zero strata. For any such $f : G \to G$, there is a filtration $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_N = G$ by f-invariant core subgraphs such that G_i is obtained from G_{i-1} by adding a single topological edge E_i whose image $f(E_i) \subset G_i$ crosses E_i exactly once and crosses $\bar{E}_i$ not at all. [Bestvina et al. 2000, Proposition 5.7.5] therefore implies that ϕ is UPG.

3.6 CTs

A CT is a particularly nice kind of homotopy equivalence $f : G \to G$ of a marked directed graph. Every rotationless ϕ, and in particular every $\phi \in \mathrm{UPG}(F_n)$, is represented by a CT; see [Feighn and Handel 2011, Theorem 4.28] or [Bestvina et al. 2000, Theorem 5.1.8]. Moreover, CTs are considerably simpler in the $\mathrm{UPG}(F_n)$-case than in the general case.

For the remainder of the section we assume that $f : G \to G$ is a CT representing an element of $\mathrm{UPG}(F_n)$ and review its properties. Complete details can be found in [Feighn and Handel 2011] (see in particular Section 4.1) and in [Feighn and Handel 2018]. The latter introduces the (Inheritance) property for a CT $f : G \to G$, which states that the restriction of f to each component of each core filtration element is also a CT, and contains an algorithm to produce CTs satisfying (Inheritance). We say that $f : G \to G$ *realizes* a chain $\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_K$ of ϕ-invariant free factor systems if each F_j is realized by an f-invariant core subgraph of G; see Section 3.3.

Theorem 3.20 [Feighn and Handel 2018, Theorem 1.1] *There is an algorithm whose input is a rotationless* $\phi \in \mathsf{Out}(F_n)$ *and whose output is a* CT $f : G \to G$ *that represents* ϕ *and satisfies* (Inheritance). *Moreover, for any chain* $\mathcal{C}$ *of* ϕ*-invariant free factor systems, one can choose* $f : G \to G$ *to realize* $\mathcal{C}$.

We assume throughout this paper that our chosen CTs satisfy (Inheritance).

The marked graph G comes equipped with an f-invariant filtration $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_N = G$ by subgraphs G_i in which each G_i is obtained from G_{i-1} by adding a single oriented edge E_i. For each E_i there is a (possibly trivial) closed path $u_i \subset G_{i-1}$ such that $f(E_i) = E_i u_i$; if u_i is nontrivial then it forms a circuit. A path or circuit has *height* i if it crosses E_i, meaning that either E_i or $\overline{E}_i$ occurs in its edge decomposition, but does not cross E_j for any $j > i$.

Example 3.1 (continued) The homotopy equivalence $f : G \to G$ is a CT with f-invariant filtration $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_7 = G$ given by adding one edge at a time in alphabetical order.

Every map α into G with domain a closed interval or S^1 and with endpoints, if any, at vertices is properly homotopic rel endpoints to a path or circuit $[\alpha]$; we say that $[\alpha]$ is obtained from α by *tightening*. If σ is a path or circuit then we usually denote $[f(\sigma)]$ by $f_\#(\sigma)$. A decomposition into subpaths $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots$ is a *splitting* if $f^k_\#(\sigma) = f^k_\#(\sigma_1) \cdot f^k_\#(\sigma_2) \cdot \ldots$ for all $k \geq 1$. In other words, $f^k(\sigma)$ can be tightened by tightening each $f^k(\sigma_i)$.

A finite path σ is a *Nielsen path* if $f_\#(\sigma) = \sigma$; it is an *indivisible Nielsen path* if it is not a fixed edge and does not split into a nontrivial concatenation of Nielsen paths. Every Nielsen path has a splitting into fixed edges and indivisible Nielsen paths. If $f^k_\#(\sigma) = \sigma$ for some $k \geq 1$ then σ is a *periodic Nielsen path*. In a CT, every periodic Nielsen path is a Nielsen path.

An edge E_i is *linear* if u_i is a nontrivial Nielsen path. The *set of oriented linear edges* is denoted by $\mathsf{Lin}(f)$ and the set obtained from $\mathsf{Lin}(f)$ by reversing orientation is denoted by $\mathsf{Lin}^{-1}(f)$. In our example, $\mathsf{Lin}(f) = \{b, p\}$.

Associated to a CT $f : G \to G$ is a finite set of nontrivial closed Nielsen paths called *twist paths*. This set is well-defined up to a change of orientation on each path. In the remainder of this paragraph we recall some useful properties of twist paths. Each twist path w determines a circuit $[w]$ in G representing a root-free[1] conjugacy class in F_n and distinct twist paths determine distinct unoriented circuits; ie circuits whose cyclic edge decompositions differ by more than a change of orientation. For each twist path w, the set $\mathsf{Lin}_w(f)$ of (necessarily linear) edges E_i such that $f(E_i) = E_i w^{d_i}$ for some $d_i \neq 0$ is nonempty and is called the *linear family associated to* w; note that $f^k_\#(E_i) = E_i w^{kd_i}$ grows linearly in k. Every linear edge belongs to one of these linear families. If $E_i \in \mathsf{Lin}_w(f)$ and $p \neq 0$ then $E_i w^p \overline{E}_i$ is an indivisible Nielsen path. All indivisible Nielsen paths have this form. If E_i and E_j are distinct edges in $\mathsf{Lin}_w(f)$ then $d_i \neq d_j$; if d_i and d_j have the same sign, then paths of the form $E_i w^p \overline{E}_j$ are *exceptional paths*

[1]Nontrivial $a \in F_n$ is *root-free* if $x \in F_n$ and $x^k = a$ implies $k = \pm 1$.

associated to w. Note that $f_\#^k(E_i w^p \bar{E}_j) = E_i w^{p+k(d_i-d_j)} \bar{E}_j$ so these paths also grow linearly under iteration. Exceptional paths have no nontrivial splittings (which would not be true if we allowed d_i and d_j to have the opposite sign).

Example 3.1 (continued) In our example, we choose our set of twist paths to be $\{a\}$, as opposed to $\{a^{-1}\}$.

A splitting $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots$ is a *complete splitting* if each σ_i is either a single edge or an indivisible Nielsen path or an exceptional path. If σ_i is not a Nielsen path then it is *a growing term*; if at least one σ_i is growing then σ is growing. We say that σ_i is *a linear term* if it is exceptional or equal to E or $\bar{E}$ for some $E \in \mathsf{Lin}(f)$. Complete splittings are unique when they exist [Feighn and Handel 2011, Lemma 4.11]. A path with a complete splitting is said to be *completely split*. For each edge E_i there is a complete splitting of $f(E_i)$ whose first term is E_i and whose remaining terms define a complete splitting of u_i. The image under $f_\#$ of a completely split path or circuit is completely split. For each path or circuit σ, the image $f_\#^k(\sigma)$ is completely split for all sufficiently large k [Feighn and Handel 2011, Lemma 4.25].

The *set of oriented nonfixed nonlinear edges* is denoted by $\mathcal{E}_f$ and the set obtained from $\mathcal{E}_f$ by reversing orientation is denoted by $\mathcal{E}_f^{-1}$. We say that an edge in $\mathcal{E}_f$ or $\mathcal{E}_f^{-1}$ has *higher order*. An easy induction argument shows that, for each $E_i \in \mathcal{E}_f$ and $k \geq 1$, $f_\#^k(E_i)$ is completely split and

$$f_\#^k(E_i) = E_i \cdot u_i \cdot f_\#(u_i) \cdot \ldots \cdot f_\#^{k-1}(u_i).$$

Thus $f_\#^{k-1}(E_i)$ is an initial segment of $f_\#^k(E_i)$ and the union

$$R_{E_i} = E \cdot u_i \cdot f_\#(u_i) \cdot f_\#^2(u_i) \cdot \ldots$$

of this nested sequence is an $f_\#$-invariant ray called the *eigenray associated to* E_i. The complete splittings of the individual $f_\#^k(u_i)$'s define a complete splitting of R_{E_i}.

Example 3.1 (continued) In our example, the edge a is fixed, the edges b and p are linear, and the other edges have higher order, ie $\mathcal{E}_f = \{c, d, e, q\}$. As an example of an eigenray,

$$R_q = q \cdot c \cdot cb \cdot cbba \cdot \ldots \cdot cbba \ldots ba^{k-1} \cdot \ldots.$$

Lemma 3.21 *If $E \in \mathcal{E}_f \cup \mathcal{E}_f^{-1}$ then E is not crossed by any Nielsen path or exceptional path. In particular, each crossing of E by a completely split path is a term in the complete splitting of that path.*

Proof Suppose that some Nielsen path σ crosses E. Since σ is a concatenation of fixed edges and indivisible Nielsen paths and since every indivisible Nielsen path has the form $E_i w^p \bar{E}_i$ for some linear edge and twist path w, E must be crossed by some Nielsen path w with height lower than σ. The obvious induction argument completes the proof. □

Remark 3.22 One can define R_E for a linear edge E in the same way that one does for a higher-order edge. If $f(E) = Ew^d$ for some twist path w then $R_E = Ew^\infty$ if $d > 0$ and $R_E = Ew^{-\infty}$ if $d < 0$. These rays play a different role in the theory than eigenrays.

3.7 Principal lifts from the CT point of view

Suppose that $f: G \to G$ is a CT representing ϕ. If Φ is a principal lift for ϕ then we say that the corresponding $\tilde{f}$ *is a principal lift of* $f: G \to G$.

Lemma 3.23 *A lift* $\tilde{f}: \tilde{G} \to \tilde{G}$ *is principal if and only if* $\mathsf{Fix}(\tilde{f}) \neq \varnothing$ *in which case* $\mathsf{Fix}(\tilde{f})$ *contains a vertex.*

Proof This follows from Corollary 3.17, Corollary 3.27 and Remark 4.9 of [Feighn and Handel 2011] and the fact that $\mathsf{Fix}(f)$ is a union of vertices and fixed edges. □

Lemma 3.24 *Suppose that* $\tilde{f}: \tilde{G} \to \tilde{G}$ *is the lift of* $f: G \to G$ *corresponding to* $\Phi \in \mathcal{P}(\phi)$ *and that* $a \in F_n$. *The following are equivalent*:

- $a \in \mathsf{Fix}(\Phi)$;
- $i_a|\partial F_n = T_a|\partial F_n$ *commutes with* $\Phi|\partial F_n = \tilde{f}|\partial F_n$;
- $T_a|\tilde{G}$ *commutes with* $\tilde{f}|\tilde{G}$;
- $T_a(\mathsf{Fix}(\tilde{f}|\tilde{G})) = \mathsf{Fix}(\tilde{f}|\tilde{G})$.

Proof This is well known. All but the equivalence of the third and fourth bullets can be found in Lemma 2.4 of [Bestvina et al. 2004]. If $T_a|\tilde{G}$ commutes with $\tilde{f}|\tilde{G}$ then it preserves the fixed-point set of $\tilde{f}|\tilde{G}$. Conversely, if $\tilde{x}, T_a(\tilde{x}) \in \mathsf{Fix}(\tilde{f}|\tilde{G})$ then $\tilde{f}|\tilde{G}$ and $T_a\tilde{f}T_a^{-1}|\tilde{G}$ both fix $T_a(\tilde{x})$ and so must be equal. This proves the equivalence of the third and fourth bullets. □

We say that lifts $\tilde{f}_1$ and $\tilde{f}_2$ are *isogredient* if they correspond to isogredient automorphisms Φ_1 and Φ_2. Equivalently, $\tilde{f}_2 = T_a\tilde{f}_1T_a^{-1}$ for some covering translation T_a. Recall that $x, y \in \mathsf{Fix}(f)$ are *Nielsen equivalent* if they are the endpoints of a Nielsen path or equivalently, if for each lift $\tilde{x}$, the unique lift $\tilde{f}$ that fixes $\tilde{x}$ also fixes some lift $\tilde{y}$ of y.

Lemma 3.25 *The map which assigns to each principal lift* $\tilde{f}: \tilde{G} \to \tilde{G}$ *the projection into* G *of* $\mathsf{Fix}(\tilde{f})$ *induces a bijection between the set of isogredience classes of principal lifts and the set of Nielsen classes for* f. *In particular, there are only finitely many isogredience classes of principal lifts.*

Proof This follows from [Feighn and Handel 2011, Lemma 3.8] and Lemma 3.23. □

Recall from Section 3.6 that for each $E \in \mathcal{E}_f$ there is a closed completely split path u such that $f(E) = E \cdot u$ is a splitting and such that the eigenray $R_E = E \cdot u \cdot f_\#(u) \cdot f_\#^2(u) \cdot \ldots$ is $f_\#$-invariant.

The following lemma is similar to [Feighn and Handel 2018, Lemma 3.10], which applies more generally but has a weaker conclusion.

Lemma 3.26 *Suppose that* $\tilde{f}$ *corresponds to* $\Phi \in \mathcal{P}(\phi)$. *If* $\tilde{E}$ *is a lift of* $E \in \mathcal{E}_f$ *and if the initial endpoint of* $\tilde{E}$ *is contained in* $\mathsf{Fix}(\tilde{f})$ *then the lift* $\tilde{R}_{\tilde{E}}$ *of* R_E *that begins with* $\tilde{E}$ *converges to a point in* $\mathsf{Fix}_+(\Phi)$. *This defines a bijection between* $\mathsf{Fix}_+(\Phi)$ *and the set of all such* $\tilde{E}$ *and also a bijection between* $\mathcal{R}(\phi)$ *and* $\mathcal{E}_f$.

Proof $\widetilde{R}_{\widetilde{E}}$ converges to some $P \in \mathsf{Fix}_{\mathsf{N}}(\Phi)$ by [Feighn and Handel 2011, Lemma 4.36(1)]. Since E is not linear, u is not a Nielsen path and hence not a periodic Nielsen path. The length of $f_\#^k(u)$ therefore goes to infinity with k. Proposition I.1 of [Gaboriau et al. 1998] implies that $P \in \mathsf{Fix}_+(\Phi)$.

Suppose that $\widetilde{E}_1$ and $\widetilde{E}_2$ are distinct edges that project into $\mathcal{E}_f$, that the initial endpoint $\tilde{x}_i$ of $\widetilde{E}_i$ is fixed by $\tilde{f}$ and that, for $i = 1, 2$, $\widetilde{R}_i$ is the lift of R_{E_i} with initial edge $\widetilde{E}_i$. The path that connects $\tilde{x}_1$ to $\tilde{x}_2$ projects to a Nielsen path $\sigma \subset G$. If $\widetilde{R}_1$ and $\widetilde{R}_2$ converge to the same point in $\mathsf{Fix}_+(\Phi)$ then σ crosses E_1 or E_2 in contradiction to Lemma 3.21. This proves that the map $\{\widetilde{E}\} \mapsto \mathsf{Fix}_+(\Phi)$ is injective; surjectivity follows from [Feighn and Handel 2011, Lemma 4.36(2)]. The second bijection is obtained from the first by projecting to the sets of F_n-orbits. □

Example 3.1 (continued) Since $\mathcal{E}_f = \{c, d, e, q\}$, $\mathcal{R}(\phi)$ has four elements, denoted by $\{r_c, r_d, r_e, r_q\}$; see Figure 2.

4 Recognizing a conjugator

Associated to each CT $f : G \to G$ representing a rotationless element $\phi \in \mathsf{Out}(F_n)$ is a finite type labeled graph $\Gamma(f)$ that realizes $\mathsf{Fix}_{\mathsf{N}}(\phi)$. We refer to $\Gamma(f)$ as the *eigengraph* for $f : G \to G$. In Section 4.1 we recall the construction and relevant properties of $\Gamma(f)$ in the case that $\phi \in \mathsf{UPG}(F_n)$. Every ϕ-invariant conjugacy class $[a]$ is represented by an oriented circuit in $\Gamma(f)$. There is a finite set $\mathcal{A}_{\mathrm{or}}(\phi)$ of such $[a]$ that are root-free and that are represented by more than one oriented circuit in $\Gamma(f)$. The "extra" circuits correspond to the linear edges in $f : G \to G$. In Section 4.2, we describe how the extra circuits can be incorporated into an invariant $\mathsf{SA}(\phi)$ of ϕ that is independent of the choice of $f : G \to G$. Moreover, certain pairs of elements of $\mathsf{SA}(\phi)$ have *twist coordinates* that can be read off from the twist coordinates on linear edges in $f : G \to G$. Section 4.3 is an application of the recognition theorem of [Feighn and Handel 2011]. Assuming that $\phi, \psi \in \mathsf{UPG}(F_n)$, we use eigengraphs, $\mathsf{SA}(\phi)$ and twist coordinates to give necessary and sufficient conditions for a given $\theta \in \mathsf{Out}(F_n)$ to conjugate ϕ to ψ.

4.1 The eigengraph $\Gamma(f)$

In this section, we recall a finite type labeled graph that captures many of the invariants of ϕ that are essential to our algorithm. For further details and more examples, see [Feighn and Handel 2018, Sections 9, 10 and 12].

A graph Γ without valence one vertices and equipped with a simplicial immersion $p : \Gamma \to G$ to a marked graph G will be called a *Stallings graph*. We label the vertices and edges of Γ by their p-images in G. Two Stallings graphs $p_1 : \Gamma_1 \to G$ and $p_2 : \Gamma_2 \to G$ are equivalent if there is a label-preserving simplicial homeomorphism $h : \Gamma_1 \to \Gamma_2$. We will not distinguish between equivalent Stallings graphs. Since all vertices have valence at least two, every edge in Γ is crossed by a line. We say that Γ has *finite type* if its core is finite and if the complement of the core is a finite union of rays.

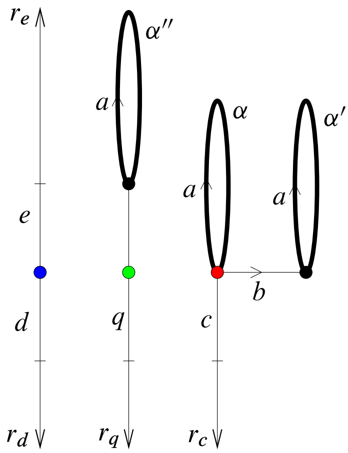


Figure 2: The eigengraph $\Gamma(f)$ of our example. Only the first edge of the eigenrays is labeled here. For example, the eigenray $R_c = cbbaba^2ba^3\ldots$ starts at the red vertex and only its first edge is labeled (by c). Other aspects of this figure are explained later.

Given a CT $f : G \to G$ representing ϕ, we construct a finite type Stallings graph $\Gamma(f)$, called the *eigengraph for* $f : G \to G$, as follows. Let $\Gamma^0(f)$ be G with the interiors of all nonfixed edges removed. In particular, $\Gamma^0(f)$ may contain isolated vertices. The labeling on $\Gamma^0(f)$ is the obvious one. For each $E \in \mathsf{Lin}(f)$, first attach an edge, say E', to $\Gamma^0(f)$ by identifying the initial endpoint of E' with the initial endpoint of E, thought of as a vertex in $\Gamma^0(f)$. The label on E' is E. Then add a path ω by attaching both of its endpoints to the terminal endpoint of E', which now has valence three. If w denotes the twist path associated to E then we label ω by w, thought of as an edge path, and subdivide ω so that each edge in ω is labeled by a single edge in G. The net effect is to add a *lollipop* to $\Gamma^0(f)$ for each edge in $\mathsf{Lin}(f)$. This labeling defines an immersion because w determines a circuit that does not cross the edge E. Finally, for each $E \in \mathcal{E}_f$, attach a ray labeled R_E (as defined in Section 3.6) by identifying the initial endpoint of this ray with the initial endpoint of E, thought of as a vertex in $\Gamma^0(f)$. We will also use the term *eigenray* for this added ray. The resulting graph is denoted by $\Gamma(f)$. This labeling maintains the immersion property because R_E is an immersed ray in G and because no other edge labeled E has initial vertex in $\Gamma^0(f)$.

Example 3.1 (continued) The eigengraph for our running example is pictured in Figure 2.

The vertices of $\Gamma(f)$ that are not in $\Gamma^0(f)$ have valence either two or three by construction. The valence of $v \in \Gamma^0(f)$ in $\Gamma(f)$ is equal to the number of fixed directions based at v in G. If v, thought of as a vertex in G, is not the terminal endpoint of an edge in $\mathcal{E}_f \cup \mathsf{Lin}(f)$, then v has the same valence in $\Gamma(f)$ that it does in G. If v is the terminal endpoint of an edge in $\mathcal{E}_f \cup \mathsf{Lin}(f)$, let E_i be the lowest such edge. Then $f(E_i) = E_i \cdot u_i$, where u_i is a closed path based at v whose ends determine distinct fixed directions at v by [Feighn and Handel 2011, Lemma 4.21]. This proves that v has valence at least two in $\Gamma(f)$ and hence that $\Gamma(f)$ is a Stallings graph. It has finite type by construction.

As noted in Section 3.6, each Nielsen path in G decomposes as a concatenation of fixed edges and indivisible Nielsen paths and each indivisible Nielsen path is a closed path. It follows that two vertices in $\mathsf{Fix}(f)$ are in the same Nielsen class if and only if they are connected by a sequence of fixed edges. In particular, the vertices in each component of $\Gamma^0(f)$ form exactly one Nielsen class in $\mathsf{Fix}(f)$. Since the inclusion of $\Gamma^0(f)$ into $\Gamma(f)$ induces a bijection of components, there is a bijection between the set of components of $\Gamma(f)$ and the set of Nielsen classes in $\mathsf{Fix}(f)$ and hence (Lemma 3.25) a bijection between the set of components of $\Gamma(f)$ and the set of isogredience classes in $\mathcal{P}(\phi)$. We denote the component of $\Gamma(f)$ corresponding to the isogredience class $[\Phi]$ by $\Gamma_{[\Phi]}(f)$ or by $\Gamma(\tilde{f})$, where $\tilde{f}$ is the lift of f that corresponds to Φ.

We say that *a line is carried by* $\Gamma_{[\Phi]}(f)$ if its realization $L \subset G$ lifts into $\Gamma_{[\Phi]}(f)$ and is *carried by* $\Gamma(f)$ if it is carried by some component $\Gamma_{[\Phi]}(f)$. The following lemma shows that the set of lines carried by $\Gamma(f)$ is independent of the choice of $f: G \to G$. We will sometimes refer to these as *principal lines.*

Lemma 4.1 *The following are equivalent for any* CT $f: G \to G$ *representing* ϕ, *any* $\Phi \in \mathcal{P}(\phi)$ *and any line* $L \subset G$.

(1) *L is carried by $\Gamma_{[\Phi]}(f)$ (resp. the core of $\Gamma_{[\Phi]}(f)$).*

(2) *There is a lift $\tilde{L} \subset \tilde{G}$ such that $\{\partial_\pm \tilde{L}\} \subset \mathsf{Fix}_{\mathsf{N}}(\Phi)$ (resp. $\{\partial_\pm \tilde{L}\} \subset \partial\,\mathsf{Fix}(\Phi)$).*

Proof It suffices to prove the unbracketed statement. Let $q: \tilde{G} \to G$ and $q_\Gamma: \tilde{\Gamma}_{[\Phi]}(f) \to \Gamma_{[\Phi]}(f)$ be the universal covering maps. The labeling map $p: \Gamma_{[\Phi]}(f) \to G$ is an immersion and so lifts to an embedding $\tilde{p}: \tilde{\Gamma}_{[\Phi]}(f) \hookrightarrow \tilde{G}$. If a line $L \subset G$ lifts to a line $L_\Gamma \subset \Gamma_{[\Phi]}(f)$ and if $\tilde{L}_\Gamma \subset \tilde{\Gamma}_{[\Phi]}(f)$ is a lift of L_Γ then $\tilde{L} := \tilde{p}(\tilde{L}_\Gamma) \subset \tilde{G}$ is a lift of L. Conversely, if $L \subset G$ lifts to $\tilde{L} \subset \tilde{G}$ and there exists $\tilde{L}_\Gamma \subset \tilde{\Gamma}_{[\Phi]}(f)$ with $\tilde{p}(\tilde{L}_\Gamma) = \tilde{L}$, then $L_\Gamma := q_\Gamma(\tilde{L}_\Gamma)$ is a lift of L. The lemma therefore follows from [Feighn and Handel 2018, Lemma 12.4], which states that $L \subset G$ lifts to $\tilde{p}(\tilde{L}_\Gamma)$ if and only if (2) is satisfied. □

An end of an immersed line in $\Gamma(f)$ can either be an end of $\Gamma(f)$ or can wrap infinitely around one of the lollipop circuits or can cross a vertex in $\Gamma^0(f)$ infinitely often. This gives the following description of lines that lift into $\Gamma(f)$.

Lemma 4.2 *A line $\sigma \subset G$ lifts into $\Gamma(f)$ if and only if it contains a (possibly trivial) subpath β that is a concatenation of fixed edges and indivisible Nielsen paths and such that the complement of β is* $0, 1$ *or* 2 *rays, each of which is either R_e for some higher-order edge e, or $Ew^{\pm\infty}$ for some twist path w and some linear edge $E \in \mathsf{Lin}_w(f)$.* □

The following lemma, in conjunction with Lemma 4.1, implies that the set of conjugacy classes determined by twist paths and their inverses is an invariant of ϕ. This set is explored further in Section 4.2.

Lemma 4.3 *Suppose that $f: G \to G$ is a* CT *representing ϕ and that $[a]$ is a root-free conjugacy class that is fixed by ϕ and that σ_a is the circuit in G representing $[a]$.*

(1) *If $[a] = [w]$ (resp. $[a] = [\overline{w}]$) for some twist path w, then for each edge $E \in \mathsf{Lin}_w(f)$ there is a lift of σ_a to the loop ω (resp. $\overline{\omega}$) in the lollipop associated to E. Additionally there is a unique lift of σ_a to a circuit in $\Gamma(f)$ that is not contained in any lollipop.*

(2) *Otherwise, there is a unique lift of σ_a to a circuit in $\Gamma(f)$.*

Proof We claim that if τ' is a path in the core of $\Gamma(f)$ that projects to a Nielsen path τ in G and if the initial vertex v'' of τ' is not in $\Gamma^0(f)$, then τ' is contained in the loop ω of some lollipop. We may assume without loss that the claim holds for paths with height less than that of τ and that τ is either a single fixed edge or an indivisible Nielsen path. Since the core of $\Gamma(f)$ is contained in the union of $\Gamma^0(f)$ with the lollipops associated to the linear edges of G, there is a lollipop composed of an edge E_1' projecting to a linear edge $E_1 \subset G$ and a loop ω projecting to its twist path w_1 such that $v'' \in \omega$. If τ is a fixed edge then τ' is disjoint from the interior of E_1' and so is contained in ω. If τ is an indivisible Nielsen path then it has the form $E_2 {w_2}^p \overline{E}_2$ for some linear edge E_2 with twist path w_2. There is an induced decomposition $\tau' = X' w_2' Y'$, where X', w_2' and Y' project to E_2, w_2^p and $\overline{E}_2$, respectively. Since $E_2 \neq \overline{E}_1$, we have $X' \subset \omega$ and the initial vertex of ${w_2'}^p$ is contained in ω. Since w_2 has height less than E_2 and so height less than τ, ${w_2'}^p \subset \omega$. Finally, since X' is contained in ω, E_2 has height less than E_1 and in particular, $\overline{E}_2 \neq \overline{E}_1$. Thus $Y' \subset \omega$. This completes the proof of the claim.

Each $\sigma = \sigma_a$ as in the statement of the lemma has a cyclic splitting $\sigma = \sigma_1 \cdot \ldots \cdot \sigma_m$ into fixed edges and indivisible Nielsen paths σ_i. The above claim shows that if $\sigma' = \sigma_1' \cdot \ldots \cdot \sigma_m'$ is a lift to $\Gamma(f)$ in which an endpoint of some σ_i' is not contained in $\Gamma(f)^0$ then σ' is entirely contained in the loop ω associated to one of the lollipops.

To complete the proof we need only show each σ has a unique lift in which the endpoints of each σ_i lift into $\Gamma^0(f)$. Since the vertices of G have unique lifts into $\Gamma^0(f)$, it suffices to show that each σ_i has a unique lift with endpoints in $\Gamma^0(f)$ and this is immediate from the construction of $\Gamma^0(f)$. □

4.2 Strong axes and twist coordinates

The following lemma describes the extent to which fixed subgroups fail to be malnormal.

Lemma 4.4 *For distinct automorphisms Φ_1 and Φ_2 representing the same outer automorphism and for any $c \in F_n$:*

(1) $\mathsf{Fix}(\Phi_1) \cap \mathsf{Fix}(\Phi_2)$ *is either trivial or a maximal cyclic subgroup.*

(2) *If $c \notin \mathsf{Fix}(\Phi_1)$ then $\mathsf{Fix}(\Phi_1) \cap (\mathsf{Fix}(\Phi_1))^c = \mathsf{Fix}(\Phi_1) \cap \mathsf{Fix}(i_c \Phi_1 i_c^{-1})$ is either trivial or a maximal cyclic subgroup.*

(3) $\mathsf{Fix}(\Phi_1)$ *is its own normalizer.*

Proof If Φ_1 and Φ_2 are distinct automorphisms representing the same outer automorphism then $\Phi_1^{-1}\Phi_2$ is a nontrivial inner automorphism and $\mathsf{Fix}(\Phi_1)\cap\mathsf{Fix}(\Phi_2)$ is a subgroup of the cyclic group $\mathsf{Fix}(\Phi_1^{-1}\Phi_2)$. Maximality of $\mathsf{Fix}(\Phi_1)\cap\mathsf{Fix}(\Phi_2)$ follows from Lemma 3.6 and the fact that $(a^k)^{\pm}=a^{\pm}$ for all nontrivial $a\in F_n$ and all $k\geq 1$. This proves (1).

For (2) note that if $c\notin\mathsf{Fix}(\Phi_1)$ then $i_c\Phi_1 i_c^{-1}=i_{c\Phi_1(c^{-1})}\Phi_1\neq\Phi_1$. Note also that $\mathsf{Fix}(i_c\Phi_1 i_c^{-1})=(\mathsf{Fix}(\Phi_1))^c$. Item (2) therefore follows from (1) applied with $\Phi_2=i_c\Phi_1 i_c^{-1}$. In proving (3) we may assume by (2) that $\mathsf{Fix}(\Phi_1)=\langle a\rangle$ for some root-free $a\in F_n$ and in this case (3) is obvious. □

The conjugacy class of a cyclic subgroup is determined by the conjugacy class of either of its generators. As we have no way to canonically choose a generator, we work, for now, with unoriented conjugacy classes. The following definition appeared as [Bestvina et al. 2004, Definition 4.6] under slightly different hypotheses and in the paragraph before [Feighn and Handel 2011, Remark 4.39] in the CT context.

Definition 4.5 Elements $a,b\in F_n$ are in the same *unoriented conjugacy class* if $a=i_c(b)$ or $a=i_c(b^{-1})$ for some $c\in F_n$. An unoriented conjugacy class $[a]_u$ of a nontrivial root-free $a\in F_n$ is an *axis* for ϕ if $\langle a\rangle=\mathsf{Fix}(\Phi_1)\cap\mathsf{Fix}(\Phi_2)$ for distinct $\Phi_1,\Phi_2\in\mathcal{P}(\phi)$. The *multiplicity of an axis* $[a]_u$ is the number of distinct $\Phi_i\in\mathcal{P}(\phi)$ that fix a. The *set of axes for* ϕ is denoted by $\mathcal{A}(\phi)$. The set $\{[a]\mid[a]_u\in\mathcal{A}(\phi)\}$ is denoted by $\mathcal{A}_{\mathrm{or}}(\phi)$.

There is a very useful description of $\mathcal{A}(\phi)$ in terms of a CT $f\colon G\to G$.

Lemma 4.6 *If $f\colon G\to G$ is a* CT *representing ϕ and $\{w_i\}$ is the set of twist paths for f, then $\mathcal{A}(\phi)=\{[w_i]_u\}$. In particular, $\mathcal{A}(\phi)$ is finite.*

Proof This follows from [Feighn and Handel 2011, Lemma 4.40]. □

Notation 4.7 If $[a]_u=[w]_u$ for some twist path w then, up to a reversal of orientation, the axis of the covering translation $T_a\colon\widetilde{G}\to\widetilde{G}$ can be viewed as an infinite concatenation $\ldots\widetilde{w}_{-1}\widetilde{w}_0\widetilde{w}_1\ldots$ of paths that project to w. There is a principal lift $\tilde{f}_{a,0}\colon\widetilde{G}\to\widetilde{G}$, called the *base principal lift for* a, that fixes the endpoints of each $\widetilde{w}_l$. The principal automorphism $\Phi_{a,0}$ corresponding to $\tilde{f}_{a,0}$ is called the *base principal automorphism* for a. If a is an element of some basis for F_n then the base principal lift for a depends on the choice of $f\colon G\to G$, and not just on ϕ.

For each edge $E^j\in\mathsf{Lin}_w(f)$, there is a principal lift $\tilde{f}_{a,j}\colon\widetilde{G}\to\widetilde{G}$ that fixes the initial endpoint of each lift $\widetilde{E}^j$ with terminal endpoint equal to the initial endpoint of some $\widetilde{w}_l$. (We write E^j rather than E_j to emphasize that j is not an indicator of height in G.) The principal automorphism corresponding to $\tilde{f}_{a,j}$ is denoted by $\Phi_{a,j}$. Note that $\Phi_{a,0}=\Phi_{a^{-1},0}$ and $\Phi_{a,j}=\Phi_{a^{-1},j}$. Further details can be found in [Feighn and Handel 2011, Lemma 4.40] and the paragraph that precedes it.

Lemma 4.8 *Suppose that $f : G \to G$ is a* CT *representing ϕ, that w is a twist path for f and that $a \in F_n$ satisfies $[a]_u = [w]_u$. Suppose also that $E^1, \dots, E^{m-1}$ are the edges in $\mathsf{Lin}_w(f)$.*

(1) *$\{\Phi_{a,0}, \Phi_{a,1}, \dots, \Phi_{a,m-1}\}$ is the set of principal automorphisms that fix a. In particular, the multiplicity of each element of $\mathcal{A}(\phi)$ is finite.*

(2) *If $f(E^j) = E^j w^{d_j}$ then $\Phi_{a,j} = i_a^{d_j} \Phi_{a,0}$ if $[a] = [w]$, and $\Phi_{a,j} = i_a^{-d_j} \Phi_{a,0}$ if $[a] = [\overline{w}]$.*

Proof This follows from Lemma 4.40 of [Feighn and Handel 2011]. □

Definition 4.9 Suppose that the group G acts on the sets X_i for $i = 1, \dots, k$, and that $x_i \in X_i$. The orbit of $(x_1, \dots, x_k)$ under the diagonal action of G on $\prod_{i=1}^k X_i$, denoted by $[x_1, \dots, x_k]_G$, is a *conjugacy k-tuple*. If $k = 2$ then we say $[x_1, x_2]_G$ is a *conjugacy pair*. We sometimes suppress the subscript, in which case $G = F_n$.

Examples 4.10 Here are some examples of conjugacy pairs where $G = F_n$.

- We will often take X_i to be the set of finitely generated subgroups of F_n or F_n itself with the action of F_n given by conjugation. If $H < F_n$ (resp. $x \in F_n$) then $[H]_{F_n}$ (resp. $[x]_{F_n}$) is the conjugacy class of H in F_n (resp. x in F_n). Conjugacy pairs formed with these X_i will play an important role in this paper, especially in Section 10.3.
- If $X = \partial F_n$ and if $x \neq y \in X$, then $(x, y) \in X \times X$ is an oriented line. The conjugacy pair $[x, y]_{F_n}$ is represents an oriented line in any marked graph.
- If X is the power set of ∂F_n and A and B are disjoint subsets of ∂F_n, then $(A, B) \in X \times X$ denotes the set of lines L with $\partial_- L \in A$ and $\partial_+ L \in B$. The conjugacy pair $[A, B]_{F_n}$ represents a set of oriented lines in any marked graph.

We now define strong axes, the first of our invariants that is expressed as a conjugacy pair.

Definition 4.11 Let $\mathcal{A}_{\mathrm{or}}(\phi)$ be the set of conjugacy classes representing elements of $\mathcal{A}(\phi)$, ie $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$ if $\{[a], [a^{-1}]\} \in \mathcal{A}(\phi)$. F_n acts on pairs (Φ, a) where $\Phi \in \mathcal{P}(\phi)$, $a \in \mathsf{Fix}(\Phi)$, and $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$ via $(\Phi, a)^g = (\Phi^{i_g}, a^g)$. The F_n-orbit, equivalently conjugacy pair, $[\Phi, a]$, is a *strong axis for ϕ*. If $\alpha_s = [\Phi, a]$ then we let $\alpha_s^{-1} := [\Phi, a^{-1}]$. The set of all strong axes for ϕ is denoted by $\mathsf{SA}(\phi)$. $\mathsf{Aut}(F_n)$ acts on pairs (Φ, a) by $\Theta \cdot (\Phi, a) = (\Theta\Phi\Theta^{-1}, \Theta(a))$. This descends to an action of $\mathsf{Out}(F_n)$ on $\mathsf{SA}(\phi)$.

We can partition $\mathsf{SA}(\phi)$ according to the second coordinate: for each $\mu \in \mathcal{A}_{\mathrm{or}}(\phi)$ let $\mathsf{SA}(\phi, \mu)$ be the subset of $\mathsf{SA}(\phi)$ consisting of elements in which some, and hence every, representative (Φ, a) satisfies $[a] = \mu$.

Lemma 4.12 *Suppose that $a \in F_n$, that $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$ and that $\Phi_{a,0}, \dots, \Phi_{a,m-1}$ are as in Notation 4.7. For each $\alpha_s \in \mathsf{SA}(\phi, [a])$, there is a unique $\Phi_{a,j}$ such that $\alpha_s = [\Phi_{a,j}, a]$. Thus*

$$\mathsf{SA}(\phi, [a]) = \{[\Phi_{a,0}, a], [\Phi_{a,1}, a], \dots, [\Phi_{a,m-1}, a]\}.$$

Proof Each $\alpha_s \in \mathsf{SA}(\phi,[a])$ is represented by (Φ, a^c) and hence by $(i_c\Phi i_c^{-1}, a)$, for some $c \in F_n$ and some $\Phi \in \mathcal{P}(\phi)$. Since $a \in \mathsf{Fix}(i_c\Phi i_c^{-1})$, there exists j such that $i_c\Phi i_c^{-1} = \Phi_j$.

For uniqueness, note that if $(\Phi_j, a) = c \cdot (\Phi_i, a)$ for some $c \in F_n$ then $c = a^p$ for some p so i_c commutes with Φ_j and $c \cdot (\Phi_i, a) = (\Phi_i, a)$. □

Remark 4.13 There is another useful description of $[\Phi_{a,j}, a] \in \mathsf{SA}(\phi,[a])$ in terms of a CT $f : G \to G$. Let w be the twist path satisfying $[a]_u = [w]_u$ and let v be the initial vertex of w. There is an automorphism $f_{v\#} \colon \pi_1(G, v) \to \pi_1(G, v)$ that sends the homotopy class of the closed path σ with basepoint v to the homotopy class of the closed path $f(\sigma)$ with basepoint v. Let τ be the element of $\pi_1(G, v)$ determined by w if $[a] = [w]$ and by $\bar{w}$ if $[a] = [\bar{w}]$. In both cases, τ is fixed by $f_{v\#}$. There is an isomorphism from $\pi_1(G, v)$ to F_n that is well-defined up to postcomposition with an inner automorphism of F_n. The pair $(f_{v\#}, \tau)$ determines a well-defined element, namely $[\Phi_{a,0}, a]$, of $\mathsf{SA}(\phi,[a])$. Similarly if v_j is the initial endpoint of $E^j \in \mathsf{Lin}_w(f)$, let τ_j be the element of $\pi_1(G, v_j)$ determined by $E^j w \bar{E}^j$ if $[a] = [w]$ and by $E^j \bar{w} \bar{E}^j$ if $[a] = [\bar{w}]$. Then $(f_{v_j\#}, \tau_j)$ determines $[\Phi_{a,j}, a]$.

Continuing with this notation, we can relate $\mathsf{SA}(\phi,[a])$ to circuits in the eigengraph $\Gamma(f)$ that are lifts of $[w]_u$. For $j \neq 0$, $[\Phi_{a,j}, a]$ corresponds to the loop at the end of the lollipop in $\Gamma(f)$ determined by E^j. By Lemma 4.3 there is one more lift of $[w]_u$ into $\Gamma(f)$, and this corresponds to $[\Phi_{a,0}, a]$.

Definition 4.14 Suppose that $\mu \in \mathcal{A}_{\mathrm{or}}(\phi)$ and that $\alpha_s, \alpha'_s \in \mathsf{SA}(\phi, \mu)$. Choose $a \in F_n$ such that $[a] = \mu$ and let $\Phi, \Phi' \in \mathcal{P}(\phi)$ be the unique elements such that $\alpha_s = [\Phi, a]$ and $\alpha'_s = [\Phi', a]$. Since Φ and Φ' both fix a there exists $\tau \in \mathbb{Z}$ such that $\Phi' = i_a^\tau \Phi$; equivalently, $\Phi'\Phi^{-1} = i_a^\tau$. We say that $\tau = \tau(\alpha'_s, \alpha_s)$ is the *twist coordinate* associated to α'_s and α_s.

Example 3.1 (continued) In our example, $\mathsf{SA}(\phi,[a])$ is represented in Figure 2 by the three circles α, α', and α'' labeled a and drawn with thicker lines. $\mathsf{SA}(\phi) = \mathsf{SA}(\phi,[a]) \cup \mathsf{SA}(\phi,[a^{-1}])$. We have for example $\tau(\alpha', \alpha) = 1$.

Lemma 4.15 *Twist coordinates are well-defined.*

Proof We have to show that $\tau(\alpha'_s, \alpha_s)$ is independent of the choice of a representing μ. If a is replaced by a^c then Φ and Φ' are replaced by $i_c\Phi i_c^{-1}$ and $i_c\Phi' i_c^{-1}$, respectively, and so $i_a^\tau = \Phi'\Phi^{-1}$ is replaced by $i_c\Phi'\Phi^{-1}i_c^{-1} = i_c i_a^\tau i_c^{-1} = i_{a^c}{}^\tau$. □

The following lemma allows us to compute twist coordinates for strong axes from a CT $f : G \to G$. It is an immediate consequence of Lemma 4.8 and the definitions.

Lemma 4.16 (1) *If $[a] = [w]$ and $E^j \in \mathsf{Lin}_w(f)$ satisfies $f(E^j) = E^j w^{d_j}$, then*

$$\tau([\Phi_{a,j}, a], [\Phi_{a,0}, a]) = d_j.$$

(2) *Suppose that* $\mu \in \mathcal{A}_{\mathrm{or}}(\phi)$ *and that* $\alpha_s, \beta_s, \gamma_s \in \mathsf{SA}(\phi, \mu)$*. Then*

(a) $\tau(\alpha_s, \beta_s) = -\tau(\beta_s, \alpha_s)$,

(b) $\tau(\alpha_s, \gamma_s) = \tau(\alpha_s, \beta_s) + \tau(\beta_s, \gamma_s)$,

(c) $\tau(\alpha_s, \beta_s) = -\tau(\alpha_s^{-1}, \beta_s^{-1})$.

The next lemma shows that $\mathcal{A}(\phi)$, $\mathsf{SA}(\phi, [a])$ and $\mathsf{SA}(\phi)$ are natural invariants.

Lemma 4.17 *Assume that* $\psi = \theta\phi\theta^{-1}$ *and that* Θ *represents* θ*.*

(1) *The correspondence* $[a]_u \leftrightarrow (\theta[a])_u$ *defines a bijection* $\mathcal{A}(\phi) \leftrightarrow \mathcal{A}(\psi)$*.*

(2) *The correspondence* $(\Phi, a) \leftrightarrow (\Theta\Phi\Theta^{-1}, \Theta(a))$ *induces a bijection* $\mathsf{SA}(\phi, [a]) \leftrightarrow \mathsf{SA}(\psi, \theta([a]))$ *that preserves twist coordinates.*

Proof If $\Phi_1, \Phi_2 \in \mathcal{P}(\phi)$ fix $a \in F_n$, then $\Psi_1 := \Theta\Phi_1\Theta^{-1}$ and $\Psi_2 := \Theta\Phi_2\Theta^{-1}$ in $\mathcal{P}(\psi)$ fix $\Theta(a)$. This proves (1).

For (2), let $\Psi = \Theta\Phi\Theta^{-1}$ and note that if $\Phi \in \mathcal{P}(\phi)$ and $a \in \mathsf{Fix}(\Phi)$ then $\Psi \in \mathcal{P}(\psi)$ and $\Theta(a) \in \mathsf{Fix}(\Psi)$. Moreover,

$$c \cdot (\Phi, a) = (i_c \Phi i_c^{-1}, a^c) \mapsto (i_{\Theta(c)} \Psi i_{\Theta(c)}^{-1}, \Theta(a)^{\Theta(c)}) = \Theta(c) \cdot (\Psi, \Theta(a)).$$

This proves that $(\Phi, a) \mapsto (\Theta\Phi\Theta^{-1}, \Theta(a))$ induces a well-defined map $\mathsf{SA}(\phi, [a]) \to \mathsf{SA}(\psi, \theta([a]))$ that is obviously invertible and is hence a bijection. If Φ_i and Ψ_i are as in the proof of (1) and if $\Phi_2 = i_a^\tau \Phi_1$, then $\Psi_2 = i_{\Theta(a)}^\tau \Psi_1$. This proves that twist coordinates are preserved. □

We conclude this section with a conjugacy class of pairs construction that is better suited to the techniques in Section 10 than the one in Definition 4.11 but is only applicable when the fixed subgroups in question have rank at least two.

Definition 4.18 Given $\phi \in \mathsf{Out}(F_n)$, consider pairs $(\mathsf{Fix}(\Phi), a)$ where $\Phi \in \mathcal{P}(\phi)$, $a \in \mathsf{Fix}(\Phi)$ and $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$. Using $\mathsf{Fix}(i_c \Phi i_c^{-1}) = i_c(\mathsf{Fix}(\Phi))$, the action of F_n on such pairs is given by $c \cdot (\mathsf{Fix}(\Phi), a) = (i_c(\mathsf{Fix}(\Phi)), i_c(a))$, giving a conjugacy pair $[\mathsf{Fix}(\Phi), a]$. Similarly, $\mathsf{Aut}(F_n)$ acts on pairs $(\mathsf{Fix}(\Phi), a)$ by $\Theta \cdot (\mathsf{Fix}(\Phi), a) = (\Theta(\mathsf{Fix}(\Phi)), \Theta(a))$. This descends to an action of $\mathsf{Out}(F_n)$ on the set of such conjugacy pairs.

Remark 4.19 Since $\mathsf{Fix}(\Phi)$ is its own normalizer (Lemma 4.4(3)), $[\mathsf{Fix}(\Phi), a] = [\mathsf{Fix}(\Phi), a']$ if and only if $a' = i_c(a)$ for some $c \in \mathsf{Fix}(\Phi)$; equivalently, a and a' are conjugate as elements of $\mathsf{Fix}(\Phi)$.

Lemma 4.20 *Suppose that* $\mathsf{Fix}(\Phi)$ *and* $\mathsf{Fix}(\Phi')$ *have rank at least two. Then*

$$[\Phi, a] = [\Phi', a'] \iff [\mathsf{Fix}(\Phi), a] = [\mathsf{Fix}(\Phi'), a'].$$

Proof By definition, $[\Phi, a] = [\Phi', a']$ if and only if there exists $c \in F_n$ such that

$$\Phi' = i_c \Phi i_c^{-1} \quad \text{and} \quad a' = i_c(a).$$

Similarly, $[\mathsf{Fix}(\Phi), a] = [\mathsf{Fix}(\Phi'), a']$ if and only if there exists $c \in F_n$ such that $\mathsf{Fix}(\Phi') = i_c \,\mathsf{Fix}(\Phi)$ and $a' = i_c(a)$. As we are assuming that $\mathsf{Fix}(\Phi)$ and $\mathsf{Fix}(\Phi')$ have rank at least two, $\Phi' = i_c \Phi i_c^{-1}$ if and only if $\mathsf{Fix}(\Phi') = i_c \,\mathsf{Fix}(\Phi)$. □

4.3 Applying the recognition theorem

The recognition theorem [Feighn and Handel 2011, Theorem 5.1] gives invariants that completely determine rotationless elements of $\mathsf{Out}(F_n)$. In this paper, via the following lemma, we use it to give a sufficient condition for two elements of $\mathsf{UPG}(F_n)$ to be conjugate in $\mathsf{Out}(F_n)$.

Lemma 4.21 *Suppose that $f: G \to G$ and $g: G' \to G'$ are* CT*s representing ϕ and ψ respectively, that $\theta \in \mathsf{Out}(F_n)$ and that a line L lifts into $\Gamma(f)$ (meaning that the realization of L in G is the image of a line in $\Gamma(f)$) if and only if $\theta(L)$ lifts into $\Gamma(g)$. Then for each $\Theta \in \mathsf{Aut}(F_n)$ representing θ:*

1. *There is a bijection $B_{\mathcal{P}}: \mathcal{P}(\phi) \to \mathcal{P}(\psi)$ such that $\mathsf{Fix}_{\mathsf{N}}(B_{\mathcal{P}}(\Phi)) = \Theta(\mathsf{Fix}_{\mathsf{N}}(\Phi)) = \mathsf{Fix}_{\mathsf{N}}(\Phi^{\Theta})$. In particular, $\mathsf{Fix}(B_{\mathcal{P}}(\Phi)) = \Theta \,\mathsf{Fix}(\Phi) = \mathsf{Fix}(\Phi^{\Theta})$.*
2. *The map $[\Phi, a] \mapsto [B_{\mathcal{P}}(\Phi), \Theta(a)]$ defines a bijection $B_{\mathsf{SA}}: \mathsf{SA}(\phi) \to \mathsf{SA}(\psi)$, independent of the choice of Θ, such that $\phi^{\theta} = \psi$ if and only if B_{SA} preserves twist coordinates.*

Proof Given $\Phi \in \mathcal{P}(\phi)$, choose a line $\widetilde{L}_1 \subset \widetilde{G}$ with both ends nonperiodic and both ends in $\mathsf{Fix}_{\mathsf{N}}(\Phi)$. (This is possible by Remark 3.10.) By Lemma 4.1, the projection $L \subset G$ lifts to the component $\Gamma_{[\Phi]}(f)$ of $\Gamma(f)$ that corresponds to $[\Phi]$. By hypothesis, the line $L_1' \subset G'$ corresponding to $\theta(L)$ lifts to a component of $\Gamma(g)$ and so by a second application of Lemma 4.1 there is a unique $\Psi \in \mathcal{P}(\psi)$ such that $\mathsf{Fix}_{\mathsf{N}}(\Psi)$ contains the endpoints $\{\Theta(\partial_{\pm}\widetilde{L}_1)\}$ of $\Theta(\widetilde{L}_1)$; moreover, L_1' lifts into $\Gamma_{[\Psi]}(g)$. To see that Ψ is independent of the choice of $\widetilde{L}_1$, suppose that we are given some other $\widetilde{L}_2$ with both ends nonperiodic and both ends in $\mathsf{Fix}_{\mathsf{N}}(\Phi)$. Let $\widetilde{L}_3$ be the line connecting the terminal endpoint of $\widetilde{L}_1$ to the initial endpoint of $\widetilde{L}_2$. Since $\widetilde{L}_1$ and $\widetilde{L}_3$ have a common endpoint, replacing $\widetilde{L}_1$ with $\widetilde{L}_3$ does not change Ψ. For the same reason, replacing $\widetilde{L}_3$ with $\widetilde{L}_2$ does not change Ψ. We conclude that $B_{\mathcal{P}}(\Phi) = \Psi$ is well-defined. This argument also shows that Θ maps each nonperiodic element of $\mathsf{Fix}_{\mathsf{N}}(\Phi)$ to a nonperiodic element of $\mathsf{Fix}_{\mathsf{N}}(\Psi)$. Since nonperiodic points in $\mathsf{Fix}_{\mathsf{N}}(\Phi)$ are dense in $\mathsf{Fix}_{\mathsf{N}}(\Phi)$, $\Theta(\mathsf{Fix}_{\mathsf{N}}(\Phi)) \subset \mathsf{Fix}_{\mathsf{N}}(\Psi)$. Reversing the roles of ϕ and ψ and replacing θ with θ^{-1}, we see that $\Theta^{-1}(\mathsf{Fix}_{\mathsf{N}}(\Psi)) \subset \mathsf{Fix}_{\mathsf{N}}(\Phi)$, which completes the proof of (1). Note that if $\Psi = B_{\mathcal{P}}(\Phi)$ then for all $c \in F_n$,

$$B_{\mathcal{P}}(i_c \Phi i_c^{-1}) = i_{\Theta(c)} \Psi i_{\Theta(c)}^{-1}$$

because

$$\Theta(\mathsf{Fix}_{\mathsf{N}}(i_c \Phi i_c^{-1})) = \Theta(i_c \,\mathsf{Fix}_{\mathsf{N}}(\Phi)) = i_{\Theta(c)} \Theta(\mathsf{Fix}_{\mathsf{N}}(\Phi)) = i_{\Theta(c)} \,\mathsf{Fix}_{\mathsf{N}}(\Psi) = \mathsf{Fix}_{\mathsf{N}}(i_{\Theta(c)} \Psi i_{\Theta(c)}^{-1}).$$

For (2), suppose that $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$, that $\Phi \in \mathcal{P}(\phi)$ fixes a and that $\Psi = B_{\mathcal{P}}(\Phi)$. Define

$$B_{\mathsf{SA}}([\Phi, a]) = [B_{\mathcal{P}}(\Phi), \Theta(a)] = [\Psi, \Theta(a)].$$

Then for all $c \in F_n$,

$$B_{\mathsf{SA}}([\Phi, a]^c) = B_{\mathsf{SA}}([i_c \Phi i_c^{-1}, i_c(a)]) = [i_{\Theta(c)} \Psi i_{\Theta(c)}^{-1}, i_{\Theta(c)}(\Theta(a))] = [\Psi, a]^{\Theta(c)},$$

so B_{SA} is well defined. By symmetry, B_{SA} is a bijection. If Θ is replaced by $i_b\Theta$ for some $b \in F_n$ then $(\Psi, \Theta(a))$ is replaced by $(i_b \Psi i_b^{-1}, i_b \Theta(a)) = b \cdot (\Psi, \Theta(a))$. This shows that B_{SA} is independent of the choice of Θ. It remains to show that $\phi^\theta = \psi$ if and only if B_{SA} preserves twist coordinates.

Let $\upsilon = \phi^\theta$. By Lemmas 3.16 and 4.17, conjugation by Θ induces

- a bijection $B'_{\mathcal{P}} \colon \mathcal{P}(\phi) \to \mathcal{P}(\upsilon)$ defined by $\Phi \mapsto \Theta\Phi\Theta^{-1}$ and satisfying $\mathsf{Fix}_{\mathsf{N}}(B'_{\mathcal{P}}(\Phi)) = \Theta\, \mathsf{Fix}_{\mathsf{N}}(\Phi)$,
- a bijection $B'_{\mathsf{SA}} \colon \mathsf{SA}(\phi) \to \mathsf{SA}(\upsilon)$ defined by $[\Phi, a] \mapsto [\Theta\Phi\Theta^{-1}, \Theta(a)]$ that preserves twist coordinates.

The bijections $B''_{\mathcal{P}} = B_{\mathcal{P}} {B'}_{\mathcal{P}}^{-1} \colon \mathcal{P}(\upsilon) \to \mathcal{P}(\psi)$ and $B''_{\mathsf{SA}} = B_{\mathsf{SA}} {B'}_{\mathsf{SA}}^{-1} \colon \mathsf{SA}(\upsilon) \to \mathsf{SA}(\psi)$ satisfy:

(a) $\mathsf{Fix}_{\mathsf{N}}(B''_{\mathcal{P}}(\Upsilon)) = \mathsf{Fix}_{\mathsf{N}}(\Upsilon)$ for all $\Upsilon \in \mathcal{P}(\upsilon)$.

(b) B''_{SA} preserves twist coordinates if and only if B_{SA} does.

Applying (b), it suffices to show that $\phi^\theta = \psi$ if and only if B''_{SA} preserves twist coordinates.

Suppose that $[b] \in \mathcal{A}_{\mathrm{or}}(\upsilon)$, that $b \in \mathsf{Fix}(\Upsilon)$ and that $\Upsilon, i_{b^d}\Upsilon \in \mathcal{P}(\upsilon)$. Let $a = \Theta^{-1}(b)$ and $\Phi = \Theta^{-1}\Upsilon\Theta$. Then

$$B''_{\mathsf{SA}}[\Upsilon, b] = B_{\mathsf{SA}}[\Phi, a] = [B_{\mathcal{P}}(\Phi), b],$$

and likewise

$$B''_{\mathsf{SA}}[i_b^d \Upsilon, b] = B_{\mathsf{SA}}[i_a^d \Phi, a] = [B_{\mathcal{P}}(i_a^d \Phi), b].$$

By definition, the twist coordinate for $[i_{b^d}\Upsilon, b]$ and $[\Upsilon, b]$ is d. It follows that B''_{SA} preserves twist coordinates if and only if

$$B_{\mathcal{P}}(i_a^d \Phi) = i_{b^d} B_{\mathcal{P}}(\Phi).$$

Since

$$B_{\mathcal{P}}(i_a^d \Phi) = B''_{\mathcal{P}}(i_b^d \Upsilon) \quad \text{and} \quad B_{\mathcal{P}}(\Phi) = B''_{\mathcal{P}}(\Upsilon),$$

we conclude that B''_{SA} preserves twist coordinates if and only if

(c) $B''_{\mathcal{P}}(i_b^d \Upsilon) = i_b^d B''_{\mathcal{P}}(\Upsilon)$.

By the recognition theorem [Feighn and Handel 2011, Theorem 5.3], (a) and (c) are equivalent to $\upsilon = \psi$. □

5 Limit lines $\Omega(r) \subset \mathcal{B}$

Each point $P \in \partial F_n$ determines a closed set of lines; see eg [Feighn and Handel 2011, Section 2.4], where the closed set of lines is called the accumulation set of P. In this section we focus on the case that $P \in \mathcal{R}(\phi)$ and analyze these lines using CTs.

Definition 5.1 For each $r \in \partial F_n/F_n$, we define the set $\Omega(r) \subset \mathcal{B}$ of *limit lines of* r as follows. Choose a lift $\tilde{r} \in \partial F_n$, a marked graph K and a ray $\tilde{R} \subset \tilde{K}$ with terminal end $\tilde{r}$. Let $R \subset K$ be the projected image of $\tilde{R}$. Then $L \in \Omega(r)$ (thought of as a line in K) if and only if the following equivalent conditions are satisfied.

(1) Each finite subpath of L occurs as a subpath of R.

(2) For each lift $\tilde{L} \subset \tilde{K}$ of $L \subset K$ there are translates $\tilde{R}_j$ of $\tilde{R}$ such that the initial endpoints of $\tilde{R}_j$ converge to the initial endpoint of $\tilde{L}$ and the terminal endpoints of $\tilde{R}_j$ converge to the terminal endpoint of $\tilde{L}$.

Let $\Omega_{\mathsf{NP}}(r)$ be the set of nonperiodic elements of $\Omega(r)$.

Lemma 5.2 *$\Omega(r)$ and $\Omega_{\mathsf{NP}}(r)$ are well-defined. Moreover, for each $\theta \in \mathsf{Out}(F_n)$, $\theta(\Omega(r)) = \Omega(\theta(r))$ and $\theta(\Omega_{\mathsf{NP}}(r)) = \Omega_{\mathsf{NP}}(\theta(r))$.*

Proof If R' is another ray with terminal end r, then R and R' have a common terminal subray R''. Let $R = \alpha R''$ and $R' = \alpha' R''$. Given a finite subpath $\tau_2 \subset K$ of a line ℓ, extend it to a finite subpath $\tau_1\tau_2 \subset K$ of ℓ, where τ_1 is longer than both α and α'. If $\tau_1\tau_2$ occurs in R then τ_2 occurs in R''. Since τ_2 was arbitrary, every finite subpath of ℓ occur in R if and only if every finite subpath of ℓ occurs in R''. The same holds for R' and R''. This proves that $\Omega(r)$ is independent of the choice of R. Independence of the choice of $\tilde{R}$ is obvious, as is the equivalence of (1) and (2).

Suppose that K' is another marked graph and that $g\colon K \to K'$ is a homotopy equivalence that preserves markings and so represents the identity outer automorphism. Let $\tilde{g}\colon \tilde{K} \to \tilde{K}'$ be a lift of g. If $\tilde{L} \subset \tilde{K}$ is a lift of L and $\tilde{R}_j \subset \tilde{K}$ is a sequence of translates of ray $\tilde{R}$ such that the initial and terminal endpoints of $\tilde{R}_j$ converge to those of $\tilde{L}$, then the same is true of $\tilde{L}' = \tilde{g}_\#(\tilde{L}) \subset \tilde{K}'$ and $\tilde{R}'_j = \tilde{g}_\#(\tilde{R}_j) \subset \tilde{K}'$. This proves that $\Omega(r)$ is independent of the choice of K.

For the moreover statement, choose a homotopy equivalence $h\colon K \to K$ that represents θ and lifts $\tilde{L} \subset \tilde{K}$ and $\tilde{h}\colon \tilde{K} \to \tilde{K}$. If $\tilde{R}_j \subset \tilde{K}$ is a sequence of translates of $\tilde{R}$ whose initial and terminal endpoints converge to those of $\tilde{L}$, then the initial and terminal endpoints of $\tilde{h}_\#(\tilde{R}_j)$ converge to those of $\tilde{h}_\#(\tilde{L}) \subset \tilde{K}$. This proves that $\theta(\Omega_{\mathsf{NP}}(r)) \subset \Omega_{\mathsf{NP}}(\theta(r))$. The reverse inclusion follows by symmetry. □

We now specialize to $r \in \mathcal{R}(\phi)$.

Notation 5.3 For $\phi \in \mathsf{UPG}(F_n)$, let

$$\Omega(\phi) = \bigcup_{r \in \mathcal{R}(\phi)} \Omega(r) \quad \text{and} \quad \Omega_{\mathsf{NP}}(\phi) = \bigcup_{r \in \mathcal{R}(\phi)} \Omega_{\mathsf{NP}}(r).$$

As an immediate consequence of Lemma 3.16 and the moreover statement of Lemma 5.2 we have:

Corollary 5.4 *Suppose that* $\theta \in \mathsf{Out}(F_n)$ *and that* $\psi = \theta\phi\theta^{-1} \in \mathsf{UPG}(F_n)$*. Then* $\theta(\Omega(\phi)) = \Omega(\psi)$ *and* $\theta(\Omega_{\mathsf{NP}}(\phi)) = \Omega_{\mathsf{NP}}(\psi)$*.* □

For the remainder of the section we assume that $f: G \to G$ *is a* CT *representing* $\phi \in \mathsf{UPG}(F_n)$*.* Our goal is to describe $\Omega(\phi)$ and $\Omega_{\mathsf{NP}}(\phi)$ in terms of $f: G \to G$. See in particular Corollary 5.17.

One advantage of working in a CT is that we can work with finite paths and not just with lines and rays.

Definition 5.5 Given a path $\sigma \subset G$, we say that a line $L \subset G$ is contained in the *accumulation set* $\mathsf{Acc}(\sigma)$ *of* σ *with respect to* f if every finite subpath of L occurs as a subpath of $f_\#^k(\sigma)$ for arbitrarily large k.

Notation 5.6 For each twist path w, we write w^∞ for both the ray that is an infinite concatenation of copies of w and the line that is a bi-infinite concatenation of copies of w, using context to distinguish between the two. We use either $\overline{w}^\infty$ or $w^{-\infty}$ for the ray or line obtained from w^∞ by reversing orientation on w.

Examples 5.7 (1) If σ is a Nielsen path, then $\mathsf{Acc}(\sigma) = \varnothing$.

(2) Suppose that $E \in \mathsf{Lin}(f)$ and $f(E) = Ew^d$.
  (a) If $d > 0$ then $\mathsf{Acc}(E) = \{w^\infty\}$ and $\mathsf{Acc}(\bar{E}) = \{\overline{w}^\infty\}$.
  (b) If $d < 0$ then $\mathsf{Acc}(E) = \{\overline{w}^\infty\}$ and $\mathsf{Acc}(\bar{E}) = \{w^\infty\}$.

(3) If $E_i, E_j \in \mathsf{Lin}(f)$ satisfy $f(E_i) = E_i w^{d_i}$ and $f(E_j) = E_j w^{d_j}$ for $d_i \neq d_j$ then for all $p \in \mathbb{Z}$, $\mathsf{Acc}(E_i w^p \bar{E}_j) = \{w^\infty\}$ if $d_i > d_j$, and $\mathsf{Acc}(E_i w^p \bar{E}_j) = \{\overline{w}^\infty\}$ if $d_i < d_j$.

Recall from Lemma 3.26 that there is a bijection between $\mathcal{R}(\phi)$ and the set $\mathcal{E}_f$ of nonfixed nonlinear edges of G and that if $r \in \mathcal{R}(\phi)$ corresponds to $E \in \mathcal{E}_f$ then the eigenray $R_E = E \cdot u_E \cdot [f(u_E)] \cdot [f^2(u_E)] \cdot \ldots \subset G$ has terminal end r. Thus, a line $L \subset G$ is an element of $\Omega(r)$ if and only if each finite subpath of L occurs as a subpath of R_E.

Limit lines of eigenrays are connected to accumulation sets as follows.

Lemma 5.8 *If* $r \in \mathcal{R}(\phi)$ *corresponds to* $E \in \mathcal{E}_f$*, and* $f(E) = E \cdot u_E$*, then*

$$\Omega(r) = \mathsf{Acc}(E) = \mathsf{Acc}(u_E \cdot f_\#(u_E)) = \mathsf{Acc}\big(u_E \cdot f_\#(u_E) \cdot \ldots \cdot f_\#^k(u_E)\big)$$

for any $k \geq 1$*.*

Proof The first equality is an immediate consequence of the definitions and the fact that $E \subset f(E) \subset f_\#^2(E) \subset \cdots$ is an increasing sequence whose union is R_E. Likewise,

$$\mathsf{Acc}(u_E \cdot f_\#(u_E)) \subset \mathsf{Acc}(u_E \cdot f_\#(u_E) \cdot \ldots \cdot f_\#^k(u_E)) \subset \mathsf{Acc}(f_\#^{k+1}(E)) = \mathsf{Acc}(E)$$

is immediate. It therefore suffices to show that $\Omega(r) \subset \mathsf{Acc}(u_E \cdot f_\#(u_E))$.

If $L \in \Omega(r)$ then every finite subpath σ of L occurs as a subpath of every subray of R_E. Since the length of $f_\#^k(u_E)$ tends to infinity with k, each occurrence of σ that is sufficiently far away from the initial endpoint of R_E is contained in some $f_\#^k(u_E) \cdot f_\#^{k+1}(u_E) = f_\#^k(u_E \cdot f_\#(u_E))$. As the occurrence of σ moves farther down the ray, $k \to \infty$. □

Notation 5.9 Define a partial order on the set $\mathcal{E}_f \cup \mathcal{E}_f^{-1}$ by $E_1 \gg E_2$ if $E_1 \neq E_2$ and if, for some $k \geq 0$, E_2 is crossed by $f_\#^k(E_1)$ and so by Lemma 3.21 is a term in the complete splitting of $f_\#^k(E_1)$. (In Notation 6.1 we define a partial order $>$ on $\mathcal{E}_f$ that does not distinguish between E and $\bar{E}$.)

As an immediate consequence of the definition, we have:

Lemma 5.10 *If $E, E' \in \mathcal{E}_f \cup \mathcal{E}_f^{-1}$ and $E \gg E'$, then the height of E' is less than the height of E, and $\mathsf{Acc}(E') \subset \mathsf{Acc}(E)$.* □

The terms μ_i in the complete splitting of $u_E \cdot f_\#(u_E)$ are Nielsen paths, exceptional paths and single edges with height strictly less than that of E. Each $\mathsf{Acc}(\mu_i)$ is a subset of $\mathsf{Acc}(E) = \Omega(r)$. If $\mu_i \in \mathcal{E}_f \cup \mathcal{E}_f^{-1}$ then $\mathsf{Acc}(\mu_i)$ can be understood inductively. The remaining $\mathsf{Acc}(\mu_i)$ are given in Examples 5.7. The work in identifying $\Omega(r) = \mathsf{Acc}(u_E \cdot f_\#(u_E))$ is to determine what additional lines must be added to $\bigcup \mathsf{Acc}(\mu_i)$.

Notation 5.11 For a path $\alpha \subset G$, we say that $f_\#^k(\alpha)$ *converges to a ray* $R \subset G$ if for all m there exists K such that the initial m-length segments of $f_\#^k(\alpha)$ and of R are equal for all $k \geq K$. Note that R is necessarily unique and $f_\#$-invariant. We sometimes write $R = f_\#^\infty(\alpha)$.

Examples 5.12 (1) Suppose that $E \in \mathsf{Lin}(f)$ and that $f(E) = Ew^d$.

(a) If $d > 0$ then $f_\#^k(E)$ converges to Ew^∞ and $f_\#^k(\bar{E})$ converges to $\bar{w}^\infty$.

(b) If $d < 0$ then $f_\#^k(E)$ converges to $E\bar{w}^\infty$ and $f_\#^k(\bar{E})$ converges to w^∞.

(2) If $E \in \mathcal{E}_f$ then $f_\#^k(E)$ converges to R_E.

(3) If $E_i, E_j \in \mathsf{Lin}(f)$ satisfy $f(E_i) = E_i w^{d_i}$ and $f(E_j) = E_j w^{d_j}$ for $d_i \neq d_j$ then for all $p \in \mathbb{Z}$, $f_\#^k(E_i w^p \bar{E}_j)$ converges to $E_i w^\infty$ if $d_i > d_j$ and to $E_i \bar{w}^\infty$ if $d_i < d_j$.

Notation 5.13 If $E \in \mathcal{E}_f$, then the first growing term of $f(\bar{E})$ has height less than that of E. It follows that there exists $M > 1$ such that if σ_i is a growing term in the complete splitting of a path σ and if $m \geq M$, then the first growing term in the complete splitting of $f_\#^m(\sigma_i)$ is not an element of $\mathcal{E}_f^{-1}$, and the last growing term in the complete splitting of $f_\#^m(\sigma_i)$ is not an element of $\mathcal{E}_f$. We refer to M as the *stabilization constant* for f.

Lemma 5.14 *Let M be the stabilization constant for f. If σ is a completely split growing path, then $f_\#^k(\sigma)$ converges to a ray $f_\#^\infty(\sigma) = \rho R$, where*

(1) *ρ is a (possibly trivial) Nielsen path and one of the following holds:*

(a) *$R = R_E$ for some $E \in \mathcal{E}_f$.*

(b) *$R = Ew^{\pm\infty}$ for some $E \in \mathsf{Lin}_w(f)$.*

(c) *$R = w^{\pm\infty}$ for some twist path w.*

(2) *If* $\sigma = \mu_1 \cdot \nu_1 \cdot \mu_2 \cdot \ldots$ *is the coarsening of the complete splitting of* σ *into maximal (possibly trivial) Nielsen paths* μ_i *and single growing terms* ν_i, *then* $f_\#^\infty(\sigma) = \mu_1 f_\#^\infty(\nu_1)$.

(3) *In case* (1c) *there exists* $E \in \mathsf{Lin}_w(f)$ *and a smallest* $k_\sigma \le M$ *such that the first growing term in the coarsened complete splitting of* $f_\#^k(\sigma)$ *is* $\bar{E}$ *for all* $k \ge k_\sigma$. *Moreover, if the first growing term in the coarsened complete splitting of* σ *is not an edge in* $\mathcal{E}_f^{-1}$ *then* $k_\sigma = 1$.

Proof There is no loss in replacing σ with its first growing term. The only case that does not follow from Examples 5.12 is that $\sigma = \bar{E} \in \mathcal{E}_f^{-1}$. This case follows from the definition of M and the obvious induction argument. □

Remark 5.15 The rays in Lemma 5.14 are finitely determined: in case (a) R is determined by the edge E, in case (b) R is determined by E, w and a choice of $\pm$, and in case (c) R is determined by w and a choice of $\pm$. From this data one can write down any finite initial subpath of R.

Lemma 5.16 *Suppose that* $\sigma \subset G$ *is a completely split path and that* $\sigma = \alpha \cdot \beta$ *is a coarsening of the complete splitting in which both* α *and* β *are growing. Let* $R^- = f_\#^\infty(\bar{\alpha})$, *let* $R^+ = f_\#^\infty(\beta)$ *and let* $\ell = (R^-)^{-1} R^+$. *Then* $\mathsf{Acc}(\sigma) = \mathsf{Acc}(\alpha) \cup \mathsf{Acc}(\beta) \cup \{\ell\}$.

Proof The inclusion $\mathsf{Acc}(\alpha) \cup \mathsf{Acc}(\beta) \subset \mathsf{Acc}(\sigma)$ follows from the fact that α and β occur as concatenation of terms in a splitting of σ. It is an immediate consequence of the definitions that $\ell \in \mathsf{Acc}(\gamma)$. It therefore suffices to assume that $L \in \mathsf{Acc}(\sigma)$ is not contained in $\mathsf{Acc}(\alpha) \cup \mathsf{Acc}(\beta)$ and prove that $L = \ell$.

Choose a finite subpath L_1 of L and $K > 0$ so that L_1 does not occur as a subpath of $f_\#^k(\alpha)$ or of $f_\#^k(\beta)$ for $k \ge K$. Extend L_1 to an increasing sequence $L_1 \subset L_2 \subset \cdots$ of finite subpaths of L whose union is L. For each $j \ge 1$, let C_j be the length of L_j. There exist arbitrarily large k so that L_j includes as a subpath of $f_\#^k(\sigma) = f_\#^k(\alpha) \cdot f_\#^k(\beta)$. The induced inclusion of L_1 in $f_\#^k(\sigma)$ must intersect both $f_\#^k(\alpha)$ and $f_\#^k(\beta)$ and so L_j is included as a subpath of the concatenation of the terminal segment of $f_\#^k(\alpha)$ of length C_j with the initial segment of $f_\#^k(\beta)$ of length C_j. If k is sufficiently large then the length C_j initial segments of R^- and of $f_\#^k(\bar{\alpha})$ agree and the length C_j initial segments of R^+ and of $f_\#^k(\beta)$ agree. Thus each L_j can be included as a subpath of ℓ. Since the induced inclusion of L_1 contains the juncture point between R^- and R^+, we may pass to a subsequence of L_j's and choose inclusions of L_j into ℓ so that induced inclusion of L_1 in ℓ is independent of j. It follows that if $i < j$ then the inclusion of L_i into ℓ is the restriction of the inclusion of L_j into ℓ and hence that there is a well-defined inclusion of L into ℓ. This inclusion is necessarily onto and so $L = \ell$. □

Corollary 5.17 *For each* $r \in \mathcal{R}(\phi)$:

(1) *Each* $L \in \Omega(r)$ *decomposes as* $L = (R^-)^{-1} \rho R^+$ *where* ρ *is a (possibly trivial) Nielsen path and* R^+ *and* R^- *satisfy* (1a), (1b) *or* (1c) *of* Lemma 5.14. *In particular, each* L *is* ϕ*-invariant and is finitely determined in the sense of* Remark 5.15, *and each periodic* L *equals* $w^{\pm\infty}$ *for some twist path* w.

(2) *$\Omega(r)$ is a finite set and the finite data that determines each of its elements can be read off from $f: G \to G$.*
(3) $\Omega_{\mathsf{NP}}(r) \neq \varnothing$.
(4) *For each $L \in \Omega(r)$ and each lift $\widetilde{L}$, there exists $\Phi \in \mathcal{P}(\phi)$ such that $\partial_- \widetilde{L}, \partial_+ \widetilde{L} \in \mathsf{Fix}_{\mathsf{N}}(\Phi)$. Equivalently, L lifts into $\Gamma(f)$.*

Proof By Lemma 5.8 we can replace $\Omega(r)$ with $\mathsf{Acc}(E)$, where $E \in \mathcal{E}_f$ corresponds to r. Lemma 5.8 also implies that $\mathsf{Acc}(E) = \mathsf{Acc}(u \cdot f_\#(u))$, where $f(E) = E \cdot u$. Let

$$u \cdot f_\#(u) = \rho_0 \cdot \sigma_1 \cdot \rho_1 \cdot \sigma_2 \ldots \sigma_q \cdot \rho_q$$

be a coarsening of the complete splitting of $u \cdot f_\#(u)$ so that each σ_i is a single growing term and so that the ρ_i are (possibly trivial) Nielsen paths. For $1 \le i \le q-1$, let $R_i^- = f_\#^\infty(\bar{\sigma}_i)$ and for $2 \le i \le q$, let $R_i^+ = f_\#^\infty(\sigma_i)$. For $1 \le i \le q-1$, define $\ell_i = (R_i^-)^{-1} \rho_i R_{i+1}^+$. Lemma 5.16 and the obvious induction argument imply that

$$\Omega(r) = \mathsf{Acc}(u \cdot f_\#(u)) = \mathsf{Acc}(\sigma_1) \cup \ell_1 \cup \mathsf{Acc}(\sigma_2) \cup \cdots \cup \ell_{q-1} \cup \mathsf{Acc}(\sigma_q).$$

Lemma 5.14(1) implies that each ℓ_i satisfies (1). If σ_i is linear then $\mathsf{Acc}(\sigma_i) = w^{\pm\infty}$ for some twist path w by Examples 5.7. The remaining σ_i have the form E' or $\bar{E}'$ for some $E' \in \mathcal{E}_f$ with height less than that of E. Downward induction on the height of E completes the proof of (1) and (2).

We now turn to (3), assuming at first that $q > 2$. If σ_2 is exceptional or an element of $\mathcal{E}_f \cup \mathsf{Lin}(f)$ then ℓ_1 is nonperiodic. Otherwise, $\bar{\sigma}_2 \in \mathcal{E}_f \cup \mathsf{Lin}(f)$ and ℓ_2 is nonperiodic. Both of these statements follow from Lemma 5.14. If $q = 2$ then σ_1 is linear. One easily checks that ℓ_1 is nonperiodic in the various cases that can occur. For example if $\sigma_1 = E_1 \in \mathsf{Lin}(f)$ then $\sigma_2 = E_1$ and $\ell_1 = \bar{w}^{\pm\infty} \rho_1 E_1 w^{\pm\infty}$. The remaining cases are left to the reader.

The equivalence of the two conditions in (4) follows from Lemma 4.1. To prove that L lifts into $\Gamma(f)$, we make use of the fact that each vertex in G lifts uniquely to $\Gamma^0(f)$ and the fact that each Nielsen path in G lifts uniquely into $\Gamma(f)$ with one, and hence both, endpoints in $\Gamma^0(f)$. These facts follow immediately from the construction of $\Gamma(f)$ and the fact that every Nielsen path is a concatenation of fixed edges and (necessarily closed) indivisible Nielsen paths. Given these facts, we may assume that $L = (R^-)^{-1} \rho R^+$ is not a concatenation of Nielsen paths and hence that the initial edge E_j of either R^- or R^+ is an element of $\mathsf{Lin}(f) \cup \mathcal{E}_f$. The two cases are symmetric so we may assume that E_j is the initial edge of R^-. Let $E'_j \subset \Gamma(f)$ be the unique lift of E_j with initial vertex $v' \in \Gamma^0(f)$ and then extend this to a lift of R^- into $\Gamma(f)$. The Nielsen path ρ lifts to a path $\rho' \subset \Gamma^0(f)$ with initial vertex v' and terminal vertex, say w'. If R^+ is a concatenation of Nielsen paths then it lifts into $\Gamma^0(f)$ with initial vertex w'. Otherwise we lift R^+ in the same way that we lifted R^-. □

Example 3.1 (continued) Recall $R_q = q \cdot c \cdot cb \cdot cbba \cdot \ldots \cdot cbba \ldots ba^{k-1} \cdot \ldots$ and so $\Omega(r_q) = \{a^\infty R_c, a^\infty b a^\infty, a^\infty\}$ and $\Omega_{\mathsf{NP}}(r_q) = \{a^\infty R_c, a^\infty b a^\infty\}$.

6 Special free factor systems

6.1 A canonical collection of free factor systems

In this section, we define a canonical partial order $<$ on $\mathcal{R}(\phi)$ and then associate a nested sequence $\vec{\mathcal{F}}(\phi, <_T) = \mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_t$ of ϕ-invariant free factor systems to each total order $<_T$ on $\mathcal{R}(\phi)$ that extends $<$. The bottom free factor system $\mathcal{F}_0$ is the smallest free factor system that carries all conjugacy classes that grow at most linearly and is independent of $<_T$. The inclusions $\mathcal{F}_{i-1} \sqsubset \mathcal{F}_i$ are all one-edge extensions. The CTs that represent ϕ with filtrations that realize $\vec{\mathcal{F}}(\phi, <_T)$ are easier to work with than generic CTs; see Lemma 6.9.

Notation 6.1 Suppose that $f : G \to G$ is a CT representing ϕ and that E_1 and E_2 are distinct elements of $\mathcal{E}_f$. If E_1 or $\bar{E}_1$ is a term of the complete splitting of $f_\#^k(E_2)$ for some $k \geq 1$, then we write $E_1 < E_2$. Lemma 3.21 implies that $<$ is a partial order on $\mathcal{E}_f$. If $E_1 < E_2$ are consecutive elements in the partial order then we write $E_1 <_c E_2$. Note that if we define $E_1 <' E_2$ to mean E_1 or $\bar{E}_1$ is a term of the complete splitting of $f(E_2)$ then $<$ is the partial order determined from $<'$ by extending transitively. Thus $<$ can be computed.

If $r_1, r_2 \in \mathcal{R}(\phi)$ and r_1 is an end of some element of $\Omega_{\mathsf{NP}}(r_2)$ then we write $r_1 < r_2$. Lemma 6.2 below implies that $<$ defines a partial order on $\mathcal{R}(\phi)$. If $r_1 < r_2$ are consecutive elements in the partial order then we write $r_1 <_c r_2$.

Example 3.1 (continued) In our example, the only relation is $r_c < r_q$.

Recall from Lemma 3.26 that the map that sends E to the end of R_E defines a bijection between $\mathcal{E}_f$ and $\mathcal{R}(\phi)$.

Lemma 6.2 *For any* CT $f : G \to G$, *the bijection between* $\mathcal{E}_f$ *and* $\mathcal{R}(\phi)$ *preserves* $<$.

Proof Suppose that $E_1, E_2 \in \mathcal{E}_f$ correspond to $r_1, r_2 \in \mathcal{R}(\phi)$, respectively.

If $E_1 < E_2$ and $f(E_2) = E_2 \cdot u_2$, then E_1 or $\bar{E}_1$ is a term in the complete splitting of $f_\#^k(u_2)$ for some, and hence all sufficiently large, k. By Lemma 5.8, there exists a completely split path γ such that $\Omega(r_2) = \mathsf{Acc}(\gamma)$ and such that the complete splitting of γ has a coarsening $\gamma = \gamma_1 \cdot \gamma_2 \cdot \gamma_3$ into three growing terms with γ_2 equal to either $\bar{E}_1$ or E_1. Lemma 5.16 therefore implies that R_{E_1} is a terminal ray of L or L^{-1} for some $L \in \Omega(r_2)$. Thus $r_1 < r_2$.

If $r_1 < r_2$ then R_{E_1} is a terminal ray of L or L^{-1} for some $L \in \Omega(r_2)$ by Corollary 5.17(1). It follows that $f_\#^k(E_2)$ crosses E_1 or $\bar{E}_1$ for all sufficiently large k. Lemma 3.21 implies that $E_1 < E_2$. □

Lemma 6.3 *If* $\psi = \theta\phi\theta^{-1}$ *then the bijection* $\mathcal{R}(\phi) \to \mathcal{R}(\psi)$ *induced by* θ (*see* Lemma 3.16) *preserves partial orders.*

Proof By Lemmas 3.16 and 5.4, we have $\theta(\mathcal{R}(\phi)) = \mathcal{R}(\psi)$ and $\theta(\Omega_{\mathsf{NP}}(r)) = \Omega_{\mathsf{NP}}(\theta(r))$ for each $r \in \mathcal{R}(\phi)$. The fact that θ preserves the partial order now follows from the definition of the partial order. □

Notation 6.4 Extend the partial order $<$ on $\mathcal{R}(\phi)$ to a total order $<_T$ and write $\mathcal{R}(\phi) = \{r_1, \dots, r_s\}$, where the elements are listed in increasing order. Given a CT $f : G \to G$ representing ϕ, transfer the total order $<_T$ on $\mathcal{R}(\phi)$ to a total order (also called) $<_T$ on $\mathcal{E}_f = \{E_1, \dots, E_s\}$ using the bijection between $\mathcal{E}_f$ and $\mathcal{R}(\phi)$ given in Lemma 3.26.

Recall from Section 4.1 that each component C of the eigengraph $\Gamma(f)$ is constructed from a component C_0 of $\mathsf{Fix}(f)$ by first adding "lollipops", one for each $E \in \mathsf{Lin}(f)$ with initial vertex in C_0, to form C_1, and then adding rays labeled R_E, one for each $E \in \mathcal{E}_f$ with initial vertex in C_0. Each $E \in \mathcal{E}_f$ contributes exactly one ray to $\Gamma(f)$ and we identify that ray with the eigenray R_E; it is the unique lift of R_E to $\Gamma(f)$. Each $E \in \mathsf{Lin}(f)$ contributes exactly one lollipop to $\Gamma(f)$. Note that C is contractible if and only if C_1 is contractible if and only if C_0 is contractible and there are no $E \in \mathsf{Lin}(f)$ with initial vertex in C_0. In this contractible case, C is obtained from a (possibly trivial) tree in $\mathsf{Fix}(f)$ by adding eigenrays and we single out the ray $R_E \subset C$ whose associated edge E is lowest with respect to $<_T$. These edges define subsets $\mathcal{R}^*(\phi) \subset \mathcal{R}(\phi)$ and $\mathcal{E}^*_f \subset \mathcal{E}_f$ that correspond under the bijection between $\mathcal{R}(\phi)$ and $\mathcal{E}_f$.

Definition 6.5 A conjugacy class $[a]$ *grows at most linearly* under iteration by ϕ if for some, and hence every, set of generators there is a linear function P such that word length of $\phi^k([a])$ with respect to those generators is bounded by $P(k)$. If $f : G \to G$ represents ϕ, then word length of $\phi^k([a])$ can be replaced by edge length of $f^k_\#(\sigma)$ in G, where $\sigma \subset G$ is the circuit representing $[a]$. The *linear growth free factor system* $\mathcal{F}_0(\phi)$ is the minimal free factor system that carries all conjugacy classes that grow at most linearly under iteration by ϕ.

Lemma 6.6 *Suppose that* $f : G \to G$ *is a* CT *representing* ϕ *and that* $<_T$ *and* $\mathcal{E}_f = \{E_1, \dots, E_s\}$ *are as in* Notation 6.4. *Let* $K_0 \subset G$ *be the subgraph consisting of all fixed and linear edges for* $f : G \to G$. *For* $1 \le j \le s$, *inductively define* $K_j = K_{j-1} \cup E_j$. *Then:*

(1) $\mathcal{F}_0(\phi) = \mathcal{F}(K_0, G)$ *(as defined at the beginning of* Section 3.3*).*

(2) *Each* K_j *is* f*-invariant.*

(3) *If* $E_j \in \mathcal{E}^*_f$, *then* $\mathcal{F}(K_j, G) = \mathcal{F}(K_{j-1}, G)$; *otherwise* $\mathcal{F}(K_{j-1}, G) \sqsubset \mathcal{F}(K_j, G)$ *is a proper one-edge extension.*

Proof In proving (1), we work with circuits $\sigma \subset G$ and edge length in G rather than conjugacy classes $[a]$ and word length with respect to a set of generators of F_n. If E is an edge of K_0, then $f(E) = E \cdot u$ for some (possibly trivial) closed Nielsen path u. Lemma 3.21 implies that $u \subset K_0$ and hence that K_0 is f-invariant. Each circuit in K_0 grows at most linearly under iteration by $f_\#$ since every edge in K_0 does. Thus $\mathcal{F}(K_0, G) \sqsubset \mathcal{F}_0(\phi)$.

After replacing σ with $f^m_\#(\sigma)$ for some $m \ge 0$, we may assume by [Feighn and Handel 2011, Lemma 4.25] that σ is completely split. Lemma 3.21 implies that if σ is not contained in K_0 then, up to reversal of orientation, some term in the complete splitting of σ is an edge $E \in \mathcal{E}_f$. In this case, σ grows at least as fast as E does. If $f(E) = E \cdot u$ then the length of $f_\#^k(u)$ goes to infinity with k and so the length of $f_\#^k(E)$ grows faster than any linear function. This proves that K_0 contains every circuit that grows at most linearly so $\mathcal{F}_0(\phi) \sqsubset \mathcal{F}(K_0, G)$. This completes the proof of (1).

For the remainder of the proof we may assume that $j \geq 1$. For $E_j \in \mathcal{E}_f$, the terms in the complete splitting of $f(E_j)$, other than E_j itself, are exceptional paths, Nielsen paths and single edges E_i or $\bar{E}_i$ that are either linear or satisfy $E_i < E_j$. Lemma 3.21 implies that the exceptional paths and Nielsen paths are contained in K_0. The single edge terms other than E_j are contained in K_{j-1} by construction. Thus $f(E_j) \subset K_j$ and K_j is f-invariant. This proves (2).

The terminal endpoint of each $E_j \in \mathcal{E}_f$ is contained in a noncontractible component of K_{j-1} because $f(E_j) = E_j u_j$ for a nontrivial closed path $u_j \subset K_{j-1}$. If $E_j \in \mathcal{E}_f^*$ with initial vertex v_j then the component of K_{j-1} that contains v_j is a contractible component of $\mathsf{Fix}(f)$. In this case every line in K_j is contained in K_{j-1} so $\mathcal{F}(K_j, G) = \mathcal{F}(K_{j-1}, G)$. Otherwise, v_j is contained in a noncontractible component of K_{j-1} so $\mathcal{F}(K_{j-1}, G) \sqsubset \mathcal{F}(K_j, G)$ is a proper inclusion. Obviously K_j is obtained from K_{j-1} by adding a single edge. □

Recall from Lemma 4.1 that the set of lines that lift to $\Gamma(f)$ is independent of the choice of CT $f : G \to G$ representing ϕ. The next lemma shows that the $\mathcal{F}(K_j, G)$ defined in Lemma 6.6 depend only on ϕ and $<_T$ and not on the choice of CT $f : G \to G$.

Lemma 6.7 *Continue with the notation of Lemma 6.6. For each $r_j \notin \mathcal{R}^*(\phi)$, there exists at least one line $L(r_j)$ that lifts to $\Gamma(f)$, whose terminal end is r_j and whose initial end is not r_l for any $l \geq j$. Moreover, for any such choice of lines, $\mathcal{F}(K_j, G)$ is the smallest free factor system that contains $\mathcal{F}_0(\phi)$ and carries $\{L(r_l) \mid l \leq j$ and $r_l \notin \mathcal{R}^*(\phi)\}$.*

Proof Let $C = C(j)$ be the component of $\Gamma(f)$ that contains R_{E_j}, let $C_0 \subset C_1 \subset C$ be as in Notation 6.4 and, for each $1 \leq q \leq s$, let $A_q \subset C$ be the union of C_1 with the rays R_{E_l} in C with $E_l \leq E_q$. By construction, and by Lemma 6.6, R_{E_l} is included in A_q if and only if $R_{E_l} \subset K_q$. Since $r_j \notin \mathcal{R}^*(\phi)$, either C_1 is noncontractible or A_{j-1} contains at least one ray R_{E_l}. In both cases, the ray $R_{E_j} \subset C$ extends by a ray in A_{j-1} to a line in A_j. The projection $L(r_j)$ of this line into K_j satisfies the conclusions of the main statement of the lemma.

The "moreover" part of the lemma is proved by induction on j, with the base case $j = 0$ following from Lemma 6.6(1). For the inductive case, let $\mathcal{F}'_j$ be the smallest free factor system that carries K_{j-1} and $L(r_j)$. Then $\mathcal{F}(K_{j-1}, G) \sqsubset \mathcal{F}'_j \sqsubset \mathcal{F}(K_j, G)$ with the first inclusion being proper and $\mathcal{F}'_j$ does not have more components than $\mathcal{F}(K_{j-1}, G)$. Lemma 3.4 implies that $\mathcal{F}'_j = \mathcal{F}(K_j, G)$. □

Notation 6.8 Let $K_0 \subset K_1 \subset \cdots \subset K_s = G$ be as in Lemma 6.6 and let $\vec{\mathcal{F}}(\phi, <_T) = \mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_t$ be the increasing sequence of distinct free factor systems determined by the K_j. (Equivalently, $\vec{\mathcal{F}}(\phi, <_T)$ is the sequence determined by those K_j with $r_j \notin \mathcal{R}^*(\phi)$.) We say that $\vec{\mathcal{F}}(\phi, <_T)$ is the *sequence of free factor systems determined by ϕ and $<_T$*. Lemma 6.7 justifies this description by showing that $\vec{\mathcal{F}}(\phi, <_T)$

depends only on ϕ and $<_T$. To simplify notation a bit, we write L_k for $L(r(j))$ where $r(j)$ is the k^{th}-lowest element of $\mathcal{R}(\phi) \setminus \mathcal{R}^*(\phi)$. Thus $\mathcal{F}_k$ is filled by $\mathcal{F}_0$ and $L_1, \dots, L_k$.

We sometimes refer to a nested sequence of free factor systems as a *chain*. A chain $\mathfrak{c} = (\mathcal{F}_0 \sqsubset \cdots \sqsubset \mathcal{F}_t)$ is *special for* ϕ if $\mathfrak{c} = \vec{\mathcal{F}}(\phi, <_T)$ for some extension $<_T$ of $<$ to a total order on $\mathcal{R}(\phi)$. A free factor system $\mathcal{F}$ is *special for* ϕ if $\mathcal{F}$ is an element of some special chain for ϕ. The set of special free factor systems for ϕ is denoted by $\mathfrak{L}(\phi)$. A free factor F or its conjugacy class is *special for* ϕ if $[F]$ is an element of some special free factor system for ϕ. A pair $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ of free factor systems is a *special one-edge extension for* ϕ if its appears as consecutive elements of some special chain for ϕ.

By applying the existence theorem for CTs given in [Feighn and Handel 2018, Theorem 1.1], we can choose a CT whose filtration realizes $\vec{\mathcal{F}}(\phi, <_T)$ for any given $<_T$. The following lemma shows that the case analysis for a CT with this property is simpler than that of a random CT.

Lemma 6.9 *Suppose that $\vec{\mathcal{F}}(\phi, <_T) = \mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_t$ and that $f : G \to G$ and $\varnothing = G_0 \subset G_1 \subset \cdots \subset G_N = G$ are a* CT *and filtration representing ϕ and realizing $\vec{\mathcal{F}}(\phi, <_T)$; ie for all $0 \le k \le t$ there is an f-invariant core subgraph G_{i_k} such that $\mathcal{F}_k = \mathcal{F}(G_{i_k}, G)$. Then $G_{i_k} \setminus G_{i_{k-1}}$ is a single topological arc A_k with both endpoints in $G_{i_{k-1}}$. Moreover, letting D_k be the element of $\mathcal{E}_f$ corresponding to $\partial_+ L_k \in \mathcal{R}(\phi)$ (as in Lemma 6.7), A_k can be oriented so that one of the following is satisfied*:

- [HH] $A_k = \bar{C}_k D_k$, *where* $C_k \in \mathcal{E}_f$.
- [LH] $A_k = \bar{C}_k D_k$, *where* $C_k \in \mathsf{Lin}(f)$.
- [H] $A_k = D_k$.

Proof By Lemma 6.6, each $\mathcal{F}_{j-1} \sqsubset \mathcal{F}_j$ is a one-edge extension. [Handel and Mosher 2020, Part II, Lemma 2.5] therefore implies that G_{i_k} is constructed from $G_{i_{k-1}}$ in one of three ways: add a single topological edge with both endpoints in $G_{i_{k-1}}$; add a single topological edge that forms a circuit that is disjoint from $G_{i_{k-1}}$; add an edge forming a disjoint circuit and then add an edge connecting that circuit to $G_{i_{k-1}}$. In the second and third cases the circuit is f-invariant in contradiction to the fact that K_0 contains all ϕ-invariant conjugacy classes. Thus G_{i_k} is obtained from $G_{i_{k-1}}$ by adding a single topological arc A_k with both endpoints in $G_{i_{k-1}}$.

The arc A_k consists of either one or two edges of G. Indeed, a "middle" edge cannot be fixed by the (Periodic Edge) property [Feighn and Handel 2011, Definition 4.7(5)] of a CT and cannot be nonfixed because in that case its terminal end would be contained in a core subgraph of $G_{i_{k-1}}$ by [Feighn and Handel 2011, Lemma 4.21]. The (Periodic Edge) property also implies that if A_k consists of two edges, then neither is fixed. To complete the proof, it suffices to show that A_k crosses D_k. We will do so by showing that D_k is not contained in $G_{i_{k-1}}$ and is contained in G_{i_k}.

A line $L \subset G$ that lifts to $\Gamma(f)$ but is not contained in K_0 either decomposes as the concatenation of a ray in K_0 and an eigenray $R_{E'}$, or decomposes as the concatenation of a finite path in K_0 and a pair of eigenrays $R_{E'}$ and $R_{E''}$. In the former case, each $E \in \mathcal{E}_f$ crossed by L satisfies $E \le E'$, and in the latter case each $E \in \mathcal{E}_f$ satisfies $E \le E'$ or $E \le E''$. It follows that every edge $E \in \mathcal{E}_f$ crossed by $\bigcup_{q=1}^{k-1} L_q$ satisfies $E <_T D_k$. Since K_0 and $L_1, \dots, L_{k-1}$ fill $\mathcal{F}_{k-1}$, we conclude that D_k is not contained in $G_{i_{k-1}}$. Since L_k lifts to $\Gamma(f)$ and $\partial_+ L_k = \partial R_{D_k}$, it follows that R_{D_k} is a terminal ray of L_k. In particular, L_k crosses D_k. Lemma 6.7 implies that $L_k \subset G_{i_k}$ and we are done. □

Lemma 6.10 *Let $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ be special for ϕ.*

1. *The types* HH, LH *or* H *of $\mathfrak{e}$ as in Lemma 6.9 are mutually exclusive and independent of the special chain $\vec{\mathcal{F}}(\phi, <_T)$ containing $\mathfrak{e}$ and the choice of* CT *$f : G \to G$ realizing $\vec{\mathcal{F}}(\phi, <_T)$.*
2. *Suppose that $\mathfrak{e}$ appears as consecutive elements $\mathcal{F}_{k-1} \sqsubset \mathcal{F}_k$ in $\vec{\mathcal{F}}(\phi, <_T)$, which is realized by the* CT *$f : G \to G$. Using terminology as in Lemma 6.9, say that $\mathfrak{e}$ is, respectively*, contractible, infinite cyclic, *or* large *depending on whether the component of the eigengraph $\Gamma_{f|\mathcal{F}_k}$ containing the eigenray R_{D_k} is contractible, has infinite cyclic fundamental group, or has fundamental group with rank at least two. The types* contractible, infinite cyclic, *or* large *of $\mathfrak{e}$ are mutually exclusive and independent of the choices of $\vec{\mathcal{F}}(\phi, <_T)$ and f.*

Proof (1) Suppose $\mathfrak{e} = (\mathcal{F}_{k-1} \sqsubset \mathcal{F}_k)$ in $\vec{\mathcal{F}}(\phi, <_T)$. The difference between the cardinality of $\mathcal{R}(\phi|\mathcal{F}_k)$ and the cardinality of $\mathcal{R}(\phi|\mathcal{F}_{k-1})$ is 2 in the [HH] case and 1 in the [LH] and [H] cases. In case [LH], either the number of axes for $\phi|\mathcal{F}_k$ is strictly larger than the number of axes for $\phi|\mathcal{F}_{k-1}$ or there is a common axis of $\phi|\mathcal{F}_k$ and $\phi|\mathcal{F}_{k-1}$ whose multiplicity in the former is strictly larger than in the latter. Neither of these happens in case [H].

(2) Here is an invariant description. Let $r \in \Delta := \mathcal{R}(\phi|\mathcal{F}_k) \setminus \mathcal{R}(\phi|\mathcal{F}_{k-1})$, for example we could take r to be determined by R_{D_k}. Either $\Delta = \{r\}$ or $\Delta = \{r, s\}$ and there is a $\phi|\mathcal{F}_k$-fixed line L whose ends represent r and s. Let $\tilde{r}$ be a lift of r to ∂F_n and let $\tilde{L} = [\tilde{r}, \tilde{s}]$ be a lift of L if $\Delta = \{r, s\}$. By definition, $\mathfrak{e}$ is contractible, infinite cyclic or large if and only if $\mathrm{Fix}(\Phi_{\tilde{r}})$ is trivial, infinite cyclic, or of rank at least two, where $\Phi_{\tilde{r}}$ is the unique representative of ϕ fixing $\tilde{r}$. We are done by noting that $\Phi_{\tilde{r}} = \Phi_{\tilde{s}}$ if $\Delta = \{r, s\}$. □

Example 3.1 (continued) If we extend the partial order $r_c < r_q$ on $\mathcal{R}(\phi)$ to the total order $r_c <_T r_d <_T r_e <_T r_q$ we get the special chain $\mathfrak{c}$ represented by the sequence of graphs in Figure 3. See the notation in the examples on pages 1700 and 1707.

Example 6.11 Consider the CT $f : G \to G$ given as follows: start with a rose with edges a and b. Define $f(a) = a$ and $f(b) = ba$. Add a new vertex v with adjacent edges c, d and define $f(c) = cb$ and $f(d) = db^2$. Add another new vertex v' with adjacent edges c' and d' with $f(c') = c'b^3$ and $f(d') = d'b^4$ finally add an f-fixed edge e with endpoints v and v'. The ϕ-fixed free factor system $\mathcal{F}$ represented by the complement of e in G is not in $\mathfrak{L}(\phi)$. Indeed, $\mathcal{F} \sqsubset \{F_n\}$ is 1-edge, but not of type H, HH or LH, contradicting Lemma 6.9.

Example 6.12 Suppose $f: G \to G$ is a CT containing a circle C with only one vertex x and such that x is the initial endpoint of an H-edge, the terminal endpoint of a linear edge in an LH extension (so that C is an axis), and there are no other edges containing x. Then $\Gamma(f)$ has no components of rank at least two containing an axis corresponding to C.

Lemma 6.13 *Referring to Notation 6.8, suppose $\mathcal{F}$, $[F]$, $\mathfrak{c}$ and $\mathfrak{e}$ are special for ϕ. If $\theta \in \mathsf{Out}(F_n)$, then*

- *$\theta(\mathcal{F}), \theta([F]), \theta(\mathfrak{c})$, and $\theta(\mathfrak{e})$ are special for ϕ^θ, and*
- *the types* H, HH *or* LH *and the types* contractible, infinite cyclic *or* large *of $\mathfrak{e}$ and $\theta(\mathfrak{e})$ are the same.*

Proof This is an immediate consequence of the fact (Lemma 6.3) that conjugation preserves partial orders and the invariant description of types given in the proof of Lemma 6.10. □

Definition 6.14 (added lines) Suppose that $\mathfrak{c}$ is a special chain for ϕ that is realized by $f: G \to G$ and that $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+) \in \mathfrak{c}$. Then $\mathcal{R}(\phi|\mathcal{F}^+) \setminus \mathcal{R}(\phi|\mathcal{F}^-)$ contains two elements if $\mathfrak{e}$ has type HH and one element otherwise. These elements are said to be *new* with respect to $\mathfrak{e}$. Similarly $\Gamma(f|\mathcal{F}^+)$ carries more lines than $\Gamma(f|\mathcal{F}^-)$. The set of *added lines with respect to* $\mathfrak{e}$, denoted by $\mathsf{L}_\mathfrak{e}(\phi)$, is a ϕ-invariant subset of these lines. In case $\mathfrak{e}$ is contractible, $\mathsf{L}_\mathfrak{e}(\phi)$ consists of all lines L in $\Gamma(f|\mathcal{F}^+)$ with $\partial_+ L$ new. If $\mathfrak{e}$ is not contractible then we also require that $\partial_- L$ is not in $\mathcal{R}(\phi)$. $\mathsf{L}_\mathfrak{e}(\phi)$ has an equivalent invariant description as follows. Set $\Phi := \Phi_{\tilde{r}^+}|F^+$, where r^+ is new, $\tilde{r}^+ \in \partial F^+ \subset \partial F_n$, and $[F^+]$ is the component of $\mathcal{F}^+$ carrying r^+. Define $\mathsf{L}_\mathfrak{e}(\phi)$ to be $[\partial \,\mathsf{Fix}(\Phi), \tilde{r}^+]$ if $\mathsf{Fix}(\Phi)$ is nontrivial; else $[\mathsf{Fix_N}(\Phi) \setminus \{\tilde{r}^+\}, \tilde{r}^+]$ if there is only one new eigenray; else the set consisting of the two lines with lifts with endpoints in $\mathsf{Fix_N}(\Phi)$. This invariant description shows that $\mathsf{L}_\mathfrak{e}(\phi)$ is independent of the special chain $\mathfrak{e} \in \mathfrak{c}$, which is why $\mathfrak{c}$ does not appear in the notation.

Example 3.1 (continued) Referring to Figures 1, 2 and 3, if $\mathfrak{e}_1 := (\{[G_2]\} \sqsubset \{[G_3]\})$ then $\mathsf{L}_{\mathfrak{e}_1}(\phi)$ consists of the infinitely many lines L in the third listed component of $\Gamma(f)$ in Figure 2 that cross the oriented edge c exactly once and c^{-1} not at all. If $\mathfrak{e}_2 := (\{[G_3]\} \sqsubset \{[G_5]\})$ then $\mathsf{L}_{\mathfrak{e}_2}(\phi)$ consists of two lines; they are represented by $(R_d)^{-1} R_e$ and its inverse. If $\mathfrak{e}_3 := (\{[G_5]\} \sqsubset \{[G_7]\})$ then $\mathsf{L}_{\mathfrak{e}_3}(\phi)$ consists of two lines; they are represented by $a^\infty p^{-1} R_q$ and $a^{-\infty} p^{-1} R_q$.

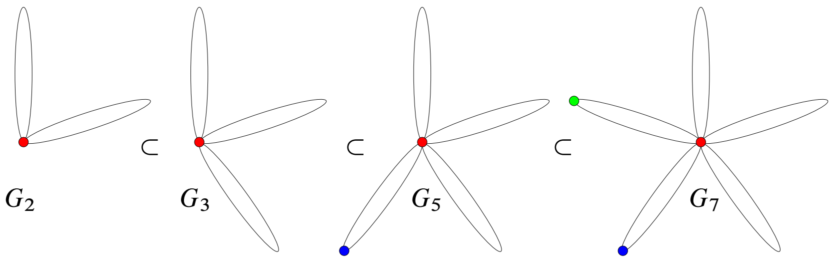


Figure 3: The special chain $\mathfrak{c} = \{[G_2]\} \sqsubset \{[G_3]\} \sqsubset \{[G_5]\} \sqsubset \{[G_7]\}$.

Lemma 6.15 *For* $\theta \in \mathsf{Out}(F_n)$, $\theta(\mathsf{L}_{\mathfrak{c}}(\phi)) = \mathsf{L}_{\theta(\mathfrak{c})}(\phi^\theta)$.

Proof By Lemma 6.13, the $\theta(\mathfrak{e}) \in \theta(\mathfrak{c})$ are special for ϕ^θ. The set of new elements of $\mathcal{R}(\phi)$ with respect to $\mathfrak{c} = (\mathcal{F}^+ \sqsubset \mathcal{F}^-)$ has the invariant description $\mathcal{R}(\phi|\mathcal{F}^+) \setminus \mathcal{R}(\phi|\mathcal{F}^-)$. In particular, θ takes the new elements with respect to ϕ to those with respect to ϕ^θ; see Lemma 3.16. The equation in the lemma then follows from the invariant definition of added lines in Definition 6.14 and the naturality results of Lemma 3.16. ◻

In the next lemma, we record some consequences of Lemmas 6.3, 6.6 and 6.7.

Lemma 6.16 (1) *A conjugacy class grows at most linearly under iteration by* ϕ *if and only if it is carried by* $\mathcal{F}_0(\phi)$.

(2) $\mathcal{F}_0(\phi) = \mathcal{F}(\mathsf{Fix}(\phi))$, *ie* $\mathcal{F}_0(\phi)$ *is the smallest free factor system carrying* $\mathsf{Fix}(\phi)$.

Proof (1) By definition $\mathcal{F}_0(\phi)$ carries all conjugacy classes that grow at most linearly. Conversely, by Lemma 6.6, $\mathcal{F}_0(\phi)$ is represented a graph K_0 consisting of linear and fixed edges. Hence every conjugacy class carried by $\mathcal{F}_0(\phi)$ grows at most linearly.

(2) By (1), $\mathcal{F}(\mathsf{Fix}(\phi)) \sqsubset \mathcal{F}_0(\phi)$. Suppose $\sqsubset$ is proper. By [Feighn and Handel 2018, Theorem 1.1], there is a CT $f : G \to G$ realizing ϕ with f-invariant core subgraphs $G_k \subsetneq G_l$ representing these two free factor systems and such that $f|G_l$ is a CT. By Lemma 6.6, every edge of G_l is fixed or linear. Let E be an edge of $G_l \setminus G_k$. There is a Nielsen circuit ρ in G_l containing E. Indeed, by the construction of eigengraphs in Section 4.1,

- every edge of a CT is the label of some edge in its eigengraph,
- the eigengraph of a linear growth CT is a compact core graph, and
- every circuit in an eigengraph is Nielsen.

The existence of ρ now follows from the defining property of a core graph that there is a circuit through every edge. The fixed conjugacy class represented by ρ is not in $\mathcal{F}(\mathsf{Fix}(\phi))$, a contradiction. ◻

6.2 The lattice of special free factor systems

This section is not needed for the rest of the paper and so could be skipped by the reader. Recall (second paragraph of Notation 6.8) that $\mathfrak{L}(\phi)$ denotes the set of special free factor systems for ϕ. The main results, Lemmas 6.18 and 6.20, are that $(\mathfrak{L}(\phi), \sqsubset)$ is a lattice that is natural with respect to $\mathsf{Out}(F_n)$ in the sense that, for $\theta \in \mathsf{Out}(F_n)$,

$$\theta(\mathfrak{L}(\phi), \sqsubset) = (\mathfrak{L}(\phi^\theta), \sqsubset).$$

In this section, $f : G \to G$ will always denote a CT for ϕ. We will conflate an element of $\mathcal{R}(\phi)$ and its image in $\mathcal{E}_f$ under the bijection $\mathcal{R}(\phi) \leftrightarrow \mathcal{E}_f$; see Lemma 3.26. A subset S of $\mathcal{R}(\phi)$ is *admissible* if it satisfies

$$(q \in S) \wedge (r < q) \implies r \in S.$$

If $S \subset \mathcal{R}(\phi)$, then mimicking Lemma 6.6 we let $K(S)$ denote the union of K_0 and the edges in S. Recall that K_0 is the union of the fixed and linear edges of G.

Lemma 6.17 *The following are equivalent*:

(1) *$\mathcal{F}$ is special for ϕ.*

(2) *$\mathcal{F} = \mathcal{F}(K(S), G)$ for some admissible $S \subset \mathcal{R}(\phi)$.*

(3) *$\mathcal{F} = \mathcal{F}(H, G)$ for some f-invariant $H \subset G$ containing K_0.*

Proof (1) $\Longrightarrow$ (2) By definition, a free factor system $\mathcal{F}$ is special if and only if there is a total order $<_T$ extending $<$ and an initial interval $[r_1, \dots, r_k]$ of $(\mathcal{R}, <_T)$ such that $\mathcal{F} = \mathcal{F}(K(\{r_1, \dots, r_k\}), G)$. Since an initial interval is admissible, we may take $S = \{r_1, \dots, r_k\}$.

(2) $\Longrightarrow$ (3) We may take $H = K(S)$.

(3) $\Longrightarrow$ (2) Let S be the set of edges in H that are not in K_0. It is enough to show that S is admissible. Let $q \in S$ and let $r \in \mathcal{R}(\phi)$ satisfy $r < q$. By definition of $<$, there is $k > 1$ so that the edge r or its inverse is a term in the complete splitting of $f_\#^k(q)$. Since the edge q is in H and H is f-invariant, the edge r is also in H.

(2) $\Longrightarrow$ (1) We claim that if S is admissible then there is an extension $<_T$ of $<$ such that S is an initial segment of $(\mathcal{R}(\phi), <_T)$. Indeed, start with any total order $<_T$ extending $<$ and iteratively interchange r and s if $r <_T s$ are consecutive, $s \in S$ and $r \notin S$. For such a $<_T$, S represents an element of $\vec{\mathcal{F}}(\phi, <_T)$. □

Lemma 6.18 *Let $\mathcal{F}$ be a special free factor system for ϕ.*

(1) *The set of admissible subsets of $\mathcal{R}(\phi)$ is a sublattice of $2^{\mathcal{R}(\phi)}$.*

(2) *There is a minimal admissible $S \subset \mathcal{R}(\phi)$ such that $\mathcal{F} = \mathcal{F}(K(S), G)$. We say that such an admissible S is **efficient for $\mathcal{F}$**. In fact, if S' is admissible then the set of edges of* core$(K(S'))$ *not in K_0 is efficient for $\mathcal{F}(K(S'), G)$.*

(3) *$S = \bigcap \{S' \mid \mathcal{F} = \mathcal{F}(K(S'), G)\}$ is efficient for $\mathcal{F}$.*

(4) *S is efficient for $\mathcal{F}$ if and only if $\mathcal{F} = \mathcal{F}(K(S), G)$ and through every edge representing an element of S there is a circuit in $K(S)$.*

(5) *If S_1 and S_2 are efficient admissible and $\mathcal{F}(K(S_1), G) \sqsubset \mathcal{F}(K(S_2), G)$, then $S_1 \subset S_2$.*

(6) *$(\mathfrak{L}(\phi), \sqsubset)$ is a lattice.*

(7) *Every maximal chain in $\mathfrak{L}(\phi)$ is special.*

(8) *Every minimal*[2] *pair $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ in $\mathfrak{L}(\phi)$ is special.*

Proof (1) This follows directly from the definition of admissible.

[2]That is, if $\mathcal{F}$ is special and $\mathcal{F}^- \sqsubset \mathcal{F} \sqsubset \mathcal{F}^+$, then $\mathcal{F}^- = \mathcal{F}$ or $\mathcal{F} = \mathcal{F}^+$.

(2) Suppose S_1, S_2 are admissible and $K(S_1), K(S_2)$ each represent $\mathcal{F}$. Hence $C := \mathrm{core}(K(S_1)) = \mathrm{core}(K(S_2))$ and, since G is a CT, C is f-invariant. (Indeed, by [Feighn and Handel 2011, Lemma 4.21], the removal of an edge with a valence one vertex from an f-invariant subgraph results in an f-invariant subgraph.) It follows that $K_0 \cup C$ is the minimal f-invariant subgraph of G representing $\mathcal{F}$ and containing K_0; see Lemma 6.17. Hence S is the set of edges of $K_0 \cup C$ not in K_0.

(3), (4), (5) These follow easily from (2).

(6) Suppose S_1 and S_2 are efficient. Then using (4), $S_1 \cup S_2$ is efficient. It follows that $\mathcal{F}(K(S_1 \cup S_2), G)$ is the smallest (with respect to $\sqsubset$) special free factor system for ϕ containing $\mathcal{F}_1 := \mathcal{F}(K(S_1), G)$ and $\mathcal{F}_2 := \mathcal{F}(K(S_2), G)$. Suppose S is efficient and $\mathcal{F}(K(S), G) \sqsubset F_i$ for $i = 1, 2$. By (5), $S \subset S_1 \cap S_2$. Since $K(S_1 \cap S_2) = K(S_1) \cap K(S_2)$, the largest special free factor system $\mathcal{F}$ for ϕ in each of $\mathcal{F}_1$ and $\mathcal{F}_2$ is represented by $K(S_1 \cap S_2)$, ie by $K(S)$, where S is efficient for $\mathcal{F}(K(S_1 \cap S_2), G)$.

(7) Let $\mathfrak{c}$ be represented by $K(S_1) \subset \cdots \subset K(S_N)$ with each S_i efficient. By (5), $S_i \subset S_{i+1}$. An argument similar to that in the proof of Lemma 6.17, (2) $\Longrightarrow$ (1), shows that there is $<_T$ extending $<$ such that each S_i is an initial interval in $(\mathcal{R}(\phi), <_T)$. Hence $\mathfrak{c}$ is special.

(8) This follows from (7) by enlarging $\mathfrak{c}$ to a maximal chain. □

Remark 6.19 $\mathfrak{L}(F_n)$ is not a sublattice of the lattice of all ϕ-invariant free factor systems. For example, reconsider Example 6.11. $S = \{c, d\}$ and $S' = \{c', d'\}$ are efficient. If $\mathcal{F} = \mathcal{F}(K(S), G)$ and $\mathcal{F}' = \mathcal{F}(K(S'), G)$ then the smallest ϕ-invariant free factor system containing $\mathcal{F}$ and $\mathcal{F}'$ is represented by the complement of the fixed edge e whereas the smallest element of $\mathfrak{L}(\phi)$ containing $\mathcal{F}$ and $\mathcal{F}'$ is F_n.

In the proof of Lemma 6.18 we noted that the union of efficient sets is efficient. The intersection need not be efficient. For example, suppose highest-order edges a, b and c share an initial vertex of valence three. Consider the complement S of a and the complement S' of b. The edge c is in $S \cap S'$ and has initial vertex of valence one in $K(S \cap S')$.

Lemma 6.20

- *If* $\mathfrak{c} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ *is minimal in* $\mathfrak{L}(\phi)$, *then* $\mathfrak{c}$ *has a well-defined type* H, HH *or* LH, *and a well-defined type* contractible, infinite cyclic *or* large.
- *For* $\theta \in \mathrm{Out}(F_n)$, *the map* $\mathcal{F} \mapsto \theta(\mathcal{F})$ *induces a lattice isomorphism*

$$(\mathfrak{L}(\phi), \sqsubset) \to (\mathfrak{L}(\phi^\theta), \sqsubset)$$

that preserves the above types.

Proof This follows Lemmas 6.18 and 6.13. □

7 More on conjugacy pairs

Recall that conjugacy pairs were introduced in Definition 4.9. In this section we define some conjugacy pairs that will be used to define invariants of elements of $\mathsf{UPG}(F_n)$ and describe their properties.

7.1 $[\partial H, \partial K]$

We will want to compare conjugacy pairs of subgroups $[H, K]$ with the set of lines $[\partial H, \partial K]$; see Examples 4.10. For this we will use the next lemma, which is a corollary of [Kapovich and Short 1996, Lemma 3.9].

Lemma 7.1 *Suppose that $H < F_n$ is finitely generated. Then the stabilizer G in F_n of $\partial H \subset \partial F_n$ is the maximal $K < F_n$ in which H has finite index.*

Corollary 7.2 *Suppose that finitely generated $H < F_n$ is root-closed, ie $a^k \in H$, $k \neq 0$ implies $a \in H$. Then H is the stabilizer in F_n of ∂H.*

Proof If $H < G$ has finite index and $H \neq G$, then H is not root-closed. □

Corollary 7.3 (1) *If $H < F_n$ is a free factor, then H is the stabilizer of ∂H.*

(2) *If $H = \mathsf{Fix}(\Phi)$ for $\Phi \in \mathsf{Aut}(F_n)$, then H is the stabilizer of ∂H.*

(3) *If $a \in F_n$ is root-free, then $A = \langle a \rangle$ is the stabilizer of the two-point set ∂A.*

Proof Free factors and the group generated by a root-free element are clearly root-closed. For (2), $\Phi(a^k) = a^k$ for $k \neq 0$ implies that $\Phi(a)$ is a k^{th} root of a^k and so equals a. □

Remark 7.4 Corollary 7.3, which contains the only cases that we need in this paper, does not require the generality of Lemma 7.1. Item (3) is elementary. Items (1) and (2) follow from (3), and:

- For $H, K < F_n$ finitely generated, $\partial H \cap \partial K = \partial(H \cap K)$. See [Kapovich and Benakli 2002, Theorem 12.2(9)] in the setting of hyperbolic groups or, for the case at hand, [Handel and Mosher 2020, Fact 1.2].
- If H is a nontrivial free factor, then $H \cap H^g = \{1\}$ unless $g \in H$.
- $\mathsf{Fix}(\Phi) \cap \mathsf{Fix}(\Phi)^g$ is cyclic unless $g \in \mathsf{Fix}(\Phi)$; see Lemma 4.4(1).

Corollary 7.5 *Suppose that $H, K < F_n$ are finitely generated and root-closed. Then $[H, K]$ determines $[\partial H, \partial K]$, and vice versa.*

Proof Suppose that $H', K' < F_n$ are finitely generated and root-closed.

If $[\partial H, \partial K] = [\partial H', \partial K']$ then there is $x \in F_n$ such that $(x\partial H, x\partial K) = (\partial H^x, \partial K^x) = (\partial H', \partial K')$. Hence

$$(H^x, K^x) = (\mathsf{Stab}(\partial H^x), \mathsf{Stab}(\partial K^x)) = (\mathsf{Stab}(\partial H'), \mathsf{Stab}(\partial K')) = (H', K').$$

So $[H, K] = [H', K']$.

Conversely, if $[H, K] = [H', K']$ then there is $x \in F_n$ such that $(H^x, K^x) = (H', K')$. Hence

$$(\partial H^x, \partial K^x) = (x\partial H, x\partial K) = (\partial H', \partial K'),$$

and so $[\partial H, \partial K] = [\partial H', \partial K']$. □

Remark 7.6 If we are in the setting of Corollary 7.5 and $\partial H \cap \partial K = \varnothing$, we will sometimes abuse notation and think of $[H, K]$ as the set of lines $[\partial H, \partial K]$ and vice versa.

7.2 Some Stallings graph algorithms

In this section we assume that G is a marked graph with marking $m\colon (R_n, *) \to (G, b)$, where $*$ is the unique vertex of the rose R_n and $b = m(*)$ is the basepoint for G. There is an induced identification of F_n with $\pi_1(G, b)$.

For each finitely generated subgroup $H < F_n$, Stallings [1983, 5.4] constructs a finite graph $\Sigma_b(H)$ with basepoint b_H and an immersion $p_H\colon (\Sigma_b(H), b_H) \to (G, b)$ such that the image of the injection $\pi_1(\Sigma_b(H), b_H) \to \pi_1(G, b)$ induced by p_H equals H. The basepoint b_H may have valence one but all other vertices of $\Sigma_b(H)$ have valence at least two. We equip $\Sigma_b(H)$ with the CW-structure whose vertex set is the preimage of the vertex set of G. The resulting edges of $\Sigma_b(H)$, sometimes called *edgelets*, are labeled by their image edges in G. The core of $\Sigma_b(H)$ is denoted by $\Sigma(H)$. The minimal edgelet-path from b_H to $\Sigma(H)$ is denoted by β_H. The terminal endpoint of β_H is denoted by $c_H \in \Sigma(H)$.

For finitely generated subgroups $K, H < F_n$, let $\mathrm{Imm}(K, H)$ be the set of immersions $J\colon \Sigma(K) \to \Sigma(H)$ that maps edgelets to edgelets and preserves labels; we say that J *preserves labels*. We do not distinguish between elements of $\mathrm{Imm}(K, H)$ that induce the same map on the set of edgelets. Thus $\mathrm{Imm}(K, H)$ is finite and can be computed by inspection. An *equivalence* is an element of $\mathrm{Imm}(K, H)$ that is a homeomorphism. Note that elements of $\mathrm{Imm}(K, H)$ that agree on a vertex of K are equal.

Lemma 7.7 *If $K < H$ are finitely generated subgroups of F_n then there is a (necessarily unique) label-preserving immersion $J_{K,H}\colon (\Sigma_b(K), b_K) \to (\Sigma_b(H), b_H)$.*

Proof We recall Stallings' construction [1983, 5.4] of $\Sigma_b(K)$. Choose closed paths $\rho_1, \dots, \rho_m \subset G$ based at b that represent generators of $K < \pi_1(G, b)$. Define Γ_1 to be a rose of rank m with unique vertex b' and define $p_1\colon (\Gamma_1, b'_1) \to (G, b)$ to be an immersion on edges, mapping the i^{th} edge to ρ_i. Subdivide Γ_1 into edgelets labeled by edges of G to obtain $p_2\colon (\Gamma_2, b'_2) \to (G, b)$. The map p_2 factors into a sequence of edgelet folds $(\Gamma_2, b'_2) \to (\Gamma_3, b'_3) \to \cdots \to (\Gamma_k, b'_k)$ followed by an immersion $p_k\colon (\Gamma_k, b'_k) \to (G, b)$. Define $(\Sigma_b(K), b_K) = (\Gamma_k, b'_k)$ and $p_H = p_k$.

Since $K < H$, each ρ_i lifts to a closed edgelet-path ρ'_i in $\Sigma_b(H)$ based at b_H. Since the i^{th} edge of Γ_2 and ρ'_i agree as labeled edgelet-paths, there is an induced label-preserving map $q_2\colon (\Gamma_2, b') \to (\Sigma_b(H), b_H)$ satisfying $p_2 = p_H q_2$. Since p_H is an immersion, the edgelets that are identified by the folding maps $\Gamma_2 \to \Gamma_3 \to \cdots \to \Gamma_k$ are also identified by q_2. Thus, there exists a map $J_{K,H}\colon (\Gamma_k, b'_k) \to (\Sigma_b(H), b_H)$ such that $p_K = p_H q_k$. Since p_H and p_K are immersions, the same is true for $J_{K,H}$. □

Note that if $a \in F_n$ and $K^a < H$ then $K^{ha} < H$ for all $h \in H$. Let $\mathrm{RC}(K, H)$ be the set of right cosets of H in F_n such that $K^a < H$ for some (each) a representing that coset.

Lemma 7.8 *There is an algorithm with output a bijection* $\mathrm{Imm}(K,H) \leftrightarrow \mathrm{RC}(K,H)$*. In particular, there is an algorithm that produces coset representatives for the elements of* $\mathrm{RC}(K,H)$*.*

Proof ($\to$) We associate a coset $Ha \in \mathrm{RC}(K,H)$ to $J \in \mathrm{Imm}(K,H)$ as follows. Choose a path $\xi \subset \Sigma(H)$ from c_H to $J(c_K)$ and note that $p_K(c_K) = p_H(J(c_K))$. Let $a \in \pi_1(G,b)$ be represented by the closed path $[p_H(\beta_H \xi) p_K(\overline{\beta}_K)] \subset G$, where $[\cdot]$ indicates tightening. Each $x \in K$ is represented in $\pi_1(G,b)$ by $p_K(\beta_K \gamma \overline{\beta}_K)$ for some closed path $\gamma \subset \Sigma(K)$ based at c_K. It follows that x^a is represented in $\pi_1(G,b)$ by

$$[p_H(\beta_H \xi) p_K(\gamma) p_H(\overline{\xi}\overline{\beta}_H)] = [p_H(\beta_H \xi) p_H(J(\gamma)) p_H(\overline{\xi}\overline{\beta}_H)] = p_H(\beta_H [\xi J(\gamma) \overline{\xi}] \overline{\beta}_H),$$

which represents an element in H. This proves that $K^a < H$. If ξ is replaced by another path ξ' connecting c_H to $J(c_K)$ then a is replaced by ha, where $h \in H$ is represented by $[p_H(\beta_H \xi' \overline{\xi} \overline{\beta}_H)]$. Thus, Ha is independent of the choice of ξ. If J is replaced with $J' \neq J$ and if $\eta \subset \Sigma(H)$ is a path connecting $J(c_K)$ to $J'(c_K)$, then ξ is replaced with $\xi\eta$ and a is replaced with $a' = da$, where d is represented in $\pi_1(G,b)$ by $[p_H(\beta_H \xi \eta \overline{\xi} \overline{\beta}_H)]$. Since η is not a closed path, d does not lift into $\Sigma(H)$ and a' does not belong to the same right coset of H as a. This shows that $J \mapsto Ha$ defines an injection from $\mathrm{Imm}(K,H)$ to $\mathrm{RC}(K,H)$.

($\leftarrow$) We begin the proof of surjectivity by constructing $\Sigma_b(K^a)$ from $\Sigma_b(K)$. Represent a in $\pi_1(G,b)$ by a closed edge-path $\alpha \subset G$ based at b and let β' be the edgelet path labeled by the path in G obtained by tightening $\alpha p_K(\beta_K)$. Define Σ' from the disjoint union of β' and a copy $(\Sigma'(K), c')$ of $(\Sigma(K), c_K)$ by identifying the terminal endpoint of β' with c'. The labeling on edgelets induces $p' : (\Sigma', b') \to (G, b)$, where b' is the initial vertex of β'. The image of the injection $\pi_1(\Sigma', b') \to \pi_1(G,b)$ induced by p' equals K^a. If p' is an immersion then $(\Sigma(K^a), \beta_{K^a}, c_{K^a}) = (\Sigma', \beta', c')$. Otherwise, $\Sigma(K^a)$ is obtained from Σ' by folding a maximal initial edgelet-subpath of $\overline{\beta}'$ with an edgelet-subpath $\mu \subset \Sigma'(K)$ that begins at c'. In this case, b_{K^a} is the folded image of b' and c_{K^a} is the terminal endpoint of μ.

Continuing with the above notation, define the equivalence $J_{K,a} \in \mathrm{Imm}(K, K^a)$ to be the identifying homeomorphism from $(\Sigma(K), c_K)$ to $(\Sigma'(K), c')$. Assuming that $K^a < H$, apply Lemma 7.7 and define $J_{a,K,H} = J_{K^a,H} | \Sigma(K^a) \circ J_{K,a} \in \mathrm{Imm}(K,H)$. By construction, $[\alpha p_K(\beta_K)] \subset G$ lifts to a the path in $\Sigma_b(H)$ from b_H to $J_{a,K,H}(c_K)$. Writing this path as $\beta_H \xi$, we have that $[p_H(\beta_H \xi) p_K(\overline{\beta}_K)] = \alpha$ and hence that (in the notation of the first paragraph of this proof) $J_{a,K,H} \mapsto Ha$. □

We will need the following well-known result.

Corollary 7.9 *If* $H < F_n$ *is a finitely generated and* $H^a < H$ *for* $a \in F_n$*, then* $H^a = H$*.*

Proof The obvious induction argument shows that $H^{a^p} < H^{a^{p-1}} < \cdots < H^a < H$ for all $p \geq 1$. Each Ha^s for $s \geq 1$ is therefore an element $\mathrm{RC}(H,H)$, which is finite by Lemma 7.8. It follows that $Ha^s = Ha^t$ for some $s \neq t$ and hence that $a^p \in H$ for some $p \geq 1$. Thus $H^{a^p} = H$, which implies that $H < H^a < H$ and hence that $H = H^a$. □

The following three algorithms are easy consequences of Lemma 7.8.

Lemma 7.10 *There is an algorithm that decides if a given pair H and K of finitely generated subgroups of F_n are conjugate, and if so produces an element $a \in F_n$ satisfying $K^a = H$.*

Proof We continue with notation from the proof of Lemma 7.8. If $K^a = H$ then $J_{K^a,H}$ is the identity and hence $J_{a,K,H} \in \mathrm{Imm}(K,H)$ is an equivalence. This shows that if $\mathrm{Imm}(K,H)$ does not contain an equivalence then K and H are not conjugate. If $\mathrm{Imm}(K,H)$ does contain an equivalence J, apply Lemma 7.8 to J and $J^{-1} \in \mathrm{Imm}(H,K)$ to produce $a, b \in F_n$ such that $K^a < H$ and $H^b < K$. From $H^{ab} < H$ and Corollary 7.9 it follows that $H^{ab} = H$ and hence that $H = (H^b)^a < K^a < H$, which implies that $K^a = H$. □

Lemma 7.11 *The normalizer $N(H)$ of a finitely generated subgroup $H < F_n$ is finitely generated. We have an algorithm that produces coset representatives $\{a_1, \dots, a_p\}$ of H in $N(H)$.*

Proof Corollary 7.9 implies that $N(H)/H = \mathrm{RC}(H,H)$. Lemma 7.8 therefore completes the proof. □

Lemma 7.12 *If $K < H < F_n$ are finitely generated subgroups then the set of subgroups of H that are F_n-conjugate to K determine finitely many H-conjugacy classes. There is an algorithm that produces representatives $K_1, \dots, K_p$ of these H-conjugacy classes.*

Proof If $K^{a_1}, K^{a_2} < H$ then K^{a_1} and K^{a_2} determine the same H-conjugacy class if and only if $ha_1 = a_2$ for some $h \in H$. Lemma 7.8, which produces representatives of the elements of $\mathrm{RC}(K,H)$, therefore completes the proof. □

7.3 Good conjugacy pairs

In addition to conjugacy classes of finitely generated subgroups of F_n, our adaptation of Gersten's algorithm will also take conjugacy pairs as input; see Notation 11.1. If H_1 and H_2 are subgroups of H and the natural map $H_1 * H_2 \to H$ is an isomorphism then we say that H is the *internal free product* of H_1 and H_2. If $A < B < F_n$ then $[A]_B$ denotes the conjugacy class of A in B. If $B = F_n$ then we sometimes suppress the subscript.

Definition 7.13 (good conjugacy pairs) For $H_1, H_2 < F_n$, the conjugacy pair $[H_1, H_2]_{F_n}$ is *good* if $\langle H_1, H_2 \rangle$ is the internal free product of H_1 and H_2.

The next lemma collects some facts about good pairs.

Lemma 7.14 *Let $H_1, H_2 < F_n$ be finitely generated.*

(1) *$[H_1, H_2]$ is good if and only if* $\mathrm{rank}(\langle H_1, H_2 \rangle) = \mathrm{rank}(H_1) + \mathrm{rank}(H_2)$.

(2) *If $[H_1, H_2]$ is good, then ∂H_1 and ∂H_2 are disjoint.*

Proof The natural map $H_1 * H_2 \to \langle H_1, H_2\rangle$ is surjective. Since finitely generated free groups are Hopfian (surjective endomorphisms are isomorphisms) [Magnus et al. 1966, Theorem 2.13], the "only if" direction of (1) follows. The "if" direction of (1) is obvious.

(2) This follows from $\partial H_1 \cap \partial H_2 = \partial(H_1 \cap H_2)$; see the first item in Remark 7.4. □

Our next goal is necessary and sufficient conditions for two good conjugacy pairs to be equal. We begin with an important special case.

Lemma 7.15 *Suppose that $K_1, K_2, L_1, L_2 < F_n$ are finitely generated, that $[K_1, K_2]_{F_n}$ and $[L_1, L_2]_{F_n}$ are good conjugacy pairs, and that $\langle K_1, K_2\rangle = \langle L_1, L_2\rangle = F_n$. Then the following are equivalent.*

(1) $[K_1, K_2]_{F_n} = [L_1, L_2]_{F_n}$.

(2) $[K_1]_{F_n} = [L_1]_{F_n}$ *and* $[K_2]_{F_n} = [L_2]_{F_n}$.

Proof (1) $\Longrightarrow$ (2) If $[K_1, K_2]_{F_n} = [L_1, L_2]_{F_n}$ then by definition there is $g \in F_n$ such that $(K_1^g, K_2^g) = (L_1, L_2)$.

(2) $\Longrightarrow$ (1) By hypothesis there are $g_i \in F_n$ such that $K_i^{g_i} = L_i$. In particular, $\Delta := (i_{g_1}|K_1) * (i_{g_2}|K_2) \in \mathsf{Aut}(F_n)$ represents an element $\delta \in \mathsf{Out}(F_n)$. Let r_i be the rank of K_i, let A_i be a rose with rank r_i whose petals are labeled by a basis for K_i and let A be the rose of rank n obtained from A_1 and A_2 by identifying their unique vertices v_1 and v_2 to a single vertex v. Blow up v to an arc. More precisely, let X be the graph obtained from the disjoint union of A_1, A_2 and a vertex w by adding oriented edges E_1 and E_2 connecting w to v_1 and v_2, respectively. Denote the arc $\overline{E}_1 E_2 \subset X$ by E and the subgraph $A_i \cup E_i \subset X$ by X_i. Identify $\pi_1(X_i, w)$ with $\pi_1(A_i, v_i) = K_i$ via the map $q_i : X_i \to A_i$ that collapses E_i to v. Let $q : X \to A$ be the map that collapses E to v. If $\alpha_i \subset A$ is the closed path based at v that represents g_i then there is a unique closed path $\beta_i \subset X$ based at w that satisfies $q_\#(\beta_i) = \alpha_i$. The map $f : X \to X$ defined by $f|A_i =$ identity and $f(E_i) = [\beta_i E_i]$ induces the automorphism Δ and so is a homotopy equivalence. Homotop f rel $A_1 \cup A_2$ to a map $f' : X \to X$ whose restriction to E is an immersion. Then f' is a topological representative of δ and [Bestvina et al. 2000, Corollary 3.2.2] implies that $f'(E) = \overline{\gamma}_1 E \gamma_2$ for some (necessarily closed) paths $\gamma_i \subset A_i$. If $k_i \in K_i$ is represented by the homotopy class of γ_i, then f' induces the automorphism $\Delta' := (i_{k_1}|K_1) * (i_{k_2}|K_2)$. There exists $h \in F_n$ such that $\Delta = i_h \Delta'$. We have $i_{g_i} = i_{hk_i}$ and hence $i_{g_1 k_1^{-1}} = i_{g_2 k_2^{-1}}$. Thus

$$\begin{aligned}[L_1, L_2]_{F_n} &= [K_1^{g_1}, K_2^{g_2}]_{F_n} = [K_1^{g_1}, (K_2^{g_1^{-1} g_2})^{g_1}]_{F_n} = [K_1, K_2^{g_1^{-1} g_2}]_{F_n} \\ &= [K_1, K_2^{k_1^{-1} k_2}]_{F_n} = [K_1^{k_1}, K_2^{k_2}]_{F_n} = [K_1, K_2]_{F_n} \qquad .\end{aligned}$$ □

For each finitely generated $L < F_n$, we define a function f_L as follows. The domain of f_L is the set of good conjugacy pairs $[K_1, K_2]$ with $K := \langle K_1, K_2\rangle$ conjugate to L. Any $g \in F_n$ such that $K^g = L$ is well-defined up to the normalizer $N(L)$ of L in F_n. That is, if $L = K^g = K^{g'}$ then $g' g^{-1} \in N(L)$.

Hence $([K_1^g]_L, [K_2^g]_L)$ is well-defined up to the diagonal action of $N(L)$ (equivalently $N(L)/L$) on the set of pairs of conjugacy classes of subgroups of L. We define $f_L([K_1, K_2])$ to be the $N(L)/L$ orbit of $([K_1^g]_L, [K_2^g]_L)$. Note that if $K = L$ and $\xi_1, \dots, \xi_r$ are coset representatives of L in $N(L)$ then $f_L([K_1, K_2]) = \{([K_1^{\xi_1}]_L, [K_2^{\xi_1}]_L), \dots, ([K_1^{\xi_r}]_L, [K_2^{\xi_r}]_L)\}$.

Remark 7.16 Suppose $x \in F_n$ and $L^x = L'$. Then $\mathsf{Domain}(f_L) = \mathsf{Domain}(f_{L'})$ and conjugation i_x by x induces a bijection (which we give the same name) $i_x \colon \mathsf{Codomain}(f_L) \to \mathsf{Codomain}(f_{L'})$ given by mapping the $N(L)$-orbit of $([L_1]_L, [L_2]_L)$ to the $N(L')$-orbit of $([L_1^x]_{L'}, [L_2^x]_{L'})$. It is an easy check that $f_{L'} = i_x \circ f_L$.

Lemma 7.17 *Suppose $K_1, K_2, L_1, L_2 < F_n$ are finitely generated and that $[K_1, K_2]_{F_n}$ and $[L_1, L_2]_{F_n}$ are good conjugacy pairs. Set $K := \langle K_1, K_2\rangle$ and $L := \langle L_1, L_2\rangle$. Then the following are equivalent*:

(1) $[K_1, K_2]_{F_n} = [L_1, L_2]_{F_n}$.

(2) *There is $g \in F_n$ such that $K^g = L$, $[K_1^g]_L = [L_1]_L$ and $[K_2^g]_L = [L_2]_L$.*

(3) $[K] = [L]$ *and* $f_L([K_1, K_2]) = f_L([L_1, L_2])$.

(4) $[K] = [L]$ *and, for some (any) $H < F_n$ with $[H] = [L] = [K]$, $f_H([K_1, K_2]) = f_H([L_1, L_2])$.*

Proof (1) $\Longrightarrow$ (2) If $[K_1, K_2]_{F_n} = [L_1, L_2]_{F_n}$ then by definition there is $g \in F_n$ such that $(K_1^g, K_2^g) = (L_1, L_2)$.

(2) $\Longrightarrow$ (1) By Lemma 7.15 applied to K_i^g and L_i with L playing the role of F_n we have $[K_1^g, K_2^g]_L = [L_1, L_2]_L$. In particular, $[K_1, K_2]_{F_n} = [L_1, L_2]_{F_n}$.

(2) $\Longrightarrow$ (3) This is clear from the definition of f_L.

(3) $\Longrightarrow$ (2) Suppose $[K] = [L]$ and $f_L([K_1, K_2]) = f_L([L_1, L_2])$. By the former there is $g' \in F_n$ such that $K^{g'} = L$ and by the latter there is $n \in N(L)$ such that $([K_1^{g'}]_L, [K_2^{g'}]_L) = ([L_1^n]_L, [L_2^n]_L)$. Take $g = n^{-1} g'$.

(3) $\Longleftrightarrow$ (4) This follows directly from Remark 7.16. □

Corollary 7.18 *There is an algorithm with input two good conjugacy pairs $[K_1, K_2]$ and $[L_1, L_2]$ of finitely generated subgroups of F_n, and output* `YES` *or* `NO` *depending on whether or not $[K_1, K_2] = [L_1, L_2]$.*

Proof Apply Lemma 7.10 to decide if K and L are F_n-conjugate. If not, then output `NO`. Otherwise, Lemma 7.10 gives $x \in F_n$ such that $K^x = L$ and we replace K_1 and K_2 by K_1^x and K_2^x so that now $K = L$. Apply Lemma 7.11 to produce coset representatives $\xi_1, \dots, \xi_r$ of L in $N(L)$. According to Lemma 7.17, $[K_1, K_2] = [L_1, L_2]$ if and only if $[K_1^{\xi_i}]_L = [L_1]_L$ and $[K_2^{\xi_i}]_L = [L_2]_L$ for some $1 \le i \le r$. This can be checked by applying Lemma 7.10 with F_n replaced by L. □

The following lemma is used in Lemma 7.21 to determine which pairs of conjugacy classes correspond to good conjugacy pairs.

Lemma 7.19 *There is algorithm with input two finitely generated subgroups $K_1, K_2 < F_n$ and output* YES *or* NO *depending on whether or not there exist $K_i' < F_n$ such that $[K_i'] = [K_i]$ and such that F_n is the internal free product of K_1' and K_2'. If* YES *then one such K_1' and K_2' are produced.*

Proof Choose any finitely generated subgroups A_i such $\mathrm{rank}(A_i) = \mathrm{rank}(K_i)$ and such that F_n is the internal free product of A_1 and A_2. K_1' and K_2' exist if and only if there is a $\theta \in \mathsf{Out}(F_n)$ such that $\theta([K_i]) = [A_i]$. The existence of such a θ can be checked using Gersten's generalization [1984] of Whitehead's theorem [Bestvina et al. 2023], which appears as Theorem 10.2 in this paper. Additionally, the algorithm produces such a θ if one exists; we take $K_i' = \Theta^{-1}(A_i)$, where $\Theta \in \theta$. □

Notation 7.20 $\mathcal{C}(F_n)$ denotes the set of conjugacy classes of finitely generated subgroups of F_n.

To aid in working with good conjugacy pairs, we relate them to ordered triples in $\mathcal{C}(F_n)$. Consider the following map from good conjugacy pairs to ordered triples in $\mathcal{C}(F_n)$:

$$[H_1, H_2] \mapsto ([H_1], [H_2], [H]), \quad \text{where } H := \langle H_1, H_2 \rangle. \tag{7-1}$$

Lemma 7.21 *We have an algorithm with input a good conjugacy pair $[H_1, H_2]$ and output a finite enumeration of the fiber F of the above map (7-1) containing $[H_1, H_2]$.*

Proof Consider the map induced by f_H from F to $\{([K_1]_H, [K_2]_H)\}/N(H)$, where K_i ranges over finitely generated subgroups of H such that H_i and K_i are conjugate in F_n. By Lemma 7.17, this map is injective. So, it remains to produce an element of F for each element of the image. The $[K_i]_H$ can be finitely enumerated by Lemma 7.12. By Lemma 7.19 we can decide if $([K_1]_H, [K_2]_H)$ represents an element of the image. Applying Lemma 7.11 and then Lemma 7.10 (with F_n replaced by H) we can decide if two pairs $([K_1]_H, [K_2]_H)$ and $([K_1']_H, [K_2']_H)$ are in the same $N(H)$-orbit and so remove redundancy from our list. □

Lemma 7.22 *We have an algorithm with input an ordered triple $([H_1], [H_2], [H])$ of elements of $\mathcal{C}(F_n)$ and output* YES *or* NO *depending on whether or not the fiber F of the above map (7-1) is empty. Further, if* NO*, the algorithm also outputs an element of F.*

Proof Our goal is to either find subgroups K_i in the same F_n conjugacy class as H_i such that $K_i < H$ and such that H is the internal free product of K_1 and K_2, or to conclude that no such K_i exist.

By Lemma 7.8, we can compute coset representatives for the elements of $\mathrm{RC}(H_i, H)$. If $\mathrm{RC}(H_1, H) = \varnothing$, then no element of the F_n-conjugacy of H_1 is a subgroup of H and we return YES. Similarly, return YES if $\mathrm{RC}(H_2, H) = \varnothing$. Otherwise, choose b_i representing a coset in $\mathrm{RC}(H_i, H)$. Replacing H_i by $H_i^{b_i}$ we may assume that $H_i < H$.

Lemma 7.11 produces coset representatives $\{a_1, \dots, a_p\}$ of H in $N(H)$. Thus a subgroup $K_i < H$ is in the same F_n conjugacy class as H_i if and only if it is in the same H-conjugacy class as $H_i^{a_j}$ for some $1 \le j \le p$. Order the pairs $(H_1^{a_j}, H_2^{a_k})$ lexicographically on $1 \le j, k \le p$. Apply Lemma 7.19 with F_n replaced by H and with (K_1, K_2) replaced by the first pair on the list $(H_1^{a_1}, H_2^{a_1})$ to either produce $K_1, K_2 < H$ such that

- K_1 is in the same H conjugacy class as $H_1^{a_1}$,
- K_2 is in the same H conjugacy class as $H_2^{a_1}$,
- H is the internal free product of K_1 and K_2,

or to conclude that no such K_1 and K_2 exist. In the former case return `NO` and $[K_1, K_2]$. In the latter case proceed on to the next pair on the list. Continue until you either return `NO` and the desired $[K_1, K_2]$, or reach the end of the list, in which case return `YES`. □

Corollary 7.23 *We have an algorithm with input an ordered triple* $([H_1],[H_2],[H])$ *of elements of* $\mathcal{C}(F_n)$ *and output a finite enumeration of the fiber* F *of the above map* (7-1) *over* $([H_1],[H_2],[H])$.

Proof Use Lemma 7.22 to determine if F is empty or not and obtain an element $[H_1', H_2'] \in F$ if not. Input $[H_1', H_2']$ into the algorithm of Lemma 7.21 to enumerate F. □

We will also use conjugacy pairs that aren't necessarily good.

Lemma 7.24 *Consider the set of conjugacy pairs of the form* $[H, A]$ *with* $A < H < F_n$ *all finitely generated and nontrivial.* (*In particular this pair is not good.*)

1. *Two such* $[H, A]$ *and* $[H', A']$ *are equal if and only if there is* $g \in F_n$ *such that* $H^g = H'$ *and* $[A^g]_{H'} = [A']_{H'}$. *In particular,* $[H, A] = [H, A']$ *if and only if* A *and* A' *are in the same orbit of the action of* $N(H)$ *on* H.
2. *The map from the set of such pairs to ordered sequences in* $\mathcal{C}(F_n)$ *given by* $[H, A] \mapsto ([H],[A])$ *has fibers that can be finitely enumerated.*

Proof (1) The "only if" direction is obvious. The "if" direction follows from the fact that if $[A^g]_{H'} = [A']_{H'}$ then there exists $h' \in H'$ such that $A^{h'g} = A'$ for some $h' \in H'$.

(2) Given finitely generated nontrivial subgroups $K, L < F_n$, compute $\mathrm{RC}(L, K)$ by applying Lemma 7.8. If $\mathrm{RC}(L, K) = \varnothing$, then L is not conjugate into K so the fiber over $([K],[L])$ is empty. Otherwise, choose b representing a coset in $\mathrm{RC}(L, K)$. Replacing L by L^b we may assume that $L < K$. Apply Lemma 7.11 to produce coset representatives $\xi_1, \dots, \xi_p$ of K in $N(K)$. The fiber containing $[K, L]$ equals $\{[K, L^{\xi_1}], \dots, [K, L^{\xi_p}]\}$ by (1). □

8 Computable G-sets

The ultimate goal of this paper is to provide an algorithm solving the conjugacy problem for $\mathsf{UPG}(F_n)$, ie Theorem 1.1. We will need other algorithms as part of our solution. In this section and the following two (Sections 8, 9 and 10), we formalize some of the algorithmic aspects present in the $\mathsf{Out}(F_n)$-setting. In particular, we provide what could be viewed as a "data structure" for the input and output of our algorithms. These sections require no knowledge of F_n and are independent of the rest of the paper.

Definition 8.1 (computable)

- A function $f: X \to Y$ is *computable* if it comes equipped with an algorithm with input $x \in X$ and output $f(x) \in Y$.
- An *enumeration* of a set X is a surjection $\mathbb{N} \to X$. A *finite enumeration* of X is a surjection $\{1, 2, \dots, N\} \to X$. The *index* of $x \in X$ is the minimal n such that $n \mapsto x$.
- A set X is *computable* if it comes equipped with a computable enumeration $\mathbb{N} \to X$ and an algorithm with input $x, x' \in X$ and output YES or NO depending on whether or not $x = x'$. By default, the empty set is computable. We sometimes write $X = (x_1, x_2, \dots)$ to indicate the enumeration. See Lemma 8.2.
- A group G is *computable* if the underlying set is computable and it comes equipped with a third algorithm with input $\theta, \theta', \theta'' \in G$ and output YES or NO depending on whether or not $\theta\theta' = \theta''$.
- A G-set X is *computable* if G and X are computable and it comes equipped with yet another algorithm with input $\theta \in G$, $x, x' \in X$ and output YES or NO depending on whether or not $\theta(x) = x'$.

Lemma 8.2 *If $X = (x_1, x_2, \dots)$ is a computable set then we have an algorithm with input $x \in X$ and output the index of x.*

Proof Starting with $i = 1$, iteratively check if $x = x_i$. □

To see how $\mathsf{Out}(F_n)$ and our $\mathsf{Out}(F_n)$-sets are enumerated and that they are computable, see Section 11.

A set Y of interest is often the quotient of a computable set X, ie $Y = X/{\sim}$ for some equivalence relation $\sim$. We want to use X to give Y the structure of a computable set. We view elements of X as *representatives* of elements of Y and always give elements $y \in Y$ as $y = [x]$, where $x \in X$ and $[x]$ is the equivalence class of x.

Lemma 8.3 *Suppose X is a computable set and $Y = X/{\sim}$ is a quotient of X. If we have an algorithm with input $x, x' \in X$ and output* YES *or* NO *depending on whether or not $[x] = [x']$, then Y is computable.*

There are the obvious generalizations for groups, etc.

Proof The computable enumeration of Y maps $i \in \mathbb{N}$ to $[x_i] \in Y$. Given input $y = [x]$, $y' = [x']$, we can use the algorithm in the hypothesis to output YES or NO depending on whether or not $[x] = [x']$, ie whether or not $y = y'$. □

Example 8.4 Suppose we are given a finite generating set for a group G. Elements of G are represented as finite words in the generators and their inverses. This set X of finite words can be computably enumerated, say using length, and X is computable. The composition of the enumeration for X and the evaluation map $e\colon X \to G$ computably enumerates G. If we have an algorithm with input $x, x' \in X$ and output `YES` or `NO` depending on whether or not $e(x) = e(x')$ then G is computable. This is the case, for example, if we are given a finite presentation for G and an algorithm solving the word problem for this presentation.

Lemma 8.5 (1) *Suppose Z is a subset of the computable set X. If we have an algorithm with input $x \in X$ and output* `YES` *or* `NO` *depending on whether or not $x \in Z$, then Z is computable.*

(2) *If X and Y are computable sets then $X \times Y$ is a computable set.*

There are the obvious generalizations for groups, etc.

Proof (1) If Z is empty then it is computable by definition. Suppose $Z \neq \varnothing$. The computable enumeration $f : \mathbb{N} \to Z$ is given as follows. Applying the algorithm in the hypothesis a finite number of times, we can find the minimal $i \in \mathbb{N}$ such that $x_i \in Z$. Define $f(j) = x_i$, for $1 \le j \le i$. If $n > i$, then $f(n) = x_n$ if $x_n \in Z$ and $f(n) = f(n-1)$ otherwise.

The proof of (2) is left to the reader. □

In the setting of Lemma 8.5(1), we view elements of Z as given to us as elements of X that are in Z. One reason for the rather odd looking enumeration f in the proof is that we have to make sure that f is defined on all of $\mathbb{N}$. (Consider the case where Z is finite.)

We now collect some basic properties of computable groups.

Lemma 8.6 *Let $G = (g_1, g_2, \dots)$ and $G' = (g'_1, g'_2, \dots)$ be computable groups.*

(1) *We have algorithms*
   (a) *with input $g \in G$ and output the index of G,*
   (b) *with input $g, h \in G$ and output the index of gh,*
   (c) *with output the index of 1, and*
   (d) *with input g and output the index of g^{-1}.*

(2) *We have an algorithm with input a finite word w in $\{g_1, g_2, \dots\}^{\pm 1}$ and output the index of w in G. In particular, we have an algorithm to solve the word problem in G.*

(3) *Suppose we are given a finite generating set $\mathcal{G} = \{h_1, \dots, h_N\} \subset \{g_1, g_2, \dots\}$ for G. Then we have an algorithm with input $g \in G$ and output a word w with letters in $\mathcal{G}$ such that $g = w$ in G.*

(4) *Suppose $f : G \to G'$ is a homomorphism. If we are given a finite generating set $\mathcal{G} \subset \{g_1, g_2, \dots\}$ for G and $f(\mathcal{G})$, then f is computable (with algorithm given in the proof).*

(5) *If $f : G \to G'$ is a computable homomorphism, then $\mathrm{Ker}(f)$ is computable.*

Proof (1) For (a), see Lemma 8.2. For (b), (c) and (d): starting with $i = 1$, iteratively use the algorithm that comes with a computable group to respectively check: (b) if $gh = g_i$; (c) if $g_i g_i = g_i$; and (d): if $gg_i = 1$.

(2) Use (d) to remove all negative exponents in w. Then use (b) to iteratively reduce the length of w by replacing consecutive letters $g_i g_j$ with a single letter g_k.

(3) Enumerate the words in $\mathcal{G}$ (say using length). Iteratively check if g is the N^{th} word.

(4) To compute $f(g_i)$, use (3) to write g_i as a word w in $\mathcal{G}$. Then $f(w)$ is a word in $f(\mathcal{G})$. Finally, use (2) to find the index of $f(w)$.

(5) This follows from Lemma 8.5 since we can algorithmically check if $f(g_i) = 1$ in G'. □

Remark 8.7 Using Lemma 8.6(1), we may rewrite a given finite subset of $\{g_1, g_2, \dots\}^{\pm 1}$ as a finite subset of $\{g_1, g_2, \dots\}$. In particular, if $\mathcal{G}$ generates G then we may algorithmically compute a finite subset of $\{g_1, g_2, \dots\}$ that is a symmetric generating set.

Example 8.8 A computable group need not be finitely generated; a finitely generated computable group need not be finitely presented; etc. Indeed, the kernel of $f : (F_2)^n \to \mathbb{Z}$ which sends each basis element to $1 \in \mathbb{Z}$ has varying finiteness properties depending on n; see [Bestvina and Brady 1997]. By Lemma 8.6(5), $\mathsf{Ker}(f)$ is computable.

Lemma 8.9 *If the G-set X is computable and $x \in X$, then the stabilizer G_x in $G = (g_1, g_2, \dots)$ of x is computable.*

Proof This follows from Lemma 8.5 applied to $G_x < G$, since we can algorithmically check whether $g_i(x) = x$. □

9 Finite presentations and finite-index subgroups

The following lemma is useful for finding presentations of finite-index subgroups of a finitely presented group. It is well-known — see eg [Lyndon and Schupp 1977, Chapter 2, Section 4, The Reidemeister–Schreier method] — but for completeness we provide a proof.

Lemma 9.1 *There is an algorithm that takes as input*

- *a finite presentation for a computable group $G = (g_1, g_2, \dots)$,*
- *the multiplication table for a finite group Q,*
- *a computable surjection $P : G \to Q$, and*
- *a subgroup Q' of Q given as a finite list of elements of Q,*

and outputs

(1) *a finite presentation for the subgroup* $G' := P^{-1}(Q')$ *of* G, *and*

(2) *finite sets of left and right coset representatives for* $G' < G$.

Remark 9.2 In some applications, G will act on a finite set S and so we have a homomorphism $G \to \mathsf{Perm}(S)$ to the permutation group of S. The group Q will be the image of this map and $P : G \to Q$ the induced map. This will allow us to compute the multiplication table for Q.

Proof of Lemma 9.1 (1) Let $G = \langle h_1, \dots, h_i \mid r_1, \dots, r_j \rangle$ be the given finite presentation for G, where the generators are in $\{g_1, \dots\}$ and the relations are words in the generators; see Lemma 8.6(2). Let X_G denote its presentation 2-complex. We assume the reader is familiar with obtaining a finite presentation for a group H from a finite based 2-complex with fundamental group H. Hence (1) is reduced to constructing the finite based cover $X_{G'}$ of X_G whose fundamental group has image in the fundamental group of X_G equal to G'.

Set $k := [G : G'] = [Q : Q']$ and note that $k \le |Q|$. Then every index k based cover of X_G has $k \cdot i$ 1-cells and $k \cdot j$ 2-cells. Further, if $|r_q|$ denotes the length of r_q as a word in $\{h_1, \dots, h_i\}$, then each 2-cell in the cover has boundary of length at most $k \cdot \max\{|r_q| \mid 1 \le q \le j\}$. Hence we can construct all based covers of X_G of index k. Examine each in turn to check whether the image K of its fundamental group in X_G is G'. This can be done by checking whether the image in X_G of a generating set for the fundamental group of the cover has P-image contained in Q'. (Indeed, if so then $K < G'$ but both K and G' have the same index in G.) Since G' has index k in G, we are guaranteed that one of these covers satisfies $K = G'$. This completes the proof of (1).

(2) We find left coset representatives, the other case being symmetric. Using the hypotheses on Q, choose a set S_Q of left cosets representatives for $Q' < Q$. Then $S_G \subset G$ is a set of left coset representatives for $G' < G$ if the restriction $P|S_G$ is injective with image S_Q. To find $g \in G$ with P-image $q \in Q$, the P-image of a set of generators for G generates Q and in Q we may write q in terms of these generators for Q. □

Given a short exact sequence $1 \to N \to G \xrightarrow{f} Q \to 1$, we are interested in finding a finite presentation for G from finite presentations for N and Q.

Lemma 9.3 *Suppose* $f : G \to Q$ *is a computable surjection between computable groups* G *and* Q.

Suppose we are given

- *a finite presentation for* $N := \mathsf{Ker}(f)$, *and*
- *a finite presentation for* Q *(for example if* Q *is finite and we are given the multiplication table for* Q *then this item is satisfied).*

Then we may find a finite presentation for G. *(In fact, one is constructed in the proof.)*

Proof Suppose that the finite presentation for N has generating set $\{g_{N,i} \mid i \in I\} \subset G$ and set of relators $\{r_{N,j} \mid j \in J\}$ and suppose that the finite presentation for Q has generating set $\{g_{Q,k} \mid k \in K\}$ and set of relators $\{r_{Q,l} \mid l \in L\}$.

For each $g_{Q,k}$, find an element $\hat{g}_{Q,k} \in G$ with image $g_{Q,k}$ in Q. This can be done algorithmically by iteratively searching for g_i such that $f(g_i) = g_{Q,k}$.

Set $\hat{r}_{Q,l} = r_{Q,l}(\hat{g}_{Q,k} \mid k \in K)$, ie $r_{Q,l}$ is a word in $\{g_{Q,k} \mid k \in K\}$ and $\hat{r}_{Q,l}$ denotes the same word in $\{\hat{g}_{Q,k} \mid k \in K\}$. The image of $\hat{r}_{Q,l}$ in Q represents 1_Q and so there is a word $n_{Q,l}$ in $\{g_{N,i} \mid i \in I\}$ such that $\hat{r}_{Q,l} = n_{Q,l}$ in G. Since N is normal, $\hat{g}_{Q,k}(g_{N,i})\hat{g}_{Q,k}^{-1} = n_{N,i,Q,k}$ for some word $n_{N,i,Q,k}$ in $\{g_{N,i} \mid i \in I\}$. Since G is computable, $n_{Q,l}$ and $n_{N,i,Q,k}$ can be found algorithmically; see Lemma 8.6(3). By [Bridson and Wilton 2011, Lemma 2.1], there is a finite presentation for G with

- generating set $\{g_{N,i}, \hat{g}_{Q,k} \mid i \in I, k \in K\}$, and
- set of relators that is the union of
  - $\{r_{N,j} \mid j \in J\}$,
  - $\{\hat{r}_{Q,l} = n_{Q,l} \mid l \in L\}$, and
  - $\{\hat{g}_{Q,k}(g_{N,i})\hat{g}_{Q,k}^{-1} = n_{N,i,Q,k} \mid i \in I, k \in K\}$. □

10 MW-algorithms

Our solution of the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$ will use a generalization of an algorithm of Gersten that in turn generalizes algorithms of Whitehead and McCool. This section is devoted to describing these generalizations.

A set equipped with an action by a group G is a *G-set*. We will only consider *computable* G-sets; see Definition 8.1.

Definition 10.1 (property MW) A computable G-set X *satisfies property* MW (for McCool and Whitehead) if it comes equipped with an algorithm that takes as input $x, y \in X$ and outputs

(M) finite presentations for $G_x := \{\theta \in G \mid \theta(x) = x\}$ and for $G_y := \{\theta \in G \mid \theta(y) = y\}$, and

(W) `YES` or `NO` depending on whether or not there exists $\theta \in G$ such that $\theta(x) = y$ together with such a θ if `YES`.

We call such an algorithm an MW-*algorithm*. Sometimes we refer to an algorithm that satisfies item M (resp. item W) as an M-*algorithm* (resp. W-*algorithm*). Recall that G_x and G_y are computable by Lemma 8.9.

Of course our interest here is in the case $G = \mathsf{Out}(F_n)$ where the second item is associated with J H C Whitehead [1936a; 1936b] whose algorithm decides if there is θ taking one finite ordered set of conjugacy classes in F_n to another and produces such a θ if one exists. The first item is associated with McCool [1975], whose algorithm produces a finite presentation for the stabilizer of a finite ordered set of conjugacy classes of elements of F_n. S Gersten generalized the algorithms of Whitehead and McCool to finite sequences in the $\mathsf{Out}(F_n)$-set $\mathcal{C}(F_n)$; see Notation 7.20. ($\mathcal{C}(F_n)$ is shown to be a computable $\mathsf{Out}(F_n)$-set at the beginning of Section 11.) We state Gersten's result in a slightly weakened form.

Theorem 10.2 ([Gersten 1984, Theorems W&M], see also [Kalajdžievski 1992] and [Bestvina et al. 2023]) *The action of* $\mathsf{Out}(F_n)$ *on the set of finite ordered sets in* $\mathcal{C}(F_n)$ *satisfies property MW.*

We will define algebraic invariants for elements of $\mathsf{UPG}(F_n)$; see Section 13. An obstruction to $\phi, \psi \in \mathsf{UPG}(F_n)$ being conjugate in $\mathsf{Out}(F_n)$ is the existence of a $\theta \in \mathsf{Out}(F_n)$ taking the algebraic invariants of ϕ to those of ψ. More specifically, if such a θ does not exist then ϕ and ψ are not conjugate. If such a θ does exist, then we replace ϕ by ϕ^θ (or ψ by $\psi^{(\theta^{-1})}$) and reduce to the case where the algebraic invariants of ϕ and ψ agree. One step in our algorithm for the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$ will be to check whether such a θ exists and to produce one if so.

Some of our invariants are elements of $\mathcal{C}(F_n)$ and so fit nicely into the setting of Gersten's theorem. We will extend Gersten's theorem so that it applies to all our invariants. These invariants are best described in terms of *iterated sets* of elements of $\mathcal{C}(F_n)$, or more generally in terms of sets with finite-to-one maps to iterated sets. Roughly, an iterated set in a G-set $\mathbb{A}$ is a finite set consisting of elements of $\mathbb{A}$ and previously produced iterated sets. The set may be ordered or not. We commonly take $\mathbb{A}$ to be $\mathcal{C}(F_n)$.

There are two main results. The first, Proposition 10.14, promotes MW-algorithms for finite ordered sets in $\mathbb{A}$ to MW-algorithms for the set $\overline{\mathsf{IS}}(\mathbb{A})$ (of equivalence classes) of iterated sets in $\mathbb{A}$. More specifically, it states that if the G-action on finite ordered subsets of $\mathbb{A}$ satisfies property MW then so does the G-action on $\overline{\mathsf{IS}}(\mathbb{A})$. The second, Corollary 10.22, is a method for enlarging $\mathbb{A}$.

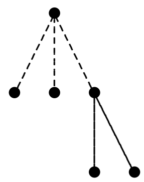
After reviewing our invariants in Section 12 and defining the algebraic invariants in Section 13, we apply our generalized Gersten's algorithm to obtain a reduction of the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$ to Proposition 14.7 in Lemma 14.8.

10.1 Iterated sets and their equivalence classes

Definition 10.3 A *rooted tree* $(T, *)$ is a finite, simplicial, directed tree T with a basepoint $*$ called the *root*. A valence 0 vertex (ie $T = \{*\}$) or a valence one vertex that is not the root is a *leaf*. The set of leaves in T is denoted by $\mathcal{L}(T)$. All edges are oriented away from the root. The set of edges exiting a vertex $x \in T$ is denoted by $\mathcal{E}_T(x)$. Paths are directed. We also may give some vertices an extra structure: a vertex x that is not a leaf is *ordered* if an order has been imposed on $\mathcal{E}_T(x)$. A vertex that is not a leaf and that is not ordered is *unordered*.

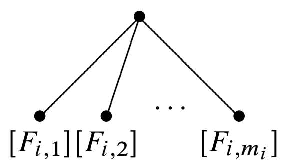
We view rooted trees $(T, *)$ as combinatorial objects. In particular, edges are specified by ordered pairs of vertices. For technical reasons having to do with computability, we require that all vertices of the trees we consider lie in a set V that we fix once and for all. (For our purposes, one can take V to be $\mathbb{N}$.)

Example 10.4 We will draw rooted trees with the root at the top. Ordered vertices are indicated by using dashed lines for its exiting edges. The imposed ordering is displayed from left to right. In the rooted tree below only the root is ordered. There are 4 leaves:

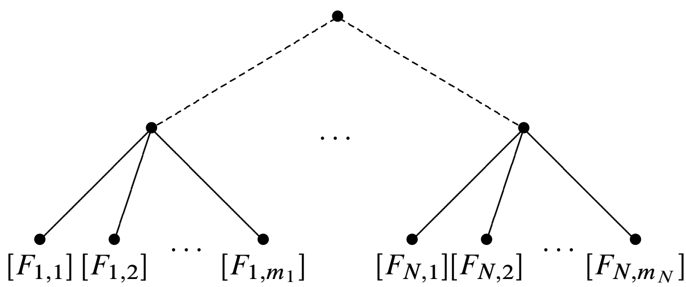


Definition 10.5 (iterated set) An *iterated set in a set* $\mathbb{A}$ is a rooted tree such that each leaf is labeled by an element of $\mathbb{A}$. Specifically, an iterated set in $\mathbb{A}$ is a pair $((T,*),\chi)$ where $(T,*)$ is a rooted tree and $\chi\colon \mathcal{L}(T)\to\mathbb{A}$ is a function. We do not assume that χ is one-to-one. We will often use sans serif capital letters for iterated sets and write, for example, $\mathsf{X}=((T,*),\chi)$. The *set of atoms of* X is $\chi(\mathcal{L}(T))$. We sometimes refer to $\mathbb{A}$ as *the set of atoms*. For $l\in\mathcal{L}(T)$, we sometimes refer to $\chi(l)$ as the *label* or *atom* of l. The set of iterated sets in $\mathbb{A}$ is denoted by $\mathsf{IS}(\mathbb{A})$.

Example 10.6 If we take $\mathbb{A}=\mathcal{C}(F_n)$, then a nested sequence $\vec{\mathcal{F}}=\mathcal{F}_1\sqsubset\cdots\sqsubset\mathcal{F}_N$ of free factor systems determines an iterated set in $\mathbb{A}$ as follows. First we identify each free factor system $\mathcal{F}_i=\{[F_{i,1}],\dots,[F_{i,m_i}]\}$ with an iterated set:



Then $\vec{\mathcal{F}}$ determines the ordered set $\{\mathcal{F}_1,\dots,\mathcal{F}_N\}$:



Definition 10.7 Let $\mathsf{X}=((T,*),\chi)$ and $\mathsf{X}'=((T',*'),\chi')$ be iterated sets in $\mathbb{A}$.

- An *order-preserving simplicial isomorphism* $f\colon (T,*)\to(T',*')$ is a simplicial isomorphism that satisfies:
  (1) A vertex x of T is ordered if and only if $f(x)$ is ordered and the induced map $\mathcal{E}_T(x)\to\mathcal{E}_{T'}(f(x))$ is order-preserving.

- An *equivalence* $\mathsf{f}\colon \mathsf{X} \to \mathsf{X}'$ is an order-preserving simplicial isomorphism $f\colon (T,*) \to (T',*')$ that additionally satisfies:

  (2) For $l \in \mathcal{L}(T)$, $\chi(l) = \chi'(f(l))$.

  Clearly equivalence induces an equivalence relation on $\mathsf{IS}(\mathbb{A})$.

- $\overline{\mathsf{IS}}(\mathbb{A})$ denotes the set of equivalence classes of iterated sets.

Remark 10.8 We will not need this, but if $\mathbb{A}$ is the set of objects of a category $\widehat{\mathbb{A}}$, then naturally so are $\mathsf{IS}(\mathbb{A})$ and $\overline{\mathsf{IS}}(\mathbb{A})$. A *morphism* $((T,*),\chi) \to ((T',*'),\chi')$ is an order-preserving simplicial isomorphism $f\colon (T,*) \to (T',*')$ together with a function $m\colon \chi(\mathcal{L}(T))$ into the morphisms of $\widehat{\mathbb{A}}$ such that, for $l \in \mathcal{L}(T)$, $m(\chi(l)) \in \mathsf{Hom}(\chi(l), \chi'(f(l)))$. An earlier version of this paper used a simplified, but more restrictive variant of this category, which was ultimately not needed.

10.2 Promoting property MW

Definition 10.9 Suppose G is a group and $\mathbb{A}$ is a G-set. Then $\mathsf{IS}(\mathbb{A})$ and $\overline{\mathsf{IS}}(\mathbb{A})$ are G-sets with actions given as follows. If $\theta \in G$ and $\mathsf{X} = ((T,*),\chi)$, then $\theta(\mathsf{X}) := ((T,*),\theta \circ \chi)$, ie $\theta(\mathsf{X})$ is obtained by relabeling $\mathcal{L}(T)$ according to θ. The G-action descends to $\overline{\mathsf{IS}}(\mathbb{A})$.

We want $\mathsf{IS}(\mathbb{A})$ and $\overline{\mathsf{IS}}(\mathbb{A})$ to be computable. This is the case if our set V of vertices and $\mathbb{A}$ are computable.

Lemma 10.10 (1) *If V is a computable set, then the set of rooted trees with vertices in V is computable.*

(2) *If additionally $\mathbb{A}$ is a computable set, then the sets $\mathsf{IS}(\mathbb{A})$ and $\overline{\mathsf{IS}}(\mathbb{A})$ are computable.*

(3) *If additionally $\mathbb{A}$ is a computable G-set, then the G-sets $\mathsf{IS}(\mathbb{A})$ and $\overline{\mathsf{IS}}(\mathbb{A})$ are computable.*

Proof (1) We view rooted trees $(T,*)$ as combinatorial objects. In particular, edges are specified by ordered pairs of vertices. To completely specify $(T,*)$ we also choose a root vertex, designate some vertices as ordered and choose an order on exiting edges of those vertices. Rooted trees T with vertices in V can be computably enumerated using, say, the sum $|T|$ of the number of vertices and largest index among the vertices of T. That is, to enumerate the set of rooted trees, first list all those with $|T| = 1$, then 2, etc. Two rooted trees are equal if and only if the they have the same vertices, edges, root, ordered vertices and same order on outgoing edges of ordered vertices.

(2) For an iterated set $\mathsf{X} = ((T,*),\chi)$, let $|\mathsf{X}|$ denote the sum of the number of vertices of $|T|$ and the largest index of a label (an element of $\chi(\mathcal{L}(T))$). $\mathsf{IS}(\mathbb{A})$ may be countably enumerated using $|\mathsf{X}|$. To enumerate $\mathsf{IS}(\mathbb{A})$, for example, list all X with $|\mathsf{X}| = 1$, then 2, etc. Two iterated sets are equal if and only if the underlying rooted trees are equal and the functions on the leaves are equal. The first condition can be checked by (1) and the second can be checked since $\mathbb{A}$ is computable. We can use the same

enumeration for $\overline{\mathsf{IS}}(\mathbb{A})$. Here if $\mathsf{X}_1 := ((T_1, *_1), \chi_1)$ and $\mathsf{X}_2 := ((T_2, *_2), \chi_2)$ represent respectively Q_1 and Q_2 in $\overline{\mathsf{IS}}(\mathbb{A})$, then $\mathsf{Q}_1 = \mathsf{Q}_2$ if and only if X_1 and X_2 are equivalent and this is a finite check. Indeed, finitely enumerate the set S of order-preserving simplicial isomorphisms $f\colon (T_1, *_1) \to (T_2, *_2)$. If S is empty then $\mathsf{Q}_1 \neq \mathsf{Q}_2$. Otherwise, if some $f \in S$ is an equivalence then $\mathsf{Q}_1 = \mathsf{Q}_2$ and if not then $\mathsf{Q}_1 \neq \mathsf{Q}_2$.

(3) Since $\mathbb{A}$ is a computable G-set, it is a finite check whether $\theta(\mathsf{X}_1) = \mathsf{X}_2$ and also whether $\theta(\mathsf{X}_1)$ and X_2 are equivalent. □

Assumption 10.11 Going forward, we assume that our fixed set V of vertices is computable; see Definition 10.3. In all applications, $\mathbb{A}$ will be computable.

Notation 10.12 Unless otherwise specified,

- $\mathbb{A}$ denotes a computable G-set,
- $\mathsf{X}, \mathsf{X}', \dots$ denote elements $((T, *), \chi), ((T', *'), \chi'), \dots$ of $\mathsf{IS}(\mathbb{A})$,
- $\mathsf{Q}, \mathsf{Q}', \dots$ denote elements of $\overline{\mathsf{IS}}(\mathbb{A})$ and are represented by $\mathsf{X}, \mathsf{X}', \dots$, and
- an equivalence $\mathsf{f}\colon \mathsf{X} \to \mathsf{X}'$ is given by $f\colon (T, *) \to (T', *')$.

Notation 10.13

- The map $\mathbb{A} \to \overline{\mathsf{IS}}(\mathbb{A})$ determined by $a \mapsto ((*, *), * \mapsto a)$ is a G-equivariant inclusion. In other words, map a to the trivial tree with vertex labeled a. Thus we may think of $\mathbb{A}$ as a subset of $\overline{\mathsf{IS}}(\mathbb{A})$.
- $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ denotes the subset of $\overline{\mathsf{IS}}(\mathbb{A})$ represented by iterated sets in which $*$ is ordered and $*$ is the initial endpoint of every edge of T. $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ is G-invariant. There is an obvious G-invariant bijection between the set of nonempty, finite, ordered sequences in $\mathbb{A}$ and $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$. We pass back and forth between these two points of view whenever convenient.
- The G-set $\mathcal{S}_{\mathrm{un}}(\mathbb{A})$ is defined analogously where $*$ is unordered. There is an obvious G-invariant bijection between the set of nonempty, finite, multisets in $\mathbb{A}$ and $\mathcal{S}_{\mathrm{un}}(\mathbb{A})$.

Proposition 10.14 (promoting MW) *Let $\mathbb{A}$ be a computable G-set. If $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ satisfies property MW, then so does $\overline{\mathsf{IS}}(\mathbb{A})$.*

Proof We follow the conventions in Notation 10.12. First we provide a W-algorithm for $\overline{\mathsf{IS}}(\mathbb{A})$; ie an algorithm that either finds $\theta \in G$ satisfying $\theta(\mathsf{Q}) = \mathsf{Q}'$ or concludes that there is no such θ. Finitely enumerate the set S of order-preserving simplicial isomorphisms $f\colon (T, *) \to (T', *')$. If S is empty then return `NO`. Otherwise, start with the first element f of S. By hypothesis there is a W-algorithm that either finds $\theta \in G$ such that $\theta(\chi(l)) = \chi'(f(l))$ for each $l \in \mathcal{L}(T)$ or concludes that no such θ exists. If θ is found then f gives an equivalence $\theta(\mathsf{X}) \to \mathsf{X}'$ and our W-algorithm returns `YES` and θ. If no such θ exists, move on to the next element of S and try again. If after considering each element of the finite set S we have not returned `YES`, then return `NO`. (Equivalently, we could make the queries indexed by S in parallel. Note that a different choice of representatives X, X' for Q, Q' would give the same queries.)

The stabilizer G_{Q} in G of Q is computable by Lemma 8.9. For the M-algorithm, we will produce a finite presentation for G_{Q} by applying Lemma 9.3 to the short exact sequence induced by $\pi: G_{\mathsf{Q}} \to \mathsf{Perm}(A)$, where A denotes the set $\chi(\mathcal{L}(\mathsf{X}))$ of atoms of X. Since the kernel of π is the subgroup of G fixing each element of A, we can use the M-algorithm for $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ to produce a finite presentation for this kernel.

To apply Lemma 9.3, it remains to produce an element of G_{Q} realizing each element of the image of π. This is done as follows. Given $\sigma \in \mathsf{Perm}(A)$, use the W-algorithm for $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ to produce $\theta \in G$ realizing σ if such exists. If there is no such θ then σ is not in the image of π. Finally, use the computability of $\overline{\mathsf{IS}}(\mathbb{A})$ (Lemma 10.10(3)) to check if $\theta(\mathsf{X})$ is equivalent to X. If it is then σ is in the image of π (and is realized by θ) and otherwise it is not. (As above with the W-algorithm, it is easy to see that the choice of representative X for Q is immaterial. Also, the choice of θ does not matter.) □

For reference we record the following consequence of the previous proof (really just definitions).

Corollary 10.15 *If* $\mathsf{X} = ((T, *), \chi)$ *represents* $\mathsf{Q} \in \overline{\mathsf{IS}}(\mathbb{A})$, *then the subgroup of* G *fixing each* $\chi(l)$, $l \in \mathcal{L}(T)$, *has finite index in the stabilizer* G_{Q} *of* Q. □

10.3 More atoms

Proposition 10.14 concludes, under conditions on $\mathbb{A}$, that $\overline{\mathsf{IS}}(\mathbb{A})$ has property MW. In this section, conclusions have the form $\overline{\mathsf{IS}}(\mathbb{A}')$ satisfies property MW, where $\mathbb{A}'$ is a G-set constructed from $\mathbb{A}$ in various ways. Intuitively, we are enlarging our collection of useful sets of atoms.

Notation 10.16 Suppose $p: \widehat{Y} \to Y$ is an equivariant map of G-sets. For $y \in Y$ [resp. $\hat{y} \in \widehat{Y}$], let G_y (resp. $G_{\hat{y}}$) denote the stabilizer of y (resp. $\hat{y}$) with respect to the action of G on Y (resp. $\widehat{Y}$). Let F_y denote the fiber $p^{-1}(y)$. If $p(\hat{y}) = y$ then by p-equivariance $G_{\hat{y}} < G_y$ and G_y acts on F_y inducing a homomorphism $\rho_y: G_y \to \mathsf{Perm}(F_y)$. (We declare the permutation group of the empty set to be trivial.)

In this setting, we say that p *has explicit finite fibers* if the G-sets Y and $\widehat{Y}$ are computable and p comes equipped with an algorithm with input $y \in Y$ and output a finite enumeration of F_y.

Lemma 10.17 *Suppose the* G*-map* $p: \widehat{Y} \to Y$ *has explicit finite fibers and that* Y *satisfies property M. Then*:

1. *There is an algorithm with input* $y \in Y$, $\theta \in G_y$ *and output* $\rho_y(\theta)$.
2. *There is an algorithm with input* $y \in Y$ *and output the multiplication table for* $\rho_y(G_y)$.
3. $\widehat{Y}$ *satisfies property M.*

Proof Let $y \in Y$. By Lemma 8.9, G_y is computable. Since p has explicit finite fibers, we can compute the finite list of elements of F_y. Since Y satisfies property M we can produce a finite presentation for G_y.

(1) Since $\widehat{Y}$ is computable, we can compute the action of θ on $F_y \subset \widehat{Y}$.

(2) This follows by applying (1) to our generators of G_y.

(3) Since $G_{\hat{y}}$ is the ρ_y-preimage of the stabilizer S of $\hat{y}$ in $\mathsf{Perm}(F_y)$, we can find a finite presentation for $G_{\hat{y}}$ by applying Lemma 9.1, taking P to be the surjective homomorphism $G_y \to \rho_y(G_y)$ and $Q' := \rho_y(G_y) \cap S$. □

Lemma 10.18 *Suppose $f: Z \to Y$ and $g: Y \to X$ each has explicit finite fibers. Then the composition $h = g \circ f : Z \to X$ has explicit finite fibers.*

Proof Since each one of f and g has explicit finite fibers, the G-sets X, Y and Z are each computable. Given $x \in X$, since g has explicit finite fibers, we can list the elements of $g^{-1}(x)$. Since f has explicit finite fibers, we can list the elements of $f^{-1}(y)$ for each $y \in g^{-1}(x)$. We are done by noting that $h^{-1}(x) = \bigsqcup\{f^{-1}(y) \mid y \in g^{-1}(x)\}$. □

Lemma 10.19 *Suppose that $p: \hat{Y} \to Y$ is a G-equivariant map of G-sets such that*

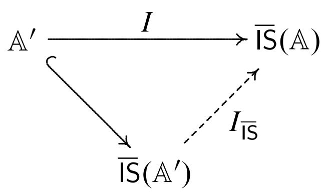
- *p has explicit finite fibers, and*
- *Y satisfies property MW.*

Then $\hat{Y}$ satisfies property MW.

Proof $\hat{Y}$ satisfies property M by Lemma 10.17(3).

For the W-algorithm, we use Notation 10.16. Suppose that $\hat{z} \in \hat{Y}$ and that $z = p(\hat{z})$. Since Y satisfies property W we can check whether or not there is θ_0 such that $\theta_0(y) = z$ and we can compute such a θ_0 if it exists. If not, then return `NO`. If yes, then $\hat{y}$ and $\theta_0^{-1}(\hat{z})$ are in F_y and there is an element of G taking $\hat{y}$ to $\hat{z}$ if and only if there is an element $\theta \in G$ (necessarily in G_y) taking $\hat{y}$ to $\theta_0^{-1}(\hat{z})$. Our goal becomes to check whether there is $\theta \in G_y$ such that $\rho_y(\theta) \in \mathsf{Perm}(F_y)$ takes $\hat{y}$ to $\theta_0^{-1}(\hat{z})$, and to produce such a θ if so. This can be done using our finite generating set for G_y and its action on F_y. If there is no such θ then return `NO`. Otherwise return `YES` and $\theta_0 \cdot \theta$. □

Construction 10.20 (canonical extension) Suppose that $\mathbb{A}$ and $\mathbb{A}'$ are G-sets and that $I: \mathbb{A}' \to \overline{\mathsf{IS}}(\mathbb{A})$ is G-equivariant. We now define a G-equivariant map $I_{\overline{\mathsf{IS}}}: \overline{\mathsf{IS}}(\mathbb{A}') \to \overline{\mathsf{IS}}(\mathbb{A})$ that restricts to I on $\mathbb{A}'$; see Notation 10.13. We call $I_{\overline{\mathsf{IS}}}$ the *canonical extension of I*:



From $\mathsf{Q}' \in \overline{\mathsf{IS}}(\mathbb{A}')$, we construct $\mathsf{Q} := I_{\overline{\mathsf{IS}}}(\mathsf{Q}') \in \overline{\mathsf{IS}}(\mathbb{A})$. We do this by constructing a representative $\mathsf{X} = ((T, *), \chi)$ for Q from a representative $\mathsf{X}' = ((T', *'), \chi')$ for Q' and, for each $l' \in \mathcal{L}(T')$, a representative $\mathsf{X}_{l'} = ((T_{l'}, *_{l'}), \chi_{l'})$ for $\mathsf{Q}_{l'} := I(\chi'(l')) \in \overline{\mathsf{IS}}(\mathbb{A})$. Let T be the tree obtained from

$T' \sqcup \left(\bigsqcup\{T_{l'} \mid l' \in \mathcal{L}(T')\}\right)$ by identifying $l' \in T'$ and $*_{l'} \in T_{l'}$. We declare the image of $*'$ in T to be the root $*$ of T. The leaves of T biject naturally with $\bigsqcup\{\mathcal{L}(T_{l'}) \mid l' \in \mathcal{L}(T')\}$ and we define $\chi\colon \mathcal{L}(T) \to \mathbb{A}$ by $\chi|\mathcal{L}(T_{l'}) := \chi_{l'}$.

We next show that Q is independent of our choices of representatives, ie that $\mathsf{Q}' \mapsto \mathsf{Q}$ is well-defined. Let $\mathsf{Y}' = ((S', \star'), \eta')$ also represent Q' and so we have a simplicial isomorphism $f'\colon T' \to S'$ inducing an equivalence $\mathsf{X}' \to \mathsf{Y}'$. In particular, $I(\chi'(l')) = I(\eta'(f(l')))$ in $\overline{\mathsf{IS}}(\mathbb{A})$. Thus if $\mathsf{Y}_{f(l')}$ is a representative of $I(\eta'(f(l')))$, then we have equivalences $\mathsf{X}_{l'} \to \mathsf{Y}_{f(l')}$ induced by simplicial isomorphisms $f_{l'}\colon (T_{l'}, *_{l'}) \to (S_{f(l')}, \star_{f(l')})$ between the underlying trees. The map f' and the $f_{l'}$ induce a map

$$T' \sqcup \left(\bigsqcup\{T_{l'} \mid l' \in \mathcal{L}(T')\}\right) \to S' \sqcup \left(\bigsqcup\{S_{f_{l'}(l')} \mid l' \in \mathcal{L}(T')\}\right),$$

which descends to a simplicial isomorphism $T \to S$ that induces an equivalence $\mathsf{X} \to \mathsf{Y}$. Hence the map $I_{\overline{\mathsf{IS}}}\colon \overline{\mathsf{IS}}(\mathbb{A}') \to \overline{\mathsf{IS}}(\mathbb{A})$ given by $\mathsf{Q}' \mapsto \mathsf{Q}$ is well-defined.

If we start the above construction with $\theta(\mathsf{X}')$ instead of X', the only difference is that χ is replaced by $\theta \circ \chi$, ie $I_{\overline{\mathsf{IS}}}$ is G-equivariant. Recall (Notation 10.13) that we identify $\mathbb{A}'$ with the subset of elements of $\overline{\mathsf{IS}}(\mathbb{A}')$ with underlying tree consisting of only the root. Thus I and $I_{\overline{\mathsf{IS}}}$ agree on $\mathbb{A}'$.

Lemma 10.21 *Let $\mathbb{A}$ and $\mathbb{A}'$ be G-sets and suppose the G-map $I : \mathbb{A}' \to \overline{\mathsf{IS}}(\mathbb{A})$ has explicit finite fibers. Then $I_{\overline{\mathsf{IS}}}\colon \overline{\mathsf{IS}}(\mathbb{A}') \to \overline{\mathsf{IS}}(\mathbb{A})$ has explicit finite fibers. If additionally $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ satisfies property MW, then $\overline{\mathsf{IS}}(\mathbb{A}')$ satisfies property MW.*

Proof Since I has explicit finite fibers, the G-set $\mathbb{A}'$ is computable. The G-set $\overline{\mathsf{IS}}(\mathbb{A}')$ is therefore computable by Lemma 10.10(3). Let $\mathsf{Q} \in \overline{\mathsf{IS}}(\mathbb{A})$ be given and let $\mathsf{X} = ((T, *), \chi) \in \mathsf{IS}(\mathbb{A})$ represent Q. Using notation as in Construction 10.20, each Q' in the fiber F of $I_{\overline{\mathsf{IS}}}$ over Q has a representative of the form $((T', *), \chi')$, where $(T', *)$ is a rooted subtree of $(T, *)$. Further, each leaf l' of T' then determines a rooted tree $(T_{l'}, l')$, where $T_{l'}$ is the subtree of T spanned by l' and all leaves l of T with a directed path from l' to l. We then get a representative $((T_{l'}, l'), \chi_{l'})$ for an element $\mathsf{Q}_{l'}$ of $\overline{\mathsf{IS}}(\mathbb{A})$, where $\chi_{l'}$ is the restriction of χ to the leaves of $T_{l'}$. If there are elements $a'_{l'} \in \mathbb{A}'$ such that $I(a'_{l'}) = \mathsf{Q}_{l'}$ and if we define $\chi'(l') := a'_{l'}$ then $((T', *), \chi')$ represents an element of F and all elements of F have this form (for some choice of T'). Since I has explicit finite fibers, we can finitely enumerate the fiber of I over $\mathsf{Q}_{l'}$ and so also find a finite list in $\mathsf{IS}(\mathbb{A}')$ of representatives for the elements of F. (It is easy to see that a different choice of representative X for Q produces F with a perhaps different enumeration.)

If additionally $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ satisfies property MW, then by Proposition 10.14 so does $\overline{\mathsf{IS}}(\mathbb{A})$. That $\overline{\mathsf{IS}}(\mathbb{A}')$ satisfies property MW is now a direct consequence of Lemma 10.19. □

Corollary 10.22 *Let $\mathbb{A}$ and $\mathbb{A}_i$ for $i = 1, \dots, k$ be G-sets with $\mathbb{A}$ computable. Suppose that $\mathcal{S}_{\mathrm{or}}(\mathbb{A})$ satisfies property MW and that $I_i : \mathbb{A}_i \to \overline{\mathsf{IS}}(\mathbb{A})$ is G-equivariant and has explicit finite fibers. Then the induced map $\bigsqcup_i \mathbb{A}_i \to \overline{\mathsf{IS}}(\mathbb{A})$ has explicit finite fibers, as does $\overline{\mathsf{IS}}\left(\bigsqcup_i \mathbb{A}_i\right) \to \overline{\mathsf{IS}}(\mathbb{A})$, and $\overline{\mathsf{IS}}\left(\bigsqcup_i \mathbb{A}_i\right)$ satisfies property MW.*

Proof It is apparent that, since I_i has explicit finite fibers, so does $\bigsqcup_i \mathbb{A}'_i \to \overline{\mathsf{IS}}(\mathbb{A})$. The rest of the corollary then follows directly from Lemma 10.21. □

Corollary 10.23 *Suppose $\mathbb{A}'$ is a computable G-set and $\mathcal{S}_{\mathrm{or}}(\mathbb{A}')$ satisfies property MW. Then*:

(1) *For $k = 2, 3, \dots,$ inductively define $\overline{\mathsf{IS}}_k(\mathbb{A}') := \overline{\mathsf{IS}}(\overline{\mathsf{IS}}_{k-1}(\mathbb{A}'))$, where $\overline{\mathsf{IS}}_1(\mathbb{A}') := \overline{\mathsf{IS}}(\mathbb{A}')$. The G-set $\overline{\mathsf{IS}}_k(\mathbb{A}')$ satisfies property MW.*

(2) *Here we use Notation 10.13. For $i = 1, 2, \dots,$ let $s_i \in \{\mathcal{S}_{\mathrm{or}}, \mathcal{S}_{\mathrm{un}}\}$. Set $S_1(\mathbb{A}') := s_1(\mathbb{A}') \subset \overline{\mathsf{IS}}(\mathbb{A}')$ and inductively define $S_k(\mathbb{A}') := s_k(S_{k-1}(\mathbb{A}')) \subset \overline{\mathsf{IS}}_k(\mathbb{A}')$. The G-set $S_k(\mathbb{A}')$ satisfies property MW. The natural map $S_k(\mathbb{A}') \subset \overline{\mathsf{IS}}_k(\mathbb{A}') \to \overline{\mathsf{IS}}_{k-1}(\mathbb{A}') \to \cdots \to \overline{\mathsf{IS}}(\mathbb{A}')$ is injective, where $\overline{\mathsf{IS}}_i(\mathbb{A}') \to \overline{\mathsf{IS}}_{i-1}(\mathbb{A}')$ is the canonical extension (Construction 10.20) of the identity map of $\overline{\mathsf{IS}}_{i-1}(\mathbb{A}')$.*

Proof (1) An application of Proposition 10.14 gives that $\overline{\mathsf{IS}}(\mathbb{A}')$ satisfies property MW. Suppose that, for $k \geq 2$, we have that $\overline{\mathsf{IS}}_{k-1}(\mathbb{A}')$ satisfies property MW. Since the identity map of $\overline{\mathsf{IS}}_{k-1}(\mathbb{A}')$ has explicit finite fibers, so does $\overline{\mathsf{IS}}_k(\mathbb{A}') \to \overline{\mathsf{IS}}_{k-1}(\mathbb{A}')$ by Lemma 10.21. By Lemma 10.19, $\overline{\mathsf{IS}}_k(\mathbb{A}')$ also satisfies property MW.

(2) The first conclusion follows from (1) since $S_k(\mathbb{A}') \subset \overline{\mathsf{IS}}_k(\mathbb{A}')$, and an MW-algorithm for $\overline{\mathsf{IS}}_k(\mathbb{A}')$ provides an MW-algorithm for $S_k(\mathbb{A}')$. To prove the injectivity statement by induction, we show that if $\mathbb{A}''$ is a G-set and $f : \mathbb{A}'' \to \overline{\mathsf{IS}}(\mathbb{A}')$ is injective, then the restriction of the canonical extension $s_i(\mathbb{A}'') \to \overline{\mathsf{IS}}(\mathbb{A}')$ is also injective for $s_i \in \{\mathcal{S}_{\mathrm{or}}, \mathcal{S}_{\mathrm{un}}\}$.

Suppose that $s_i = \mathcal{S}_{\mathrm{or}}$ and that $(a_1, \dots, a_r), (b_1, \dots, b_s) \in s_i(\mathbb{A}'')$. If the images $(f(a_1), \dots, f(a_r))$ and $(f(b_1), \dots f(b_s))$ under the canonical map are equal, then $f(a_k) = f(b_k)$, $k = 1, \dots, m = n$. Since f is injective, $a_k = b_k$. The case where $s_i = \mathcal{S}_{\mathrm{un}}$ is similar. The inductive proof is left to the reader. □

Remark 10.24 Via the natural map in Corollary 10.23(2), we sometimes view $S_k(\mathbb{A}')$ as a subset of $\overline{\mathsf{IS}}(\mathbb{A}')$.

Example 10.25 Let $\mathbb{A} = \mathcal{C}(F_n)$. A free factor system is a finite unordered set of free factors and so may be interpreted as an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A})$; cf Example 10.6. A special chain is an ordered set of free factor systems and so gives an element of $\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A})$. The set of special chains associated to ϕ gives an element of $\mathcal{S}_{\mathrm{un}}\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A})$. We may view $\mathcal{S}_{\mathrm{un}}(\mathbb{A})$, $\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A})$, and $\mathcal{S}_{\mathrm{un}}\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A})$ all as subsets of $\overline{\mathsf{IS}}(\mathbb{A})$ (Remark 10.24).

11 Our atoms

In this section we apply Section 10.3 to enlarge the set $\mathcal{C}(F_n)$ (Notation 7.20) to the sets of atoms that we will need for the remainder of the paper. See Lemma 11.2 and also Section 14.

To start, we need to know that some familiar sets are computable. Some proofs refer to the Stallings graph associated to a finitely generated subgroup of F_n; see Stallings [1983, 5.4] and Section 7.2. The proofs also use Lemmas 8.3 and 8.5, sometimes without explicit mention. Let $\mathcal{B}$ be a basis for F_n. The following objects are computable.

- **The group F_n** Elements of F_n are represented by words in $\mathcal{B} \sqcup \mathcal{B}^{-1}$ and can be enumerated using length. Multiplication is given by concatenation. An element is represented uniquely by a reduced word, which can be found by iteratively canceling a letter and its inverse. See Example 8.4.
- **The set of conjugacy classes of elements of F_n** The enumeration of F_n serves also as an enumeration of the set of conjugacy classes. An element is represented uniquely by a cyclically reduced word.
- **The set of finitely generated subgroups of F_n** An enumeration is given by an enumeration of finite sets of representatives of elements of F_n. Two such sets are equal if and only if they determine the same based Stallings graph with labels in $\mathcal{B}$, which can be found using the Stallings folding algorithm.
- **$\mathcal{C}(F_n)$** The enumeration of the set of finitely generated subgroups of F_n also serves as an iteration for $\mathcal{C}(F_n)$. Two representatives are equal in $\mathcal{C}(F_n)$ if and only if they determine the same (unbased) Stallings graph with labels in $\mathcal{B}$; see Lemma 7.10.
- **The group $\mathsf{Aut}(F_n)$** The set of endomorphisms of F_n is bijective with $(F_n)^{\mathcal{B}}$ under the map $\Theta \mapsto (b \mapsto \Theta(b))$. Using the Hopf property of F_n [Magnus et al. 1966, Theorem 2.13], an endomorphism Θ is an isomorphism if and only if it is surjective if and only if the based Stallings graph of $\langle \Theta(\mathcal{B}) \rangle$ is the rose with petals labeled by the elements of $\mathcal{B}$. Thus an enumeration of $\mathsf{Aut}(F_n)$ can be obtained using the enumeration for endomorphisms.
- **The group $\mathsf{Out}(F_n)$** Our enumeration of automorphisms serves also as an enumeration of outer automorphisms. Two automorphisms represent the same outer automorphisms if and only if they have the same action on conjugacy classes of words of length at most two in $\mathcal{B} \sqcup \mathcal{B}^{-1}$; see [Serre 1980].
- **The $\mathsf{Aut}(F_n)$-set F_n** Representative endomorphisms act on representative words.
- **The $\mathsf{Out}(F_n)$-set of conjugacy classes of elements of F_n** Representative endomorphisms act on representative words.
- **The $\mathsf{Aut}(F_n)$-set of finitely generated subgroups of F_n** Representative endomorphisms act on representative finite subsets of F_n.
- **The $\mathsf{Out}(F_n)$-set $\mathcal{C}(F_n)$** Representative endomorphisms act on representative finite subsets of F_n.
- **The $\mathsf{Out}(F_n)$-set $\overline{\mathsf{IS}}(\mathcal{C}(F_n))$** See Lemma 10.10(3).
- **The $\mathsf{Out}(F_n)$-sets $\mathcal{S}_{\mathrm{or}}(\mathcal{C}(F_n))$ and $\mathcal{S}_{\mathrm{un}}(\mathcal{C}(F_n))$** See Corollary 10.23.

Notation 11.1 In the remainder of the paper, $G = \mathsf{Out}(F_n)$. We now define $\mathsf{Out}(F_n)$-sets $\mathbb{A}_0, \mathbb{A}_1, \dots, \mathbb{A}_6$ that will be used to express our algebraic invariants for elements of $\mathsf{UPG}(F_n)$. At the same time we show each $\mathbb{A}_i$ is computable and admits a map that has explicit finite fibers to a previously defined set.

- $\mathbb{A}_0$ denotes $\mathcal{C}(F_n)$, is computable (see the above itemized list), and has been identified (Notation 10.13) as an $\mathsf{Out}(F_n)$-subset of $\overline{\mathsf{IS}}(\mathbb{A}_0)$.

- $\mathbb{A}_1$ denotes the set of good conjugacy pairs of nontrivial finitely generated subgroups of F_n. We saw above that the set of finitely generated subgroups of F_n is computable, hence (Lemma 8.5(2)) so is its square. By Lemma 7.14(1), a pair (K_1, K_2) represents a good $[K_1, K_2]$ if and only if $\operatorname{rank}(\langle K_1, K_2\rangle) = \operatorname{rank}(K_1) + \operatorname{rank}(K_2)$, and this can be checked using the Stallings graphs of $[K_1]$, $[K_2]$ and $[\langle K_1, K_2\rangle]$. Hence (Lemma 8.5(1)) the subset of pairs representing good conjugacy pairs is computable. By Corollary 7.18, there is an algorithm deciding whether good conjugacy pairs are equal. Hence, by Lemma 8.3, $\mathbb{A}_1$ is computable.

By Corollary 7.23, the map $\mathbb{A}_1 \to \mathcal{S}_{\mathrm{or}}(\mathbb{A}_0) \hookrightarrow \overline{\mathsf{IS}}(\mathbb{A}_0)$ has explicit finite fibers where $\mathbb{A}_1 \to \mathcal{S}_{\mathrm{or}}(\mathbb{A}_0)$ is given by $[H_1, H_2] \mapsto ([H_1], [H_2], [\langle H_1, H_2\rangle])$.

- $\mathbb{A}_2$ denotes the set of conjugacy pairs $[H, a]$ where H is a nontrivial finitely generated subgroup of F_n, $a \in F_n$ is nontrivial, and $[H, \langle a\rangle]$ is good. It is clear that $[H, a] = [H', a']$ if and only if $[H, \langle a\rangle] = [H', \langle a'\rangle]$ and a and a' are conjugate. In particular, $\mathbb{A}_2$ is computable and the map $\mathbb{A}_2 \to \mathbb{A}_1$ given by $[H, a] \mapsto [H, \langle a\rangle]$ has explicit finite fibers with fibers of size zero or two.

- $\mathbb{A}_3$ denotes the set of conjugacy pairs $[a, H]$ where H is a nontrivial finitely generated subgroup of F_n, $a \in F_n$ is nontrivial, and $[H, \langle a\rangle]$ is good. That $\mathbb{A}_3$ is computable and $\mathbb{A}_3 \to \mathbb{A}_1$ given by $[a, H] \mapsto [\langle a\rangle, H]$ has explicit finite fibers with fibers of size zero or two follows exactly as with $\mathbb{A}_2$.

- $\mathbb{A}_4$ is the set of conjugacy pairs $[a, b]$ of elements of F_n where $\langle a, b\rangle$ has rank 2. We have $[a, b] = [a', b']$ if and only if $[\langle a\rangle, \langle b\rangle] = [\langle a'\rangle, \langle b'\rangle]$, a and a' are conjugate, and b and b' are conjugate. It follows that $\mathbb{A}_4$ is computable and $\mathbb{A}_4 \to \mathbb{A}_1$ given by $[a, b] \mapsto [\langle a\rangle, \langle b\rangle]$ has explicit finite fibers with fibers of size zero or four.

- $\mathbb{A}_5$ is the set of conjugacy classes $[a]$ of nontrivial elements $a \in F_n$. We saw earlier in this subsection that $\mathbb{A}_5$ is computable. The map $\mathbb{A}_5 \to \mathbb{A}_0$ given by $[a] \mapsto [\langle a\rangle]$ has explicit finite fibers with fibers of size zero or two.

- $\mathbb{A}'_6$ is the set of conjugacy pairs $[H, A]$ with $A < H < F_n$ all finitely generated and nontrivial. (In particular, $[H, A]$ is not good.) Using Lemma 7.24(1), the proof that $\mathbb{A}'_6$ is computable is similar to the proof that $\mathbb{A}_1$ is computable. By Lemma 7.24(2), $\mathbb{A}'_6 \to \mathcal{S}_{\mathrm{or}}(\mathbb{A}_0) \hookrightarrow \overline{\mathsf{IS}}(\mathbb{A}_0)$ given by $[H, A] \mapsto ([H], [A])$ has explicit finite fibers. $\mathbb{A}'_6$ is only used to define $\mathbb{A}_6$.

- $\mathbb{A}_6$ is the set of conjugacy pairs $[H, a]$ where H is a nontrivial finitely generated subgroup of F_n and $a \neq 1$ is in H. (In particular, $[H, \langle a\rangle]$ is not good.) $[H, a] = [H', a']$ if and only if $[H, \langle a\rangle] = [H', \langle a'\rangle]$ and a is conjugate to a'. Hence $\mathbb{A}_6 \to \mathbb{A}'_6$ has explicit finite fibers with fibers of size zero or two.

- $\mathbb{A}_\bullet := \mathbb{A}_0 \sqcup \mathbb{A}_1 \sqcup \mathbb{A}_2 \sqcup \mathbb{A}_3 \sqcup \mathbb{A}_4 \sqcup \mathbb{A}_5 \sqcup \mathbb{A}_6$.

Lemma 11.2 *Using Notation 11.1, the $\mathsf{Out}(F_n)$-set $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ satisfies property MW.*

Proof Since $\mathcal{S}_{\mathrm{or}}(\mathbb{A}_0)$ satisfies property MW by Theorem 10.2, it follows from Corollary 10.22 that it is enough to show that, for each i, we have that $\mathbb{A}_i$ admits G-equivariant map to $\overline{\mathsf{IS}}(\mathbb{A}_0)$ that has explicit finite fibers. Using Notation 11.1, we see that each $\mathbb{A}_i$ admits a map to $\overline{\mathsf{IS}}(\mathbb{A}_0)$ that is a composition of two maps, each of which has explicit finite fibers. We are done by Lemma 10.18. □

12 List of dynamical invariants

In Section 13 we define algebraic invariants of $\phi \in \mathsf{UPG}(F_n)$ that are derived from the dynamical invariants of ϕ established in the first five sections of this paper. For the convenience of the reader, we list those dynamical invariants here and provide pointers to the relevant sections of the paper. Here $f: G \to G$ always denotes a CT for ϕ and $\Gamma(f)$ its eigengraph; see Section 4.1. We also use the notation of conjugacy pairs; recall Definition 4.9, Examples 4.10 and Section 7.

- $\mathcal{P}(\phi)$ denotes the set of principal automorphisms for ϕ (Definition 3.8) and $[\mathcal{P}(\phi)]$ denotes the set of isogredience classes in $\mathcal{P}(\phi)$ (Definition 3.11). $[\mathcal{P}(\phi)]$ parametrizes the components of $\Gamma(f)$. $\mathsf{Fix}(\phi) = \{[\mathsf{Fix}(\Phi)] \mid [\Phi] \in [\mathcal{P}(\phi)]\}$. Since $[\mathcal{P}(\phi)]$ is finite, $\mathsf{Fix}(\phi)$ is a finite multiset of (possibly trivial) conjugacy classes of finitely generated subgroups of F_n. Geometrically it is the core of $\Gamma(f)$. $\mathsf{Fix}_{\geq 2}(\phi) := \{[\mathsf{Fix}(\Phi)] \mid [\Phi] \in [\mathcal{P}(\phi)], \mathrm{rank}(\mathsf{Fix}(\Phi)) \geq 2\}$. See Sections 3.4, 3.7 and 4.2.
- We use $\mathfrak{c} = \vec{\mathcal{F}}(\phi, <_T)$ to denote a special chain for ϕ as in Notation 6.8. It is a set of free factor systems naturally ordered by $\sqsubset$. We usually work with a prechosen $\mathfrak{c}$. For example, the filtration of our CT $f: G \to G$ will usually realize $\mathfrak{c}$. If $\mathcal{F} \in \mathfrak{c}$ (resp. $[F] \in \mathcal{F} \in \mathfrak{c}$) and if the filtration of $f: G \to G$ realizes $\mathfrak{c}$ then $f|\mathcal{F}$ (resp. $f|[F]$) denotes the restriction of f to the core filtration element representing $\mathcal{F}$ (resp. the component of the core filtration element representing $[F]$). The corresponding eigengraph is denoted by $\Gamma(f|\mathcal{F})$ (resp. $\Gamma(f|[F])$).
- A free factor system is special if it is in some special chain. $\mathfrak{L}(\phi)$ denotes the set of special free factor systems of ϕ; see Notation 6.8. Each element of $\mathfrak{L}(\phi)$ is a free factor system and so is a set of conjugacy classes of free factors in F_n. If $[F] \in \mathcal{F} \in \mathfrak{L}(\phi)$ then F and $[F]$ are also said to be special. The unique minimal (with respect to $\sqsubset$) element of $\mathfrak{L}(\phi)$, denoted by $\mathcal{F}_0(\phi)$, is the linear free factor system of ϕ. It is represented by the core of the subgraph of G that is the union of fixed and linear edges. An invariant description of $\mathcal{F}_0(\phi)$ is $\mathcal{F}(\mathsf{Fix}(\phi))$, ie the smallest free factor system carrying $\mathsf{Fix}(\phi)$; see Lemma 6.16.
- Define
$$\mathcal{R}(\phi) := \left\{[P] \;\middle|\; P \in \bigcup_{i=1}^{m} \mathsf{Fix}_+(\Phi_i)\right\} \subset \partial F_n / F_n,$$
where the Φ_i are representatives of the isogredience classes in $\mathcal{P}(\phi)$. In other words, $\mathcal{R}(\phi)$ is the set of conjugacy classes of points in ∂F_n that are isolated fixed points for some principal lift of ϕ. See Section 3.4. In any CT $f: G \to G$ representing ϕ there is a bijection $r \leftrightarrow E$ between $\mathcal{R}(\phi)$ and the set $\mathcal{E}_f$ of higher-order edges of G. The eigenray R_E has terminal end r.

- Let $\mathfrak{e} \in \mathfrak{c}$ denote a special 1-edge extension in $\mathfrak{c}$, ie $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ is a pair of consecutive elements of $\mathfrak{c}$. Suppose $f : G \to G$ realizes $\mathfrak{c}$. $\Gamma(f|\mathcal{F}^+) \setminus \Gamma(f|\mathcal{F}^-)$ has one or two ends; these represent the *new* (*with respect to* $\mathfrak{e}$) elements of $\mathcal{R}(\phi)$. The 1-edge extension $\mathfrak{e}$ has type H, HH or LH. There are two new elements if and only if $\mathfrak{e}$ has type HH. A new element is often denoted by r^+. Further, $\mathfrak{e}$ can be contractible, infinite cyclic, or large. See Section 6.1.

- Continuing the previous bullet point, if the filtration of $f : G \to G$ realizes $\mathfrak{c}$, then $\Gamma(f|\mathcal{F}^+)$ carries more lines than $\Gamma(f|\mathcal{F}^-)$. The set of *added lines with respect to* $\mathfrak{e}$, denoted by $\mathsf{L}_{\mathfrak{e}}(\phi)$, is a ϕ-invariant subset of these lines. See Definition 6.14.

- $\Omega(\phi) = \bigcup_{r \in \mathcal{R}(\phi)} \Omega(r)$ denotes the finite set of limit lines for ϕ. See Section 5. Here $\Omega(r)$ denotes the accumulation set of r or equivalently of the eigenray in $\Gamma(f)$ representing r. The elements of $\Omega(\phi)$ are all represented as lines in $\Gamma(f)$. $\Omega_{\mathsf{NP}}(\phi) \subset \Omega(\phi)$ is the subset of nonperiodic lines.

- $\mathcal{A}_{\mathrm{or}}(\phi)$ denotes the set of oriented axes of ϕ, where a root-free conjugacy class $[a]$ of an element of F_n is an axis if it has more than one representation in $\Gamma(f)$. An axis has an invariant description: $[a]$ is an axis if there are $\Phi_1, \Phi_2 \in \mathcal{P}(\phi)$ such that $a \in \mathsf{Fix}(\Phi_1) \cap \mathsf{Fix}(\Phi_2)$ and $\Phi_1 \neq \Phi_2$. In this case, we say the conjugacy pair $[\Phi, a]$ is a strong axis; see Definition 4.11. It is represented geometrically as a lift to $\Gamma(f)$ of $[a]$. The set of strong axes is denoted by $\mathsf{SA}(\phi)$. Associated to each pair of strong axes $\alpha_1 = [\Phi_1, a], \alpha_2 = [\Phi_2, a]$ is a twist coordinate $\tau(\alpha_1, \alpha_2)$ in $\mathbb{Z}$. See Definition 4.14.

13 Algebraic data associated to invariants

In this section we define algebraic versions of some of our dynamical invariants. We also explain how the algebraic versions can be computed and viewed as an element of $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$; see Notation 11.1. The algebraic invariants are typically weaker than their dynamic versions. However they have the advantage that they are iterated sets and so fit into the framework of Section 10. Some of our invariants, for example chains, are already algebraic in nature and so need no modification.

All of our algebraic invariants for $\phi \in \mathsf{UPG}$ will be computed using a CT $f : G \to G$ for ϕ; see Section 3.6. Additionally, the core of the eigengraph $\Gamma(f)$ can be computed from $f : G \to G$; see Section 4.1. In fact, since $\Gamma(f)$ is obtained from its core by adding the eigenrays of f and the eigenrays have a simple form (Section 3.6), we can compute arbitrarily large neighborhoods of the core in $\Gamma(f)$.

13.1 Special chains

Recall from Notation 6.1 and Lemma 6.2 that there is a canonical partial order $(\mathcal{R}(\phi), <)$ that can be computed from any CT for ϕ. Hence all extensions of $<$ to a total order $<_T$ can also be computed. The special chain $\vec{\mathcal{F}}(\phi, <_T)$ for ϕ can also be computed from any CT for ϕ; see Notation 6.8. A special chain is an element of $\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A}_0) \subset \overline{\mathsf{IS}}(\mathbb{A}_0) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$; see Example 10.25 and Corollary 10.23(2). Similarly, the set of all special chains for ϕ and the set $\mathfrak{L}(\phi)$ of all special free factor systems for ϕ can be computed from any CT for ϕ. Note that the former set is in $\mathcal{S}_{\mathrm{un}}\mathcal{S}_{\mathrm{or}}\mathcal{S}_{\mathrm{un}}(\mathbb{A}_0) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ and $\mathfrak{L}(\phi) \in \mathcal{S}_{\mathrm{un}}\mathcal{S}_{\mathrm{un}}(\mathbb{A}_0) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. We will tacitly use Corollary 10.23(2) throughout the rest of Section 13.

- Throughout the rest of Section 13, $\phi \in \mathsf{UPG}$, $\mathfrak{c}$ denotes a special chain for ϕ, and $f: G \to G$ denotes a CT that represents ϕ, satisfies (Inheritance), and realizes $\mathfrak{c}$.

13.2 Fix(ϕ)

The multiset $\mathsf{Fix}(\phi) := \{[\mathsf{Fix}(\Phi)] \mid [\Phi] \in [\mathcal{P}(\phi)]\}$ is already algebraic and is an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A}_0) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. As reviewed in Section 12, $\mathsf{Fix}(\phi)$ is represented by the core of $\Gamma(f)$ and so can be computed.

13.3 Axes

The set $\mathcal{A}_{\mathrm{or}}(\phi)$ of oriented axes of ϕ (Definition 4.5) is already algebraic and is an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A}_5) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. In terms of $f: G \to G$, $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$ if and only if either $[a]$ or $[a^{-1}]$ is represented by a twist path, which can be found by inspecting the linear edges of G; see Section 3.6. In terms of $\Gamma(f)$, $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$ if and only if a is root-free and represented by more than one circuit in the core of $\Gamma(f)$ with at least one representative embedded.

13.4 Algebraic rays

Remark 13.1 If F is a free factor and $\tilde{r} \in \partial F$ then we say that $[F]$ *carries* r. Equivalently, if G is a marked graph and $H \subset G$ is a core subgraph representing $[F]$, then there is a ray in $\widetilde{G}$ that converges to $\tilde{r}$ and projects into H. In the case that concerns us, $r \in \mathcal{R}(\phi)$ corresponds to some $E \in \mathcal{E}_f$ (see Lemma 3.26) and $[F]$ is a component of a free factor system in $\mathfrak{c}$. If C is the component of the core filtration element of G corresponding to $[F]$ then $[F]$ carries r if and only if some subray of R_E is contained in C. By construction, $R_E = E \cdot u \cdot f_\#(u) \cdot \ldots$, where the closed path u satisfies $f(E) = E \cdot u$. The height of $f_\#^k(u)$ is independent of k, so r is carried by $[F]$ if and only if $u \subset C$.

- (algebraic rays) For $r \in \mathcal{R}(\phi)$, $F_\mathfrak{c}(r)$ denotes the minimal special free factor $[F] \in \mathcal{F} \in \mathfrak{c}$ carrying r. If $\tilde{r} \in \partial F_n$ is a lift of r then we also write $F_\mathfrak{c}(\tilde{r}) := F$, where F is the unique representative of $F_\mathfrak{c}(r)$ that contains $\tilde{r}$. An algebraic ray $F_\mathfrak{c}(r)$ is an element of $\mathbb{A}_0 \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$.

Remark 13.2 Continuing Remark 13.1, $F_\mathfrak{c}(r)$ is represented by the minimal component C of a core filtration element of G containing u. In particular, we can compute $F_\mathfrak{c}(r)$ from our CT f. In our running example (see pages 1700, 1707, 1725 and 1729), $\mathcal{R}(\phi) = \{r_c, r_d, r_e, r_q\}$, the only relation is $r_c < r_q$, and the choice of total order is $r_c <_T r_d <_T r_e <_T r_q$. We have $[F_\mathfrak{c}(r_q)] = [\langle a, b, c\rangle]$.

Remark 13.3 We could work with all chains and define $F(r)$ to be the minimal special free factor $[F] \in \mathcal{F} \in \mathfrak{L}(\phi)$ carrying r. This would cause some extra work later in Lemma 17.19.

13.5 Algebraic lines

Recall that $[\cdot,\cdot]$ denotes a conjugacy pair (see Definition 4.9, Examples 4.10 and Section 7) and, for nontrivial $a \in F_n$, $a^+ \in \partial F_n$ (resp. a^-) denotes the attractor (resp. repeller) of $i_a|\partial F_n$; see the beginning of Section 3.4. Recall also that a line L is principal with respect to ϕ if there is a lift $\widetilde{L}$ whose endpoints are contained in $\mathsf{Fix}_\mathsf{N}(\Phi)$ for some $\Phi \in \mathcal{P}(\phi)$. Equivalently, L lifts into the eigengraph $\Gamma(f)$; see Lemma 4.1. In this section, we define an algebraic version $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ for a certain principal line L and associate to $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ a set of lines containing L that in turn determines $\mathsf{H}_{\phi,\mathfrak{c}}(L)$.

Definition 13.4 (algebraic lines) Suppose that L is a nonperiodic principal line for ϕ and that the nonperiodic ends of L are contained in $\mathcal{R}(\phi)$. There are four possibilities.

- [P-P] L has type P-P if some (hence every) lift $\widetilde{L}$ has the form (a^-, b^+) for some root-free $a, b \in F_n$ with $a \neq b^{\pm 1}$ in F_n. In particular, a and b are nontrivial. $\mathsf{H}_{\phi,\mathfrak{c}}(L) := [a,b]$. To $[a,b]$ we associate $\{L\}$. We also define $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}) := (a,b)$ and associate to it $\{(a^-, b^+)\}$. In this case $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ determines L.
- [P-NP] L has type P-NP if some (hence every) lift $\widetilde{L}$ has the form $(a^-, \widetilde{r})$ for some root-free $a \in F_n$ and a lift $\widetilde{r}$ of some $r \in \mathcal{R}(\phi)$. $\mathsf{H}_{\phi,\mathfrak{c}}(L) := [a, F_\mathfrak{c}(\widetilde{r})]$. To $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ we associate the set of lines $[a^-, \partial F_\mathfrak{c}(\widetilde{r})]$. $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}) := (a, F_\mathfrak{c}(\widetilde{r}))$ and has the associated set of lines $(a^-, \partial F_\mathfrak{c}(\widetilde{r}))$.
- [NP-P] L has type NP-P if some (hence every) lift $\widetilde{L}$ has the form $(\widetilde{r}, b^+)$ for some root-free $b \in F_n$ and a lift $\widetilde{r}$ of some $r \in \mathcal{R}(\phi)$. $\mathsf{H}_{\phi,\mathfrak{c}}(L) := [F_\mathfrak{c}(\widetilde{r}), b]$. To $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ we associate the set of lines $[\partial F_\mathfrak{c}(\widetilde{r}), b^+]$. $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}) := (F_\mathfrak{c}(\widetilde{r}), b)$ with associated set of lines $(\partial F_\mathfrak{c}(\widetilde{r}), b^+)$.
- [NP-NP] L has type NP-NP if some (hence every) lift $\widetilde{L}$ has the form $(\widetilde{r}, \widetilde{s})$ for lifts $\widetilde{r}$ of $r \in \mathcal{R}(\phi)$ and $\widetilde{s}$ of $s \in \mathcal{R}(\phi)$. $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}) := [F_\mathfrak{c}(\widetilde{r}), F_\mathfrak{c}(\widetilde{s})]$. To $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L})$ we associate the set of lines $[\partial F_\mathfrak{c}(\widetilde{r}), \partial F_\mathfrak{c}(\widetilde{s})]$. $\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}) := (F_\mathfrak{c}(\widetilde{r}), F_\mathfrak{c}(\widetilde{s}))$ with associated set of lines $(\partial F_\mathfrak{c}(\widetilde{r}), \partial F_\mathfrak{c}(\widetilde{s}))$.

Lemma 13.5 *Suppose that L lifts to $\Gamma(f)$ and has one of the types* P-P, P-NP, NP-P, *or* NP-NP. *Then with notation as above*:

- [P-P] $[\langle a\rangle, \langle b\rangle]$ *is a good conjugacy pair.*
- [P-NP] $[\langle a\rangle, F_\mathfrak{c}(\widetilde{r})]$ *is a good conjugacy pair.*
- [NP-P] $[F_\mathfrak{c}(\widetilde{r}), \langle b\rangle]$ *is a good conjugacy pair.*
- [NP-NP] $[F_\mathfrak{c}(\widetilde{r}), F_\mathfrak{c}(\widetilde{s})]$ *is a good conjugacy pair.*

Proof [P-P] Since $a \neq b^{\pm 1}$ and a and b are root-free, $\langle a, b\rangle$ is a free group of rank 2. In particular, $[\langle a\rangle, \langle b\rangle]$ is good by Lemma 7.14(1).

[P-NP] Suppose L has the lift $\widetilde{L} = (a^-, \widetilde{r})$. Set $A = \langle a\rangle$. Since L lifts to $\Gamma(f)$, $L = \alpha^\infty \sigma R_E$ for some $E \in \mathcal{E}_f$, where α is a circuit in the core of $\Gamma(f)$ representing $[a]$ (Lemma 4.2). We choose σ to have minimal length. Let $\star$ be the terminal vertex of E and let $\widetilde{\star}$ be the terminal vertex of the unique lift $\widetilde{E}$ of E in $\widetilde{L}$.

The based labeled graph $(C,\star)$ as in Remark 13.2 immerses to $(G,\star)$, and similarly the based labeled graph $(G_A,\star)$ that is a lollipop formed by the union of a circle labeled α and a segment labeled σE immerses to $(G,\star)$. If we define $(H,\star)$ to be the one-point union of $(G_A,\star)$ and $(C,\star)$ then by construction, the immersions of $(G_A,\star)$ and $(C,\star)$ to $(G_A,\star)$ induce a map of $H\to G$ that does not admit any Stallings folds and so, by [Stallings 1983, Proposition 5.3], induces an injection on the level of fundamental groups. We now have an identification of $(A, F_{\mathfrak{c}}(\tilde r), \langle A, F_{\mathfrak{c}}(\tilde r)\rangle)$ and $(\pi_1(G_A,\star), \pi_1(C,\star), \pi_1(H,\star))$. By Van Kampen, we see that $\langle A, F_{\mathfrak{c}}(\tilde r)\rangle$ is the internal free product of A and $F_{\mathfrak{c}}(\tilde r)$.

The cases [NP-P] and [NP-NP] are similar. □

Remark 13.6 $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ is an element of $\mathbb{A}_1\sqcup\mathbb{A}_2\sqcup\mathbb{A}_3\sqcup\mathbb{A}_4\subset\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. Not all lines that lift to $\Gamma(f)$ are assigned a type. For each L that has a type, $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ can be recovered from its associated set of lines. In the NP-NP case, this is a direct application of Corollary 7.5, Remark 7.6 and Lemma 7.14(2). The obvious modification needed for the other cases where $\langle a\rangle$ is replaced by a is left to the reader. We often conflate $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ with its associated set of lines.

Example 3.1 (continued) If L is the upward line represented by the contractible component of $\Gamma(f)$ in Figure 2 then L has type NP-NP and $\mathsf{H}_{\phi,\mathfrak{c}}(L)=[\langle a,b\rangle, \langle a,b\rangle^{d^{-1}e}]$ consists of the set of lines in the graph in Figure 4 that cross $d^{-1}e$ once and $e^{-1}d$ not at all.

Lemma 13.7 *Suppose that $\mathfrak{c}$ is a special chain for ϕ, that L is a nonperiodic principal line for ϕ whose nonperiodic ends are contained in $\mathcal{R}(\phi)$ and that $\theta\in\mathsf{Out}(F_n)$. Then $\theta(\mathfrak{c})$ is a special chain for ϕ^θ, $\theta(L)$ is a nonperiodic principal line for ϕ^θ whose nonperiodic ends are contained in $\mathcal{R}(\phi^\theta)$ and $\theta(\mathsf{H}_{\phi,\mathfrak{c}}(L))=\mathsf{H}_{\phi^\theta,\theta(\mathfrak{c})}(\theta(L))$.*

Proof We will do the case P-NP; the others are similar. Lemmas 6.13 and 3.16 imply that $\theta(\mathfrak{c})$ is a special chain for ϕ^θ and that $\theta(L)$ is a principal line for ϕ^θ. If $\Theta\in\theta$ and $\tilde L=(a^-,\tilde r)$, then $\Theta(\tilde L)=(\Theta(a),\Theta(\tilde r))$ and

$$\Theta(\mathsf{H}_{\phi,\mathfrak{c}}(\tilde L))=[\Theta(a),\Theta(F_{\mathfrak{c}}(\tilde r))]=[\Theta(a),F_{\Theta(\mathfrak{c})}(\Theta(\tilde r))]=\mathsf{H}_{\phi^\theta,\theta(\mathfrak{c})}(\Theta(\tilde L)).$$ □

Remark 13.8 For applications, it follows from definitions that if $\theta(\mathfrak{c})=\mathfrak{c}$ and $\mathfrak{e}\in\mathfrak{c}$ then $\theta(\mathfrak{e})=\mathfrak{e}$.

In the next lemma we abuse notation and identify $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ with its associated set of lines; see Remark 13.6.

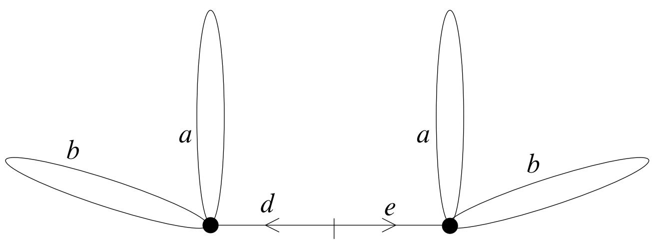


Figure 4

Lemma 13.9 *Suppose $f : G \to G$ is a CT for ϕ that realizes $\mathfrak{c}$. If L is either an element of $\Omega_{\mathsf{NP}}(\phi)$ or an element of $\mathsf{L}_{\mathfrak{c}}(\phi)$ such that $\mathfrak{c}$ is not* large, *then L is the only line in $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ that lifts to $\Gamma(f)$.*

Proof We assume at first that $L \in \Omega_{\mathsf{NP}}(\phi)$. By Corollary 5.17(4), all elements of $\Omega_{\mathsf{NP}}(\phi)$ lift to $\Gamma(f)$ and further, by Corollary 5.17(1), have one of the types P-P, P-NP, NP-P or NP-NP. In particular, $\mathsf{H}_{\phi,\mathfrak{c}}(L)$ is defined. We will do the case that $L = [a^-, \tilde{r}]$ is P-NP, the others being similar. By Corollary 5.17(1), L has the form $(R_1)^{-1} R_E$, where R_1 consists of only linear and fixed edges, and where $E \in \mathcal{E}_f$. In particular, E is the highest edge of L.

Every line $L' \in \mathsf{H}_{\phi,\mathfrak{c}}(L)$ has a lift of the form $\tilde{L}' = (a^-, \tilde{s})$, where $\tilde{s} \in F_{\mathfrak{c}}(\tilde{r})$. Since $f : G \to G$ realizes $\mathfrak{c}$, $F_{\mathfrak{c}}(r)$ is represented by a subgraph of G whose edges are all lower than E. If $\tilde{r} = \tilde{s}$ then $L' = L$ and we are done. We may therefore assume $\tilde{L}'' := (\tilde{r}, \tilde{s})$ is a line with both endpoints in $\partial F_{\mathfrak{c}}(\tilde{r})$. Thus L'' only crosses lifts of edges that are lower than E. It follows that $L' = (R_1)^{-1} E R_2$, where the ray R_2 only crosses edges that are lower than E. If L' lifts to $\Gamma(f)$ then E is the first higher-order edge it crosses and so $E R_2 = R_E$ and $L' = L$. This completes the proof when $L \in \Omega_{\mathsf{NP}}(\phi)$.

We now assume that $L \in \mathsf{L}_{\mathfrak{c}}(\phi)$ and that $\mathfrak{c}$ is not large. If $\mathfrak{c}$ is contractible then, because L lifts to $\Gamma(f)$, L has the form $[\tilde{r}, \tilde{s}] = R_{E_1}^{-1} \rho R_{E_2}$, where ρ is a Nielsen path. By definition $L' \in \mathsf{H}_{\phi,\mathfrak{c}}(L)$ has a representative $\tilde{L}' = (\tilde{r}_1, \tilde{r}_2)$ such that either $\tilde{r}_1 = \tilde{r}$ or $(\tilde{r}_1, \tilde{r})$ is a line with both endpoints in $\partial F_{\mathfrak{c}}(\tilde{r})$ and such that either $\tilde{r}_2 = \tilde{s}$ or $(\tilde{s}, \tilde{r}_2)$ is a line with both endpoints in $\partial F_{\mathfrak{c}}(\tilde{s})$. We argue as above to conclude that $L' = L$ if L' lifts to $\Gamma(f)$. The case where $\mathfrak{c}$ is infinite cyclic is similar. □

Lemma 13.10 *Suppose that L has one of the types* P-P, P-NP, NP-P *or* NP-NP *and that $\theta(\mathsf{H}_{\phi,c}(L)) = \mathsf{H}_{\phi,\mathfrak{c}}(L)$. If $\tilde{L}$ is a lift of L then there is a unique $\Theta \in \theta$ such that $\Theta(\mathsf{H}_{\phi,\mathfrak{c}}(\tilde{L})) = \mathsf{H}_{\phi,\mathfrak{c}}(\tilde{L})$.*

Proof The existence of at least one such Θ follows from the definitions.

(P-P) Suppose $\tilde{L} = (a^-, b^+)$. If $\Theta_1 \neq \Theta_2$ represent θ and leave (a^-, b^+) invariant then the difference $\Theta_1 \Theta_2^{-1}$ has the form i_x for some $x \neq 1$ in F_n and leaves (a^-, b^+) invariant. It follows that a, b and x share a power. This is impossible since L is not periodic.

(NP-P) Suppose $\tilde{L} = (\tilde{r}, b^+)$. If $\Theta_1 \neq \Theta_2$ in θ leave $(F_{\mathfrak{c}}(\tilde{r}), b^+)$ invariant, then $\Theta_1 \Theta_2^{-1} = i_x$ for some $x \neq 1$ in F_n, $i_x\big(F_{\mathfrak{c}}(\tilde{r})\big) = F_{\mathfrak{c}}(\tilde{r})$, and $i_x(b) = b$. Hence $x \in F_{\mathfrak{c}}(\tilde{r}) \cap \langle b \rangle$, which is impossible since $\partial F_{\mathfrak{c}}(\tilde{r}) \cap \partial \langle b \rangle = \varnothing$ by Lemma 13.5.

The cases P-NP and NP-NP are similar. □

13.6 Algebraic strong axes

- (algebraic strong axes) If $[\Phi, a]$ is a strong axis and $\mathsf{Fix}(\Phi)$ is not cyclic (equivalently $\mathrm{rank}(\mathsf{Fix}(\Phi)) > 1$), then $[\mathsf{Fix}(\Phi), a]$ is the associated algebraic strong axis.

Remark 13.11 In (algebraic strong axes) above we are focusing only on strong axes for the restriction of ϕ to its linear free factor system $\mathcal{F}_0(\phi)$; see the proof of Lemma 17.1. By definition of strong axes, $a \in \mathsf{Fix}(\Phi)$. In particular, $[\mathsf{Fix}(\Phi), a]$ is an element of $\mathbb{A}_6 \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ and is not good. The set of algebraic strong axes of ϕ is an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A}_6) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. By Lemma 4.20, $[\mathsf{Fix}(\Phi), a]$ determines $[\Phi, a]$.

Remark 4.13 can be used to compute the algebraic strong axes from our CT $f: G \to G$. Using the notation there, the set $\mathsf{SA}(\phi)$ of strong axes is in one-to-one correspondence with the set of nontrivial circuits in the core of $\Gamma(f)$ representing elements of $\mathcal{A}_{\mathrm{or}}(\phi)$. The strong axis $[\Phi_{a,0}, a]$ corresponding to $(f_{v_\#}, \tau)$ has algebraic invariant represented by $[\mathsf{Fix}(f_{v\#}), a]$. $\mathsf{Fix}(f_{v\#})$ is the image in $\pi_1(G, v)$ of $\pi_1(\Gamma(f), v)$, where $v \in \Gamma^0(f)$ is viewed as an element of G; see the construction of $\Gamma(f)$ in Section 4.1. The case for $(f_{v_j\#}, v_j)$ is similar.

In our running example (see page 1716), the algebraic invariants of α and α' are respectively represented by the pairs $(\langle a, a^b\rangle, a)$ and $(\langle a, a^b\rangle, a^b)$. The strong axis α'' doesn't have an algebraic invariant since the corresponding fix set is cyclic. This ends Remark 13.11.

13.7 Algebraic added lines

- (algebraic added lines with respect to $\mathfrak{e}$) We use notation as in Definition 6.14. $\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)$ is defined to be $\{\mathsf{H}_{\phi,\mathfrak{c}}(L) \mid L \in \mathsf{L}_{\mathfrak{e}}(\phi)\}$ if $\mathsf{L}_{\mathfrak{e}}(\phi)$ is finite, and is defined to be the singleton $\{[\mathsf{Fix}(\Phi), F_{\mathfrak{c}}(\tilde{r}^+)]\}$ otherwise. The proof that $[\mathsf{Fix}(\Phi), F_{\mathfrak{c}}(\tilde{r}^+)]$ is good is similar to the proof of Lemma 13.5. $\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)$ is an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A}_1 \sqcup \mathbb{A}_2 \sqcup \mathbb{A}_3) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$.

In terms of our CT $f: G \to G$, algebraic added lines can be computed as follows. We use the notation of Definition 6.14. Since we already know how to compute algebraic lines, in the cases where the number of added lines is finite, it is enough to describe the added lines. This is done in Definition 6.14 using the eigengraph $\Gamma(f|F^+)$. The case where there are infinitely many added lines is the intersection of [H] and large in Lemma 6.10. Let v be the initial vertex of the edge $E \in \mathcal{E}_f$ in $\Gamma(f|F^+)$ corresponding to r^+. Then $[\mathsf{Fix}(\Phi), F_{\mathfrak{c}}(\tilde{r}^+)]$ is represented by $(\mathsf{Fix}(f_{v\#}|F^+), \pi_1(C^E, v))$, where C is the connected, core subgraph of G representing $F_{\mathfrak{c}}(r^+)$ and C^E denotes the one-point union at the terminal endpoint of E of C and an edge labeled E (so that C^E immerses to G). See our running example on page 1730, where $\mathfrak{e}_2$ is contractible, $\mathfrak{e}_3$ is infinite cyclic, $\mathfrak{e}_1$ is large and

$$\mathsf{H}_{\mathfrak{e}_1\in\mathfrak{c}}(\phi) = \{[\langle a, a^b\rangle, \langle a, b\rangle^c]\},$$
$$\mathsf{H}_{\mathfrak{e}_2\in\mathfrak{c}}(\phi) = \{[\langle a, b\rangle, \langle a, b\rangle^{d^{-1}e}], [\langle a, b\rangle, \langle a, b\rangle^{e^{-1}d}]\},$$
$$\mathsf{H}_{\mathfrak{e}_3\in\mathfrak{c}}(\phi) = \{[a^{-1}, \langle a, b, c\rangle^{p^{-1}q}], [a, \langle a, b, c\rangle^{p^{-1}q}]\}.$$

13.8 Algebraic limit lines

- $\{\mathsf{H}_{\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}$ is the set of algebraic limit lines.

The set of algebraic limit lines is an element of $\mathcal{S}_{\mathrm{un}}(\mathbb{A}_1 \sqcup \mathbb{A}_2 \sqcup \mathbb{A}_3 \sqcup \mathbb{A}_4) \subset \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. As in the case of added lines, to compute algebraic limit lines, we only need to compute $\Omega_{\mathsf{NP}}(\phi)$ from our CT $f : G \to G$. This is done in Section 5; see Corollary 5.17. Referring to our running example (see page 1724), $\Omega_{\mathsf{NP}}(\phi) = \Omega_{\mathsf{NP}}(r_q) = \{a^\infty R_c, a^\infty b a^\infty\}$ and so $\mathsf{H}_{\mathfrak{c}}(\phi) = \{[a^{-1}, \langle a, b\rangle^c], [a^{-1}, a^b]\}$.

13.9 Naturality

We will need the following naturality statements.

Lemma 13.12 *Suppose* $\mathfrak{e} \in \mathfrak{c}$ *and* $\Theta \in \theta \in \mathsf{Out}(F_n)$. *Then*:

(1) $\theta(\{\mathsf{H}_{\phi,\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}) = \{\mathsf{H}_{\phi^\theta,\theta(\mathfrak{c})}(L') \mid L' \in \Omega_{\mathsf{NP}}(\phi^\theta)\}$.

(2) $\theta(\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)) = \mathsf{H}_{\theta(\mathfrak{e})\in\theta(\mathfrak{c})}(\phi^\theta)$.

(3) $[\mathsf{Fix}(\Phi), a] \leftrightarrow \theta([\mathsf{Fix}(\Phi), a]) = ([\mathsf{Fix}(\Phi^\Theta), \Theta(a)]$ *defines a bijection between the algebraic strong axes for* ϕ *and the algebraic strong axes for* ϕ^θ.

Proof (1) We have

$$\begin{aligned}\theta(\{\mathsf{H}_{\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}) &\overset{\text{def}}{=} \{\theta(\mathsf{H}_{\mathfrak{c}}(L)) \mid L \in \Omega_{\mathsf{NP}}(\phi)\} \\ &= \{\mathsf{H}_{\theta(\mathfrak{c})}(\theta(L)) \mid L \in \Omega_{\mathsf{NP}}(\phi)\} \quad \text{by Lemma 13.7,} \\ &= \{\mathsf{H}_{\theta(\mathfrak{c})}(L') \mid L' \in \Omega_{\mathsf{NP}}(\phi^\theta)\} \quad \text{by Corollary 5.4.}\end{aligned}$$

(2) If $\mathsf{L}_{\mathfrak{e}}(\phi)$ is finite,

$$\begin{aligned}\theta(\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)) &\overset{\text{def}}{=} \theta(\{\mathsf{H}_{\mathfrak{c}}(L) \mid L \in \mathsf{L}_{\mathfrak{e}}(\phi)\}) \\ &= \{\theta(\mathsf{H}_{\mathfrak{c}}(L)) \mid L \in \mathsf{L}_{\mathfrak{e}}(\phi)\}) \\ &= \{\mathsf{H}_{\theta(\mathfrak{c})}(\theta(L)) \mid L \in \mathsf{L}_{\mathfrak{e}}(\phi)\} \quad \text{by Lemma 13.7,} \\ &= \{\mathsf{H}_{\theta(\mathfrak{c})}(L') \mid L' \in \mathsf{L}_{\theta(\mathfrak{e})}(\phi^\theta)\} \quad \text{by Lemma 6.15,} \\ &\overset{\text{def}}{=} \mathsf{H}_{\theta(\mathfrak{e})\in\theta(\mathfrak{c})}(\phi^\theta).\end{aligned}$$

If $\mathsf{L}_{\mathfrak{e}}(\phi)$ is infinite,

$$\begin{aligned}\theta(\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)) &\overset{\text{def}}{=} \theta(\{[\mathsf{Fix}(\Phi), F_{\mathfrak{c}}(\tilde{r}^+)]\}) \\ &= \{[\Theta(\mathsf{Fix}(\Phi)), \Theta(F_{\mathfrak{c}}(\tilde{r}^+))]\} \\ &= \{[\mathsf{Fix}(\Phi^\Theta), F_{\theta(\mathfrak{c})}(\Theta(\tilde{r}^+))\} \quad \text{by Lemma 6.13,} \\ &\overset{\text{def}}{=} \mathsf{H}_{\theta(\mathfrak{e})\in\theta(\mathfrak{c})}(\phi^\theta).\end{aligned}$$

(3) Lemmas 4.17(2) and 3.16(2) imply that $(\Phi, a) \leftrightarrow (\Phi^\Theta, \Theta(a))$ induces a bijection $\mathsf{SA}(\phi, [a]) \leftrightarrow \mathsf{SA}(\psi, \theta([a]))$ and that the ranks of $\mathsf{Fix}(\Phi)$ and $\mathsf{Fix}(\Phi^\Theta)$ are equal. (3) therefore follows from

$$\theta([\mathsf{Fix}(\Phi), a]) \overset{\text{def}}{=} [\Theta(\mathsf{Fix}(\Phi)), \Theta(a)] = [\mathsf{Fix}(\Phi^\Theta), \Theta(a)].$$ □

13.10 The algebraic invariant of ϕ rel $\mathfrak{c}$

In this subsection we collect our algebraic invariants into a single master algebraic invariant.

Definition 13.13 Fix a special chain $\mathfrak{c}$ for $\phi \in \mathsf{UPG}$.

(1) The *algebraic invariant of* ϕ *rel* $\mathfrak{c}$ is the element of $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ that is the ordered set $\mathsf{I}_\mathfrak{c}(\phi)$ consisting of
- $\mathfrak{c}$,
- $\mathsf{Fix}(\phi)$,
- $(\mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi) \mid \mathfrak{e} \in \mathfrak{c})$, where the special 1-edge extensions $\mathfrak{e}$ are ordered using $\mathfrak{c}$,
- $\{\mathsf{H}_{\phi,\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}$,
- $\mathcal{A}_{\mathrm{or}}(\phi)$, and
- the set of algebraic strong axes for ϕ.

The six elements in the ordered list $\mathsf{I}_\mathfrak{c}(\phi)$ are the *components of* $\mathsf{I}_\mathfrak{c}(\phi)$.

(2) Order (noncanonically) the elements of the union of the six sets defining $\mathsf{I}_\mathfrak{c}(\phi)$. The resulting element of $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ is denoted by J.

Remark 13.14 In light of Theorem 3.20, we have seen in this section that the set of special chains, $\mathsf{I}_\mathfrak{c}(\phi)$, and J can be computed. We stress that they take values in $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$, which satisfies property MW by Lemma 11.2.

14 Stabilizers of algebraic invariants

At this point, it is reasonably straightforward to reduce the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$ to the problem of deciding whether $\phi, \psi \in \mathsf{UPG}(F_n)$ with $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$ are conjugate by some θ in the stabilizer of $\mathsf{I}_\mathfrak{c}(\phi)$. We want a little more. Namely, we want to restrict the set of potential conjugators to those elements that stabilize $\mathsf{I}_\mathfrak{c}(\phi)$ and induce trivial permutations on the components of $\mathsf{I}_\mathfrak{c}(\phi)$. Continuing with the notation of the previous section, we make this precise as follows.

Definition 14.1 ($\mathcal{X}_\mathfrak{c}(\phi)$) Let $\phi \in \mathsf{UPG}(F_n)$, $\mathfrak{c}$ be a special chain for ϕ, and let J be in Definition 13.13(2). Let $\mathcal{X}_\mathfrak{c}(\phi)$ denote the stabilizer $\mathsf{Out}_\mathsf{J}(F_n)$ of J in $\mathsf{Out}(F_n)$.

Unraveling definitions, $\mathcal{X}_\mathfrak{c}(\phi)$ also has a description as the subgroup of $\mathsf{Out}(F_n)$ fixing each element in the union of the following six sets:

(1) $\{[F] \mid [F] \in \mathcal{F} \in \mathfrak{c}\}$,

(2) $\mathsf{Fix}(\phi)$,

(3) $\bigcup_{\mathfrak{e}\in\mathfrak{c}} \mathsf{H}_{\mathfrak{e}\in\mathfrak{c}}(\phi)$,

(4) $\{\mathsf{H}_{\phi,\mathfrak{c}}(L) \mid L \in \Omega_{\mathsf{NP}}(\phi)\}$,

(5) $\mathcal{A}_{\mathrm{or}}(\phi)$, and

(6) $\{[\mathsf{Fix}(\Phi), a] \mid [\Phi, a] \in \mathsf{SA}(\phi), \operatorname{rank} \mathsf{Fix}(\Phi) \geq 2\}$.

Remark 14.2 As noted in Definition 13.13(2), the construction of J was noncanonical. That is, there were choices in its construction. Every choice has the same stabilizer and so $\mathcal{X}_{\mathfrak{c}}(\phi)$ is independent of choice.

In passing and for future use, we have the expected:

Lemma 14.3 $$\phi \in \mathcal{X}_{\mathfrak{c}}(\phi).$$

Proof We have to check that ϕ fixes each of the sets (1)–(6) above elementwise. In (1), (2), (5) and (6) this is immediate from definitions. For set (4), this is because $\phi(L) = L$ for all $L \in \Omega_{\mathsf{NP}}(\phi)$ (Corollary 5.17(4)) and Lemma 13.7. That the elements of set (3) are fixed follows from definitions and Lemmas 6.15 and 13.12(2). □

Definition 14.4 A group G *is of type* F if it has a finite Eilenberg–Mac Lane space. G *is of type* VF if it has a finite-index subgroup of type F.

Proposition 14.5 *The stabilizer* $\mathsf{Out}_{\mathsf{Y}}(F_n)$ *of an element* $\mathsf{Y} \in \overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ *is of type* VF.

Proof We saw in the proof of Lemma 11.2 that there is a map $\overline{\mathsf{IS}}(\mathbb{A}_\bullet) \to \overline{\mathsf{IS}}(\mathbb{A}_0)$ that has explicit finite fibers (Notation 10.16). If $\overline{\mathsf{Y}}$ is the image of Y, then $\mathsf{Out}_{\mathsf{Y}}(F_n)$ has finite index in $\mathsf{Out}_{\overline{\mathsf{Y}}}(F_n)$. By Corollary 10.15, the subgroup G of $\mathsf{Out}(F_n)$ fixing each label of $\overline{\mathsf{Y}}$ has finite index in $\mathsf{Out}_{\overline{\mathsf{Y}}}(F_n)$. Also, the subgroup G has type VF by [Bestvina et al. 2023, Theorem 1.1]. $\mathsf{Out}_{\mathsf{Y}}(F_n)$, being commensurate with a group of type VF, also has type VF. □

As usual, naturality will be important.

Lemma 14.6 *Suppose* $\xi \in \mathsf{Out}(F_n)$.

(1) $(\mathcal{X}_{\mathfrak{c}}(\phi))^{\xi} = \mathcal{X}_{\xi(\mathfrak{c})}(\phi^{\xi})$.

(2) $\xi(\mathsf{I}_{\mathfrak{c}}(\phi)) = \mathsf{I}_{\xi(\mathfrak{c})}(\phi^{\xi})$.

Proof (1) By Lemma 6.13, $\xi(\mathfrak{c})$ is special for ϕ^{ξ} and so the statement makes sense. The lemma follows easily from the naturality of the quantities appearing in Definition 14.1. For example, we verify that if $\theta \in \mathcal{X}_{\mathfrak{c}}(\phi)$, then $\theta^{\xi}(\mathsf{H}_{\phi^{\xi},\xi(\mathfrak{c})}(L)) = \mathsf{H}_{\phi^{\xi},\xi(\mathfrak{c})}(L)$ for all $L \in \Omega_{\mathsf{NP}}(\phi^{\xi})$. Indeed,

$$\xi\theta\xi^{-1}(\mathsf{H}_{\phi^{\xi},\xi(\mathfrak{c})}(L)) = \xi\theta(\mathsf{H}_{\phi,\mathfrak{c}}(\xi^{-1}(L))) = \xi(\mathsf{H}_{\phi,\mathfrak{c}}(\xi^{-1}(L))) = \mathsf{H}_{\phi^{\xi},\xi(\mathfrak{c})}(L),$$

where the first and third equalities use Lemma 13.7 and the second uses Corollary 5.4. The remainder of the proof consists of similar checks and is left to the reader.

The proof of (2) is similar. □

We next reduce the proof of the main result (Theorem 1.1) of this paper, ie the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$, to the proof of Proposition 14.7 stated immediately below.

Proposition 14.7 *There is an algorithm that takes as input* $\phi, \psi \in \mathsf{UPG}(F_n)$ *and a chain* $\mathfrak{c}$ *such that*

- $\mathfrak{c}$ *is special for both* ϕ *and* ψ, *and*
- $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$,

and that outputs `YES` *or* `NO` *depending whether or not there is a* $\theta \in \mathcal{X}_\mathfrak{c}(\phi)$ *conjugating* ϕ *to* ψ. *Further, if* `YES` *then such a* θ *is produced.*

Proposition 14.7 is proved in Sections 16 and 17 below.

Lemma 14.8 *Proposition 14.7 implies Theorem 1.1. That is, an algorithm that satisfies the conclusions of Proposition 14.7 can be used to produce an algorithm that satisfies the conclusions of Theorem 1.1.*

Proof Assume Proposition 14.7 holds and $\phi, \psi \in \mathsf{UPG}(F_n)$.

View the multiset $\mathsf{I}(\phi) := \{\mathsf{I}_\mathfrak{c}(\phi) \mid \mathfrak{c} \text{ is a special chain for } \phi\}$ as an element of $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ as described in Section 13.1. By Lemma 14.6(2), $\mathsf{I}(\phi)$ is a conjugacy invariant of ϕ. That is, if $\phi^\theta = \psi$ then $\theta(\mathsf{I}(\phi)) = \mathsf{I}(\cdot)$. We may compute $\mathsf{I}(\phi)$ and $\mathsf{I}(\psi)$; see Remark 13.14.

Since $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$ satisfies property MW (Lemma 11.2), we can algorithmically check if there is $\theta' \in \mathrm{Out}(F_n)$ such that $\theta'(\mathsf{I}(\phi)) = \mathsf{I}(\psi)$. If there is no such θ' then ϕ and ψ are not conjugate; return `NO`. If there is such a θ', then one is produced by the M-algorithm for $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$. Note that ϕ and ψ are conjugate in $\mathrm{Out}(F_n)$ if and only if ϕ and $\psi' := \theta'^{-1}\psi\theta'$ are conjugate in $\mathrm{Out}(F_n)$ if and only if ϕ and ψ' are conjugate in the stabilizer $G := \mathrm{Out}_{\mathsf{I}(\phi)}(F_n)$ of $\mathsf{I}(\phi)$.

G acts by permutation on the set of labels of $\mathsf{I}(\phi)$. Let $G' < G$ denote the subgroup fixing each label. By the M-algorithm for $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$, we may construct a finite presentation for G. Using our finite set of generators for G, we may construct the image Q of G in our permutation group. By Lemma 9.1, we can compute a finite set θ_i such that $G = \bigsqcup_i \theta_i G'$. Hence, ϕ and ψ' are conjugate in $\mathrm{Out}(F_n)$ if and only if ϕ and some $\theta_i^{-1}\psi'\theta_i$ are conjugate in $\mathrm{Out}(F_n)$ if and only if ϕ and some $\theta_i^{-1}\psi'\theta_i$ are conjugate in G'. Since $G' < \mathcal{X}_\mathfrak{c}(\phi) < \mathrm{Out}(F_n)$ for all $\mathfrak{c}$, ϕ and ψ' are conjugate in $\mathrm{Out}(F_n)$ if and only if ϕ and some $\theta_i^{-1}\psi'\theta_i$ are conjugate in $\mathcal{X}_\mathfrak{c}(\phi)$. We may use the supposed algorithm of Proposition 14.7 to decide whether or not this is the case and return a conjugator if it is. The returned conjugator allows us to compute a conjugator for ϕ and ψ. □

15 Staple pairs

15.1 Limit lines $\Omega_{\mathsf{NP}}(\phi, \tilde{r}) \subset \tilde{\mathcal{B}}$

In Section 5, we associated a finite set $\Omega_{\mathsf{NP}}(r) \subset \mathcal{B}$ of ϕ-invariant nonperiodic lines to each $r \in \mathcal{R}(\phi)$. In this section we associate, to each lift $\tilde{r}$ of r, a subset $\Omega_{\mathsf{NP}}(\phi, \tilde{r}) \subset \tilde{\mathcal{B}}$ of the full preimage of $\Omega_{\mathsf{NP}}(r)$ and then establish properties of $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ that will be needed later in the paper.

Definition 15.1 Choose a marked graph K. For each lift $\tilde{r} \in \partial F_n$ of $r \in \mathcal{R}(\phi)$, let $\Phi_{\tilde{r}}$ be the unique lift of ϕ that fixes $\tilde{r}$ and let $\tilde{R} \subset \tilde{K}$ be a ray with terminal end $\tilde{r}$. If $\tilde{L}$ is a lift of $L \in \Omega_{\mathsf{NP}}(r)$ then L is ϕ-invariant by Corollary 5.17(1) and so each $\Phi^j_{\tilde{r}}(\tilde{L})$ is a translate of $\tilde{L}$, say $\Phi^j_{\tilde{r}}(\tilde{L}) = T_j(\tilde{L})$ for some unique T_j. Define $\tilde{L}$ to be in $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ if for every finite subpath $\tilde{\beta}$ of $\tilde{L}$ there exists $J(\tilde{\beta})$ such that $T_j(\tilde{\beta}) \subset \tilde{R}$ — equivalently, $\tilde{\beta} \subset T_j^{-1}(\tilde{R})$ — for all $j \geq J(\tilde{\beta})$.

Remark 15.2 As defined, $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ depends on $\Phi_{\tilde{r}}$ and hence on ϕ, in contrast to $\Omega_{\mathsf{NP}}(r)$, which is independent of ϕ.

Lemma 15.3 *$\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ is well-defined and $\Phi_{\tilde{r}}$-invariant. Moreover, if $\psi = \theta\phi\theta^{-1}$ for some $\theta \in \mathsf{Out}(F_n)$ and if Θ is a lift of θ, then $\Theta(\Omega_{\mathsf{NP}}(\phi, \tilde{r})) = \Omega_{\mathsf{NP}}(\psi, \partial\Theta(\tilde{r}))$.*

Proof Replacing $\tilde{R}$ by a subray does not change $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$. Since any two rays with terminal end $\tilde{r}$ share a common subray, $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ is independent of the choice of $\tilde{R}$.

As defined above, $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ depends on the marked graph K so we write $\Omega_{\mathsf{NP}}(\phi, \tilde{r}, K)$ to make this explicit. We will prove:

(∗) $\Theta(\Omega_{\mathsf{NP}}(\phi, \tilde{r}, K)) = \Omega_{\mathsf{NP}}(\theta\phi\theta^{-1}, \Theta(\tilde{r}), K')$ *for any marked graphs K and K' and any $\Theta \in \mathsf{Aut}(F_n)$ representing any $\theta \in \mathsf{Out}(F_n)$.*

Applied with $\Theta =$ identity, (∗) proves that $\Omega_{\mathsf{NP}}(\phi, \tilde{r}, K)$ is independent of K and hence that $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ is well defined. The moreover statement is equivalent to (∗) and $\Phi_{\tilde{r}}$-invariance of $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ is an immediate consequence of the definitions. Thus the proof of the lemma will be complete once we prove (∗).

Assume the notation of Definition 15.1. Let $\tilde{r}' = \Theta(\tilde{r})$ and $\psi = \theta\phi\theta^{-1}$; note that $\Psi_{\tilde{r}'} = \Theta\Phi_{\tilde{r}}\Theta^{-1}$. Choose a homotopy equivalence $g\colon K \to K'$ of marked graphs that represents θ when $\pi_1(K)$ and $\pi_1(K')$ are identified with F_n via their markings. Let $\tilde{g}\colon \tilde{K} \to \tilde{K}'$ be the lift satisfying $\tilde{g}|\partial F_n = \Theta|\partial F_n$, let $\tilde{R}' = \tilde{g}_\#(\tilde{R}) \subset \tilde{K}'$, let $\tilde{L}' = \Theta(\tilde{L}) = \tilde{g}_\#(\tilde{L})$ and let $T'_j\colon \tilde{K}' \to \tilde{K}'$ be the covering translation satisfying $T'_j|\partial F_n = (\Theta T_j \Theta^{-1})|\partial F_n$. Then

$$\Psi^j_{\tilde{r}'}(\tilde{L}') \cap \tilde{R}' = \Psi^j_{\tilde{r}'}(\tilde{g}_\#(\tilde{L})) \cap \tilde{g}_\#(\tilde{R}) = \tilde{g}_\#(\Phi^j_{\tilde{r}}(\tilde{L})) \cap \tilde{g}_\#(\tilde{R}).$$

By [Cooper 1987] (see also [Bestvina et al. 1997, Lemma 3.1]), there is a constant C, depending only on g, such that $\tilde{g}_\#(\Phi^j_{\tilde{r}}(\tilde{L})) \cap \tilde{g}_\#(\tilde{R})$ contains the subpath of $\tilde{g}_\#(\Phi^j_{\tilde{r}}(\tilde{L}) \cap \tilde{R})$ obtained by C-trimming (ie removing the first and last C edges) and so contains the subpath of $\tilde{g}_\#(T_j(\tilde{\beta})) = T'_j\tilde{g}_\#(\tilde{\beta})$ obtained by C-trimming for any chosen $\tilde{\beta}$ and all $j \geq J(\tilde{\beta})$. Given a finite subpath $\tilde{\beta}'$ of $\tilde{L}'$ choose a finite subpath $\tilde{\beta}$ of $\tilde{L}$ such that the C-trimmed subpath of $\tilde{g}_\#(\tilde{\beta})$ contains $\tilde{\beta}'$. Then $\Psi^j_{\tilde{r}'}(\tilde{L}') \cap \tilde{R}' \supset T'_j(\tilde{\beta}')$ for all $j \geq J(\tilde{\beta})$. Letting $J(\tilde{\beta}') = J(\tilde{\beta})$, we conclude that $\tilde{L}' \in \Omega_{\mathsf{NP}}(\phi, \tilde{r})$. By symmetry, we have proved (∗). □

Our goal in the remainder of this subsection is to understand $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ from the CT point of view.

Notation 15.4 Choose $r \in \mathcal{R}(\phi)$ and a CT $f: G \to G$ representing ϕ; let $E \in \mathcal{E}_f$ correspond to r as in Lemma 3.26. Following the proof of Corollary 5.17, let

$$R_E = E \cdot \rho_0 \cdot \sigma_1 \cdot \rho_1 \cdot \sigma_2 \cdot \ldots$$

be the *coarsened complete splitting* of R_E, where each σ_i is a single growing term in the complete splitting of R_E and each ρ_i is a (possibly trivial) Nielsen path. For future reference, note that if $f(E) = E \cdot u$ then Eu is an initial subpath of R_E whose terminal endpoint is a splitting vertex in the complete splitting of R_E and hence is contained in some ρ_p.

Following Notation 5.11 and Lemma 5.14, define, for all $i \geq 1$,

$$R_i^- = f_\#^\infty(\bar{\sigma}_i), \quad R_i^+ = f_\#^\infty(\sigma_i), \quad \ell_i = (R_i^-)^{-1} \rho_i (R_{i+1}^+).$$

Choose a lift $\tilde{r}$ of r, let $\Phi_{\tilde{r}}$ be the automorphism representing ϕ that fixes $\tilde{r}$ and let $\tilde{f}: \tilde{G} \to \tilde{G}$ be the lift corresponding to $\Phi_{\tilde{r}}$. Let $\tilde{R}_{\tilde{E}}$ be the lift of R_E whose terminal end converges to $\tilde{r}$ and whose initial edge is denoted by $\tilde{E}$, let

$$\tilde{R}_{\tilde{E}} = \tilde{E} \cdot \tilde{\rho}_0 \cdot \tilde{\sigma}_1 \cdot \tilde{\rho}_1 \cdot \tilde{\sigma}_2 \cdot \ldots$$

be the induced decomposition and let $\tilde{\ell}_i$ be the lift of ℓ_i in which ρ_i lifts to $\tilde{\rho}_i$. Thus

$$\tilde{\ell}_i = (\tilde{R}_i^{\,-})^{-1} \tilde{\rho}_i \tilde{R}_{i+1}^+,$$

where $\tilde{\sigma}_{i+1}$ and $\tilde{R}_{i+1}^+$ have the same initial endpoint and likewise for $\tilde{\sigma}_i^{-1}$ and $\tilde{R}_i^-$. We say that lines $\tilde{\ell}_1, \tilde{\ell}_2, \ldots$ are *visible* in $\tilde{R}_{\tilde{E}}$.

Lemma 15.5 *Assume Notation 15.4.*

(1) *Each ℓ_i is an element of $\Omega(r)$; see Definition 5.1.*

(2) *If $\ell_i \in \Omega_{\mathrm{NP}}(r)$, then $\tilde{\ell}_i \in \Omega_{\mathrm{NP}}(\phi, \tilde{r})$.*

Proof Item (1) follows from Lemmas 5.8 and 5.16 applied with $\alpha = \sigma_i$ and $\beta = \rho_i \sigma_{i+1}$.

When verifying that $\tilde{\ell}_i$ satisfies Definition 15.1, it suffices to consider finite subpaths $\tilde{\beta} = \tilde{\mu}^{-1} \tilde{\rho}_i \tilde{\nu}$ of $\tilde{\ell}_i$ with projections $\beta = \mu^{-1} \rho_i \nu$, where μ is an initial segment of $R_i^- = f_\#^\infty(\bar{\sigma}_i)$ that is a concatenation of terms in the coarsened complete splitting of R_i^- and ν is an initial segment of $R_{i+1}^+ = f_\#^\infty(\sigma_{i+1})$ that is a concatenation of terms in the coarsened complete splitting of R_{i+1}^+. It follows from the definition of $f_\#^\infty$ (Notation 5.11) that for all sufficiently large j, the lift of ρ_i to $\tilde{f}_\#^j(\tilde{\rho}_i)$ extends to a lift of β to a path

$$\tilde{\beta}_j \subset \tilde{f}_\#^j(\tilde{\sigma}_i) \cdot \tilde{f}_\#^j(\tilde{\rho}_i) \cdot \tilde{f}_\#^j(\tilde{\sigma}_{i+1}) = \tilde{f}_\#^j(\tilde{\sigma}_i \cdot \tilde{\rho}_i \cdot \tilde{\sigma}_{i+1}) \subset \tilde{R}_{\tilde{E}}.$$

Since $f_\#^j$ preserves ρ_i, R_i^- and R_{i+1}^+, there is a covering translation T_j such that

$$T_j(\tilde{\rho}_i) = \tilde{f}_\#^j(\tilde{\rho}_i), \quad T_j(\tilde{R}_i^-) = \tilde{f}_\#^j(\tilde{R}_i^-), \quad T_j(\tilde{R}_{i+1}^+) = \tilde{f}_\#^j(\tilde{R}_{i+1}^+),$$

and so

$$T_j(\widetilde{\ell}_i) = \widetilde{f}^j_\#(\widetilde{\ell}_i).$$

From $T_j(\widetilde{\rho}_i) = \widetilde{f}^j_\#(\widetilde{\rho}_i)$ we conclude that $T_j(\widetilde{\beta}) = \widetilde{\beta}_j$ and so $T_j(\widetilde{\beta}) \subset \widetilde{R}_{\widetilde{E}}$. This completes the proof of (2). □

Our next result is a weak converse of Lemma 15.5(2), namely that if $\widetilde{L} \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ then $\widetilde{L}$ is in the $\Phi_{\widetilde{r}}$-orbit of some $\widetilde{\ell}_i$.

Proposition 15.6 *Assume Notation 15.4.*

1. *For each $\widetilde{L} \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ there exists K such that $\widetilde{f}^k_\#(\widetilde{L}) \in \{\widetilde{\ell}_i\}$ for all $k \ge K$. Moreover, $\Omega_{\mathsf{NP}}(\phi, \widetilde{r}) = \bigcup \Phi^m_{\widetilde{r}}(\widetilde{\ell}_i)$, where the union varies over all $\widetilde{\ell}_i \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ and all $m \in \mathbb{Z}$.*
2. *For each $L \in \Omega_{\mathsf{NP}}(r)$ there is a lift $\widetilde{L} \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.*

We delay the proof of Proposition 15.6 for two needed lemmas.

Lemma 15.7 *Given a CT $f : G \to G$, there exists $M \ge 1$ so that the following holds for each twist path w, each nonfixed edge E and each $k \ge 0$: If $|m| \ge M$ and $\alpha_0 = w^m$ is a subpath of $f^k_\#(E)$ then α_0 extends to a subpath α_1 of $f^k_\#(E)$ satisfying the following two properties*:

1. *$\alpha_1 = E'w^q$ or $\alpha_1 = w^q\bar{E}'$ for some $E' \in \mathsf{Lin}_w(f)$.*
2. *α_1 is not contained in any Nielsen subpath of $f^k_\#(E)$.*

Proof Let us first note that if the conclusions of the lemma hold for a subpath of $\alpha_0 = w^m$ of the form w^t then they also hold for α_0. We may therefore shorten α_0 whenever it is convenient. After replacing E by $\bar{E}$ if necessary, we may assume that $E \in \mathcal{E}_f \cup \mathsf{Lin}(f)$.

Choose $M'' > 0$ so that if $w_1 \ne w_2$ are twist paths then $w_1^{M''}$ is not a subpath of w_2^m for any $m \in \mathbb{Z}$. Items (1) and (2) hold for $M = M''$ and $E \in \mathsf{Lin}(f)$. We therefore assume that $E \in \mathcal{E}_f$ and that there exists $M' \ge M''$ such that (1) and (2) hold for $M = M'$ and for all edges $E' \in \mathcal{E}_f$ with height less than that of E.

There is a path u with height less than that of E and a complete splitting

$$u = \tau_1 \cdot \ldots \cdot \tau_s \quad \text{such that} \quad f^k(E) = E \cdot u \cdot f_\#(u) \cdot \ldots \cdot f^{k-1}_\#(u)$$

for all $k \ge 1$. Assuming without loss that M' is greater than the length of any τ_j, choose $M_1 \ge sM'$.

As a special case, we prove the lemma when $|m| \ge M_1$ and when $\alpha_0 = w^m$ is contained in some $f^l_\#(u)$. In this case there exists $1 \le j \le s$ and $|m'| \ge M'$ and a subpath $\alpha'_0 = w^{m'}$ of α_0 such that $\alpha'_0 \subset f^l_\#(\tau_j)$ for some $1 \le l \le k-1$. As observed above, we can replace α_0 with α'_0. Since the length of τ_j is less than M' and the length of $f^l_\#(\tau_j)$ is at least M', τ_j is not a Nielsen path and so is either exceptional or an edge E' with height less than that of E. If τ_j is exceptional then its linear edges must be in the family determined by w and we take α_1 to be all of τ_j except for the terminal edge. If $\tau_j = E'$, then the inductive hypothesis implies that α_0 extends to a subpath α_1 of $f_\#^l(E')$ that is not contained in a Nielsen subpath of $f_\#^l(E')$ and that satisfies (1). The hard splitting property of a complete splitting (Lemma 4.11(2) of [Feighn and Handel 2011]) implies that an indivisible Nielsen path in a completely split path is contained in a single term of that splitting. Thus α_1 is not contained in a Nielsen subpath of $f_\#^k(E)$ and so (2) is satisfied and we have completed the proof of the special case.

Now choose M so large that if $|m| \geq M$ and $\alpha_0 = w^m$ is a subpath of $f_\#^k(E)$ then there is a subpath $\alpha_0' = w^{m'}$ of α_0 with $m' \geq M_1$ so that $\alpha_0' \subset f_\#^l(u)$ for some l. The existence of M follows from the fact that the length of $f_\#^l(u)$ goes to infinity with l. Replacing α_0 with α_0', we are reduced to the special case. □

We choose a "central" subpath τ_L of $L \in \Omega_{\mathsf{NP}}(r)$ as follows. By Corollary 5.17 $L = (R^-)^{-1} \cdot \rho \cdot R^+$, where $R^\pm$ satisfy (1a), (1b) or (1c) of Lemma 5.14. In all three cases we will choose $\tau = \tau_L = \tau_-^{-1}\rho\tau_+$, where $\tau_\pm$ is an initial segment of $R^\pm$. Let M be the constant from Lemma 15.7.

- In the case (1a), $R^+ = R_{E'}$ for some $E' \in \mathcal{E}_f$ and we take $\tau_+ = E'$.
- In the case (1b), $R^+ = E'w^{\pm\infty}$ for some $E' \in \mathsf{Lin}_w(f)$ and we take $\tau_+ = E'w^{\pm M}$.
- In the case (1c), $R^+ = w^{\pm\infty}$ and we take $\tau_+ = w^{\pm M}$.

The subpath τ_- is defined symmetrically.

Lemma 15.8 *Assume the notation of Notation 15.4 and of the previous paragraph. Suppose that $\widetilde{\tau}_L \subset \widetilde{L}$ is a lift of $\tau_L \subset L$ and that $\widetilde{\tau}_L \subset \widetilde{R}_{\widetilde{E}}$. Then $\widetilde{L} = \widetilde{\ell}_i$ for some i.*

Proof As a first case, suppose that $\tau_+ = E' \in \mathcal{E}_f$ and so $R^+ = R_{E'}$. Then $\widetilde{\tau}_+$ is a term $\widetilde{\sigma}_{i+1}$ in the coarsened complete splitting of $\widetilde{R}_{\widetilde{E}}$ by Lemma 3.21 and $R^+ = f_\#^\infty(\sigma_{i+1})$ by Examples 5.12. There are three subcases to consider, the first being that $\tau_- = E'' \in \mathcal{E}_f$. In this subcase, $\widetilde{\tau}_-^{-1}$ is also a term $\widetilde{\sigma}_j$ in coarsened complete splitting. Since $\widetilde{\tau}_-$ is separated from $\widetilde{\sigma}_{i+1} = \widetilde{\tau}_+$ by the Nielsen subpath $\widetilde{\rho}$, we have $\widetilde{\tau}_-^{-1} = \widetilde{\sigma}_i$ and $\widetilde{\rho} = \widetilde{\rho}_i$. Thus $f_\#^\infty(\overline{\sigma}_i) = R^-$ and $\widetilde{L} = \widetilde{\ell}_i$.

The second subcase is that $\tau_- = E''w^{\pm M}$, where $E'' \in \mathsf{Lin}_w(f)$. By Lemma 15.7(2), τ_-^{-1} is not contained in a Nielsen subpath of R_E. It follows that the terminal edge $\widetilde{E''}^{-1}$ of $\widetilde{\tau}_-^{-1}$ is contained in a $\widetilde{\sigma}_j$ that is either a single edge or an exceptional path. As in the previous subcase, $j = i$. Also as in the previous subcase, $\widetilde{\rho} = \widetilde{\rho}_i$, $f_\#^\infty(\overline{\sigma}_i) = R^-$ and $\widetilde{\ell}_i = \widetilde{L}$.

The third and final subcase is that $\tau_- = w^{\pm M}$. Since $\widetilde{\tau}_-^{-1}$ is followed in $\widetilde{R}_{\widetilde{E}}$ by $\widetilde{\rho}\widetilde{E}'$, it is not contained in a subpath of $\widetilde{R}_{\widetilde{E}}$ of the form $\widetilde{w}^m\widetilde{E}_1^{-1}$, where $E_1 \in \mathsf{Lin}_w(f)$. Items (1) and (2) of Lemma 15.7 imply that $\widetilde{\sigma}_i = \widetilde{E}_1$, where $E_1 \in \mathsf{Lin}_w(f)$, and that $\widetilde{\rho}_i = \widetilde{w}^t\widetilde{\rho}$ for some t. Examples 5.12 implies that $f_\#^\infty(\overline{\sigma}_i) = w^{\pm\infty} = R^-$ and so $\widetilde{\ell}_i = \widetilde{L}$. We have now completed the proof in the case that $\tau_+ = E' \in \mathcal{E}_f$. Symmetric arguments apply in the case that $\tau_- = E' \in \mathcal{E}_f$.

Our next case is that $\tau_+ = E' w^{\pm M}$, where $E' \in \mathsf{Lin}_w(f)$ and $R^+ = E' w^{\pm\infty}$. Lemma 15.7(2) implies that the initial edge $\widetilde{E}'$ of $\widetilde{\tau}_+$ is not contained in a Nielsen subpath of $\widetilde{R}_{\widetilde{E}}$ and so is either equal to some $\widetilde{\sigma}_{i+1}$ or is the first edge in some $\widetilde{\sigma}_{i+1}$ that projects to an exceptional path. In either case $f_\#^\infty(\sigma_{i+1}) = E' w^{\pm\infty} = R^+$. The remainder of the proof in this second case is exactly the same as in the first case. Symmetric arguments apply in the case that $\tau_- = E' w^{\pm M}$ with $E' \in \mathsf{Lin}_w(f)$.

We are now reduced to the case that $\tau_+ = w_2^{\pm M}$, $R_+ = w_2^{\pm\infty}$, $\tau_- = w_1^{\pm M}$ and $R_- = w_1^{\pm\infty}$. Thus $\widetilde{\tau} = \widetilde{w}_1^{\mp M} \widetilde{\rho} \widetilde{w}_2^{\pm M}$. If $w_1 = w_2$ then ρ is not an iterate of $w_1 = w_2$ because L is not periodic. Lemma 15.7(1) implies that $\widetilde{\tau}$ extends to a subpath $\widetilde{E}_1 \widetilde{w}_1^p \widetilde{\rho} \widetilde{w}_2^q \widetilde{E}_2^{-1}$, where $E_i \in \mathsf{Lin}_{w_i}(f)$ and $p, q \in \mathbb{Z}$. It follows that $\widetilde{E}_1 = \widetilde{\sigma}_i$, $\widetilde{\rho}_i = \widetilde{w}_1^p \widetilde{\rho} \widetilde{w}_2^q$ and $\widetilde{E}_2^{-1} = \widetilde{\sigma}_{i+1}$ for some choice of i. As in the previous cases, $\widetilde{\ell}_i = \widetilde{L}$. □

Proof of Proposition 15.6 The first statement of (1) follows from Lemma 15.8 and the definition of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$. The moreover statement of (1) follows from the first statement and $\Phi_{\widetilde{r}}$-invariance of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.

For (2), let $E \in \mathcal{E}_f$ correspond to r. Let $\tau_L \subset L$ be as in Lemma 15.8. Since $L \in \Omega_{\mathsf{NP}}(r)$, τ_L lifts to a subpath $\widetilde{\tau}_L \subset \widetilde{R}_{\widetilde{E}}$. Lemma 15.8 implies that the lift of τ_L to $\widetilde{\tau}_L$ extends to a lift of L to an element of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$. □

Lemma 15.9 *Continue with Notation 15.4.*

(1) *For all $i \ge 1$, there exists $j = j(i) > i$ such that $\widetilde{f}_\#(\widetilde{\ell}_i) = \widetilde{\ell}_j$. More precisely, there exists $j > i$ such that $\widetilde{f}_\#(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$ and $\widetilde{f}_\#(\widetilde{\ell}_i) = \widetilde{\ell}_j$ and there is a covering translation T such that $T(\widetilde{\rho}_i) = \widetilde{f}_\#(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$ and $T(\widetilde{\ell}_i) = \widetilde{\ell}_j$.*

(2) *The assignment $i \mapsto j(i)$ is order-preserving and $j(1) > p$.*

Proof It suffices to prove (2) and the "more precisely" statement of (1). We begin with the latter. Since $\widetilde{f}_\#(\widetilde{\rho}_i)$ is a Nielsen path that is a concatenation of terms in the complete splitting of $\widetilde{R}_{\widetilde{E}}$, there exists j such that $\widetilde{f}_\#(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$. Let T be the unique covering translation satisfying $T(\widetilde{\rho}_i) = \widetilde{f}_\#(\widetilde{\rho}_i)$. It suffices to prove that $\widetilde{f}_\#(\widetilde{\ell}_i) = \widetilde{\ell}_j$ and $T(\widetilde{\ell}_i) = \widetilde{\ell}_j$.

From $\widetilde{f}_\#(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$ it follows that
$$\widetilde{\rho}_j = \widetilde{\alpha} \cdot \widetilde{f}_\#(\widetilde{\rho}_i) \cdot \widetilde{\beta},$$
where $\widetilde{\alpha}$ and $\widetilde{\beta}$ are, possibly trivial, Nielsen paths. Since $\widetilde{\sigma}_i$ and $\widetilde{\sigma}_{i+1}$ are growing,
$$\widetilde{\sigma}_j \cdot \widetilde{\rho}_j \cdot \widetilde{\sigma}_{j+1} = \widetilde{\sigma}_j \cdot \widetilde{\alpha} \cdot \widetilde{f}_\#(\widetilde{\rho}_i) \cdot \widetilde{\beta} \cdot \widetilde{\sigma}_{j+1} \subset \widetilde{f}_\#(\widetilde{\sigma}_i) \cdot \widetilde{f}_\#(\widetilde{\rho}_i) \cdot \widetilde{f}_\#(\widetilde{\sigma}_{i+1}).$$
By Lemma 5.14(2),
$$R_{i+1}^+ = f_\#^\infty(\sigma_{i+1}) = \beta f_\#^\infty(\sigma_{j+1}) = \beta \cdot R_{j+1}^+ \quad \text{and} \quad R_i^- = f_\#^\infty(\bar{\sigma}_i) = \bar{\alpha} f_\#^\infty(\bar{\sigma}_j) = \bar{\alpha} \cdot R_j^-,$$
which implies that
$$\widetilde{f}_\#(\widetilde{\ell}_i) = \widetilde{f}_\#((\widetilde{R}_i^-)^{-1} \cdot \widetilde{\rho}_i \cdot \widetilde{R}_{i+1}^+) = (\widetilde{R}_j^-)^{-1} \cdot \widetilde{\alpha} \cdot \widetilde{f}_\#(\widetilde{\rho}_i) \cdot \widetilde{\beta} \cdot \widetilde{R}_{j+1}^+ = (\widetilde{R}_j^-)^{-1} \cdot \widetilde{\rho}_j \cdot \widetilde{R}_{j+1}^+ = \widetilde{\ell}_j.$$
This completes the proof that $\widetilde{f}_\#(\widetilde{\ell}_i) = \widetilde{\ell}_j$.

The covering translation T that carries $\widetilde{\rho}_i$ to $\widetilde{f}_\#(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$ also carries $\widetilde{R}^+_{i+1}$ to $\widetilde{\beta} \cdot \widetilde{R}^+_{j+1}$ and R^-_i to $\widetilde{\alpha}^{-1} \cdot \widetilde{R}^-_j$. Thus $T(\widetilde{\ell}_i) = \widetilde{\ell}_j$.

Finally, note that $j(i+1) - j(i)$ is equal to the number of growing terms in $\widetilde{f}_\#(\widetilde{\sigma}_i)$. This implies (2) and hence also $j(i) > j$. Since $\widetilde{E} \cdot \widetilde{f}_\#(\widetilde{u}) \cdot \widetilde{f}_\#(\widetilde{\rho}_0) \cdot f_\#(\widetilde{\sigma}_1) \cdot \widetilde{f}_\#(\widetilde{\rho}_1)$ is an initial segment of $\widetilde{R}_E$ and since $\widetilde{f}_\#(\widetilde{\rho}_0) \subset \widetilde{\rho}_p$, it follows that $j(1) > p$. □

We conclude this subsection by defining a total order on $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.

Definition 15.10 Continue with Notation 15.4. Given distinct $\widetilde{L}_1, \widetilde{L}_2 \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$, choose $k \geq 0$ so that $\widetilde{f}^k_\#(\widetilde{L}_1) = \widetilde{\ell}_i$ and $\widetilde{f}^k_\#(\widetilde{L}_2) = \widetilde{\ell}_j$ for some $i \neq j$. (The existence of k is guaranteed by Proposition 15.6.) Define $\widetilde{L}_1 \prec \widetilde{L}_2$ if $i < j$.

Lemma 15.11 *The relation $\prec$ is a well-defined, $\Phi_{\widetilde{r}}$-invariant total order on $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ that is independent of the choice of $f: G \to G$ representing ϕ. Moreover, if $\psi = \theta\phi\theta^{-1}$ for some $\theta \in \mathsf{Out}(F_n)$ and if Θ is a lift of θ, then $\Theta: \Omega_{\mathsf{NP}}(\phi, \widetilde{r}) \to \Omega_{\mathsf{NP}}(\psi, \Theta(\widetilde{r}))$ preserves $\prec$.*

We delay the proof of Lemma 15.11 to state and prove a technical lemma that allows us to redefine $\prec$ with less dependence on the location of $\widetilde{\rho}_i$ and $\widetilde{\rho}_j$ in $\widetilde{R}_{\widetilde{E}}$.

Lemma 15.12 *Continue with Notation 15.4. Suppose that $\widetilde{\ell}_i$ and $\widetilde{\ell}_j$ are distinct nonperiodic visible lines. For all $k \geq 0$, let $\widetilde{\ell}_{i_k} = \widetilde{f}^k_\#(\widetilde{\ell}_i)$ and $\widetilde{\ell}_{j_k} = \widetilde{f}^k_\#(\widetilde{\ell}_j)$, and let $\widetilde{y}_{i,k}$ and $\widetilde{y}_{j,k}$ be the terminal endpoints of $\widetilde{\ell}_{i_k} \cap \widetilde{R}_{\widetilde{E}}$ and $\widetilde{\ell}_{j_k} \cap \widetilde{R}_{\widetilde{E}}$, respectively. Then $i < j$ if and only if the following two conditions are satisfied:*

(1) $\widetilde{\ell}_i \notin \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_j)$.

(2) *One of the following is satisfied:*

  (a) $\widetilde{\ell}_j \in \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_i)$.

  (b) $y_{j,k} - y_{i,k} \to \infty$, *where* $y_{j,k} - y_{i,k} = \pm$ *the number of edges in the subpath connecting* $\widetilde{y}_{i,k}$ *to* $\widetilde{y}_{j,k}$, *and the sign is* $+$ *if and only if* $\widetilde{y}_{,jk} > \widetilde{y}_{i,k}$ *in the orientation on* $\widetilde{R}_{\widetilde{E}}$.

Proof For the "only if" direction, assume that $i < j$. Lemma 15.9, and an obvious induction argument imply that $i_k < j_k$ and that there are unique covering translations T_k satisfying

$$T_k(\widetilde{\rho}_i) \subset \widetilde{\rho}_{i_k} \quad \text{and} \quad T_k(\widetilde{\ell}_i) = \widetilde{f}^k_\#(\widetilde{\ell}_i) = \widetilde{\ell}_{i_k},$$

and S_k satisfying

$$S_k(\widetilde{\rho}_j) \subset \widetilde{\rho}_{j_k} \quad \text{and} \quad S_k(\widetilde{\ell}_j) = \widetilde{f}^k_\#(\widetilde{\ell}_j) = \widetilde{\ell}_{j_k}.$$

Note that $\widetilde{h}_k := S_k^{-1} \widetilde{f}^k$ is the lift of f^k that preserves $\widetilde{\ell}_j$ and so corresponds to the automorphism $\Phi^k_{\partial_+\widetilde{\ell}_j}$. Note also that $S_k^{-1} T_k(\widetilde{\ell}_i) = S_k^{-1} \widetilde{f}^k_\#(\widetilde{\ell}_i) = (\widetilde{h}_k)_\#(\widetilde{\ell}_i)$, and $T_k(\widetilde{\rho}_i) \subset \widetilde{\rho}_{i_k}$ is disjoint from $\widetilde{\rho}_{j_k} \widetilde{R}^+_{j_k+1}$. The latter implies that $S_k^{-1} T_k(\widetilde{\rho}_i)$ is disjoint from $\widetilde{\rho}_j \widetilde{R}^+_{j+1}$. It now follows from the definition that $\widetilde{\ell}_i \notin \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_j)$. This completes the proof of (1).

For (2), we assume that $\widetilde{\ell}_j \notin \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_i)$ and prove that $\widetilde{y}_{j,k} - \widetilde{y}_{i,k} \to \infty$. Continuing with the above notation, $\widetilde{g}_k := T_k^{-1} \widetilde{f}^k$ corresponds to $\Phi^k_{\partial_+\widetilde{\ell}_i}$ and $T_k^{-1} S_k(\widetilde{\ell}_j) = \widetilde{g}_k(\widetilde{\ell}_j)$. We claim that there is a finite subpath $\widetilde{\beta} \subset \widetilde{\ell}_j$ so that for all $k \geq 0$, $T_k^{-1} S_k(\widetilde{\beta}) \not\subset \widetilde{R}^+_{i+1}$ and hence $S_k(\widetilde{\beta}) \not\subset T_k(\widetilde{R}^+_{i+1})$. If $\ell_j \notin \Omega_{\mathsf{NP}}(\partial_+\ell_i)$ then this follows from Definition 5.1 and the fact that $T_k^{-1} S_k(\widetilde{\ell}_j)$ is a lift of ℓ_j. If $\ell_j \in \Omega_{\mathsf{NP}}(\partial_+\ell_i)$ then this follows from Lemmas 15.8 and 15.5.

On the other hand, $S_k(\widetilde{\beta}) \subset \widetilde{R}_{\widetilde{E}}$ for all sufficiently large k. It follows that $\widetilde{y}_{i,k}$ precedes the terminal endpoint of $S_k(\widetilde{\beta})$ in $\widetilde{R}_{\widetilde{E}}$. Since the number of edges in $S_k(\widetilde{R}^+_{j+1}) \cap \widetilde{R}_{\widetilde{E}}$ goes to infinity with k, $\widetilde{y}_{j,k} - \widetilde{y}_{i,k} \to \infty$.

For the "if" direction, we assume that $j < i$ and we prove that either (1) or (2) fails. From the "only if" direction we know that (1) with i and j reversed is satisfied. Thus (2a) fails. Similarly, either (2a) or (2b), with the roles of i and j reversed, is satisfied. If the former holds then (1) fails and we are done. Suppose then that (2b) with the roles of i and j reversed is satisfied. Then $\widetilde{y}_{i,k} - \widetilde{y}_{j,k} \to \infty$ so (2b) fails. □

Proof of Lemma 15.11 To make the dependence of $\prec$ on $f: G \to G$ explicit we will write $\prec_f$. Lemma 15.9 implies that $\prec_f$ is a well-defined, $\Phi_{\widetilde{r}}$-invariant total order on $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.

Suppose that θ, ψ and Θ are as in the "moreover" statement, that $f': G' \to G'$ is a CT representing ψ, that $g: G \to G'$ is a homotopy equivalence representing θ and that $\widetilde{g}: \widetilde{G} \to \widetilde{G}'$ is the lift corresponding to Θ. Letting $\widetilde{r}' = \Theta(\widetilde{r})$, we have $\Theta\Phi_{\widetilde{r}} = \Psi_{\widetilde{r}'}\Theta$. By Lemma 15.3, $\Theta(\Omega_{\mathsf{NP}}(\phi, \widetilde{r})) = \Omega_{\mathsf{NP}}(\psi, \widetilde{r}')$. Let $\widetilde{f}'$ be the lift of f' corresponding to $\Psi_{\widetilde{r}'}$.

Given $\widetilde{L}_1, \widetilde{L}_2 \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ such that $\widetilde{L}_1 \prec_f \widetilde{L}_2$, we must show that $\widetilde{L}'_1 \prec_{f'} \widetilde{L}'_2$, where $\widetilde{L}'_1 = \Theta(\widetilde{L}_1) = \widetilde{g}_\#(\widetilde{L}_1)$ and $\widetilde{L}'_2 = \Theta(\widetilde{L}_2) = \widetilde{g}_\#(\widetilde{L}_2)$. We may replace $\widetilde{L}_1$ and $\widetilde{L}_2$ with $\Phi^k_{\widetilde{r}}\widetilde{L}_1$ and $\Phi^k_{\widetilde{r}}\widetilde{L}_2$ for any $k \geq 1$. This follows from the $\Phi_{\widetilde{r}}$-invariance of $\prec_f$, the $\Psi_{\widetilde{r}'}$-invariance of $\prec_{f'}$ and the fact that $\Theta\Phi_{\widetilde{r}} = \Psi_{\widetilde{r}'}\Theta$. In particular, we may assume that there exists $i < j$ and $i' \neq j'$ such that $\widetilde{L}_1 = \widetilde{\ell}_i, \widetilde{L}_2 = \widetilde{\ell}_j, \widetilde{L}'_1 = \widetilde{\ell}'_i$ and $\widetilde{L}'_2 = \widetilde{\ell}'_j$, where the $\widetilde{\ell}'_i$ and $\widetilde{\ell}'_j$ are visible lines determined for $\Omega_{\mathsf{NP}}(\psi, \widetilde{r}')$ defined with respect to f'.

To prove that $i' < j'$, and thereby complete the proof of the lemma, we will verify items (1) and (2) of Lemma 15.12 in the prime system, which we will call (1)′ and (2)′. Items (1)′ and (2a)′ follow from (1), (2a) and Lemma 15.3. Item (2b)′ follows from (2b) and the bounded cancellation lemma applied to g. □

We conclude this section with a result that will be used in Lemma 15.45.

Lemma 15.13 *Suppose that F is a free factor, that $\widetilde{r} \in \partial F$ is a lift of $r \in \mathcal{R}(\phi)$ and that $[F]$ is ϕ-invariant. Then each endpoint of each $\widetilde{\ell} \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ is contained in ∂F.*

Proof Choose a CT $f: G \to G$ representing ϕ in which F is realized by a component C of a core filtration element H. By assumption, there is a ray R in C with terminal end r. For each $\ell \in \Omega_{\mathsf{NP}}(r)$, each finite subpath of ℓ is contained in C and hence ℓ is contained in C. Let $\widetilde{C}$ be the unique lift of C whose boundary contains $\widetilde{r}$ and note that $\partial\widetilde{C} = \partial F$. Let $\widetilde{R} \subset \widetilde{C}$ be the lift of R with terminal endpoint $\widetilde{r}$

and let $\Phi_{\tilde r}$ be the automorphism representing ϕ that fixes $\tilde r$. From uniqueness of $\widetilde C$, it follows that $\partial \widetilde C$ is $\Phi_{\tilde r}$-invariant. For all sufficiently large j, $\Phi_{\tilde r}^j(\tilde\ell) \cap \widetilde R \neq \varnothing$. Since $\ell \subset C$ and distinct lifts of C are disjoint, $\Phi_{\tilde r}^j(\tilde\ell) \subset \widetilde C$. It follows that the endpoints of $\Phi_{\tilde r}^j(\tilde\ell)$, and hence the endpoints of $\tilde\ell$, are contained in ∂F. □

15.2 Topmost lines, translation numbers and offset numbers

We continue with Notation 15.4 and with the partial orders $<$ on $\mathcal{R}(\phi)$ and $\mathcal{E}_f$ given in Notation 6.1 and Lemma 6.2.

Definition 15.14 An element $L \in \Omega_{\mathrm{NP}}(r)$ is *ϕ-topmost* if one of the following mutually exclusive properties is satisfied for the partial order $<$ on $\mathcal{R}(\phi)$:

(1) r is minimal in the partial order $<$.

(2) L has an end $r_1 \in \mathcal{R}(\phi)$ such that $r_1 <_c r$.

If $\widetilde L \in \Omega_{\mathrm{NP}}(\phi, \tilde r)$ projects to a ϕ-topmost element of $\Omega_{\mathrm{NP}}(r)$, then $\widetilde L$ is a *topmost* element of $\Omega_{\mathrm{NP}}(\phi, \tilde r)$. Let $\mathcal{T}_{\phi,\tilde r}$ be the *set of topmost elements of* $\Omega_{\mathrm{NP}}(\phi, \tilde r)$.

Lemma 15.15 *$\mathcal{T}_{\phi,\tilde r}$ is nonempty and $\Phi_{\tilde r}$-invariant.*

Proof Lemma 15.3 implies that $\Omega_{\mathrm{NP}}(\phi, \tilde r)$ is $\Phi_{\tilde r}$-invariant. Since each element of $\mathcal{R}(\phi)$ is ϕ-invariant, $\Phi_{\tilde r}$-invariance of $\mathcal{T}_{\phi,\tilde r}$ follows from the definitions. If r is minimal with respect to $<$ then every element of $\Omega_{\mathrm{NP}}(\phi, \tilde r)$ is topmost and we are done. Otherwise, apply Lemma 6.2 to choose $E' \in \mathcal{E}_f$ such that $E' <_c E$. Either E' or $\overline{E}'$ occurs as a term σ_j in the coarsened complete splitting of R_E. In the former case, $\tilde\ell_{j-1}$ is topmost in $\Omega_{\mathrm{NP}}(\phi, \tilde r)$; in the latter case $\tilde\ell_j$ is topmost in $\Omega_{\mathrm{NP}}(\phi, \tilde r)$. □

Lemma 15.16 *There is an algorithm that lists the ϕ-topmost elements of $\Omega_{\mathrm{NP}}(r)$.*

Proof Recall that the elements of $\Omega_{\mathrm{NP}}(r)$ can be enumerated by Corollary 5.17(2) and that the partial order on $\Omega_{\mathrm{NP}}(r)$ can be computed by Notation 6.1. If r is minimal then every element of $\Omega_{\mathrm{NP}}(r)$ is topmost. Otherwise inspect the elements of $\Omega_{\mathrm{NP}}(r)$ to see which ones satisfy 15.14(2). □

Recall from Notation 15.4 that p is chosen so that $\tilde f_\#(\rho_0) \subset \tilde\rho_p$.

Lemma 15.17 *Each $\widetilde L \in \mathcal{T}_{\phi,\tilde r}$ is in the $\Phi_{\tilde r}$-orbit of $\tilde\ell_j$ for some $1 \le j \le p$.*

Proof By Proposition 15.6 and Lemma 15.15, we may assume that $\widetilde L = \tilde\ell_i$ for some $i > p$. By Lemma 15.9, it suffices to show that there exist $1 \le j \le p$ and $k \ge 1$ such that $\tilde f_\#^k(\tilde\rho_j) \subset \tilde\rho_i$. If this fails then there exists $1 \le j' \le p$ and $k' \ge 1$ such that $\tilde\rho_i$ separates $\tilde f_\#^{k'}(\tilde\rho_{j'-1})$ from $\tilde f_\#^{k'}(\tilde\rho_{j'})$. Assuming this we argue towards a contradiction by showing that neither (1) nor (2) in Definition 15.14 is satisfied. First note that $\tilde\sigma_i \tilde\rho_i \tilde\sigma_{i+1} \subset \tilde f_\#^{k'}(\tilde\sigma_{j'})$. It follows that $\sigma_{j'}$ is not a linear term and so $\sigma_{j'} = E'$ or $\overline{E}'$ for some $E' \in \mathcal{E}_f$. Since $E' < E$, (1) is not satisfied. If an end r'' of $\tilde\ell_i$ corresponds to an element $E'' \in \mathcal{E}_f$ then $E'' < E' < E$ and so (2) is not satisfied. □

Notation 15.18 The total order $\prec$ on $\Omega_{\mathsf{NP}}(\phi,\tilde r)$ given in Definition 15.10 induces a total order (also called) $\prec$ on $\mathcal{T}_{\phi,\tilde r}$. Let $\tilde L_1,\dots,\tilde L_{\tau(\phi,\tilde r)}$ be, in order, the elements of $\{\tilde\ell_1,\dots,\tilde\ell_p\}\cap\mathcal{T}_{\phi,\tilde r}$. For $k\in\mathbb{Z}$ and $1\le j\le\tau(\phi,\tilde r)$, define $\tilde L_{j+k\tau(\phi,\tilde r)}=\Phi_{\tilde r}^k(\tilde L_j)$.

The following lemma allows us to change our notation from $\tau(\phi,\tilde r)$ to $\tau(\phi,r)$.

Lemma 15.19 *The number $\tau(\phi,\tilde r)$ depends only on ϕ and r and not on the choice of $\tilde r$.*

Proof The definition of $\tau(\phi,\tilde r)$ uses the lift $\tilde f:\tilde G\to\tilde G$ corresponding to $\Phi_{\tilde r}$, the lift $\tilde R_{\tilde E}$ of R_E whose terminal endpoint is $\tilde r$, the lines $\{\tilde\ell_i\}$ determined by $\tilde R_{\tilde E}$ as described in Notation 15.4 and the integer p, which depends only on E and f. If $a\in F_n$ and $T_a:\tilde G\to\tilde G$ is the corresponding covering translation, then the data associated to $\tilde r'=a\tilde r$ is $\tilde f'=T_a\tilde f T_a^{-1}$, $\tilde R_{\tilde E'}=T_aR_{\tilde E}$, $\tilde\ell_i'=T_a\tilde\ell_i$ and p. Since $\tilde\ell_i$ and $\tilde\ell_i'$ are lifts of the same line, $\tilde\ell_i\in\mathcal{T}_{\phi,\tilde r}\iff\tilde\ell_i'\in\mathcal{T}_{\phi,\tilde r'}$. This proves that $\tau(\phi,\tilde r)=\tau(\phi,\tilde r')$, as desired. □

Lemma 15.20 *With notation as above,*

(1) *$s\mapsto\tilde L_s$ defines an order-preserving bijection between $\mathbb{Z}$ and $\mathcal{T}_{\phi,\tilde r}$,*

(2) *$\Phi_{\tilde r}(\tilde L_s)=\tilde L_{s+\tau(\phi,r)}$ for all s, and*

(3) *$\tilde L_s$ is visible if and only if $s\ge 1$.*

Proof The map $s\mapsto\tilde L_s$ is surjective by Lemma 15.17 and is order-preserving (and hence injective) because $\tilde f_\#$ preserves $\prec$ and because $\tilde L_1\prec\tilde L_2\prec\cdots\prec\tilde L_{\tau(\phi,r)}\prec\tilde f_\#(\tilde L_1)$, where the last inequality follows Lemma 15.9, which implies that $\tilde f_\#(\tilde L_1)=\tilde\ell_j$ for some $j>p$. Item (2) follows from the definitions. Item (3) follows from Lemma 15.9. □

For the next lemma, we must choose a CT $f':G'\to G'$ representing ψ and then define $\mathcal{T}_{\psi,\tilde r'}$ and $\tau(\psi,r')$ with respect to $f':G'\to G'$.

Lemma 15.21 *Suppose that $\theta\in\mathsf{Out}(F_n)$ conjugates ϕ to ψ, that $\theta(r)=r'\in\mathcal{R}(\psi)$, that $\tilde r,\tilde r'\in\partial F_n$ represent r and r', respectively, and that Θ is the lift of θ such that $\Theta(\tilde r)=\tilde r'$. Then*

(1) *$\tau(\phi,r)=\tau(\psi,r')$, and*

(2) *there is an integer $\mathsf{offset}(\theta,r)$ such that $\Theta(\tilde L_s)=\tilde L'_{s+\mathsf{offset}(\theta,r)}$ for all s.*

Proof Lemmas 15.3 and 15.11 imply that Θ induces a $\prec$-preserving bijection between $\Omega_{\mathsf{NP}}(\phi,\tilde r)$ and $\Omega_{\mathsf{NP}}(\psi,\tilde r')$. Lemma 6.3 implies that this bijection restricts to a bijection between $\mathcal{T}_{\phi,\tilde r}$ and $\mathcal{T}_{\psi,\tilde r'}$. Since the only order-preserving bijections of $\mathbb{Z}$ are translations, there is an integer $\mathsf{offset}(\theta,\tilde r,\tilde r')$ such that $\Theta(\tilde L_s)=\tilde L'_{s+\mathsf{offset}(\theta,\tilde r,\tilde r')}$ for all s. If we replace $\tilde r$ by another lift $\tilde r^*=i_a\tilde r$ then Θ is replaced by $\Theta^*=\Theta i_a^{-1}$ and $\tilde L_i\in\mathcal{T}_{\phi,\tilde r}$ is replaced by $\tilde L_i^*=i_a\tilde L_i\in\mathcal{T}_{\phi,\tilde r^*}$; see the proof of Lemma 15.19. It follows that $\Theta^*(L_i^*)=\Theta(\tilde L_i)$ and hence that $\mathsf{offset}(\theta,\tilde r,\tilde r')$ is independent of the choice of lift $\tilde r$. The symmetric argument implies that $\mathsf{offset}(\theta,\tilde r,\tilde r')$ is also independent of the choice of $\tilde r'$. This completes the proof of (2).

Item (1) therefore follows from

$$\widetilde{L}'_{s+\tau(\phi,r)+\mathsf{offset}(\theta,r)} = \Theta\Phi_{\tilde r}\widetilde{L}_s = \Psi_{\tilde r'}\Theta\widetilde{L}_s = \widetilde{L}'_{s+\mathsf{offset}(\theta,r)+\tau'(\psi,r')}. \qquad \square$$

Remark 15.22 The bijection between $\mathbb{Z}$ and $\mathcal{T}_{\phi,\tilde r}$ depends on the notion of visible lines and so depends on the choice of CT. On the other hand, Lemma 15.20(2) implies that $\tau(\phi, r)$ depends only on ϕ and r and not on the choice of a CT. As such it can be computed from any CT for ϕ. The integer $\mathsf{offset}(\theta, r)$ depends on the choices of CTs.

15.3 Staple pairs

We continue with Notation 15.4. We set further notation as follows.

Notation 15.23 If E_i and E_j are distinct elements of $\mathsf{Lin}_w(f)$ then there exist nonzero $d_i \neq d_j$ such that $f(E_i) = E_i w^{d_i}$ and $f(E_j) = E_j w^{d_j}$. Recall that a path of the form $E_i w^p \overline{E}_j$ is called *exceptional* if d_i and d_j have the same sign. If d_i and d_j have different signs then we say $E_i w^p \overline{E}_j$ is *quasi-exceptional*.

Notation 15.24 We write $L \in \mathcal{S}(\phi)$ and say that L is a *staple* if $L \in \Omega_{\mathsf{NP}}(\phi)$ has at least one periodic end; if both ends of L are periodic then L is a *linear staple*. If $\widetilde{L} \in \Omega_{\mathsf{NP}}(\phi, \tilde r)$ projects to an element of $\mathcal{S}(\phi)$ for $r \in \mathcal{R}(\phi)$ and lift $\tilde r$, then we write $\widetilde{L} \in \mathcal{S}(\phi, \tilde r)$ and $L \in \mathcal{S}(\phi, r)$ and we say that L and $\widetilde{L}$ *occur* in r and $\tilde r$, respectively.

For each $r \in \mathcal{R}(\phi)$, an ordered pair $b = (L_1, L_2)$ of elements of $\mathcal{S}(\phi, r)$ is a *staple pair* if there are lifts $\widetilde{L}_1, \widetilde{L}_2 \in \mathcal{S}(\phi, \tilde r)$ and a periodic line $\widetilde{A}$ such that $\{\partial_+\widetilde{L}_1, \partial_-\widetilde{L}_2\} \subset \{\partial_-\widetilde{A}, \partial_+\widetilde{A}\}$. We write $b \in \mathcal{S}_2(\phi, r)$ and $\tilde b = (\widetilde{L}_1, \widetilde{L}_2) \in \mathcal{S}_2(\phi, \tilde r)$ and say that b and $\tilde b$ *occur* in r and $\tilde r$ respectively and that $\widetilde{A}$ is the *common axis* of $\tilde b$. By Corollary 5.17, $\widetilde{A}$ corresponds to an element of $\mathcal{A}(\phi)$. Define $\mathcal{S}_2(\phi) = \cup\mathcal{S}_2(\phi, r)$, where the union is taken over all $r \in \mathcal{R}(\phi)$.

Lemma 15.25 *Each $b \in \mathcal{S}_2(\phi, r)$ is ϕ-invariant. The set $\mathcal{S}_2(\phi, \tilde r)$ is $\Phi_{\tilde r}$-invariant.*

Proof The first statement follows from the second and the fact (Corollary 5.17(1)) that each element of $\Omega_{\mathsf{NP}}(\phi)$ is ϕ-invariant. The second follows from the $\Phi_{\tilde r}$-invariance of $\Omega_{\mathsf{NP}}(\phi, \tilde r)$ (Lemma 15.3) and the definition of $\mathcal{S}_2(\phi, \tilde r)$. $\square$

Example 3.1 (continued) In our example,

$$\mathcal{S}(\phi) = \{a^\infty R_c, a^\infty b a^\infty\} \quad \text{and} \quad \mathcal{S}_2(\phi) = \{(a^\infty b a^\infty, a^\infty b a^\infty), (a^\infty b a^\infty, a^\infty R_c)\}.$$

Throughout this section, M is the stabilization constant defined in Notation 5.13.

Our next lemma explains how staple pairs occur in an eigenray $\widetilde{R}_{\widetilde{E}}$.

Lemma 15.26 *Assume Notation 15.4.*

(1) *If $\sigma_i\rho_i\sigma_{i+1}$ is quasi-exceptional, then $(\tilde\ell_{i-1},\tilde\ell_{i+1}) \in \mathcal{S}_2(\phi,\tilde r)$ with common axis $\tilde\ell_i$.*

(2) *If one of the following holds:*
   (a) *σ_i is exceptional, or*
   (b) *$\sigma_i \in \mathsf{Lin}_w(f)$ and ℓ_i is not periodic, or*
   (c) *$\bar\sigma_i \in \mathsf{Lin}_w(f)$ and ℓ_{i-1} is not periodic;*
   then $(\tilde\ell_{i-1},\tilde\ell_i) \in \mathcal{S}_2(\phi,\tilde r)$.

(3) *If $\tilde\ell_i$ is periodic and neither σ_i nor $\bar\sigma_{i+1}$ is in $\mathcal{E}_f$, then $\sigma_i\rho_i\sigma_{i+1}$ is quasi-exceptional and so $(\tilde\ell_{i-1},\tilde\ell_{i+1}) \in \mathcal{S}_2(\phi,\tilde r)$ with common axis $\tilde\ell_i$. See also Remark 15.43.*

(4) *For each $\tilde b \in \mathcal{S}_2(\phi,\tilde r)$, there exists $K = K(\tilde b)$ such that $\Phi^k_{\tilde r}(\tilde b)$ is as in (1) or (2) for all $k \ge K$. Moreover, in case (b), $R_i^- = w^{\pm\infty}$ and in case (c), $R_{i-1}^+ = w^{\pm\infty}$.*

Proof If $\sigma_i\rho_i\sigma_{i+1}$ is quasi-exceptional then there are a twist curve w and edges $E', E'' \in \mathsf{Lin}_w(f)$ such that $\sigma_i = E'$, $\rho_i = w^q$ for some $q \in \mathbb{Z}$ and $\sigma_{i+1} = \bar E''$. Moreover, $f(E') = E'w^{d'}$ and $f(E'') = E''w^{d''}$, where d' and d'' have opposite signs. If $\tilde\sigma_i = \tilde E'$, let $\tilde w$ be the lift of w that begins with the terminal endpoint $\tilde x$ of $\tilde E'$. Extend $\tilde w$ to a periodic line $\tilde A$ that projects bi-infinitely to the circuit determined by w and is oriented consistently with $\tilde w$. Let $\tilde y \in \tilde A$ be the terminal endpoint of the lift of w^q that begins at $\tilde x$ and let $\tilde E''$ be the lift of E'' that ends at $\tilde y$. Then $\tilde R_i^+$ is the concatenation of $\tilde E'$ and a ray in $\tilde A$ beginning at $\tilde x$ and terminating at $\partial_+\tilde A$ if $d' > 0$ and at $\partial_-\tilde A$ if $d' < 0$. Similarly, $\tilde R_{i+1}^-$ is the concatenation of $\tilde E''$ and a ray in $\tilde A$ beginning at $\tilde y$ and terminating at $\partial_+\tilde A$ if $d'' > 0$ and at $\partial_-\tilde A$ if $d'' < 0$. Neither $\tilde\ell_{i-1}$ nor $\tilde\ell_{i+1}$ is periodic. Up to a change of orientation, $\tilde\ell_i = \tilde A$. Thus $(\tilde\ell_{i-1},\tilde\ell_{i+1}) \in \mathcal{S}(\phi,\tilde r)$ with common axis $\tilde\ell_i$, and (1) is proved.

If σ_i is exceptional, then $\sigma_i = E'w^q\bar E''$, where E', w and E'' are as above except that d' and d'' have the same sign. Following the above notation, $\tilde R_i^+$ begins with $\tilde E'$, $\tilde R_i^-$ begins with $\tilde E''$ and both rays terminate at the same endpoint of $\tilde A$. Neither $\tilde\ell_{i-1}$ nor $\tilde\ell_i$ is periodic. This completes the proof of (2a).

If $\sigma_i = E' \in \mathsf{Lin}_w(f)$, then following the above notation, $\tilde\ell_{i-1}$ is nonperiodic (because it crosses $\tilde E'$) with terminal endpoint in $\{\partial_-\tilde A, \partial_+\tilde A\}$ and $\tilde R_i^-$ has terminal endpoint in $\{\partial_-\tilde A, \partial_+\tilde A\}$. If $\tilde\ell_i$ is nonperiodic then $(\tilde\ell_{i-1},\tilde\ell_i) \in \mathcal{S}_2(\phi,\tilde r)$. This completes the proof of (2b). The proof of (2c) is similar.

Suppose that $\tilde\ell_i$ is as in (3). If $\bar\sigma_i \in \mathcal{E}_f$ then $\tilde R_i^-$ is not asymptotic to a periodic line, in contradiction to the assumption that $\tilde\ell_i$ is periodic. If $\bar\sigma_i \in \mathsf{Lin}(f)$ or if σ_i is exceptional then $R_i^- = E_i w^{\pm\infty}$, where $E_i \in \mathsf{Lin}_w(f)$, again in contradiction to the assumption that $\tilde\ell_i$ is periodic. We conclude that $\sigma_i = E' \in \mathsf{Lin}(f)$. The symmetric argument shows that $\sigma_{i+1} = \bar E''$ for some $E'' \in \mathsf{Lin}(f)$. Thus ℓ_i has the form $(w')^{\pm\infty}\rho_i(w'')^{\pm\infty}$, where w' is the twist path for E' and w'' is the twist path for E''. Since $\tilde\ell_i$ is a periodic line, $w' = w''$ and $\rho_i = (w')^q$ for some $q \in \mathbb{Z}$. This proves that $\sigma_i\rho_i\sigma_{i+1}$ is quasi-exceptional, which in conjunction with (1), completes the proof of (3).

For (4), suppose that $\widetilde{b} \in \mathcal{S}_2(\phi, \widetilde{r})$. After replacing $\widetilde{b}$ with some $\Phi_{\widetilde{r}}^k(\widetilde{b})$, we may assume by Proposition 15.6 that $\widetilde{b} = (\widetilde{\ell}_{i-1}, \widetilde{\ell}_j)$ for some $i-1 \neq j$. After replacing $\widetilde{b}$ with $\Phi_{\widetilde{r}}^M(\widetilde{b})$, we may assume that $\overline{\sigma}_i \notin \mathcal{E}_f$. (To see this note that if $\widetilde{f}_\#^M(\widetilde{\rho}_{i-1}) \subset \widetilde{\rho}_{s-1}$ then $\widetilde{f}_\#^M(\widetilde{\ell}_{i-1}) = \widetilde{\ell}_{s-1}$ and $\widetilde{\sigma}_s$ is the first growing term of $\widetilde{f}_\#^M(\widetilde{\sigma}_i)$.) By assumption, $\partial_+ \widetilde{R}_i^+$ is an endpoint of the common axis $\widetilde{A}$ of $\widetilde{b}$. Lemma 5.14 therefore implies that $\sigma_i \notin \mathcal{E}_f$ and hence that σ_i is linear. In other words, $\sigma_i = E_i$ or $\sigma_i = \overline{E}_i$ or $\sigma_i = E_i w_i^* \overline{E}_l$ for some twist path w_i and for some $E_i, E_l \in \mathsf{Lin}_{w_i}(f)$. In all three cases, the terminal endpoint of $\widetilde{E}_i$ is contained in $\widetilde{A}$. For the same reasons, we may assume that $\sigma_j = E_j$ or $\sigma_j = \overline{E}_j$ or $\sigma_j = E_m w_j^* \overline{E}_j$ for some twist path w_j and for some $E_j, E_m \in \mathsf{Lin}_{w_j}(f)$; moreover, the terminal endpoint of $\widetilde{E}_j$ is in $\widetilde{A}$.

The proof now proceeds by a case analysis. If $\sigma_i = E_i w_i^* \overline{E}_l$ then the midpoint of $\widetilde{E}_i$ (resp. $\widetilde{E}_l^{-1}$) separates $\widetilde{A}$ from $\widetilde{\sigma}_q$ for all $q < i$ (resp. $q > i$) so $j = i$ and we are in case (a). The same argument, with the same conclusion, applies if $\sigma_j = E_m w_j^* \overline{E}_j$. We may now assume that σ_i is either E_i or $\overline{E}_i$ and that σ_j is either E_j or $\overline{E}_j$. By considering the midpoints of σ_i and σ_j as in the previous case we see that:

(a) If $\sigma_i = E_i$, then $j \geq i$.

(b) If $\sigma_i = \overline{E}_i$, then $j \leq i$.

(c) If $\sigma_j = E_j$, then $i \geq j$.

(d) If $\sigma_j = \overline{E}_j$, then $i \leq j$.

If (a) and (c) are satisfied then $j = i$, we are in case (b) and $R_i^- = w^{\pm\infty}$. Similarly, if (b) and (d) are satisfied then $j = i$, we are in case (c) and $R_{i-1}^+ = w^{\pm\infty}$. Suppose next that (a) and (d) are satisfied. In this case, $j \geq i$, $w_i = w_j$ and the interval $\widetilde{\tau}$ of $\widetilde{A}$ bounded by the terminal endpoints of $\widetilde{E}_i$ and $\widetilde{E}_j$ equals $\widetilde{w}_i^q$ for some $q \in \mathbb{Z}$; in particular, τ is a Nielsen path. It must be that $\tau = \rho_i$ and $j = i+1$, which is (1). Finally, suppose that (b) and (c) are satisfied. Then $j \leq i$, $w_i = w_j$ and the interval $\widetilde{\tau}$ of $\widetilde{A}$ bounded by the terminal endpoints of $\widetilde{E}_j$ and $\widetilde{E}_i$ equals $\widetilde{w}_i^q$ for some $q \in \mathbb{Z}$; in particular, τ is a Nielsen path. It must be that $\tau = \rho_{i-1}$ and $j = i-1$, which contradicts the fact that $i-1 \neq j$. Thus this last case does not happen, and we are done. □

Notation 15.27 We say that the staple pairs $(\widetilde{\ell}_{i-1}, \widetilde{\ell}_{i+1})$ and $(\widetilde{\ell}_{i-1}, \widetilde{\ell}_i)$ that occur in items (1) and (2) of Lemma 15.26 are *visible with index* i or just *visible* if the index is not explicitly given. Note that if $\widetilde{b}$ is visible then $\Phi_{\widetilde{r}}^k(\widetilde{b})$ is visible for all $k \geq 0$.

Corollary 15.28 *The set of visible elements of* $\mathcal{S}_2(\phi, \widetilde{r})$ *is infinite.*

Proof From $\Phi_{\widetilde{r}}$-invariance of $\mathcal{S}_2(\phi, \widetilde{r})$ (Lemma 15.25) and Lemma 15.9, we need only show that $\mathcal{S}_2(\phi, \widetilde{r})$ contains a visible element. There are always linear edges crossed by R_E. We are therefore reduced, by Lemma 15.26, to the case that some $\widetilde{\ell}_i$ is periodic. If $\widetilde{f}_\#^M(\widetilde{\rho}_i) \subset \widetilde{\rho}_j$ then $\widetilde{\sigma}_j$ is the last growing term in $\widetilde{f}_\#^M(\widetilde{\sigma}_i)$ and so $\sigma_j \notin \mathcal{E}_f$. Similarly, $\widetilde{\sigma}_{j+1}$ is the first growing term in $\widetilde{f}_\#^M(\widetilde{\sigma}_{i+1})$ and so $\overline{\sigma}_{j+1} \notin \mathcal{E}_f$. Lemma 15.26(3) implies that $(\widetilde{\ell}_{j-1}, \widetilde{\ell}_{j+1}) \in \mathcal{S}_2(\phi, \widetilde{r})$ and we are done. □

Recall from Notation 15.4 that the $\widetilde{\ell}_i$ are said to be *visible*.

Lemma 15.29 *Suppose that $\widetilde{b} = (\widetilde{L}_1, \widetilde{L}_2) \in \mathcal{S}_2(\phi, \widetilde{r})$ with common axis $\widetilde{A}$ and that one of the following two conditions are satisfied.*

(a) *Either $\widetilde{L}_1$ or $\widetilde{L}_2$ is visible and there exist $k \geq 0$ such that $\Phi^k_{\widetilde{r}}(\widetilde{L}_1, \widetilde{L}_2) = (\widetilde{\ell}_{i-1}, \widetilde{\ell}_i)$ for some i.*

(b) *Either $\widetilde{L}_1, \widetilde{L}_2$ or the common axis of $\widetilde{A}$ is visible and there exist $k \geq 0$ such that $\Phi^k_{\widetilde{r}}(\widetilde{L}_1, \widetilde{L}_2) = (\widetilde{\ell}_{i-1}, \widetilde{\ell}_{i+1})$ with common axis $\widetilde{\ell}_i$ for some i.*

Then $\Phi^{2M}_{\widetilde{r}}(\widetilde{b})$ is visible.

Proof We begin by establishing the following properties for each visible line $\widetilde{\ell}_j$.

(1) Suppose that $\partial_+ \widetilde{\ell}_j$ is periodic and that $\widetilde{\ell}_s = \Phi^M_{\widetilde{r}}(\widetilde{\ell}_j)$. Then $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_s)$ and $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_{s+1})$ are consecutive (ie their indices differ by 1) for all $m \geq 0$.

(2) Suppose that $\partial_- \widetilde{\ell}_j$ is periodic and that $\widetilde{\ell}_s = \Phi^M_{\widetilde{r}}(\widetilde{\ell}_j)$. Then $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_{s-1})$ and $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_s)$ are consecutive for all $m \geq 0$.

For (1), Lemma 15.9 implies that $\widetilde{f}^M_\#(\widetilde{\rho}_j) \subset \widetilde{\rho}_s$ and our choice of M implies that $\widetilde{\sigma}_{s+1} \notin \mathcal{E}_f^{-1}$. Since $\partial_+ \widetilde{\ell}_j$ is periodic, the same is true for $\partial_+ \widetilde{\ell}_s$ and so $\widetilde{\sigma}_{s+1} \notin \mathcal{E}_f$. We conclude that σ_{s+1} is linear. In particular, $\widetilde{f}^m_\#(\widetilde{\sigma}_{s+1})$ has exactly one growing term. If $\widetilde{f}^m_\#(\widetilde{\rho}_s) \subset \widetilde{\rho}_a$ then $\widetilde{f}^m_\#(\widetilde{\rho}_{s+1}) \subset \widetilde{\rho}_{a+1}$. Lemma 15.9 implies that $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_s) = \widetilde{\ell}_a$ and $\Phi^m_{\widetilde{r}}(\widetilde{\ell}_{s+1}) = \widetilde{\ell}_{a+1}$. This completes the proof of (1). Item (2) is proved by the symmetric argument.

We now apply (1) and (2) to prove the lemma, assuming without loss that $k > M$. In case (a), we will show that $\Phi^M_{\widetilde{r}}(\widetilde{b})$ is visible. If $\widetilde{L}_1$ is visible let $\widetilde{\ell}_{s-1} = \Phi^M_{\widetilde{r}}(\widetilde{L}_1)$. Since $\Phi^{k-M}_{\widetilde{r}}(\widetilde{\ell}_{s-1}) = \Phi^k_{\widetilde{r}}(\widetilde{L}_1) = \widetilde{\ell}_{i-1}$, property (1), applied with $m = k - M$, implies that $\Phi^{k-M}_{\widetilde{r}}(\widetilde{\ell}_s) = \widetilde{\ell}_i$. Since $\Phi^{k-M}_{\widetilde{r}}(\Phi^M_{\widetilde{r}}(\widetilde{L}_2)) = \widetilde{\ell}_i$, we have $\Phi^M_{\widetilde{r}}(\widetilde{L}_2) = \widetilde{\ell}_s$. Thus $\Phi^M_{\widetilde{r}}(\widetilde{b}) = (\Phi^M_{\widetilde{r}}(\widetilde{L}_1), \Phi^M_{\widetilde{r}}(\widetilde{L}_2)) = (\widetilde{\ell}_{s-1}, \widetilde{\ell}_s)$ is visible. This completes the proof when $\widetilde{L}_1$ is visible. When $\widetilde{L}_2$ is visible, a symmetric argument, using (2) instead of (1), shows that $\Phi^M_{\widetilde{r}}(\widetilde{L}_1)$ and hence $\Phi^M_{\widetilde{r}}(\widetilde{b})$ is visible.

In case (b), note that $\partial_+ \widetilde{L}_1, \partial_- \widetilde{L}_2$ and both ends of $\widetilde{A}$ are periodic. If $\widetilde{L}_1$ is visible then the above argument shows that the common axis of $\Phi^M_{\widetilde{r}}(\widetilde{b})$ is visible and a second application shows that $\Phi^{2M}_{\widetilde{r}}(\widetilde{L}_2)$ is visible. The other cases are similar. □

Notation 15.30 Suppose that $\widetilde{b} = (\widetilde{L}_1, \widetilde{L}_2) \in \mathcal{S}_2(\phi, \widetilde{r})$ projects to $b \in \mathcal{S}_2(\phi, r)$. If $\widetilde{\ell}_j \prec \widetilde{L}_1$ (see Definition 15.10) then we write $\widetilde{\ell}_j \prec \widetilde{b}$. We say that b and $\widetilde{b}$ are *topmost* elements of $\mathcal{S}_2(\phi, r)$ and $\mathcal{S}_2(\phi, \widetilde{r})$, respectively, if for all $r_1 < r$ (see Notation 6.1) neither b nor $b^{-1} := (L_2^{-1}, L_1^{-1})$ is an element of $\mathcal{S}_2(\phi, r_1)$. Since $\widetilde{b}$ and $\Phi_{\widetilde{r}}(\widetilde{b})$ project to the same element of $\mathcal{S}_2(\phi, r)$ and since $\mathcal{S}_2(\phi, r)$ is $\Phi_{\widetilde{r}}$-invariant, it follows that the set of topmost element of $\mathcal{S}_2(\phi, \widetilde{r})$ is $\Phi_{\widetilde{r}}$-invariant.

Lemma 15.31 *The set of topmost elements of $\mathcal{S}_2(\phi, \widetilde{r})$ is the union of a finite number of $\Phi_{\widetilde{r}}$-orbits. Moreover, there exists a computable $B(r) > 0$ such that each of these orbits has a visible representative with index at most $B(r)$.*

Proof Define $B(r) > 2M$ by $\Phi_{\tilde r}^{2M}(\tilde\ell_p) = \tilde\ell_{B(r)}$.

Suppose that $\tilde b$ is a topmost element of $\mathcal{S}_2(\phi,\tilde r)$. After replacing $\tilde b$ with some $\Phi_{\tilde r}^k(\tilde b)$, we may assume by Lemma 15.26(4) that $\tilde b = (\tilde\ell_{i-1},\tilde\ell_i)$ or $\tilde b = (\tilde\ell_{i-1},\tilde\ell_{i+1})$ with common axis $\tilde l_i$. We consider the $\tilde b = (\tilde\ell_{i-1},\tilde\ell_i)$ case first, assuming without loss that $i > 2M$. The proof below is similar to that of Lemma 15.17.

Suppose that there exists $1 \le j' \le p$ and $k' > 0$ so that $\tilde\sigma_{i-1}\tilde\rho_{i-1}\tilde\sigma_i\tilde\rho_i\tilde\sigma_{i+1} \subset \tilde f_\#^{k'}(\tilde\sigma_{j'})$. Then $\sigma_{j'} = E'$ or $\bar E'$ for some $E' \in \mathcal{E}_f$ so either $\sigma_{i-1}\rho_{i-1}\sigma_i\rho_i\sigma_{i+1}$ or $\bar\sigma_{i+1}\bar\rho_i\bar\sigma_i\bar\rho_{i-1}\bar\sigma_{i-1}$ occurs as a concatenation of terms in the coarsening of the complete splitting of $R_{E'}$. Letting $r' \in \mathcal{R}(\phi)$ correspond to E', Lemma 15.26 implies that either b or b^{-1} is an element of $\mathcal{S}_2(\phi,r')$ in contradiction to the assumption that b is topmost in $\mathcal{S}_2(\phi,r)$. Thus no such j' and k' exist. It follows that there exists $1 \le j \le p$ and $k > 0$ such that $\tilde f_\#^k(\tilde\rho_j)$ is contained in either $\tilde\rho_{i-1}$ or $\tilde\rho_i$. Equivalently, $\Phi_{\tilde r}^k(\tilde\ell_j)$ is equal to either $\tilde\ell_{i-1}$ or $\tilde\ell_i$. Since $\tilde\ell_j$ is one of the lines comprising the pair $\Phi_{\tilde r}^{-k}(\tilde b)$, Lemma 15.29 implies that $\Phi_{\tilde r}^{2M-k}(\tilde b) = \Phi_{\tilde r}^{2M}(\Phi_{\tilde r}^{-k}(\tilde b))$ is visible with index at most $B(r)$.

In the remaining case, $\tilde b = (\tilde\ell_{i-1},\tilde\ell_{i+1})$ with common axis $\tilde\ell_i$. Arguing as in the first case, we conclude that there exists $k \ge 0$ and $1 \le j \le p$ such that $\Phi_{\tilde r}^k(\tilde\ell_j)$ is equal to either $\tilde\ell_{i-1}$ or $\tilde\ell_i$ or $\tilde\ell_{i+1}$. The proof then concludes as in the first case. □

Remark 15.32 The $\Phi_{\tilde r}$-image of a visible topmost staple pair is a visible topmost staple pair. It follows that if a topmost staple pair $\tilde b$ occurs in $\tilde r$ and if $\tilde\ell_{B(r)} \prec \tilde b$ then $\tilde b$ is visible.

Remark 15.33 The set of topmost elements of $\mathcal{S}_2(\phi,r)$ could be empty.

Lemma 15.34 *If $\hat r < r$ and $b \in \mathcal{S}_2(\phi,\hat r)$, then either $b \in \mathcal{S}_2(\phi,r)$ or $b^{-1} \in \mathcal{S}_2(\phi,r)$.*

Proof The proof is similar to that of Lemma 15.31. Let $\hat E$ be the higher-order edge corresponding to $\hat r$, let $R_{\hat E} = \hat\rho_0\cdot\hat\sigma_1\cdot\hat\rho_1\cdot\ldots$ be the coarsening of the complete splitting into single growing terms and Nielsen paths and let $\hat\ell_1,\hat\ell_2,\ldots$ be the associated visible lines. By Lemmas 15.26(4) and 15.25, there exists $i \ge 1$ such that $b = (\hat\ell_{i-1},\hat\ell_i)$ or $b = (\hat\ell_{i-1},\hat\ell_{i+1})$. Since $\hat r < r$, there exists $j > 1$ such that either $\sigma_j = \hat E$ or $\sigma_j = \hat E^{-1}$. The cases are symmetric so we assume that $\sigma_j = \hat E^{-1}$ and leave the $\sigma_j = \hat E$ case to the reader. Since $\ell_j \in \Omega_{\mathsf{NP}}(r)$, the inverse of every finite subpath of $R_{\hat E}$ occurs as subpath of R_E. In particular, the inverse of $\hat\rho_{i-2}\cdot\hat\sigma_{i-1}\ldots\cdot\hat\sigma_{i+2}\hat\rho_{i+2}$ occurs as a concatenation of terms in R_E. Lemma 15.26 therefore implies that $b^{-1} \in \mathcal{S}_2(\phi,r)$. □

Lemma 15.35 *Suppose that $\theta \in \mathsf{Out}(F_n)$ conjugates ϕ to ψ, that $\theta(r) = r' \in \mathcal{R}(\psi)$, that $\tilde r,\tilde r' \in \partial F_n$ represent r and r', respectively, and that Θ is the lift of θ such that $\Theta(\tilde r) = \tilde r'$. Then Θ induces a bijection $\mathcal{S}_2(\phi,\tilde r) \mapsto \mathcal{S}_2(\psi,\tilde r')$ that restricts to a bijection on topmost elements.*

Proof This follows from Lemma 15.3, which provides a bijection between $\Omega_{\mathsf{NP}}(\phi,\tilde r)$ and $\Omega_{\mathsf{NP}}(\psi,\tilde r')$, and the definitions. □

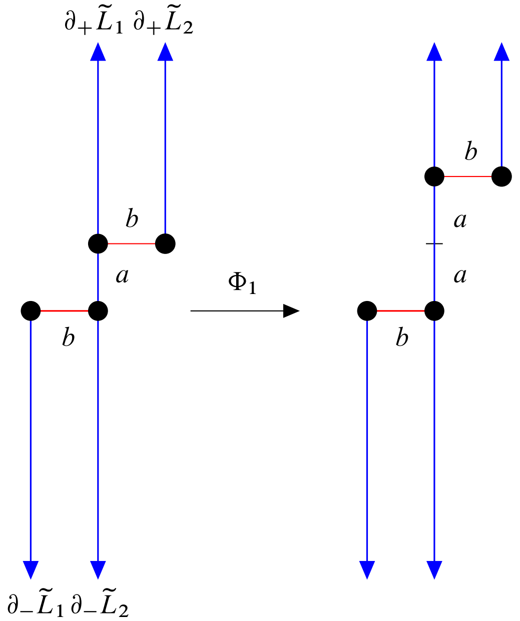


Figure 5: The graphs here are parts of $\widetilde{G}$ with each horizontal segment a lift of the edge b and where vertical segments project into a^∞. $\widetilde{L}_1$ and $\widetilde{L}_2$ are lifts of the staple $a^\infty b a^\infty$ and (L_1, L_2) is a staple pair. Φ_1 is the lift of ϕ that fixes $\widetilde{L}_1$. Intuitively $\widetilde{L}_2$ slides away from $\widetilde{L}_1$ under the action of Φ_1 by 1 period and so $m_{(L_1,L_2)}(\phi) = 1$

.

Definition 15.36 Given $b = (L_1, L_2) \in \mathcal{S}_2(\phi, r)$, choose lifts $\widetilde{L}_1, \widetilde{L}_2$ and a periodic line $\widetilde{A}$ such that $\{\partial_+\widetilde{L}_1, \partial_-\widetilde{L}_2\} \subset \{\partial_-\widetilde{A}, \partial_+\widetilde{A}\}$. Orient $\widetilde{A}$ to be consistent with the twist path w to which it projects and let $a \in F_n$ be the root-free element of F_n that stabilizes $\widetilde{A}$ and satisfies $a^+ = \partial_+\widetilde{A}$. Each $\theta \in \mathcal{X}_\mathfrak{c}(\phi)$ (Definition 14.1) satisfies $\theta(\mathsf{H}_{\phi,\mathfrak{c}}(L_i)) = \mathsf{H}_{\phi,\mathfrak{c}}(L_i)$ for $i = 1, 2$. Lemma 13.10 therefore implies that there are unique $\Theta_i \in \theta$ such that $\Theta_i(\mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}_i)) = \mathsf{H}_{\phi,\mathfrak{c}}(\widetilde{L}_i)$. Since both Θ_1 and Θ_2 represent θ and fix a there exists $m_b(\theta) \in \mathbb{Z}$ such that $\Theta_1 = i_a^{m_b(\theta)}\Theta_2$.

Example 3.1 (continued) See Figure 5.

Lemma 15.37 *For each $b = (L_1, L_2) \in \mathcal{S}_2(\phi)$, the map $m_b : \mathcal{X}_\mathfrak{c}(\phi) \to \mathbb{Z}$ is a well-defined homomorphism.*

Proof We first check that $m_b(\theta)$ is independent of the choice of $\widetilde{L}_1$ and $\widetilde{L}_2$ and so is well-defined. Suppose that $\widetilde{L}'_1$ and $\widetilde{L}'_2$ are another choice with corresponding $\widetilde{A}'$, a' and Θ'_i. Choose $c \in F_n$ such that $i_c(a) = a'$. For $i = 1, 2$, $\widetilde{L}'_i$ and $i_c(\widetilde{L}_i)$ are lifts of L_i with an endpoint in $\{a'^-, a'^+\}$ and so there exists an n_i such that

$$\widetilde{L}'_i = i_{a'}^{n_i} i_c(\widetilde{L}_i) = i_c i_a^{n_i}(\widetilde{L}_i).$$

By uniqueness,

$$\Theta'_i = (i_c i_a^{n_i})\Theta_i(i_c i_a^{n_i})^{-1} = i_c i_a^{n_i}\Theta_i i_a^{-n_i} i_c^{-1} = i_c\Theta_i i_c^{-1},$$

so

$$\Theta'_1\Theta'^{-1}_2 = i_c\Theta_1\Theta_2^{-1}i_c^{-1} = i_c i_a^{m_b(\theta)} i_c^{-1} = i_{a'}^{m_b(\theta)},$$

as desired.

To prove that $m_b(\theta)$ defines a homomorphism, suppose that $\psi \in \mathcal{X}_{\mathsf{c}}(\phi)$ and Ψ_i satisfies $\Psi_i(\mathsf{H}_{\phi,\mathsf{c}}(\widetilde{L}_i)) = \mathsf{H}_{\phi,\mathsf{c}}(\widetilde{L}_i)$. Then $\Psi_i\Theta_i(\mathsf{H}_{\phi,\mathsf{c}}(\widetilde{L}_i)) = \mathsf{H}_{\phi,\mathsf{c}}(\widetilde{L}_i)$ and

$$i_a^{m_b(\psi\theta)} = \Psi_1\Theta_1\Theta_2^{-1}\Psi_1^{-1} = \Psi_1 i_a^{m_b(\theta)}\Psi_2^{-1} = \Psi_1\Psi_2^{-1} i_a^{m_b(\theta)} = i_a^{m_b(\psi)} i_a^{m_b(\theta)} = i_a^{m_b(\psi)+m_b(\theta)},$$

so $m_b(\psi\theta) = m_b(\psi) + m_b(\theta)$. □

Remark 15.38 The same proof shows that m_b defines a homomorphism on both

$$\{\theta \in \mathsf{Out}(F_n) \mid \theta(L_i) = L_i \text{ for } i = 1, 2\} \quad \text{and} \quad \{\theta \in \mathsf{Out}(F_n) \mid \theta(\mathsf{H}_{\phi,\mathsf{c}}(L_i)) = \mathsf{H}_{\phi,\mathsf{c}}(L_i) \text{ for } i = 1, 2\}.$$

The former is the stabilizer of b and the latter can be thought of as the "weak stabilizer" of b.

The next lemma relates $m_b(\phi)$ to the twist coefficients of ϕ.

Lemma 15.39 *Suppose that $b = (L_1, L_2) \in \mathcal{S}_2(\phi, r)$, where $L_1 = (R_1^-)^{-1}\rho_1 R_1^+$ and $L_2 = (R_2^-)^{-1}\rho_2 R_2^+$ are the decompositions of Corollary 5.17(1). Suppose also that w is a twist path and that $E', E'' \in \mathsf{Lin}_w(f)$ satisfy $f(E') = E'w^{d'}$ and $f(E'') = E''w^{d''}$.*

1. *If $R_1^+ = E'w^{\pm\infty}$ and $R_2^- = E''w^{\pm\infty}$, then $m_b(\phi) = d' - d''$.*
2. *If $R_1^+ = E'w^{\pm\infty}$ and $R_2^- = w^{\pm\infty}$, then $m_b(\phi) = d'$.*
3. *If $R_1^+ = w^{\pm\infty}$ and $R_2^- = E''w^{\pm\infty}$, then $m_b(\phi) = -d''$.*

In particular, $m_b(\phi) \neq 0$ for all $b \in \mathcal{S}_2(\phi)$.

Proof Choose lifts $\widetilde{L}_1 = (\widetilde{R}_1^-)^{-1}\tilde{\rho}_1\widetilde{R}_1^+$, $\widetilde{L}_2 = (\widetilde{R}_2^-)^{-1}\tilde{\rho}_2\widetilde{R}_2^+$ and $\widetilde{A} = \tilde{w}^\infty$ so that $\partial\widetilde{R}_1^+, \partial\widetilde{R}_2^- \in \{\partial_-\widetilde{A}, \partial_+\widetilde{A}\}$. Denote the initial endpoints of $\widetilde{R}_1^+$ and $\widetilde{R}_2^-$ by $\tilde{x}$ and $\tilde{y}$, respectively. There exist $\Phi_1, \Phi_2 \in \phi$ such that Φ_1 fixes the endpoints of $\widetilde{L}_1$ and Φ_2 fixes the endpoints of $\widetilde{L}_2$. The corresponding lifts $\tilde{f}_1$ and $\tilde{f}_2$ fix $\tilde{x}$ and $\tilde{y}$, respectively. In particular, $\tilde{f}_1(\tilde{y}) = i_a^{m_b(\phi)}\tilde{f}_2(\tilde{y}) = i_a^{m_b(\phi)}\tilde{y}$. In case (1), the path $\tilde{\tau}$ connecting $\tilde{x}$ to $\tilde{y}$ equals $\widetilde{E}'\tilde{w}^p(\widetilde{E}'')^{-1}$ for some $p \in \mathbb{Z}$ and $(\tilde{f}_1)_\#(\widetilde{E}'\tilde{w}^p(\widetilde{E}'')^{-1}) = \widetilde{E}'\tilde{w}^{p+d'-d''}(\widetilde{E}'')^{-1}$. It follows that $\tilde{f}_1(\tilde{y}) = i_a^{d'-d''}\tilde{y}$ and hence that $m_b(\phi) = d' - d''$. In case (2), $\tilde{\tau} = \widetilde{E}'\tilde{w}^p$ and $(\tilde{f}_1)_\#(\widetilde{E}'\tilde{w}^p) = \widetilde{E}'\tilde{w}^{p+d'}$. Thus, $\tilde{f}_1(\tilde{y}) = i_a^{d'}\tilde{y}$ and $m_b(\phi) = d'$. Case (3) is proved similarly.

Lemma 15.26 implies that if b is as in case (1) then either $b = (\ell_{i-1}, \ell_{i+1})$, where $\sigma_i\rho_i\sigma_{i+1}$ is quasi-exceptional, or $b = (\ell_{i-1}, \ell_i)$, where σ_i is exceptional. In either case, $E' \neq E''$ so $d' \neq d''$. This completes the proof that $m_b(\phi) \neq 0$ and hence the proof of the lemma. □

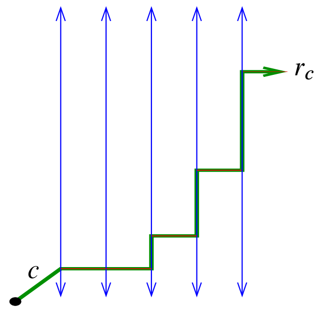


Figure 6: An end of $R_c = cbbaba^2ba^3\dots$ is in a union of staple pairs, cf Figure 5. The relation between staple pairs and R_E for E of higher than quadratic growth is more complicated.

15.4 Spanning staple pairs

We continue with the notation of the preceding subsections; in particular, see Notation 15.4. In addition, we let $\widetilde{\mu}_0, \widetilde{\mu}_1 \dots$ be the sequence of lines obtained from $\widetilde{\ell}_0, \widetilde{\ell}_1, \dots$ by removing all periodic lines. In other words, $\widetilde{\mu}_0, \widetilde{\mu}_1 \dots$ is the set of visible lines in $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.

If E has quadratic growth (equivalently, each σ_i is linear) then every $(\widetilde{\mu}_t, \widetilde{\mu}_{t+1})$ is an element of $\mathcal{S}_2(\phi, \widetilde{r})$ by Lemma 15.26; see Figure 6. This is not the case when E has higher order. We now define a related but weaker property that does hold for every $(\widetilde{\mu}_t, \widetilde{\mu}_{t+1})$. Its utility is illustrated in the proof of Lemma 15.45, which is applied in the proof of Lemma 17.9.

Definition 15.40 We say that an ordered pair $(\widetilde{\eta}_1, \widetilde{\eta}_2)$ of elements of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ is *spanned by a staple pair* $\widetilde{b} = (\widetilde{L}_1, \widetilde{L}_2) \in \mathcal{S}_2(\phi, \widetilde{r})$ if the following two conditions are satisfied:

- Either $\widetilde{L}_1 = \widetilde{\eta}_1$ or $\widetilde{L}_1 \in \Omega_{\mathsf{NP}}(\phi, \partial_+ \widetilde{\eta}_1)$.
- Either $\widetilde{L}_2 = \widetilde{\eta}_2$ or $\widetilde{L}_2^{-1} \in \Omega_{\mathsf{NP}}(\phi, \partial_- \widetilde{\eta}_2)$.

Note that if $(\widetilde{\eta}_1, \widetilde{\eta}_2)$ is spanned by a staple pair then $(\Phi^k_{\widetilde{r}} \widetilde{\eta}_1, \Phi^k_{\widetilde{r}} \widetilde{\eta}_2)$ is spanned by a staple pair for all $k \in \mathbb{Z}$.

Our next result uses techniques from the proofs of Lemmas 15.12 and 15.31.

Lemma 15.41 (1) *Suppose that $\sigma_i \in \mathcal{E}_f$ and that $\widetilde{r}_i = \partial_+ \widetilde{\ell}_{i-1} = \partial \widetilde{R}_i^+$. If $\widetilde{\sigma}_j \widetilde{\rho}_j \widetilde{\sigma}_{j+1}$ is a subpath of $\widetilde{f}_\#^m(\widetilde{\sigma}_i)$ for $m > 0$ and if $\widetilde{\ell}_j$ is nonperiodic, then $\Phi_{\widetilde{r}}^{-m}(\widetilde{\ell}_j) \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r}_i)$.*

(2) *Suppose that $\sigma_i \in \mathcal{E}_f^{-1}$ and that $\widetilde{r}_i = \partial_- \widetilde{\ell}_i = \partial \widetilde{R}_i^-$. If $\widetilde{\sigma}_j \widetilde{\rho}_j \widetilde{\sigma}_{j+1}$ is a subpath of $\widetilde{f}_\#^m(\widetilde{\sigma}_i)$ for $m > 0$ and if $\widetilde{\ell}_j$ is nonperiodic, then $\Phi_{\widetilde{r}}^{-m}(\widetilde{\ell}_j^{-1}) \in \Omega_{\mathsf{NP}}(\phi, \widetilde{r}_i)$.*

Proof The two cases are symmetric so we prove (1) and leave (2) to the reader. Assuming that $\sigma_i \in \mathcal{E}_f$, let $T : \widetilde{G} \to \widetilde{G}$ be the covering translation that carries $\widetilde{\sigma}_i$ to the initial edge of $\widetilde{f}_\#^m(\widetilde{\sigma}_i)$ and hence satisfies

$T(\widetilde R_i^+) = \tilde f_\#^m(\widetilde R_i^+) = \Phi_{\tilde r}^m(\widetilde R_i^+)$. Then $T^{-1}\tilde f^m$ is the lift of f^m that preserves the terminal endpoint $\tilde r_i$ of $\widetilde R_i^+$ and so $i_c^{-1}\Phi_{\tilde r}^m = \Phi_{\tilde r_i}^m$, where i_c is the inner automorphism corresponding to T. Note that $\tilde f_\#^m(\tilde\sigma_i)$ is a concatenation of terms in the complete splitting of $\widetilde R_{\widetilde E}$ whose first edge $T(\tilde\sigma_i)$ projects into $\mathcal{E}_f$. Thus $\tilde f_\#^m(\tilde\sigma_i) = \tilde\sigma_a\cdot\tilde\rho_a\cdot\ldots\cdot\tilde\rho_{b-1}\cdot\tilde\sigma_b\cdot\tilde\tau$ for some $a\le j<b$ and some (possibly trivial) Nielsen path $\tilde\tau$ that is an initial segment of $\tilde\rho_b$. Note also that $T^{-1}\tilde f_\#^m(\tilde\sigma_i)$ is a concatenation of terms in the complete splitting of $\widetilde R_i^+$. It follows that $T^{-1}\tilde\sigma_a\cdot T^{-1}\tilde\rho_a\cdot\ldots\cdot T^{-1}\tilde\rho_{b-1}\cdot T^{-1}\tilde\sigma_b$ is a concatenation of terms in the coarsened complete splitting of $\widetilde R_i^+$. In particular, $T^{-1}\tilde\sigma_j\cdot T^{-1}\tilde\rho_j\cdot T^{-1}\tilde\sigma_{j+1}$ is a concatenation of terms in the coarsened complete splitting of $\widetilde R_i^+$. Since $\tilde\ell_j$ is nonperiodic, the same is true for $T^{-1}(\tilde\ell_j)$ and we conclude that $T^{-1}(\tilde\ell_j)\in\Omega_{\mathsf{NP}}(\phi,\tilde r_i)$. Proposition 15.6 implies that $\Phi_{\tilde r}^{-m}(\tilde\ell_j) = \Phi_{\tilde r_i}^{-m} i_c^{-1}(\tilde\ell_j) = \Phi_{\tilde r_i}^{-m}(T^{-1}(\tilde\ell_j))$ is an element of $\Omega_{\mathsf{NP}}(\phi,\tilde r_i)$. □

Proposition 15.42 *Each ordered pair* $(\tilde\mu_t,\tilde\mu_{t+1})$ *is spanned by an element* $(\widetilde L_1,\widetilde L_2)\in\mathcal{S}_2(\phi,\tilde r)$*. If* $\partial_+\tilde\mu_t$ *(resp.* $\partial_-\tilde\mu_{t+1}$*) is periodic, then* $\widetilde L_1=\tilde\mu_t$ *(resp.* $L_2=\tilde\mu_{t+1}$*).*

Proof Set $\Phi=\Phi_{\tilde r}$. We first show that if $\tilde\ell_i$ is periodic then $\tilde\ell_{i-1}$ and $\tilde\ell_{i+1}$ are nonperiodic and $(\tilde\mu_t,\tilde\mu_{t+1}) := (\tilde\ell_{i-1},\tilde\ell_{i+1})$ satisfies the conclusions of the lemma. Let M be the stabilization constant for f (Notation 5.13). If $\tilde f_\#^M(\tilde\rho_i)\subset\tilde\rho_j$, then $\tilde\sigma_j$ is the last growing term in $\tilde f_\#^M(\tilde\sigma_i)$ and $\tilde\sigma_{j+1}$ is the first growing term in $\tilde f_\#^M(\tilde\sigma_{i+1})$. Moreover, Lemma 15.9 implies that $\tilde\ell_j=\tilde f_\#^M(\tilde\ell_i)$ and so $\tilde\ell_j$ is periodic. By our choice of M, $\tilde\sigma_j\notin\mathcal{E}_f$ and $\tilde\sigma_{j+1}\notin\mathcal{E}_f^{-1}$. Lemma 15.26(3) implies that $\tilde\sigma_j\tilde\rho_j\tilde\sigma_{j+1}$ is quasi-exceptional and $(\tilde\ell_{j-1},\tilde\ell_{j+1})\in\mathcal{S}_2(\phi,\tilde r)$. Since $\sigma_j\in\mathsf{Lin}(f)$, it follows that either $\sigma_i\in\mathsf{Lin}(f)$ or $\sigma_i\in\mathcal{E}_f$. In the former case, $\tilde f_\#^M(\tilde\rho_{i-1})\subset\tilde\rho_{j-1}$ and $\tilde f_\#^M(\tilde\ell_{i-1})=\tilde\ell_{j-1}$ so $\Phi^{-M}(\tilde\ell_{j-1})=\tilde\ell_{i-1}$; in particular, $\tilde\ell_{i-1}$ is nonperiodic and $\partial_+\tilde\ell_{i-1}$ is periodic. In the latter case, $\tilde\sigma_{j-1}\tilde\rho_{j-1}\tilde\sigma_j$ is a terminal subpath of $\tilde f_\#^M(\tilde\sigma_i)$ so Lemma 15.41 implies that $\Phi^{-M}(\tilde\ell_{j-1})\in\Omega_{\mathsf{NP}}(\phi,\partial_+\tilde\ell_{i-1})$; note that in this latter case, $\partial_+\tilde\mu_t=\partial_+\tilde\ell_{i-1}$ is not periodic. A symmetric argument shows that either $\Phi^{-M}(\tilde\ell_{j+1})=\tilde\ell_{i+1}$ or $\Phi^{-M}(\tilde\ell_{j+1}^{-1})\in\Omega_{\mathsf{NP}}(\phi,\partial_-\tilde\ell_{i+1})$. Thus $(\tilde\ell_{i-1},\tilde\ell_{i+1})$ is spanned by $\Phi^{-M}(\tilde\ell_{j-1},\tilde\ell_{j+1})$, and the "if" statement of the lemma is satisfied.

Remark 15.43 The above argument includes a proof that if $\tilde\ell_i$ is periodic and if $\tilde\ell_j=\tilde f_\#^M(\tilde\ell_i)$, then $\tilde\sigma_j\tilde\rho_j\tilde\sigma_{j+1}$ is quasi-exceptional and $\tilde\ell_j$ is the common axis of $(\tilde\ell_{j-1},\tilde\ell_{j+1})\in\mathcal{S}_2(\phi,\tilde r)$.

Continuing with the proof, we now know that for each $(\tilde\mu_t,\tilde\mu_{t+1})$ there exists i such that $(\tilde\mu_t,\tilde\mu_{t+1})$ is equal to either $(\tilde\ell_{i-1},\tilde\ell_i)$ or $(\tilde\ell_{i-1},\tilde\ell_{i+1})$. Moreover, the conclusions of the lemma hold in the latter case so we may assume that $(\tilde\mu_t,\tilde\mu_{t+1})=(\tilde\ell_{i-1},\tilde\ell_i)$. Lemma 15.26(1) implies that $\sigma_i\rho_i\sigma_{i+1}$ is not quasi-exceptional. If σ_i is linear (ie σ_i is exceptional or $\sigma_i\in\mathsf{Lin}(f)$ or $\bar\sigma_i\in\mathsf{Lin}(f)$) then $(\tilde\mu_t,\tilde\mu_{t+1}) = (\tilde\ell_{i-1},\tilde\ell_i)\in\mathcal{S}_2(\phi,\tilde r)$ by Lemma 15.26(2).

It remains to consider the $\sigma_i\in\mathcal{E}_f$ and $\sigma_i\in\mathcal{E}_f^{-1}$ cases. These are symmetric so we assume that $\sigma_i\in\mathcal{E}_f$, and leave the $\sigma_i\in\mathcal{E}_f^{-1}$ case to the reader. For the "if" statement, note that $\partial_+\tilde\ell_{i-1}$ is nonperiodic. As above, there exists $j>i$ such that $\tilde f_\#^M(\tilde\rho_i)\subset\tilde\rho_j$ and $\tilde\ell_j=\tilde f_\#^M(\tilde\ell_i)$; in particular, $\tilde\ell_j$ is not periodic. Since $\tilde\sigma_j$ is the last growing term in $\tilde f_\#^M(\tilde\sigma_i)$ and since $\tilde f_\#^M(\tilde\sigma_i)$ contains at least two growing terms, Lemma 15.41 implies that if $\widetilde{\ell}_{j-1}$ is nonperiodic then $\Phi^{-M}(\widetilde{\ell}_{j-1}) \in \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_{i-1})$. If σ_j is exceptional or $\sigma_j \in \mathsf{Lin}(f)$ then another application of Lemma 15.26(2) shows that $(\widetilde{\ell}_{j-1}, \widetilde{\ell}_j) \in \mathcal{S}_2(\phi, \widetilde{r})$. In this case, $(\widetilde{\ell}_{i-1}, \widetilde{\ell}_i)$ is spanned by $\Phi^{-M}(\widetilde{\ell}_{j-1}, \widetilde{\ell}_j) = (\Phi^{-M}(\widetilde{\ell}_{j-1}), \widetilde{\ell}_i)$ and we are done. The same argument works if $\overline{\sigma}_j \in \mathsf{Lin}(f)$ and $\widetilde{\ell}_{j-1}$ is not periodic. Suppose then that $\overline{\sigma}_j \in \mathsf{Lin}(f)$ and $\widetilde{\ell}_{j-1}$ is periodic. There exists $s > j$ such that $\widetilde{f}^M_\#(\widetilde{\rho}_j) \subset \widetilde{\rho}_s$ and $\widetilde{\ell}_s = \widetilde{f}^M_\#(\widetilde{\ell}_j) = \widetilde{f}^{2M}_\#(\widetilde{\ell}_i)$. Since $\widetilde{\sigma}_j$ is linear, $\widetilde{f}^M_\#(\widetilde{\sigma}_j)$ contains a single growing term and so $\widetilde{f}^M_\#(\widetilde{\rho}_{j-1}) \subset \widetilde{\rho}_{s-1}$. Thus $\widetilde{\ell}_{s-1} = \widetilde{f}^M_\#(\widetilde{\ell}_{j-1})$ is periodic. Remark 15.43 implies that $\widetilde{\sigma}_{s-1}\widetilde{\rho}_{s-1}\widetilde{\sigma}_s$ is quasi-exceptional and $\widetilde{\ell}_{s-1}$ is the common axis of $(\widetilde{\ell}_{s-2}, \widetilde{\ell}_s) \in \mathcal{S}_2(\phi, \widetilde{r})$. Moreover, $\widetilde{\sigma}_{s-2}\widetilde{\rho}_{s-2}\widetilde{\sigma}_{s-1}\widetilde{\rho}_{s-1}\widetilde{\sigma}_s \subset f^{2M}_\#(\widetilde{\sigma}_i)$. The proof now concludes as in the previous case with $(\widetilde{\ell}_{i-1}, \widetilde{\ell}_i)$ spanned by $\Phi^{-2M}(\widetilde{\ell}_{s-2}, \widetilde{\ell}_s) = (\Phi^{-2M}(\widetilde{\ell}_{s-2}), \widetilde{\ell}_i)$.

We may now assume that σ_j has higher order and so $\sigma_j \in \mathcal{E}_f^{-1}$ by our choice of M. In particular, $\partial_-\widetilde{\ell}_j$, and hence $\partial_-\widetilde{\ell}_i$, is nonperiodic.

Choose $k > 0$ so that the coarsened complete splitting of $\widetilde{f}^k_\#(\widetilde{\sigma}_j)$ has at least one linear term $\widetilde{\sigma}_s$ that is neither the first nor second term nor the last or next to last growing term in that splitting. Thus $\widetilde{\sigma}_{s-2} \cdot \widetilde{\rho}_{s-2} \cdot \ldots \cdot \widetilde{\rho}_{s+1} \cdot \widetilde{\sigma}_{s+2}$ is a subpath of $\widetilde{f}^k_\#(\widetilde{\sigma}_j)$ and hence also a subpath of $\widetilde{f}^{M+k}_\#(\widetilde{\sigma}_i)$. By Lemma 15.26, two of the three lines $\widetilde{\ell}_{s-1}, \widetilde{\ell}_s, \widetilde{\ell}_{s+1}$ form an element of $\mathcal{S}_2(\phi, \widetilde{r})$ that we denote by $(\widetilde{L}_1, \widetilde{L}_2)$. Lemma 15.41 implies that

$$\Phi^{-M-k}(\widetilde{L}_1) \in \Omega_{\mathsf{NP}}(\phi, \partial_+\widetilde{\ell}_{i-1}) \quad \text{and} \quad \Phi^{-k}(\widetilde{L}_2^{-1}) \in \Omega_{\mathsf{NP}}(\phi, \partial_-\widetilde{\ell}_j).$$

It follows that

$$\Phi^{-M-k}(\widetilde{L}_2^{-1}) \in \Omega_{\mathsf{NP}}(\phi, \partial_-\widetilde{\ell}_i),$$

and hence that $\Phi^{-M-k}((\widetilde{L}_1, \widetilde{L}_2))$ spans $(\widetilde{\ell}_{i-1}, \widetilde{\ell}_i)$. □

In Lemma 15.45 below we use Proposition 15.42 to give conditions on $\upsilon \in \mathsf{Out}(F_n)$ which imply that υ fixes r. The proof of Lemma 15.45 is inductive and it is useful in the induction step to know that υ strongly fixes r in the following sense.

Definition 15.44 We say that $\upsilon \in \mathsf{Out}(F_n)$ *strongly fixes* $r \in \mathcal{R}(\phi)$ if for some (and hence every) lift $\widetilde{r}$ there is a lift $\Upsilon \in \upsilon$ that fixes each element of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$.

Lemma 15.45 *Suppose that $[F]$ is a ϕ-invariant free factor conjugacy class, that $r \in \mathcal{R}(\phi)$ is carried by $[F]$ and that $\upsilon \in \mathsf{Out}(F_n)$ is such that*

(1) *$[F]$ is υ-invariant,*

(2) *υ fixes each element of $\Omega_{\mathsf{NP}}(r)$ and each $r' < r$* (as defined in Notation 6.1),

(3) *the restriction $\upsilon|F$ commutes with $\phi|F$, and*

(4) *$m_b(\upsilon) = 0$ for all $b \in \mathcal{S}_2(\phi, r)$* (see Definition 15.36 and Remark 15.38),

then υ strongly fixes r.

Proof Given a lift $\tilde{r}$, we continue with the $\tilde{\sigma}_i$, $\tilde{\rho}_i$, $\tilde{\ell}_i$, $\tilde{\mu}_t$ and $\Phi_{\tilde{r}}$ notation. We may assume without loss of generality that $\tilde{r} \in \partial F$. By uniqueness, F is $\Phi_{\tilde{r}}$-invariant. By Lemma 15.13 each $\tilde{\mu}_t$ has endpoints in ∂F.

For each $t \geq 1$, item (2) implies the existence of a (necessarily unique) lift Υ_t of υ that fixes $\tilde{\mu}_t$. We show below that Υ_t is independent of t, say $\Upsilon_t = \Upsilon$ for all t. Assuming this for now, the proof concludes as follows. Since the endpoints of the $\tilde{\mu}_t$ limit on $\tilde{r}$, we have $\Upsilon(\tilde{r}) = \tilde{r}$. From this and (1) it follows that $\Upsilon(F) = F$. Item (3) implies that the commutator $[\Phi_{\tilde{r}}|F, \Upsilon|F]$ is inner. Since the commutator $[\Phi_{\tilde{r}}|F, \Upsilon|F]$ fixes $\tilde{r}$, it must be trivial. Thus, $\Phi_{\tilde{r}}|F$ and $\Upsilon|F$ commute and the same is true for $\Phi_{\tilde{r}}|\partial F$ and $\Upsilon|\partial F$. Given $\tilde{L} \in \Omega_{\mathsf{NP}}(\phi, \tilde{r})$, there exist $m \geq 0$ and $t \geq 1$ such that $\Phi_{\tilde{r}}^m(\tilde{L}) = \tilde{\mu}_t$ by Proposition 15.6. Since $\tilde{\mu}_t$ has endpoints in ∂F, the same is true for $\tilde{L}$. Thus

$$\Upsilon(\tilde{L}) = (\Phi_{\tilde{r}}^{-m} \Upsilon \Phi_{\tilde{r}}^m)(\tilde{L}) = \Phi_{\tilde{r}}^{-m} \Upsilon(\Phi_{\tilde{r}}^m(\tilde{L})) = \Phi_{\tilde{r}}^{-m} \Upsilon(\tilde{\mu}_t) = \Phi_{\tilde{r}}^{-m}(\tilde{\mu}_t) = \tilde{L},$$

as desired.

It remains to prove that $\Upsilon_t = \Upsilon_{t+1}$ for all $t \geq 1$.

The proof is by induction on the height of r in the partial order $<$ on $\mathcal{R}(\phi)$. In the base case, r is a minimal element of $\mathcal{R}(\phi)$ so each σ_i is linear by Lemmas 6.2 and 3.21. In this case, each $\tilde{\mu}_t \in \mathcal{S}(\phi, \tilde{r})$ and each $(\tilde{\mu}_t, \tilde{\mu}_{t+1}) \in \mathcal{S}_2(\phi, \tilde{r})$ by Lemma 15.26. Item (4) completes the proof.

For the inductive step, we use Proposition 15.42. Let $\mu_t = \ell_i$ and $\mu_{t+1} = \ell_j$. As a first case, suppose that $\partial_+ \tilde{\ell}_i = \tilde{r}'$ for some $r' \in \mathcal{R}(\phi)$. Let $E' \in \mathcal{E}_f$ be the higher-order edge corresponding to r'. Then $r > r'$ and Lemmas 6.2 and 3.21 imply that either E' or $\bar{E}'$ occurs as a term in the complete splitting of some $f_\#^k(E)$. Thus either $\mathsf{Acc}(E') \subset \mathsf{Acc}(E)$ or $\mathsf{Acc}(\bar{E}') \subset \mathsf{Acc}(E)$. Lemma 5.8 therefore implies that if $L' \in \Omega_{\mathsf{NP}}(r')$ then either $L' \in \Omega_{\mathsf{NP}}(r)$ or $L'^{-1} \in \Omega_{\mathsf{NP}}(r)$. From (2) we see that υ fixes each element of $\Omega_{\mathsf{NP}}(r')$. Lemma 15.34 and (4) imply that $m_b(\upsilon) = 0$ for all $b \in \mathcal{S}_2(\phi, r')$ so r' and $[F]$ satisfy the hypotheses of this lemma. By the inductive hypothesis, there is a (necessarily unique) lift Υ' that fixes each element of $\Omega_{\mathsf{NP}}(\phi, \tilde{r}')$. As noted above, it follows that Υ' fixes $\tilde{r}'$ and so $\Upsilon' = \Upsilon_t$.

There are two subcases. The first is that $\partial_- \tilde{\ell}_j = \tilde{r}''$ for some $r'' \in \mathcal{E}_f$. Arguing as in the previous paragraph we see that Υ_{t+1} fixes each element of $\Omega_{\mathsf{NP}}(\phi, \tilde{r}'')$. By Proposition 15.42, there exist $\tilde{L}' \in \Omega_{\mathsf{NP}}(\phi, \tilde{r}')$ and $\tilde{L''}^{-1} \in \Omega_{\mathsf{NP}}(\phi, \tilde{r}'')$ such that $(\tilde{L}', \tilde{L}'') \in \mathcal{S}_2(\phi, \tilde{r})$. Item (4) implies that $\Upsilon_t = \Upsilon_{t+1}$.

The second subcase is that $\partial_- \tilde{\ell}_j$ is the end of a periodic line and hence $\ell_j \in \mathcal{S}(\phi, r)$. By Proposition 15.42, there exists $\tilde{L}' \in \Omega_{\mathsf{NP}}(\phi, \tilde{r}')$ such that $(\tilde{L}', \tilde{\ell}_j) \in \mathcal{S}_2(\phi, \tilde{r})$. Once again, (4) implies that $\Upsilon_t = \Upsilon_{t+1}$. This completes the proof when $\partial_+ \tilde{\ell}_i$ projects into $\mathcal{R}(\phi)$.

A symmetric argument handles the case that $\partial_- \tilde{\ell}_j$ projects into $\mathcal{R}(\phi)$. The remaining case is that both $\partial_+ \tilde{\ell}_i$ and $\partial_- \tilde{\ell}_j$ are endpoints of periodic lines. Proposition 15.42 implies that $(\tilde{\ell}_i, \tilde{\ell}_j) \in \mathcal{S}_2(\phi, \tilde{r})$ so (4) completes the proof as in previous cases. □

16 The homomorphism $\bar{Q}$

Recall that our main theorem is reduced to Proposition 14.7 by Lemma 14.8.

For the rest of the paper, we assume the hypotheses of Proposition 14.7, ie $\phi, \psi \in \mathsf{UPG}(F_n)$, $\mathfrak{c}$ is a special chain for ϕ and ψ, and $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$. Our goal is to find a conjugator $\theta \in \mathcal{X}_\mathfrak{c}(\phi)$ or prove that no such conjugator exists.

Because ϕ and $\mathfrak{c}$ are fixed for the rest of the paper, we will often write $\mathcal{X}$ for $\mathcal{X}_\mathfrak{c}(\phi)$. In fact, we will often suppress $\mathfrak{c}$ when it appears as a decoration.

Lemma 16.1 *For each $\mathcal{F}_i \in \mathfrak{c}$ and each $r \in \mathcal{R}(\phi|\mathcal{F}_i)$ there exists $r' \in \mathcal{R}(\psi|\mathcal{F}_i)$ such that $\theta(r) = r'$ for each $\theta \in \mathcal{X}$ whose restriction $\theta|\mathcal{F}_i$ conjugates $\phi|\mathcal{F}_i$ to $\psi|\mathcal{F}_i$.*

Proof Let $\mathfrak{e} = \mathcal{F}^- \sqsubset \mathcal{F}^+ \in \mathfrak{c}$ be the one-edge extension with respect to which r is new (Definition 6.14). In other words, $r \in \mathcal{R}^+(\mathfrak{e}, \phi) := \mathcal{R}(\phi|\mathcal{F}^+) \setminus \mathcal{R}(\phi|\mathcal{F}^-)$. Since θ fixes $\mathfrak{c}$, it follows that $\theta|\mathcal{F}^\pm$ conjugates $\phi|\mathcal{F}^\pm$ to $\psi|\mathcal{F}^\pm$ and so θ induces a bijection between $\mathcal{R}^+(\mathfrak{e}, \phi)$ and $\mathcal{R}^+(\mathfrak{e}, \psi)$. This completes the proof if r is the only element of $\mathcal{R}^+(\mathfrak{e}, \phi)$. Otherwise $\mathcal{R}^+(\mathfrak{e}, \phi) = \{r, s\}$ and $\mathcal{R}^+(\mathfrak{e}, \psi) = \{r', s'\}$ and we are in case HH. By definition, $\mathsf{L}_\mathfrak{e}(\phi) = \{L, L^{-1}\}$, where $\partial_- L = r$ and $\partial_+ L = s$. By Lemma 6.15, $\theta(L) \in \mathsf{L}_\mathfrak{e}(\psi)$. Since $\theta \in \mathcal{X}$, we get $\theta(\mathsf{H}_{\phi,\mathfrak{c}}(L)) = \mathsf{H}_{\phi,\mathfrak{c}}(L) = \mathsf{H}_{\psi,\mathfrak{c}}(\theta(L))$. By Lemma 13.9, there is a unique $L' \in \mathsf{L}_\mathfrak{e}(\psi)$ that is in $\mathsf{H}_{\psi,\mathfrak{c}}(\theta(L))$. Hence $\theta(L) = L'$ and $\theta(r) = r' := \partial_- L'$. $\square$

We continue with the notation of Section 15 and also assume that a CT $f' \colon G' \to G'$ representing ψ has been chosen that realizes $\mathfrak{c}$. We use prime notation when working with ψ and r'; for example, $E' \in \mathcal{E}_{f'}$ is the edge corresponding to r' and $\tilde{\ell}'_1, \tilde{\ell}'_2, \dots$ are the visible lines in $\tilde{R}'_{\tilde{E}'}$ and $\Psi_{\tilde{r}'}$ is the lift $\Psi \in \psi$ that fixes $\tilde{r}'$.

Definition 16.2 Recall from Corollary 5.17, Definition 15.36 and Lemma 15.39 that $\mathcal{S}_2(\phi)$ is finite, that for all $b \in \mathcal{S}_2(\phi)$ there is a homomorphism $m_b \colon \mathcal{X} \to \mathbb{Z}$ and that $m_b(\phi) \neq 0$. Define a homomorphism $Q^\phi \colon \mathcal{X} \to \mathbb{Q}^{\mathcal{S}_2(\phi)}$ by letting the b-coordinate of $Q^\phi(\theta)$ be $Q^\phi_b(\theta) = m_b(\theta)/m_b(\phi)$.

Definition 16.3 Let $\sim$ be the equivalence relation on $\mathcal{S}_2(\phi)$ generated by $b \sim b'$ if b and b' occur in the same $r \in \mathcal{R}(\phi)$ (as defined in Notation 15.24), and let

$$\mathcal{S}_2(\phi) = \mathcal{S}_2^1(\phi) \sqcup \mathcal{S}_2^2(\phi) \sqcup \cdots$$

be the decomposition of $\mathcal{S}_2(\phi)$ into $\sim$-equivalence classes. For each i, consider the diagonal action of $\mathbb{Z}$ on $\mathbb{Q}^{\mathcal{S}_2^i(\phi)}$, ie $k\vec{s} = \vec{s} + k(1, 1, \dots)$. Let $\bar{Q}^\phi$ denote the homomorphism

$$\mathcal{X}_\mathfrak{c}(\phi) \xrightarrow{Q^\phi} \mathbb{Q}^{\mathcal{S}_2(\phi)} \to \bar{\mathbb{Q}}^{\mathcal{S}_2(\phi)} := (\mathbb{Q}^{\mathcal{S}_2^1(\phi)}/\mathbb{Z}) \oplus (\mathbb{Q}^{\mathcal{S}_2^2(\phi)}/\mathbb{Z}) \oplus \cdots .$$

For the rest of the paper, Q and $\bar{Q}$ will always denote Q^ϕ and $\bar{Q}^\phi$.

We can now state the second reduction of the conjugacy problem for $\mathsf{UPG}(F_n)$ in $\mathsf{Out}(F_n)$.

Proposition 16.4 *There is an algorithm that takes as input $\phi, \psi \in \mathsf{UPG}(F_n)$ and a chain $\mathfrak{c}$ such that*

- *$\mathfrak{c}$ is a special chain for ϕ and ψ, and*
- *$\mathsf{I}_{\mathfrak{c}}(\phi) = \mathsf{I}_{\mathfrak{c}}(\psi)$,*

and that outputs `YES` *or* `NO` *depending on whether or not there is a $\theta \in \mathsf{Ker}(\bar{Q}^{\phi})$ conjugating ϕ to ψ. Further, if* `YES`*, then such a θ is produced.*

Lemma 16.5 *Proposition 16.4 implies Proposition 14.7 and hence Theorem 1.1.*

Proposition 16.4 is proved in Section 17.

Lemma 16.5 is proved by applying the following technical proposition, whose proof takes up the rest of this section.

Proposition 16.6 *We have an algorithm that produces a finite set $\{\eta_i\} \subset \mathcal{X}$ so that the union of the cosets of $\mathsf{Ker}(\bar{Q})$ determined by the η_i contains each $\theta \in \mathcal{X}$ that conjugates ϕ to ψ.*

Proof of Lemma 16.5, assuming Proposition 16.6 Let $\{\eta_i\}$ be the finite set produced by Proposition 16.6 and let $\psi_i = \psi^{(\eta_i^{-1})}$. It follows that $\phi^{\theta} = \psi$ if and only if $\phi^{\theta_i'} = \psi_i$, where $\theta_i' = \eta_i^{-1}\theta$, and that θ is in the coset represented by η_i if and only if $\theta_i' \in \mathsf{Ker}(\bar{Q})$. Thus, by applying Proposition 16.4 to ϕ and ψ_1, we can decide if there exists θ in the coset represented by η_1 that conjugates ϕ to ψ. If `YES` then return `YES` and one such θ. Otherwise move on to η_2 and repeat. If `NO` for each η_i, then return `NO`. □

The following two lemmas are proved in Sections 16.1 and 16.3, respectively. In the remainder of this section we use them to prove Proposition 16.6. The definition of topmost staple pair appears in Notation 15.30. The definition of $\mathsf{offset}(\theta, r)$ is given in Lemma 15.21(2). The partial order $<$ on $\mathcal{R}(\phi)$ is defined in Notation 6.1.

Lemma 16.7 *Suppose that $b \in \mathcal{S}_2(\phi, r)$ is topmost and that $\theta \in \mathcal{X}$ conjugates ϕ to ψ. Then given an upper bound for $|\mathsf{offset}(\theta, r)|$ one can compute an upper bound for $|m_{\theta}(b)|$.*

Lemma 16.8 *Suppose that $\theta \in \mathcal{X}$ conjugates ϕ to ψ and that $r, r_1 \in \mathcal{R}(\phi)$ satisfy $r_1 < r$. Then given an upper bound for $|\mathsf{offset}(\theta, r)|$ one can compute an upper bound for $|\mathsf{offset}(\theta, r_1)|$.*

Proof of Proposition 16.6, assuming Lemmas 16.7 and 16.8 We begin by computing $D = D(\phi, \psi)$ so that $|Q_{b_1}(\theta) - Q_{b_2}(\theta)| < D$ for all $\theta \in \mathcal{X}$ that conjugate ϕ to ψ and all $b_1, b_2 \in \mathcal{S}_2(\phi)$ satisfying $b_1 \sim b_2$.

Given $r \in \mathcal{R}(\phi)$ we will find D_r such that $|Q_{b_1}(\theta) - Q_{b_2}(\theta)| < D_r$ for all $\theta \in \mathcal{X}$ that conjugate ϕ to ψ and all $b_1, b_2 \in \mathcal{S}_2(\phi, r)$. We then take $D = |\mathcal{R}(\phi)| \max\{D_r\}$, where the $|\mathcal{R}(\phi)|$ factor allows us to consider equivalent staple pairs that do not occur in the same ray.

For all $s \in \mathbb{Z}$, $\theta\phi^s$ is an element of $\mathcal{X}$ and conjugates ϕ to ψ; see Lemma 14.3. The translation number $\tau(\phi, r)$ is defined in Notation 15.18. By definition and by Lemma 15.21 we have

$$\mathsf{offset}(\theta\phi^s, r) = \mathsf{offset}(\theta, r) + \tau(\phi^s, r) = \mathsf{offset}(\theta, r) + s\tau(\phi, r).$$

Since

$$Q_{b_1}(\theta\phi^s) - Q_{b_2}(\theta\phi^s) = (Q_{b_1}(\theta) + s) - (Q_{b_2}(\theta) + s) = Q_{b_1}(\theta) - Q_{b_2}(\theta),$$

we may assume without loss of generality that

$$0 \le \mathsf{offset}(\theta, r) \le \tau(\phi, r).$$

Using only this inequality we will produce an upper bound D_0 for $|m_\theta(b)|$ when $b \in \mathcal{S}_2(\phi, r)$. This determines an upper bound for $|Q_b(\theta)|$ when $b \in \mathcal{S}_2(\phi, r)$, which when doubled gives the desired upper bound D_r for $|Q_{b_1}(\theta) - Q_{b_2}(\theta)|$ when $b_1, b_2 \in \mathcal{S}_2(\phi, r)$.

If b is topmost in r then Lemma 16.7 gives us D_0. Otherwise, choose $r_1 < r$ so that b is topmost in r_1. Apply Lemma 16.8 to find an upper bound for $|\mathsf{offset}(\theta, r_1)|$ and then apply Lemma 16.7 to b and r_1 to produce D_0 and hence D.

To complete the proof of Proposition 16.6, define

$$\mathcal{X}(D) := \{\theta \in \mathcal{X} \mid |Q_{b_1}(\theta) - Q_{b_2}(\theta)| < D \text{ for all } b_1 \sim b_2 \in \mathcal{S}_2(\phi)\}.$$

Our choice of D guarantees that $\mathcal{X}(D)$ contains all $\theta \in \mathcal{X}$ that conjugate ϕ to ψ. For each i, the image of $\mathcal{X}(D)$ by

$$Q^i : \mathcal{X} \xrightarrow{Q} \mathbb{Q}^{\mathcal{S}_2(\phi)} \to \mathbb{Q}^{\mathcal{S}_2^i(\phi)}$$

is discrete, $\mathbb{Z}$-invariant, and contained in a bounded neighborhood of the diagonal in $\mathbb{Q}^{\mathcal{S}_2^i(\phi)}$. Hence the image of $\mathcal{X}(D)$ by

$$\bar{Q}^i : \mathcal{X} \xrightarrow{Q^i} \mathbb{Q}^{\mathcal{S}_2^i(\phi)} \to \mathbb{Q}^{\mathcal{S}_2^i(\phi)}/\mathbb{Z}$$

is finite and $\mathcal{X}(D)$ is contained in finitely many cosets of $\mathsf{Ker}(\bar{Q}^i)$ and so also in finitely many cosets of $\mathsf{Ker}(\bar{Q})$.

To get representatives of these cosets we must find, for each $\bar{q} \in \bar{Q}(\mathcal{X}(D))$, an element of $\mathcal{X} \cap \bar{Q}^{-1}(\bar{q})$. For this, it is enough to express $\bar{q}$ as a word in the $\bar{Q}$-image of the finite generating set $\mathcal{G}_{\mathcal{X}}$ for $\mathcal{X} = \mathsf{Out}_{\mathsf{J}}(F_n)$ supplied by Lemma 11.2. To accomplish this, we find a finite subset $S \subset \mathbb{Q}^{\mathcal{S}_2(\phi)}$ whose image in $\bar{\mathbb{Q}}^{\mathcal{S}_2(\phi)}$ covers $\bar{Q}(\mathcal{X}(D))$ and then express the elements of S in terms of the $Q(\mathcal{G}_{\mathcal{X}})$. To find S, we first find finite $S^i \subset \mathbb{Q}^{\mathcal{S}_2^i(\phi)}$ whose image in $\mathbb{Q}^{\mathcal{S}_2^i(\phi)}/\mathbb{Z}$ covers $\bar{Q}^i(\mathcal{X}(D))$ and then take for S the direct sum of the S^i, ie

$$S := \{q \in \mathbb{Q}^{\mathcal{S}_2(\phi)} \mid \text{the projection of } q \text{ to } \mathbb{Q}^{\mathcal{S}_2^i(\phi)} \text{ is in } S^i\}.$$

We now find S^i. By definition of Q, the denominators of the coordinates of the image of Q are bounded by $\max\{m_b(\phi) \mid b \in \mathcal{S}_2(\phi)\}$. For convenience, we assume we have cleared denominators and all coordinates in the image of Q are integers. Each $\bar{q}_i$ in the image of $\bar{Q}^i$ is represented by $q_i \in \mathbb{Q}^{\mathcal{S}_2^i(\phi)}$ with first coordinate equal to 0. Hence we may then take S^i to be the set of vectors in $\mathbb{Q}^{\mathcal{S}_2^i(\phi)}$ with integer coordinates of absolute value at most D and S to be the set of vectors in $\mathbb{Q}^{\mathcal{S}_2(\phi)}$ with integer coordinates of absolute value at most D.

The desired set of coset representatives can then be taken to be $\{\theta_s \mid s \in S \cap Q(\mathcal{X}(\phi))\}$ where by definition θ_s is a choice of element of $\mathcal{X}(\phi)$ satisfying $Q(\theta_s) = s$. We compute $S \cap Q(\mathcal{X}(\phi))$ and θ_s as follows. First compute $Q(\mathcal{G}_{\mathcal{X}})$. It remains to check which elements of S can be expressed as $\mathbb{Z}$-linear combinations of elements of this $Q(\mathcal{G}_{\mathcal{X}})$ and to produce such a $\mathbb{Z}$-linear combination if it exists. For this, recall that given a finite set of vectors in $\mathbb{Z}^N$ it is standard (see for example [Veblen and Franklin 1921]) to find compatible bases B_0 for the free $\mathbb{Z}$-submodule they generate and B for $\mathbb{Z}^N$. (B_0 and B are *compatible* if there is a subset $\{b_m\}$ of B and integers n_m such that $B_0 = \{n_m \cdot b_m\}$.) Without concern for efficiency, write each element of S in terms of B and check using divisibility of coordinates if it can be written in terms of B_0. □

16.1 Proof of Lemma 16.7

Lemma 16.9 *Assume that:*

(1) *$\theta \in \mathcal{X}$ conjugates ϕ to ψ.*

(2) *$b \in \mathcal{S}_2(\phi, r)$ and $b' = \theta(b) \in \mathcal{S}_2(\psi, r')$, where $r' = \theta(r)$.*

(3) *We are given*

  (a) *a lift $\tilde r$ of r and a lift $\tilde b$ of b that is visible in $\tilde r$ with index i, and*

  (b) *a lift $\tilde r'$ of r' and a lift $\tilde b'$ of b' that is visible in $\tilde r'$ with index i'*

  such that $\Theta(\tilde b) = \tilde b'$, where Θ is the unique automorphism representing θ and satisfying $\Theta(\tilde r) = \tilde r'$.

Then one can compute $m_b(\theta)$ up to an additive constant that is independent of θ.

Proof We give a formula for $m_b(\theta)$ up to an error of at most one in terms of quantities s and s' (defined below) and then show how to compute s and s', up to a uniform additive constant, from i and i'.

Let $\tilde b = (\tilde L_1, \tilde L_2)$, where $\tilde L_1 = \tilde\ell_{i-1}$ and $\tilde L_2 = \tilde\ell_i$ or $\tilde\ell_{i+1}$ and let $\tilde A$ be the common axis for $\tilde b$. By Corollary 5.17, $\tilde A$ projects to a twist path w and we assume that the orientation on $\tilde A$ agrees with that of w. Similarly, $\tilde b' = (\tilde L_1', \tilde L_2')$, where $\tilde L_1' = \tilde\ell'_{i'-1}$ and $\tilde L_2' = \tilde\ell'_{i'}$ or $\tilde\ell'_{i'+1}$, and $\tilde A'$ is the common axis for $\tilde b'$. Let $\tilde x_1$ be the nearest point on $\tilde A$ to the initial end $\partial_-\tilde L_1$. (See Figure 7.) If L_1 is not a linear staple then $\mathsf{H}_{\phi,\mathfrak{c}}(L_1) = [F(\partial_-\tilde L_1), \partial_+\tilde L_1]$. In this case, the ray from $\tilde x_1$ to $\partial_-\tilde L_1$ contains an edge $\tilde\sigma_{i-1}$ of height greater than that of $F(\partial_-\tilde L_1)$ and so $\tilde x_1$ is the nearest point on $\tilde A$ to any point in $F(\partial_-\tilde L_1)$.

By hypothesis, $\theta(L_1) = L_1'$. Since $\theta \in \mathcal{X}$, it follows that $L_1' \in \theta(\mathsf{H}_{\phi,\mathfrak{c}}(L_1)) = \mathsf{H}_{\phi,\mathfrak{c}}(L_1)$. Choose a homotopy equivalence $h\colon G' \to G$ that preserves markings. If L_1 is linear then $\mathsf{H}_{\phi,\mathfrak{c}}(L_1) = \{L_1\}$ so $L_1' = L_1$. In this case, we let $\tilde h^1\colon \tilde G' \to \tilde G$ be the lift of h satisfying $\tilde h^1(\tilde L_1') = \tilde L_1$. If L_1 is not linear then $\mathsf{H}_{\phi,\mathfrak{c}}(L_1) = [F_{\mathfrak{c}}(\partial_-\tilde L_1), \partial_+\tilde L_1]$. In this case, we let $\tilde h^1\colon \tilde G' \to \tilde G$ be the lift of h satisfying $\tilde h^1_\#(\tilde L_1') \in (\partial F_{\mathfrak{c}}(\partial_-\tilde L_1), \partial_+\tilde L_1)$. Let Θ_1 be the unique lift of θ satisfying $\Theta_1(\tilde L_1) = \tilde h^1_\#(\tilde L_1') \in \mathsf{H}_{\phi,\mathfrak{c}}(\tilde L_1)$.

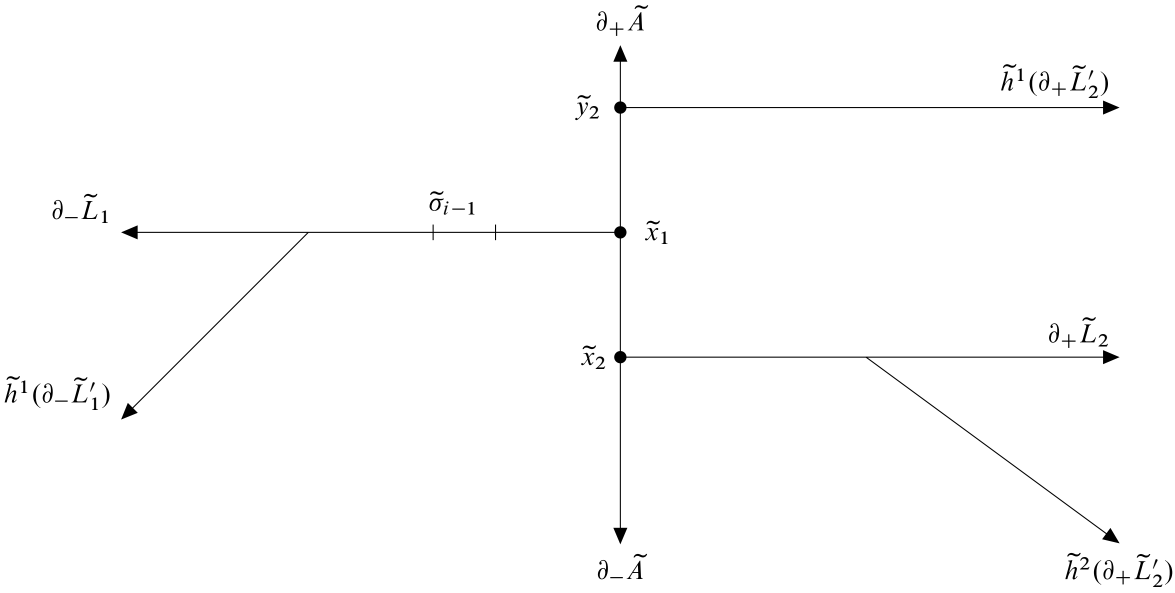


Figure 7

Let $\widetilde{x}_2$ and $\widetilde{y}_2$ be the nearest points on $\widetilde{A}$ to the terminal ends $\partial_+\widetilde{L}_2$ and $\widetilde{h}^1_\#(\partial_+\widetilde{L}'_2)$, respectively. Arguing as above, there is a lift $\widetilde{h}^2\colon \widetilde{G}' \to \widetilde{G}$ such that $\widetilde{h}^2_\#(\widetilde{A}') = \widetilde{A}$ and such that $\widetilde{x}_2$ is the nearest point to $\widetilde{h}^2(\partial_+\widetilde{L}'_2)$. Moreover, there is a lift Θ_2 of θ such that $\widetilde{h}^2_\#(\widetilde{L}'_2) = \Theta_2(\widetilde{L}_2) \in (\partial_- L_2, \partial F_{\mathfrak{c}}(\partial_+\widetilde{L}_2))$. It follows from Definition 15.36 that the oriented path $\widetilde{\alpha} \subset \widetilde{A}$ from $\widetilde{x}_2$ to $\widetilde{y}_2$ has the form $\widetilde{w}^{m_b(\theta)}$.

Let $\widetilde{\beta}$ and $\widetilde{\beta}'$ be the paths in $\widetilde{A}$ connecting $\widetilde{x}_2$ to $\widetilde{x}_1$ and $\widetilde{x}_1$ to $\widetilde{y}_2$, respectively. Let s and s' be the number of complete copies of $\widetilde{w}$ (counted with orientation) crossed by the paths $\widetilde{\beta}$ and $\widetilde{\beta}'$, respectively. Then $|m_b(\theta) - (s' + s)| \le 1$.

Determining s from the index i is straightforward. We consider the cases of Lemma 15.26. In case (1), $\sigma_i \rho_i \sigma_{i+1}$ is quasi-exceptional and $\rho_i = w^s$, where w is the twist path for σ_i and $\overline{\sigma}_{i+1}$. In case (a), $\sigma_i = E' w^s \overline{E}''$ for some $E', E'' \in \mathsf{Lin}_w(f)$. In case (b), $\sigma_i = E'$ is linear with twist path w and ℓ_i is not periodic. If ρ_i has an initial segment of the form w^j for some $j > 0$ then s is the maximal such j; otherwise $-s$ is the maximal $j \ge 0$ such that ρ_i has an initial segment of the form w^{-j}. In case (c), $\overline{\sigma}_{i+1} = E'$ is linear with twist path w and ℓ_{i-1} is not periodic. In this case s is determined by the maximal initial segment of $\overline{\rho}_i$ of the form $w^{\pm j}$ as in the case (b).

Let $\widetilde{x}'_1$ and $\widetilde{x}'_2$ be the nearest points on $\widetilde{A}'$ to $\partial_-\widetilde{L}'_1$ and $\partial_+\widetilde{L}'_1$, respectively. Let $\widetilde{w}'$ be a fundamental domain for the natural action of $\mathbb{Z}$ on $\widetilde{A}'$. Arguing as above, using G' in place of G, we can compute the number t' of complete copies of $\widetilde{w}'$ (counted with orientation) crossed by the path connecting $\widetilde{x}'_2$ to $\widetilde{x}'_1$. We can also compute the bounded cancellation constant C' for h; see [Cooper 1987], also [Bestvina et al. 1997, Lemma 3.1]. Since $|s' - t'| < 2C'$, $m_b(\theta) = t' + s$ up to the additive constant $C = 2C' + 1$. □

Proof of Lemma 16.7 By Lemma 15.31 and Remark 15.32 applied to ψ, $b' = \theta(b)$ and $r' = \theta(r)$ we can find B' so that each lift $\widetilde{b}' \in \mathcal{S}_2(\psi, \widetilde{r}')$ of b' that satisfies $\widetilde{\ell}'_{B'} \prec \widetilde{b}'$ is visible in $\widetilde{r}'$. After increasing B' if necessary, we may assume that $\widetilde{\ell}'_{B'}$ is topmost in $\widetilde{r}'$. Now apply Lemma 15.31 to find a lift $\widetilde{b}_0 \in \mathcal{S}_2(\phi, \widetilde{r})$ of b. Using the given upper bound C on $|\mathsf{offset}(\theta, r)|$, choose $q \geq 0$ so that $\widetilde{\ell}'_{B'} \prec \theta(\Phi^q \widetilde{b}_0)$. Let $\widetilde{b} = \Phi^q \widetilde{b}_0$. From C and q we can compute an upper bound I' for the index of $\theta(\widetilde{b})$. By Lemma 15.31, we can list all visible $\widetilde{b}'$ with index at most I' and so have finitely many candidates for $\theta(\widetilde{b})$. Applying Lemma 16.9 to b and each of these candidates gives us the desired upper bound for $m_b(\theta)$. □

16.2 Stabilizing a ray

Suppose that E_i is the unique edge of height $i > 0$ and that $\sigma \subset G$ is a path with height i whose endpoints, if any, are not contained in the interior of E_i. Recall from Definition 4.1.3 and Lemma 4.1.4 of [Bestvina et al. 2000] that σ has a unique splitting, called the *highest-edge splitting of* σ, whose splitting vertices are the initial endpoints of each occurrence of E_i in σ and the terminal vertices of each occurrence of $\overline{E}_i$ in σ. In particular, each term in the splitting has the form $E_i \gamma \overline{E}_i$, $E_i \gamma$, $\gamma \overline{E}_i$ or γ for some $\gamma \subset G_{i-1}$.

The following lemma is used in the proof of Lemma 16.8. We make implicit use of [Feighn and Handel 2011, Lemma 4.6] which states that if $f : G \to G$ is completely split and a path $\sigma \subset G$ is completely split then $f^k_\#(\sigma)$ is completely split for all $k \geq 0$.

Lemma 16.10 *Suppose that $f : G \to G$ is a* CT *representing ϕ, that the edge E corresponds to some $r \in \mathcal{R}(\phi)$, that ξ is a finite subpath with endpoints at vertices and that $R = [\xi R_E]$. Equivalently, $R = \tau R_1$ for some finite path τ with endpoints at vertices and some subray R_1 of R_E. Then there exists a computable $k \geq 0$ such that $f^k_\#(R)$ is completely split.*

Proof The proof is by induction on the height h of R, with the base case being vacuous because the lowest stratum in the filtration is a fixed loop.

We are free to replace R by an iterate $f^l_\#(R)$ whenever it is convenient. We also have a less obvious replacement move.

(1) If there is a splitting $R = \nu \cdot R'$ where ν has endpoints at vertices then we may replace R by R'.

This follows from:

- [Feighn and Handel 2011, Lemma 4.25] For any finite path ν with endpoints at vertices, $f^k_\#(\nu)$ is completely split for all sufficiently large k.
- [Feighn and Handel 2011, Lemma 4.11] If a path σ has a decomposition $\sigma = \sigma_1\sigma_2$ with σ_1 and σ_2 completely split and the turn $(\overline{\sigma}_1, \sigma_2)$ legal then $\sigma = \sigma_1\sigma_2$ is a complete splitting.
- One can check if a given finite path with endpoints at vertices has a complete splitting (because there are only finitely many candidate decompositions).

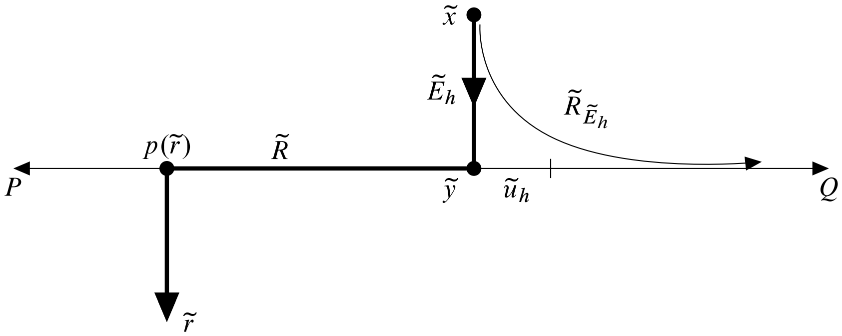


Figure 8

Let h_1 be the height of R_1. Each splitting vertex v for the highest-edge splitting of R_1 is also a splitting vertex for the complete splitting of R_E and so determines a splitting of R_1 into a finite initial subpath followed by a completely split terminal ray γ. If $h = h_1$, then v determines a splitting of R into a finite initial subpath followed by γ. In this case, an application of (1) completes the proof.

We may therefore assume that $h_1 < h$ and so the highest-edge splitting of R is finite. Applying (1), we may assume that the highest-edge splitting of R has just one term. Thus $R = E_h \mu R_1$, where E_h is the unique edge with height h and μ has height less than h. Let $h_2 < h$ be the height of $R_2 = \mu R_1$. (At various stages of the proof, we will let R_2 be the ray obtained from R by removing its initial edge. The exact edge description of R_2 will vary with the context.)

Let u_h be the path satisfying $f(E_h) = E_h \cdot u_h$. If the height of u_h is $> h_2$ then $f_\#(R) = E_h \cdot [u_h f_\#(R_2)]$ is a splitting so we may replace R by $[u_h f_\#(R_2)]$, which has height less than h. In this case the induction hypothesis completes the proof. If the height of u_h is $< h_2$ and $R_2 = \sigma_1 \cdot \sigma_2 \cdot \ldots$ is the highest-edge splitting of R_2, then $R = [E_h u_h \sigma_1] \cdot \sigma_2 \cdot \ldots$ is a splitting and the same argument completes the proof. We are now reduced to the case that the height of u_h is h_2, and we make this assumption for the rest of the proof.

We claim that there exists $k \geq 0$ so that $E_h u_h f_\#(u_h)$ is an initial segment of $f_\#^k(R)$. (Note that for any given k, one can check if $E_h u_h f_\#(u_h)$ is an initial segment of $f_\#^k(R)$ and so k with this property can be computed once one knows that it exists.) Choose a lift $\widetilde{E}_h \subset \widetilde{G}$ of E_h, let Γ be the component of the full preimage of G_{h_2} that contains the terminal $\widetilde{y}$ endpoint of $\widetilde{E}_h$ and let $\widetilde{f}\colon \widetilde{G} \to \widetilde{G}$ be the lift of f that fixes the initial endpoint $\widetilde{x}$ of $\widetilde{E}_h$. Then Γ is $\widetilde{f}$-invariant and the lift of R whose first edge is $\widetilde{E}_h$ decomposes as $\widetilde{R} = \widetilde{E}_h \widetilde{R}_2$, where $\widetilde{R}_2 \subset \Gamma$ is a lift of R_2. Let $\widetilde{u}_h$ be the lift of u_h with initial endpoint $\widetilde{y}$. Then $\widetilde{f}(\widetilde{E}_h) = \widetilde{E}_h \cdot \widetilde{u}_h$ and $\widetilde{R}_{\widetilde{E}_h} \setminus \widetilde{E}_h = \widetilde{u}_h \cdot f_\#(\widetilde{u}_h) \cdot f_\#^2(\widetilde{u}_h) \cdot \ldots$ is a ray of height h_2 that converges to an attracting fixed point $Q \in \partial\Gamma$ for the action of $\widetilde{f}$ on $\partial\Gamma$. By Lemma 2.8(ii) of [Bestvina et al. 2004] there is another fixed point $P \neq Q \in \partial\Gamma$ for the action of $\partial\widetilde{f}$. The line $\overrightarrow{PQ} \subset \Gamma$ from P to Q is $\widetilde{f}_\#$-invariant and has height h_2. Let $\mathcal{V}$ be the set of highest-edge splitting vertices of $\overrightarrow{PQ}$ with the order induced by the orientation on $\overrightarrow{PQ}$. Then $\widetilde{f}_\#$ preserves the highest-edge splitting of $\overrightarrow{PQ}$ and so $\widetilde{f}$ induces an order-preserving bijection of $\mathcal{V}$. Our choice of Q guarantees that $\tilde{f}$ moves points in $\mathcal{V}$ away from P and towards Q. Since $\tilde{f}$ induces an order-preserving injection of the set $\mathcal{V}'$ of highest-edge splitting vertices of $\tilde{R}_{\tilde{E}_h} \setminus \tilde{E}_h$ into itself, it follows that $\mathcal{V}' \subset \mathcal{V}$. To see this, note that for each $\tilde{v}' \in \mathcal{V}'$ and all sufficiently large m, $\tilde{f}^m(\tilde{v}')$ is a highest-edge splitting vertex for the common terminal ray of $\overrightarrow{PQ}$ and $R_{\tilde{E}_h} \setminus \tilde{E}_h$ and so $\tilde{f}^m(\tilde{v}') \in \mathcal{V}$. Since the restriction of $\tilde{f}^m$ to the vertex set of Γ and the restriction of $\tilde{f}^m$ to $\mathcal{V}$ are bijections, $\tilde{v}' \in \mathcal{V}$.

Since $\tilde{r}$ is an attractor for $\Phi_{\tilde{r}}$, we get $\tilde{r} \neq P$. If $\tilde{r} = Q$ then the lemma is obvious so we may assume that the nearest-point projection $p(\tilde{r})$ of $\tilde{r}$ to $\overrightarrow{PQ}$ is well-defined. The line $\overrightarrow{\tilde{r}Q}$ intersects $\overrightarrow{PQ}$ in the ray $\overrightarrow{p(\tilde{r})Q}$. The set of highest-edge splitting vertices of $\overrightarrow{p(\tilde{r})Q}$ equals the intersection of the set of highest-edge splitting vertices of $\overrightarrow{PQ}$ and the set of highest-edge splitting vertices of $\overrightarrow{\tilde{r}Q}$. It follows that the set of highest-edge splitting vertices of $\overrightarrow{p(\tilde{f}_\#(\tilde{r}))Q}$ is the $\tilde{f}_\#$-image of the set of highest-edge splitting vertices of $\overrightarrow{p(\tilde{r})Q}$. Thus $p(f_\#^k(\tilde{r})) \to Q$ and, after replacing R by some $\tilde{f}_\#^k(R)$, we may assume that $p(\tilde{r})$ is contained in $f_\#^2(\tilde{u}_h) \cdot f_\#^3(\tilde{u}_h) \cdot \ldots$. This completes the proof of the claim.

We now fix k satisfying the conclusions of the above claim and replace R by $f_\#^k(R)$. Thus $R = E_h u_h f_\#(u_h) \cdots$ and we let $R_2 = u_h f_\#(u_h) \cdots$ be the terminal ray of R obtained by removing its initial edge. We will prove that the decomposition of R determined by the highest-edge splitting vertices of R_2 is a splitting of R. The proof then concludes as in previous cases.

We continue with the notation established in the proof of the claim. Choose $\tilde{v} \in \mathcal{V} \cap \tilde{u}_h$ and decompose $\tilde{R}$ as $\tilde{R} = \tilde{\alpha}\tilde{\beta}\tilde{\gamma}$, where

$$\tilde{\alpha} = \overrightarrow{\tilde{x}\tilde{v}}, \quad \tilde{\beta} = \overrightarrow{\tilde{v}\tilde{f}(\tilde{v})} \quad \text{and} \quad \tilde{\gamma} = \overrightarrow{\tilde{f}(\tilde{v})\tilde{r}}.$$

Since $\tilde{\alpha}\tilde{\beta}$ is a subpath of $\tilde{E}_h \cdot \tilde{u}_h \cdot f_\#(\tilde{u}_h) \cdot f_\#^2(\tilde{u}_h) \cdot \ldots$, no edges of height h_2 are canceled when $\tilde{f}(\tilde{\alpha}\tilde{\beta})$ is tightened to $\tilde{f}_\#(\tilde{\alpha}\tilde{\beta})$. Similarly, no edges of height h_2 are canceled when $\tilde{f}(\tilde{\beta}\tilde{\gamma})$ is tightened to $\tilde{f}_\#(\tilde{\beta}\tilde{\gamma})$ because $\tilde{\beta}\tilde{\gamma}$ is a concatenation of terms in the highest-edge splitting of $\tilde{R}_2$. Since $\tilde{\beta}$ contains at least one edge of height h_2, it follows that no edges of height h_2 are canceled when $\tilde{f}(\tilde{R}) = \tilde{f}(\tilde{\alpha}\tilde{\beta}\tilde{\gamma})$ is tightened to $\tilde{f}_\#(R)$. This proves that the highest-edge splitting of $\tilde{R}_2$ is a splitting of $\tilde{R}$, as desired. □

16.3 Proof of Lemma 16.8

Recall from Notation 15.18 and Lemma 16.1 that $\mathcal{T}_{\phi,\tilde{r}}$ is the set of topmost elements of $\Omega_{\mathsf{NP}}(\phi, \tilde{r})$ and that $r' = \theta(r)$ and $r_1' = \theta(r_1)$ are independent of $\theta \in \mathcal{X}$ that conjugates ϕ to ψ.

Suppose that $\tilde{r}_1$ and $\tilde{r}_1'$ are lifts of r_1 and r_1', respectively, and that Θ is the lift of θ satisfying $\Theta(\tilde{r}_1) = \tilde{r}_1'$. If $\Theta(\tilde{L}) = \tilde{L}'$, where $\tilde{L} \in \mathcal{T}_{\phi,\tilde{r}_1}$ has index s and $\tilde{L}' \in \mathcal{T}_{\psi,\tilde{r}_1'}$ has index s', then $\mathsf{offset}(\theta, r_1) = s' - s$. We will not be able to find $\tilde{L}$ and $\tilde{L}'$ whose indices we know exactly but we will be able to find $\tilde{L}$ and $\tilde{L}'$ whose indices we know up to a uniform bound, and this is sufficient.

Before beginning the formal proof, we introduce a way to find distinguished elements of $\mathcal{T}_{\phi,\tilde r_1}$.

Notation 16.11 Suppose $r_1 <_c r$ (Notation 6.1) and that $\tilde r_1$ and $\tilde r$ are lifts such that $\mathcal{T}_{\phi,\tilde r_1} \cap \Omega_{\mathsf{NP}}(\phi,\tilde r) \neq \varnothing$. The $(\tilde r,\tilde r_1)$*-extreme* line is the element of $\mathcal{T}_{\phi,\tilde r_1} \cap \Omega_{\mathsf{NP}}(\phi,\tilde r)$ that is maximal in the order on $\mathcal{T}_{\phi,\tilde r_1}$.

The next lemma states that extreme lines behave well with respect to conjugation.

Lemma 16.12 *Suppose that θ conjugates ϕ to ψ, that $\Theta \in \theta$, that $\tilde r$ and $\tilde r_1$ are lifts of $r >_c r_1$ and that $\tilde L_2 \in \mathcal{T}_{\phi,\tilde r_1}$ is $(\tilde r,\tilde r_1)$-extreme. Then $\Theta(\tilde L_2)$ is $(\Theta(\tilde r),\Theta(\tilde r_1))$-extreme.*

Proof This follows from Lemmas 15.21 and 15.3, which imply that Θ maps $\mathcal{T}_{\phi,\tilde r_1}$ to $\mathcal{T}_{\psi,\Theta(\tilde r_1)}$ preserving order, and maps $\Omega_{\mathsf{NP}}(\phi,\tilde r)$ to $\Omega_{\mathsf{NP}}(\psi,\Theta(\tilde r))$. □

Proof of Lemma 16.8 If C is an upper bound for $|\mathsf{offset}(\theta,r)|$, it suffices to find, for each $|c| \le C$, an upper bound $C_{1,c}$ for $|\mathsf{offset}(\theta,r_1)|$ assuming that $\mathsf{offset}(\theta,r) = c$. The desired upper bound C_1 for $|\mathsf{offset}(\theta,r_1)|$ is then $\max\{C_{1,c}\}$. Going forward we may therefore assume that we know $|\mathsf{offset}(\theta,r)|$ exactly.

There is no loss of generality in assuming $r_1 <_c r$. Let E and E_1 be the elements of $\mathcal{E}_f$ corresponding to r and r_1, respectively. We will assume that E_1 occurs in R_E; the remaining case, in which $\bar E_1$ but not E_1 itself occurs in R_E, is similar and is left to the reader. Recall from Notation 15.18 that the visible elements of $\mathcal{T}_{\phi,\tilde r}$ are enumerated $\tilde L_1,\tilde L_2,\dots$. For $j \ge 0$, define $q_j \ge 0$ by $\tilde L_j = \tilde \ell_{q_j}$ and so $\tilde L_j = (\tilde R^-_{q_j})^{-1}\tilde\rho_{q_j}\tilde R^+_{q_j+1}$.

The first step of the proof is to show that:

(a) There is a computable $J > 0$ so that if $j \ge J$ and if $\tilde r_{1,j} := \partial_+\tilde L_j$ is a lift of r_1 (equivalently, $\sigma_{q_j+1} = E_1$ and $R^+_{q_j+1} = R_{E_1}$), then the line S_j connecting $\tilde r = \partial_+\tilde R_{\tilde E}$ to $\tilde r_{1,j}$ is completely split. See Figure 9.

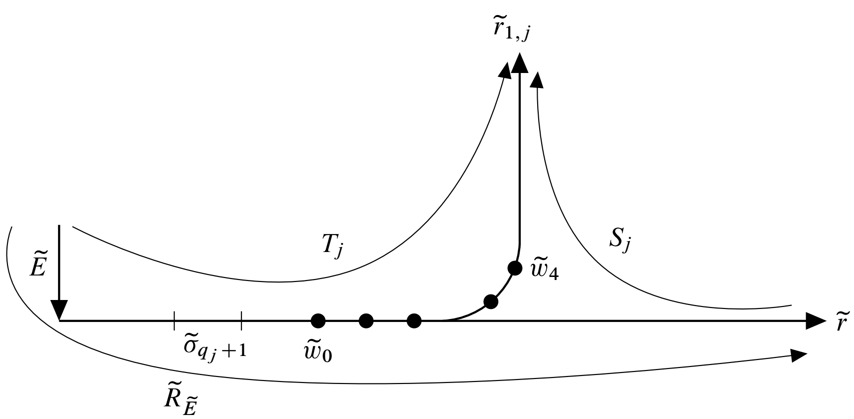


Figure 9

Lemma 15.20 implies that $\tilde{f}_\#^k(\tilde{L}_j) = \tilde{L}_{j+k\tau(\phi,r)}$ and hence $\tilde{f}_\#^k(S_j) = S_{j+k\tau(\phi,r)}$. It therefore suffices to show that for each $0 \le j \le \tau(\phi, r)$, there is a computable $K \ge 0$ so that $\tilde{f}_\#^K(S_j)$ (and hence $\tilde{f}_\#^k(S_j)$ for all $k \ge K$) is completely split.

The line S_j decomposes as a concatenation of (the inverse of) a ray in $\tilde{R}_{\tilde{E}} \setminus \tilde{E}$ and a proper subray of a lift of R_{E_1}. The height of the former is at least that of E_1 and the height of the latter is at most that of E_1. Moreover, $\tilde{R}_{\tilde{E}_1} \setminus \tilde{E}_1$ has height less than that of E_1. It follows that $\tilde{R}_{\tilde{E}} \setminus \tilde{E}$ and S_j have the same height and that each splitting vertex $\tilde{v}$ for the highest-edge splitting of S_j is contained in $\tilde{R}_{\tilde{E}} \setminus \tilde{E}$ and is a splitting vertex for $\tilde{R}_{\tilde{E}}$. Splitting S_j at one such $\tilde{v}$ we write $S_j = \tilde{A}_j^{-1} \cdot \tilde{B}_j$, where A_j is a concatenation of terms in the complete splitting of R_E, and B_j has a subray in common with R_{E_1}. For all $k \ge 0$, $S_{j+k\tau(\phi,r)} = \tilde{f}_\#^k(S_j) = \tilde{f}_\#^k(\tilde{A}_j^{-1}) \cdot \tilde{f}_\#^k(\tilde{B}_j)$. By Lemma 16.10, we can find K so that $f_\#^K(B_j)$ is completely split. It follows — see the second bullet point in the proof of Lemma 16.10 — that $\tilde{f}_\#^K(S_j)$ is completely split. This completes the first step.

Let $\mathcal{T}_{\psi,\tilde{r}'} = \{\tilde{L}'_1, \tilde{L}'_2, \ldots\}$ be the set of topmost elements of $\Omega_{\mathsf{NP}}(\psi, \tilde{r}')$. By definition, $\Theta(\tilde{L}_j) = \tilde{L}'_{j+\mathsf{offset}(\theta,r)}$. The following ψ and r' analogue of (a) is verified by the same arguments given for (a):

- (b) There is a computable $J' > 0$ such that if $j \ge J'$ and if $\tilde{r}'_{1,j} := \partial_+ \tilde{L}'_j$ projects to r'_1, then the line S'_j connecting $\tilde{r}'$ to $\tilde{r}'_{1,j}$ is completely split.

Note also that:

- (c) For all $j \ge 1$, the line T_j connecting the initial vertex of $\tilde{R}_{\tilde{E}}$ to $\tilde{r}_{1,j}$ is completely split, and similarly for the line T'_j connecting the initial vertex of $\tilde{R}'_{\tilde{E}'}$ to $\partial_+ \tilde{L}'_j$.

For $j \ge 0$, let $\mathcal{V}_j$ be the set of highest-edge splitting vertices of $\tilde{R}^+_{q_j+1} \setminus \tilde{\sigma}_{q_j+1}$ (which is a terminal ray of $\tilde{L}_j$) and $\mathcal{V}'_j$ be the set of highest-edge splitting vertices of $\tilde{R}'^+_{q'_j+1} \setminus \tilde{\sigma}'_{q'_j+1}$. The second step of the proof is to choose an index j so that the following four properties are satisfied:

- (i) S_j is completely split.
- (ii) There exist $\tilde{w} \in \mathcal{V}_j$ such that $\tilde{w}, \tilde{f}(\tilde{w}), \tilde{f}^2(\tilde{w}) \in \tilde{R}^+_{q_j+1} \cap \tilde{R}_{\tilde{E}}$.
- (iii) Letting $j' = j + \mathsf{offset}(\theta, r)$, the line $S'_{j'}$ is completely split.
- (iv) There exist $\tilde{w}' \in \mathcal{V}'_{j'}$ such that $\tilde{w}', \tilde{g}(\tilde{w}'), \tilde{g}^2(\tilde{w}') \in \tilde{R}'^+_{q'_{j'}+1} \cap \tilde{R}'_{\tilde{E}'}$.

Items (i) and (iii) hold for all $j \ge \max\{J, J' - \mathsf{offset}(\theta, r)\}$. For (ii), write $j = a_j + c_j\tau(\phi, r)$, where $0 \le a_j < \tau(\phi, r)$. Then $\tilde{L}_j = \tilde{f}_\#^{c_j}(\tilde{L}_{a_j})$, and $\tilde{L}_j \cap \tilde{R}_{\tilde{E}}$ contains an initial segment of $\tilde{R}^+_{q_j+1}$ whose length goes to infinity with j. If c_j is sufficiently large then (ii) is satisfied. Item (iv) is established in the same way, completing the second step.

We have $\Theta(\tilde{r}) = \tilde{r}'$ and $\Theta(\tilde{L}_j) = \tilde{L}'_{j'}$. The latter implies that $\Theta(\tilde{r}_{1,j}) = \tilde{r}'_{1,j'}$. Lemma 16.12 implies that Θ maps the $(\tilde{r}, \tilde{r}_{1,j})$-extreme line to the $(\tilde{r}', \tilde{r}'_{1,j'})$-extreme line. Let s_j be the index of the $(\tilde{r}, \tilde{r}_{1,j})$-extreme line (as an element of $\mathcal{T}_{\tilde{r}_{1,j}}$) and let $s'_{j'}$ be the index of the $(\tilde{r}', \tilde{r}'_{1,j'})$-extreme line (as an element of $\mathcal{T}_{\tilde{r}'_{1,j'}}$). Then $\mathsf{offset}(\theta, r_1) = s'_{j'} - s_j$. We will complete the proof by finding $a_j \le s_j \le b_j$ such that

$b_j - a_j \le 3\tau(\phi, r_1)$, and $a'_{j'} \le s'_{j'} \le b'_{j'}$ such that $b'_{j'} - a'_{j'} \le 3\tau(\psi, r'_1) = 3\tau(\phi, r_1)$. These allow us to compute $\mathsf{offset}(\theta, r_1)$ with an error at most $6\tau(\phi, r_1)$ and hence compute an upper bound for $\mathsf{offset}(\theta, r_1)$.

Let h_2 be the height of $R_{E_1} \setminus E_1$ (which is the same as the height of $\widetilde{R}^+_{q_j+1} \setminus \widetilde{\sigma}_{q_j+1}$) and let E_2 be the unique edge of height h_2. We claim that:

(d) Each $\widetilde{w} \in \mathcal{V}_j \cap \widetilde{R}_{\widetilde{E}}$ is a splitting vertex for the complete splittings of T_j and $\widetilde{R}_{\widetilde{E}}$.

It suffices to show that $\widetilde{w}$ is not contained in the interior of a term $\widetilde{\mu}$ in one of those splittings. Such a $\widetilde{\mu}$ would be an indivisible Nielsen path or exceptional path with height $\ge h_2$ and whose first edge is contained in $\widetilde{R}^+_{q_j+1} \setminus \widetilde{\sigma}_{q_j+1}$ (because $\widetilde{\sigma}_{q_j+1}$ is a term in both splittings) and so has height at most h_2. Thus E_2 would be a linear edge with twist path w_2 and μ would have one of the following forms: $E_2 w_2^p \bar{E}_2$, $E_2 w_2^p \bar{E}_3$ or $E_3 w_2^p \bar{E}_2$, where $p \ne 0$ and where $E_3 \ne E_2$ is a linear edge of height $< h_2$ with twist path w_2. In none of these cases does the interior of μ contain a vertex that is the initial endpoint of E_2 or the terminal endpoint of $\bar{E}_2$. This completes the proof of (d).

A similar analysis shows that:

(e) Each $\widetilde{w} \in \mathcal{V}_j$ that is disjoint from $\widetilde{R}_{\widetilde{E}}$ is a splitting vertex for the complete splittings of S_j and T_j.

Let $\widetilde{w}_0$ be the last element of $\mathcal{V}_j$ such that $\widetilde{w}_1 = \widetilde{f}(\widetilde{w}_0)$ and $\widetilde{w}_2 = \widetilde{f}^2(\widetilde{w}_0)$ are contained in $\widetilde{R}_{\widetilde{E}}$ (and hence contained in $\mathcal{V}_j \cap \widetilde{R}_{\widetilde{E}}$). Item (d) implies that the path $\widetilde{\alpha}$ connecting $\widetilde{w}_0$ to $\widetilde{w}_2$ inherits the same complete splitting from $\widetilde{R}_{\widetilde{E}}$ and from $\widetilde{R}^+_{q_j+1}$. Thus the lift $\widetilde{\sigma}_a$ of E_2 or $\bar{E}_2$ with endpoint $\widetilde{w}_1$ determines an element $\widetilde{L}^1$ of $\mathcal{T}_{\phi,\widetilde{r}_{1,j}} \cap \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$. (If $\widetilde{\sigma}_a$ is a lift of E_2 then $\widetilde{w}_1$ is the initial endpoint of $\widetilde{\sigma}_a$ and $\widetilde{L}^1 = \widetilde{\ell}_{a-1}$; if $\widetilde{\sigma}_a$ is a lift of $\bar{E}_2$ then $\widetilde{w}_1$ is the terminal endpoint of $\widetilde{\sigma}_a$ and then $\widetilde{L}^1 = \widetilde{\ell}_{a_1}$.) In particular, the index s_j of the $(\widetilde{r}, \widetilde{r}_{1,j})$-extreme line (as an element of $\mathcal{T}_{\phi,\widetilde{r}_{1,j}} \cap \Omega_{\mathsf{NP}}(\phi, \widetilde{r})$) is at least as big as that of $\widetilde{L}^1$.

Let $\widetilde{w}_3 = \widetilde{f}^3(\widetilde{w}_0)$ and $\widetilde{w}_4 = \widetilde{f}^4(\widetilde{w}_0)$, neither of which is contained in $\widetilde{R}_{\widetilde{E}}$. The lift of E_2 or $\bar{E}_2$ with endpoint $\widetilde{w}_4$ determines an element $\widetilde{L}^4$ in $\mathcal{T}_{\phi,\widetilde{r}_{1,j}}$. Item (e) and the hard splitting property of a complete splitting (Lemma 4.11 of [Feighn and Handel 2011]) implies that no point in the terminal ray of $\widetilde{R}^+_{q_j+1}$ that begins with $\widetilde{w}_4$ is ever identified, under iteration by $\widetilde{f}$, with a point in $\widetilde{R}_{\widetilde{E}}$. It follows that $\widetilde{L}^4$ is not an element of $\Omega_{\mathsf{NP}}(\phi, \widetilde{r})$ and so s_j is less than the index of $\widetilde{L}^4$.

Combining the inequalities established in the preceding two paragraphs we are able to compute s_j with an error of at most $3\tau(\phi, r_1)$. The parallel argument allows us to compute the index s'_j of the $(\widetilde{r}', \widetilde{r}'_{1,j'})$-extreme line (as an element of $\mathcal{T}_{\phi,\widetilde{r}_{1,j}}$) with an error of at most $3\tau(\psi, r'_1) = 3\tau(\phi, r_1)$. As noted above, this completes the proof. □

17 Proof of Proposition 16.4

Some of our arguments are by induction up through the elements $\mathcal{F}_k$ of the chain $\mathfrak{c}$. We write $\phi|\mathcal{F}_k = \psi|\mathcal{F}_k$ if $\phi|[F] = \psi|[F]$ for each component $[F]$ of $\mathcal{F}_k$. Similarly, we say $\theta|\mathcal{F}_k$ conjugates $\phi|\mathcal{F}_k$ to $\psi|\mathcal{F}_k$

if $\phi^\theta|\mathcal{F}_k = \psi|\mathcal{F}_k$. If G_s is the core filtration element corresponding to $\mathcal{F}_k$ and if C is a component of G_s with rank one, then $[C]$ is a component of $\mathcal{F}_0$ and we define $\Gamma(f|C) = C$. With this convention, $\Gamma(f|G_s)$ is the disjoint union $\bigsqcup \Gamma(f|C_i)$ as C_i varies over the components of G_s. (See Section 4.1.)

We show below that Proposition 16.4 is a consequence of the following lemma and proposition. The former addresses the restrictions to $\mathcal{F}_0$ and the latter provides the inductive step for the higher-order one-edge extensions.

Lemma 17.1 *Suppose that $\phi, \psi \in \mathsf{UPG}(F_n)$ share the special chain $\mathfrak{c}$ and satisfy $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$. Let $\mathcal{F}_0 = \mathcal{F}_0(\phi) = \mathcal{F}_0(\psi)$. Then*

1. *$\theta(L) = L$ for each $\theta \in \mathcal{X}$ and each $L \in \Omega(\phi)$ that is carried by $\mathcal{F}_0$, and*
2. *if there exists $\theta_0 \in \mathcal{X}$ such that $\phi^{\theta_0}|\mathcal{F}_0 = \psi|\mathcal{F}_0$, then $\phi|\mathcal{F}_0 = \psi|\mathcal{F}_0$ and $\phi^\theta|\mathcal{F}_0 = \psi|\mathcal{F}_0$ for all $\theta \in \mathcal{X}$.*

Proposition 17.2 *Suppose that $\phi, \psi \in \mathsf{UPG}(F_n)$ share the special chain $\mathfrak{c}$ and satisfy $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$, and that the special one-edge extension $\mathfrak{e} = (\mathcal{F}^- \sqsubset \mathcal{F}^+)$ in $\mathfrak{c}$ satisfies*

1. *$\phi|\mathcal{F}^- = \psi|\mathcal{F}^-$,*
2. *$\{L \in \Omega(\phi) \mid L \subset \mathcal{F}^-\} = \{L' \in \Omega(\psi) \mid L' \subset \mathcal{F}^-\}$.*

Then there is an algorithm to decide if there exists $\theta \in \mathsf{Ker}(\bar{Q}) < \mathcal{X}$ such that the following are satisfied:

3. *$\phi^\theta|\mathcal{F}^+ = \psi|\mathcal{F}^+$,*
4. *$\theta(\{L \in \Omega(\phi) \mid L \subset \mathcal{F}^+\}) = \{L' \in \Omega(\psi) \mid L' \subset \mathcal{F}^+\}$.*

Moreover, *if such an element θ exists*, *then one is produced.*

Before proving Lemma 17.1 and Proposition 17.2, we use them to prove Proposition 16.4.

Proof of Proposition 16.4, assuming Lemma 17.1 and Proposition 17.2 If $\phi|\mathcal{F}_0 \neq \psi|\mathcal{F}_0$, then no element of $\mathcal{X}$ conjugates ϕ to ψ by Lemma 17.1(2) so we return NO and STOP. Otherwise, $\phi^\theta|\mathcal{F}_0 = \psi|\mathcal{F}_0$ for all $\theta \in \mathcal{X}(\phi)$, and we define $\theta_0 =$identity and $\psi_0 = \psi$.

Suppose $\mathfrak{c} = (\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_t)$. Apply Proposition 17.2 with $(\phi, \psi_0, \mathcal{F}_0, \mathcal{F}_1)$ in place of $(\phi, \psi, \mathcal{F}^-, \mathcal{F}^+)$. If the 17.2-algorithm returns NO then there is no θ as in the conclusion of Proposition 16.4 because any such θ would satisfy items (3) and (4) of Proposition 17.2; we return NO and STOP. Otherwise, Proposition 17.2 gives us an element $\theta_1 \in \mathsf{Ker}(\bar{Q})$. Letting $\psi_1 = \psi_0^{(\theta_1^{-1})}$ we have that $\phi|\mathcal{F}_1 = \psi_1|\mathcal{F}_1$ and $\{L \in \Omega(\phi) \mid L \subset \mathcal{F}_1\} = \{L' \in \Omega(\psi_1) \mid L' \subset \mathcal{F}_1\}$. From $\theta_1 \in \mathcal{X}$ and Lemma 14.6, it follows that $\mathsf{I}(\phi) = \mathsf{I}(\psi_1)$.

Apply Proposition 17.2 with $(\phi, \psi_1, \mathcal{F}_1, \mathcal{F}_2)$ in place of $(\phi, \psi, \mathcal{F}^-, \mathcal{F}^+)$. Suppose that the 17.2-algorithm returns NO. Then there are no elements of $\mathsf{Ker}(\bar{Q})$ that conjugate $\phi|\mathcal{F}_2$ to $\psi_1|\mathcal{F}_2$, and so also no elements of $\mathsf{Ker}(\bar{Q})$ that conjugate ϕ to ψ_1. It follows also then that there are no elements θ of $\mathsf{Ker}(\bar{Q})$ that conjugate ϕ to ψ. Indeed for such a θ, $\theta_1^{-1}\theta$ would conjugate ϕ to ψ_1. We therefore return NO and STOP. Otherwise, Proposition 17.2 gives us an element $\theta_2 \in \mathsf{Ker}(\bar{Q})$. Letting $\psi_2 = \psi_1^{(\theta_2)^{-1}}$ we have that $\phi|\mathcal{F}_2 = \psi_2|\mathcal{F}_2$ and $\{L \in \Omega(\phi) : L \subset \mathcal{F}_2\} = \{L' \in \Omega(\psi_2) : L' \subset \mathcal{F}_2\}$. As in the previous case, $\mathsf{I}(\phi) = \mathsf{I}(\psi_2)$. Repeat this until either some application of Proposition 17.2 returns NO or until we reach $\psi_t = \psi^{(\theta_1 \dots \theta_t)^{-1}}$ satisfying $\phi = \phi|\mathcal{F}_t = \psi_t|\mathcal{F}_t = \psi_t$. In the former case there is no θ as in the conclusion of Proposition 16.4 and we return NO and STOP. In the latter case $\theta = \theta_1 \dots \theta_t$ conjugates ϕ to ψ and is an element of $\mathsf{Ker}(\bar{Q})$; we return YES and θ and then STOP. □

Proof of Lemma 17.1 If $L \in \Omega(\phi)$ is carried by $\mathcal{F}_0$, then the ends of L are periodic. If L is periodic then $\widetilde{L} = a^\infty$ for some $[a] \in \mathcal{A}(\phi)$; see Corollary 5.17(1). By definition of $\mathcal{X}$, $\theta([a]) = [a]$ and so $\theta(L) = L$. Otherwise, $L \in \Omega_{\mathsf{NP}}(\phi)$ has type P-P, in which case $\mathsf{H}(L)$ determines L; see Section 13. Again by definition of $\mathcal{X}$, $\theta(\mathsf{H}(L)) = \mathsf{H}(L)$ and so $\theta(L) = L$. This verifies (1).

It suffices to show that if a free factor F represents a component of $\mathcal{F}_0$ then either $\phi^\theta|F = \psi|F$ for all $\theta \in \mathcal{X}(\phi)$ (and in particular for $\theta =$ identity) or $\phi^\theta|F = \psi|F$ is satisfied by no element of $\mathcal{X}(\phi)$.

Let $\phi_F = \phi|F$ and $\psi_F = \psi|F$. If F has rank one, then ϕ_F and ψ_F are both the identity because ϕ and ψ are rotationless. We may therefore assume that F has rank at least two. Since $\mathcal{R}(\phi_F) = \varnothing$, Lemma 3.9 implies that $\mathsf{Fix}_{\mathsf{N}}(\Phi_F) = \partial\,\mathsf{Fix}(\Phi_F)$ for each $\Phi_F \in \mathcal{P}(\phi_F)$. Also, $\mathsf{Fix}_{\mathsf{N}}(\Phi_F)$ contains at least three points, so $\mathsf{Fix}(\Phi_F)$ has rank at least two and $\mathsf{Fix}(\Phi_F) \neq \mathsf{Fix}(\Phi'_F)$ for $\Phi_F \neq \Phi'_F \in \mathcal{P}(\phi_F)$ by Lemma 4.4.

There is a unique $\Phi \in \mathcal{P}(\phi)$ such that $\Phi_F = \Phi|F$. From $\mathsf{I}(\phi) = \mathsf{I}(\psi)$, it follows that there exists $\Psi \in \mathcal{P}(\psi)$ such that $\mathsf{Fix}(\Phi) = \mathsf{Fix}(\Psi)$. Since $\theta \in \mathcal{X}$, there exists Θ representing θ such that $\mathsf{Fix}(\Phi)$ is Θ-invariant. It follows that $\Theta(F) \cap F$ is nontrivial and hence that $\Theta(F) = F$ (because F is a free factor and θ preserves $[F]$). Letting $\Psi_F = \Psi|F$ and $\Theta_F = \Theta|F$, we have that $\mathsf{Fix}_{\mathsf{N}}(\Phi_F) = \partial\,\mathsf{Fix}(\Phi_F) = \partial\,\mathsf{Fix}(\Psi_F) = \mathsf{Fix}_{\mathsf{N}}(\Psi_F)$ is Θ_F-invariant.Lemma 4.1 implies that the eigengraphs for ϕ_F and for ψ_F carry the same lines and that θ preserves this set of lines. Thus ϕ_F, ψ_F and θ_F satisfy the hypotheses of Lemma 4.21.

If $a \in F$ is fixed by distinct $\Phi_F, \Phi'_F \in \mathcal{P}(\phi_F)$ then $[\Phi_F, a]$ is an element of $\mathsf{SA}(\phi_F)$ and $[\Phi, a]$ is an element of $\mathsf{SA}(\phi)$. Lemma 4.21 implies that

$$[\Phi_F, a] \mapsto [\Psi_F, \Theta_F(a)]$$

defines a bijection $\mathcal{B}_{\mathsf{SA},F} : \mathsf{SA}(\phi_F) \to \mathsf{SA}(\psi_F)$ that is independent of the choice of Θ_F representing θ_F and preserving $\mathsf{Fix}_{\mathsf{N}}(\Phi_F)$. Since $\theta \in \mathcal{X}$, by Definition 14.1(6) we have $[\mathsf{Fix}(\Phi), a] = \theta([\mathsf{Fix}(\Phi), a]) = [\Theta(\mathsf{Fix}(\Phi)), \Theta(a)] = [\mathsf{Fix}(\Phi), \Theta(a)]$. Equivalently, there exists $c \in F_n$ such that $i_c(\mathsf{Fix}(\Phi)) = \mathsf{Fix}(\Phi)$ and $i_c\Theta(a) = a$. Thus $c \in \mathsf{Fix}(\Phi)$ and after replacing Θ by $i_c\Theta$ we may assume that $\Theta(a) = a$ and hence that $\Theta_F(a) = a$. We conclude that $\mathcal{B}_{\mathsf{SA},F}$ is independent of θ.

Check by inspection if $\mathcal{B}_{\mathsf{SA},F}$ preserves twist coordinates. If it does then Lemma 4.21 implies that each $\theta \in \mathcal{X}(\phi)$ conjugates ϕ_F to ψ_F; if not, then no element of $\mathcal{X}(\phi)$ conjugates ϕ_F to ψ_F. □

The rest of the paper is dedicated to the proof of Proposition 17.2.

Set $\mathfrak{c} = (\mathcal{F}_0 \sqsubset \mathcal{F}_1 \sqsubset \cdots \sqsubset \mathcal{F}_t)$ and thus $\mathfrak{e} \in \mathfrak{c}$ has the form $\mathcal{F}^- \sqsubset \mathcal{F}^+$, where $\mathcal{F}^- = \mathcal{F}_{k-1}$ and $\mathcal{F}^+ = \mathcal{F}_k$ for some $1 \le k \le t$. (We will use these notations interchangeably depending on the context.)

Definition 17.3 For $\epsilon = \pm$, $\mathcal{X}^\epsilon$ is the set of $\theta \in \mathcal{X}$ such that

(a) $\theta \in \mathsf{Ker}(\bar{Q})$, and

(b) $\theta|\mathcal{F}^\epsilon$ conjugates $\phi|\mathcal{F}^\epsilon$ to $\psi|\mathcal{F}^\epsilon$.

By the next lemma, our goal is to produce an element of $\mathcal{X}^+$ or deduce that $\mathcal{X}^+$ is empty.

Lemma 17.4 *An element $\theta \in \mathsf{Ker}(\bar{Q})$ satisfies items (3) and (4) of Proposition 17.2 if and only if $\theta \in \mathcal{X}^+$.*

Proof Comparing the definitions, it suffices to show that each $\theta \in \mathcal{X}^+$ satisfies Proposition 17.2(4); namely, $\theta(\{L \in \Omega(\phi) \mid L \subset \mathcal{F}^+\}) = \{L' \in \Omega(\psi) \mid L' \subset \mathcal{F}^+\}$. By symmetry, it suffices to show that if $L \in \Omega(\phi)$ is carried by $\mathcal{F}^+$ then $\theta(L) \in \Omega(\psi)$ is carried by $\mathcal{F}^+$. Since $\mathcal{F}^+$ is θ-invariant, it suffices to show that $\theta(L) \in \Omega(\psi)$. If L is periodic then $\widetilde{L} = a^\infty$ for some $[a] \in \mathcal{A}(\phi)$ by Corollary 5.17(1). Since $\theta \in \mathcal{X}$, one has that $\theta([a]) = [a]$ and $\theta(L) = L \in \Omega(\psi)$. Otherwise $L \in \Omega_{\mathsf{NP}}(\phi)$ and, as $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$, there exists $L' \in \Omega_{\mathsf{NP}}(\psi)$ such that $\mathsf{H}(L) = \mathsf{H}(L')$. Since $\theta|\mathcal{F}^+$ conjugates $\phi|\mathcal{F}^+$ to $\psi|\mathcal{F}^+$, Corollary 5.17(4) and Lemmas 4.1 and 3.16 imply that $\theta(L)$ lifts into $\Gamma(g_{u'})$. Also, $\theta(L) \in \theta\big(\mathsf{H}(L)\big) = \mathsf{H}(L) = \mathsf{H}(L')$ because $\theta \in \mathcal{X}$. Lemma 13.9 implies that $\theta(L) = L' \in \Omega_{\mathsf{NP}}(\psi)$, and we are done. □

By [Feighn and Handel 2018, Theorem 7.4] we can choose CTs $f: G \to G$ and $g: G' \to G'$ representing ϕ and ψ, respectively, such that each $\mathcal{F}_i$ is realized by a core filtration element and such that the core filtration elements of G and G' realizing $\mathcal{F}^- = \mathcal{F}_{k-1}$ are identical as marked graphs and that after identifying them to a common subgraph G_s, the restrictions $f_s = f|G_s$ and $g_s = g|G'_s$ are equal. In particular,

(1) $\Gamma(f_s) = \Gamma(g_s)$.

Before describing f_s and g_s in more detail, we record some useful properties of $\mathcal{X}^-$. We define $\mathcal{R}(\phi|\mathcal{F}_-) = \cup\mathcal{R}(\phi|[F])$ as $[F]$ varies over the components of $\mathcal{F}_-$.

Lemma 17.5 *Each $\theta \in \mathcal{X}^-$ satisfies the following properties*:

(1) *$\theta|\mathcal{F}^-$ commutes with $\phi|\mathcal{F}^- = \psi|\mathcal{F}^-$.*

(2) *θ fixes each element of $\Omega(\phi)$ that is carried by $\mathcal{F}^-$.*

(3) *θ fixes each element of $\mathcal{R}(\phi|\mathcal{F}^-)$.*

Proof Property (1) follows from Definition 17.3(b) and the hypothesis that $\phi|\mathcal{F}^- = \psi|\mathcal{F}^-$. For (2), note that if $L \in \Omega(\phi)$ is carried by $\mathcal{F}^-$ then L lifts to $\Gamma(f_s)$ by Corollary 5.17(4) so (1) and Lemmas 4.1 and 3.16 imply that $\theta(L) \in \Gamma(g_s)$. Since $\theta(L) = L$ if L is periodic and otherwise $\theta(L) \in \theta(\mathsf{H}(L)) = \mathsf{H}(L)$, (2) follows from (1) and Lemma 13.9. By (1) and Lemma 16.1, $\theta(r)$ is independent of $\theta \in \mathcal{X}^-$. Item (3) therefore follows from the fact that $\mathcal{X}^-$ contains the identity. □

Suppose that $G_u \subset G$ and $G'_{u'} \subset G'$ are the core filtration elements realizing $\mathcal{F}^+ = \mathcal{F}_k$. Let $f_u = f|G_u$ and $g_{u'} = g|G'_{u'}$. Since $\mathcal{F}^- \sqsubset \mathcal{F}^+$ is a special one-edge extension, G_u is obtained from G_s by adding a single topological arc E which is either a single edge D or is the union $E = \bar{C}D$ of a pair of edges C and D with common initial endpoint not in G_s. (We have previously denoted edges in G by E and now we are using C and D instead and using E for a topological arc. This is more convenient for the current argument and should not cause confusion.) By Lemma 6.9, there are three possibilities. In each case, there is one component $\Gamma_*(f_u)$ of $\Gamma(f_u)$ that is not a component of $\Gamma(f_s)$.

- [HH] ($E = \bar{C}D$ consists of two higher-order edges) $\Gamma(f_u)$ is obtained from $\Gamma(f_s)$ by adding a new component $\Gamma_*(f_u)$ which is a line labeled $R_C^{-1} \cdot R_D$.
- [LH] ($E = \bar{C}D$ where C is linear and D is higher-order) $\Gamma(f_u)$ is obtained from $\Gamma(f_s)$ by adding a new component $\Gamma_*(f_u)$ which is a one-point union of a lollipop corresponding to C and a ray labeled R_D.
- [H] ($E = D$, a higher-order edge) $\Gamma(f_u)$ is a one-point union of $\Gamma(f_s)$ and a ray labeled R_D. $\Gamma_*(f_u)$ is the one-point union of a component $\Gamma_*(f_s)$ of $\Gamma(f_s)$ and a ray labeled R_D.

Similarly, we can orient the topological arc E' that is added to $G_s = G'_s$ to form $G'_{u'}$ so that $\Gamma(g_{u'})$ is obtained from $\Gamma(g_s)$ in one of these three ways.

By Lemmas 6.10 and 6.13, we may assume:

(2) The extensions $\Gamma(f_s) \subset \Gamma(f_u)$ and $\Gamma(g_s) \subset \Gamma(g_{u'})$ have the same type HH, LH or H.

For if not, then $\phi|\mathcal{F}^+$ and $\psi|\mathcal{F}^+$ are not conjugate by an element of $\mathcal{X}$ so we return `NO` and `STOP`.

Remark 17.6 A vertex v in G that is new in an HH extension, is not incident to any fixed or linear edge. It therefore follows from the construction of $\Gamma(f)$ given at the beginning of Section 4.1 that the component $\Gamma(f, v)$ of $\Gamma(f)$ corresponding to v is obtained from the disjoint union of eigenrays R_E, one for each $E \in \mathcal{E}(f)$ with initial vertex v, by identifying their initial vertices. Similarly, if v is new in an LH extension, then $\Gamma(f, v)$ is the one-point union of the lollipop associated to the unique linear edge with v as initial vertex and the eigenrays R_E associated to $E \in \mathcal{E}_f$ with v as initial vertex.

Lemma 17.7 *Suppose that $\mathfrak{e}$ has type* H *and that $\mathcal{X}^+ \neq \varnothing$. Then $\Gamma_*(f_s) = \Gamma_*(g_s)$.*

Proof Assume that $\theta \in \mathcal{X}^+$. Denote the set of lines that lift into a Stallings graph Γ by $\Lambda(\Gamma)$. Lemmas 3.16 and 4.1 imply that $\theta(\Lambda(\Gamma_*(f_u))) = \Lambda(\Gamma_*(f_{u'}))$. By construction,

$$\Lambda(\Gamma_*(f_s)) = \{L \in \Lambda(\Gamma_*(f_u)) \mid L \subset G_s\} \quad \text{and} \quad \Lambda(\Gamma_*(g_s)) = \{L \in \Lambda(\Gamma_*(g_u)) \mid L \subset G_s\}.$$

Thus $\theta(\Lambda(\Gamma_*(f_s))) = \Lambda(\Gamma_*(f_s))$.

The proof now divides into cases. If $\Gamma_*(f_s)$ contains a ray corresponding to some $r \in \mathcal{R}(\phi)$ then $\Lambda(\Gamma_*(f_s))$ contains a line that ends at r. Lemma 17.5(3) then implies that $\Lambda(\Gamma_*(g_s))$ contains a line that ends at r and hence that $\Gamma_*(f_s)$ contains a ray corresponding to $r \in \mathcal{R}(\phi)$. This proves that $\Gamma_*(f_s) = \Gamma_*(g_s)$.

We may now assume that $\Gamma_*(f_s)$ is compact. If $\Gamma_*(f_s)$ has rank at least two then $\pi_1(\Gamma_*(f_s))$ is a component of $\mathsf{Fix}(\phi)$ and is hence θ-invariant. In this case, $\pi_1(\Gamma_*(f_s)) = \pi_1(\Gamma_*(g_s))$. Lemma 4.4(1) implies that $\Gamma_*(f_s) = \Gamma_*(g_s)$. The final case is that $\Gamma_*(f_s)$ has rank one and so is a topological circle labeled by a component Y of G_0 consisting of a single edge e. In this case, $\Lambda(\Gamma_*(f_s)) = \{e^{\infty}, e^{-\infty}\}$, which is θ-invariant. It follows that $\Lambda(\Gamma_*(g_s)) = \{e^{\infty}, e^{-\infty}\}$ and hence that $\Gamma_*(f_s) = \Gamma_*(g_s)$. □

We may therefore assume that:

(3) In the case H, $\Gamma_*(f_s) = \Gamma_*(g_s)$.

We next apply the recognition theorem to give criteria for an element in $\mathcal{X}^-$ to be in $\mathcal{X}^+$.

Lemma 17.8 *The following are equivalent for each $\theta \in \mathcal{X}^-$:*

(1) *$\theta \in \mathcal{X}^+$; equivalently, $\theta|\mathcal{F}^+$ conjugates $\phi|\mathcal{F}^+$ to $\psi|\mathcal{F}^+$.*

(2) (a) *A line L lifts into $\Gamma(f_u)$ if and only if $\theta(L)$ lifts into $\Gamma(g_{u'})$.*

(b) *If $\mathcal{F}^- \sqsubset \mathcal{F}^+$ has type* LH *then the twist index for C with respect to f equals the twist index for C' with respect to g. Equivalently, if $f(C) = Cw^d$, then $g(C') = C'w^d$.*

Proof (1) implies (a) by Lemmas 4.1 and 3.16. We may therefore assume that (a) is satisfied and prove that (1) is equivalent to (b).

If $[F]$ is a component of $\mathcal{F}^+$ that is also a component of $\mathcal{F}^-$ then $\theta|F$ conjugates $\phi|F$ to $\psi|F$ because $\theta \in \mathcal{X}^-$. We may therefore restrict our attention to the unique component of $\mathcal{F}^+$ that is not also a component of $\mathcal{F}^-$. In other words, we may assume that G_u is connected and so may assume that $G_u = G$ and $\mathcal{F}^+ = \{[F_n]\}$.

By Lemma 4.21, there is a bijection $B\colon \mathsf{SA}(\phi) \to \mathsf{SA}(\psi)$ that preserves twist coordinates if and only if θ conjugates ϕ to ψ. By definition of $\mathcal{X}$, θ preserves each element of $\mathcal{A}_{\mathrm{or}}(\phi)$. We are therefore reduced to showing that (b) is satisfied if and only if the following is satisfied for each $[a] \in \mathcal{A}_{\mathrm{or}}(\phi)$:

$(*)$ The restricted bijection $B\colon \mathsf{SA}(\phi, [a]) \to \mathsf{SA}(\psi, [a])$ preserves twist coordinates.

Since $(*)$ is satisfied for $[a]$ if and only if it is satisfied for $[\overline{a}]$, we may assume that the twist path w for $[a]_u$ satisfies $[a]=[w]$. Extending Notation 4.7, we define

$$\mathcal{P}(\phi,a):=\{\Phi_{a,0},\dots,\Phi_{a,m-1}\}.$$

In particular, $\Phi_{a,0}$ is the base principal lift for a (with respect to f) and there is an order-preserving bijection between the set $\{E^1,\dots,E^{m-1}\}$ of linear edges with axis $[a]$ and $\{\Phi_{a,1},\dots,\Phi_{a,m-1}\}$. For $1\le j\le m-1$, there exist distinct twist indices $d_j\neq 0$ such that $f(E^j)=E^j w^{d_j}$. Define $d_0=0$. Lemmas 4.8 and 4.12 imply that

$$\mathsf{SA}(\phi,[a])=\{[\Phi_{a,0},a],\dots,[\Phi_{a,m-1},a]\}$$

and that the twist coefficient for the pair $([\Phi_{a,i},a],[\Phi_{a,j},a])$ is d_i-d_j.

We consider two cases. In the first, we assume that either

- $\mathcal{F}^-\sqsubset\mathcal{F}^+$ has type LH and $C\notin\{E^1,\dots,E^{m-1}\}$, or
- $\mathcal{F}^-\sqsubset\mathcal{F}^+$ does not have type LH,

and we prove that $(*)$ is satisfied.

In this case,

$$\mathcal{P}(\psi,a)=\{\Psi_{a,0},\dots,\Psi_{a,m-1}\}\quad\text{and}\quad\mathsf{SA}(\psi,[a])=\{[\Psi_{a,0},a],\dots,[\Psi_{a,m-1},a]\},$$

with the same sequence of linear edges $\{E^1,\dots,E^{m-1}\}$ and the same sequence of twist indices $\{d_0,\dots,d_{m-1}\}$. The bijection $B:\mathsf{SA}(\phi)\to\mathsf{SA}(\psi)$ induces a permutation π of $\{0,\dots,m-1\}$ satisfying $B([\Phi_{a,i},a])=[\Psi_{a,\pi(i)},a]$. We will show that π is the identity and hence that $B:\mathsf{SA}(\phi,[a])\to\mathsf{SA}(\psi,[a])$ preserves twist coordinates.

Choose an automorphism Θ representing θ and fixing a. By Lemma 4.21,

$$\Theta(\mathsf{Fix}_\mathsf{N}(\Phi_{a,i}))=\mathsf{Fix}_\mathsf{N}(\Psi_{a,\pi(i)}).$$

Let C_s be the component of G_s that contains w, and hence contains each E^i, and let $[F]$ be the corresponding component of $\mathcal{F}^-$; we may assume without loss of generality that $a\in F$. Applying Notation 4.7 to $\phi|F=\psi|F$ represented by the CT $f|C_s$, we see that

$$\begin{aligned}\mathcal{P}(\psi|F,a)=\ \mathcal{P}(\phi|F,a)\ &=\{\Phi_{a,0}|F,\dots,\Phi_{a,m-1}|F\},\\ \mathsf{SA}(\psi|F,[a])=\mathsf{SA}(\phi|F,[a])&=\{[\Phi_{a,0}|F,a],\dots,[\Phi_{a,m-1}|F,a]\},\end{aligned}$$

with the same sequence of linear edges $\{E^1,\dots,E^{m-1}\}$ and the same sequence of twist indices $\{d_0,\dots,d_{m-1}\}$. Since C_s is f-invariant and $[F]$ is θ-invariant, (a) implies that the set of lines that lift to $\Gamma(f_s)$ is θ-invariant. Applying Lemma 4.21 produces a permutation B_F of $\mathsf{SA}(\phi|F,[a])$ and an induced permutation π_F of $\{0,\dots,m-1\}$. Since $\theta|F$ commutes with $\phi|F$, B_F preserves twist-coordinates. Thus, $d_i - d_j = d_{\pi_F(i)} - d_{\pi_F(j)}$ for all i and j. The only possibility is that π_F is the identity and so

$$\mathsf{Fix}_{\mathsf{N}}(\Psi_{a,\pi(i)}) \cap \partial F = \Theta(\mathsf{Fix}_{\mathsf{N}}(\Phi_{a,i})) \cap \partial F = (\Theta|F)(\mathsf{Fix}_{\mathsf{N}}(\Phi_{a,i}|F)) = \mathsf{Fix}_{\mathsf{N}}(\Psi_{a,i}|F) = \mathsf{Fix}_{\mathsf{N}}(\Psi_{a,i}) \cap \partial F.$$

It follows that $\mathsf{Fix}_{\mathsf{N}}(\Psi_{a,\pi(i)}) \cap \mathsf{Fix}_{\mathsf{N}}(\Psi_{a,i})$ contains $\mathsf{Fix}_{\mathsf{N}}(\Psi_{a,i}) \cap \partial F$, which has cardinality at least three. Lemma 3.7 implies that $\pi(i) = i$, as desired. This completes the first case.

For the second case, we assume that:

- $\mathcal{F}^- \sqsubset \mathcal{F}^+$ has type LH and $C \in \{E^1,\dots,E^{m-1}\}$,

and prove that $(*)$ is equivalent to (b).

Assuming without loss of generality that $C = E^{m-1}$, the sequence of linear edges for ψ is given by $\{E^1,\dots,E^{m-2},C'\}$ with twist indices $\{d_0,\dots,d_{m-2},d'_{m-1}\}$. Thus (b) is the statement that $d_{m-1} = d'_{m-1}$ and we are reduced to showing that π is the identity.

If $m > 2$ then $\mathcal{P}(\phi|F,a) = \mathcal{P}(\psi|F,a)$ is indexed by $\{E^1,\dots,E^{m-2}\}$ and the above analysis applies to show that π restricts to the trivial permutation of $\{0,\dots,m-2\}$. It then follows that π must fix the one remaining element $m-1$ of $\{0,\dots,m-1\}$.

We are now reduced to the case that $m = 2$. In particular, $[a] \notin \mathcal{A}(\phi|F)$. By construction, the base lift $\Phi_{a,0}$ restricts to an element of $\mathcal{P}(\phi|F,a)$. It follows that $\Phi_{a,1}$ does not restrict to an element of $\mathcal{P}(\phi|F,a)$. The same holds for $\Psi_{a,0}$ and $\Psi_{a,1}$. Since conjugation by $\theta|F$ preserves $\mathcal{P}(\phi|F,a)$, it must be that $\Phi^{\Theta}_{a,0} = \Psi_{a,0}$ and $\Phi^{\Theta}_{a,1} = \Psi_{a,1}$. This completes the proof of the lemma. □

The next step in the algorithm is to check if the following condition is satisfied:

(4) If $\mathcal{F}^- \sqsubset \mathcal{F}^+$ has type LH then the twist index for C with respect to f equals the twist index for C' with respect to g.

If not, return `NO` and `STOP`. This is justified by Lemma 17.8.

Lemma 17.9 *It holds that $\theta(r) = r$ for all $\theta \in \mathcal{X}^-$ and $r \in \Delta\mathcal{R}(\phi) = \mathcal{R}(\phi|\mathcal{F}^+) \setminus \mathcal{R}(\phi|\mathcal{F}^-)$.*

Proof Since $\theta \in \mathsf{Ker}(\bar{Q})$, there exists $p \in \mathbb{Z}$ such that $Q_b(\theta) = p$ for all $b \in \mathcal{S}_2(\phi)$ that occur in r. Letting $\upsilon = \theta^{-1}\phi^p$, it follows that $Q_b(\upsilon) = 0$ for all $b \in \mathcal{S}_2(\phi)$ that occur in r. Thus υ satisfies Lemma 15.45(4). Lemma 15.45(1) is obvious and the two remaining items in the hypotheses of that lemma follow from Lemma 17.5. We may therefore apply Lemma 15.45 to conclude that $\upsilon(r) = r$ and hence that $\theta(r) = \theta\upsilon(r) = \phi^p(r) = r$. □

Corollary 17.10 *If $\mathcal{X}^+ \neq \varnothing$, then $\Delta\mathcal{R}(\phi) = \Delta\mathcal{R}(\psi)$.*

Proof If $\theta \in \mathcal{X}^+$ then $\Delta\mathcal{R}(\phi) = \theta(\Delta\mathcal{R}(\phi)) = \Delta\mathcal{R}(\psi)$, where the first equality follows from $\mathcal{X}^+ \subset \mathcal{X}^-$ and Lemma 17.9 and the second equality follows from Definition 17.3(b) and Lemma 3.16(3). □

Notation 17.11 One has that $f(D) = D \cdot \sigma$ for some completely split path $\sigma \subset G_s$, and letting $S_D = \sigma \cdot f_\#(\sigma) \cdot \ldots \cdot f_\#^j(\sigma) \cdot \ldots$, the eigenray R_D determined by D decomposes as $R_D = DS_D$. In the HH case, S_C is defined analogously and $R_C = CS_C$. The rays $R'_{D'}$, $S'_{D'}$, $R'_{C'}$ and $S'_{C'}$ are defined similarly using $g\colon G' \to G'$ in place of $f : G \to G$.

Each element of $\Delta\mathcal{R}(\phi)$ is represented by $R_D = DS_D$ or $R_C = CR_C$ and similarly for each element of $\Delta\mathcal{R}(\psi)$. [Feighn and Handel 2018, Lemma 6.3] therefore supplies an algorithm to decide if a given $r \in \Delta\mathcal{R}(\phi)$ and a given $r' \in \Delta\mathcal{R}(\psi)$ are equal. Applying this up to three times, we can decide if $\Delta\mathcal{R}(\phi) = \Delta\mathcal{R}(\psi)$. If $\Delta\mathcal{R}(\phi) \neq \Delta\mathcal{R}(\psi)$ then $\mathcal{X}^+ = \varnothing$ by Corollary 17.10; return `NO` and `STOP`. We may therefore assume that:

(5) $\Delta\mathcal{R}(\phi) = \Delta\mathcal{R}(\psi)$. Denote this common set by $\Delta\mathcal{R}$. In the H and LH cases, the unique element of $\Delta\mathcal{R}$ corresponds to D and D' and is denoted by r_D. In the HH case, $\Delta\mathcal{R} = \{r_C, r_D\}$, where r_C corresponds to C and C', and r_D corresponds to D and D'; this may require reversing the orientation on E'. Remark 13.1 implies that S_D and $S'_{D'}$ are contained in the core filtration element G_p that realizes $F(r_D)$ and that, in the HH case, S_C and $S'_{C'}$ are contained in the core filtration element G_q realizing $F(r_C)$.

[Feighn and Handel 2018, Lemma 6.3] also gives us initial subpaths of S_D and $S'_{D'}$ whose terminal complements are equal. We may therefore assume:

(6) There is a finite path $\kappa_D \subset G_p$ such that $S'_{D'}$ is obtained from $\kappa_D S_D$ by tightening. Similarly, in the case HH, there is a finite path $\kappa_C \subset G_q$ such that $S'_{C'}$ is obtained from $\kappa_C S_C$ by tightening.

We record the following for convenient referencing.

Lemma 17.12 *Suppose that $f : G \to G$ is a* CT *representing ϕ and realizing $\mathfrak{c}$ and that either L is an element of $\Omega(\phi)$ or L is an element of $\mathsf{L}_\mathfrak{e}(\phi)$, where $\mathfrak{e} \in \mathfrak{c}$ is not* large. *Let σ be a line in $\mathsf{H}(L)$. Then one of the following (mutually exclusive) properties is satisfied:*

- *L does not cross any higher-order edges; $\sigma = L$.*
- *$L = \beta^{-1} R_e$ (resp. $R_e^{-1}\beta$) for some higher-order edge e and ray β that does not cross any higher-order edges; $\sigma = \beta^{-1} e \tau$ (resp. $\tau^{-1}\bar{e}\beta$), where τ is a ray in the core filtration element that realizes $F(r_e)$.*
- *$L = R_{e_1}^{-1} \rho R_{e_2}$, where e_1, e_2 are higher-order edges and ρ is a Nielsen path; $\sigma = \tau_1^{-1} e_1^{-1} \rho e_2 \tau_2$, where τ_1 is a ray in the core filtration element that realizes $F(r_{e_1})$ and τ_2 is a ray in the core filtration element that realizes $F(r_{e_2})$.*

Proof The description of L comes from Lemma 4.2 and the fact (Lemma 13.9) that each such L is carried by $\Gamma(f)$. The description of σ is immediate from the definitions of $\mathsf{H}(L)$. ◻

Notation 17.13 Represent the trivial element of $\mathsf{Out}(F_n)$ by a homotopy equivalence $h\colon G \to G'$ that restricts to the identity on G_s. We may assume that $h(G_u) = G'_{u'}$ because G_u and $G'_{u'}$ are core graphs that represent $\mathcal{F}^+$. Recall from (5) that G_p is the core filtration element that realizes $F(r_D)$ and that in the HH case, G_q is the core filtration element realizing $F(r_C)$.

Remark 17.14 If the endpoint set of E is equal to the endpoint set of E', then G and G' differ only by a marking change so one can view h, combinatorially, as a homotopy equivalence from G to G (that does not preserve markings). In this case, [Bestvina et al. 2000, Corollary 3.2.2] implies that $h(E) = \bar{\mu} E' \nu$ or $h(E) = \bar{\mu}\bar{E}'\nu$ for some paths $\mu, \nu \subset G_s$. The same conclusion holds if the endpoint sets of E and E' are not equal, because one can fold initial and terminal segments of E' into G_s to arrange that the endpoint set of E is equal to the endpoint set of E'.

The next step in the algorithm is to check if the following statement is satisfied:

(7) If $\mathfrak{e}$ has type HH, then $h(E) = \bar{\mu} E' \nu$ for some paths $\mu \subset G_q$ and $\nu \subset G_p$.

If (7) fails then we return `NO` and `STOP`. We justify this by the following lemma.

Lemma 17.15 *If $\mathfrak{e}$ has type* HH *and $\mathcal{X}^+ \neq \varnothing$, then $h(E) = \bar{\mu} E \nu$ for some paths $\mu, \nu \subset G_s$. Moreover, $\mu \subset G_q$ and $\nu \subset G_p$.*

Proof By Remark 17.14, $h(E) = \bar{\mu} E' \nu$ or $h(E) = \bar{\mu}\bar{E}'\nu$ for some paths $\mu, \nu \subset G_s$, so for the main statement, we just want to rule out the latter possibility. Let $L = R_C^{-1} R_D$ and $L' = R_{C'}^{-1} R_{D'}$. Then $\mathsf{L}_{\mathfrak{e}}(\phi) = \{L, L^{-1}\}$ and $\mathsf{L}_{\mathfrak{e}}(\psi) = \{L', L'^{-1}\}$. Since $\mathcal{X}^+ \neq \varnothing$, there exists $\theta_0 \in \mathcal{X}^-$ such that $\theta_0 | \mathcal{F}^+$ conjugates $\phi|\mathcal{F}^+$ to $\psi|\mathcal{F}^+$. Thus $\theta_0(\mathsf{L}_{\mathfrak{e}}(\phi)) = \mathsf{L}_{\mathfrak{e}}(\psi)$ and, by Lemma 17.9, L and $\theta_0(L)$ have the same initial ends and the same terminal ends. It follows that $\theta_0(L) = L'$. Since h represents an element of $\mathcal{X}^-$, $h_\#(L) \in \mathsf{H}(L')$. In particular, $h_\#(R_C^{-1} R_D)$ does not cross $\bar{E}'$, which implies that $h(E)$ does not cross $\bar{E}'$. This completes the proof of the main statement.

From $h_\#(L) = h_\#(S_C^{-1} \bar{C} D S_D) = [\bar{S}_C \bar{\mu}] \bar{C}' D' [\nu S_D]$ it follows that $[\nu S_D] \subset G_p$ and $[\mu S_C] \subset G_q$. Thus $\nu \subset G_p$ and $\mu \subset G_q$. □

The remainder of the proof of Lemma 16.5 is the construction of an element $\theta^+ \in \mathcal{X}^+$. By Lemma 17.8 and (4), it suffices to find $\theta \in \mathcal{X}^-$ that induces a bijection between lines that lift to $\Gamma(f_u)$ and lines that lift to $\Gamma(g_{u'})$.

The next lemma states that the conclusions of Lemma 17.15 are satisfied in the H and LH cases without the assumption that $\mathcal{X}^+ \neq \varnothing$.

Lemma 17.16 *It holds that $h(E) = \bar{\mu} E' \nu$ for some $\mu \subset G_s$ and $\nu \subset G_p$. In the case* HH, *$\mu \subset G_q$.*

Proof The HH case follows from (7) so we consider only the H and LH cases.

By Remark 17.14, $h(E) = \bar{\mu} E' \nu$ or $h(E) = \bar{\mu} \bar{E}' \nu$ for some paths $\mu, \nu \subset G_s$. Each $L \in \mathsf{H}_{\mathfrak{e}}(\phi)$ (realized in G) decomposes as $L = \bar{\alpha} E \beta$ for some rays $\alpha \subset G_s$ and $\beta \subset G_p$. Likewise each $L' \in \mathsf{H}_{\mathfrak{e}}(\psi)$ (realized in G') decomposes as $L' = \bar{\alpha}' E' \beta'$ for some rays $\alpha' \subset G_s$ and some $\beta' \subset G_p$. Since h represents an element of $\mathcal{X}$, Lemma 13.12 implies that $h_\#(L) \in \mathsf{H}_{\mathfrak{e}}(\psi)$. It follows that $h_\#(L)$ does not cross $\bar{E}'$ and hence that $h(E)$ does not cross $\bar{E}'$. This proves that $h(E) \neq \bar{\mu} \bar{E}' \nu$ and so $h(E) = \bar{\mu} E' \nu$. Note also that $h_\#(L) = [\bar{\alpha} \bar{\mu}] E' [\nu \beta]$, which implies that $[\nu \beta] \subset G_p$ and hence that $\nu \subset G_p$. □

Lemma 17.17 *In the case* H, *the path* μ *is a Nielsen path for* $f|G_s = g|G_s$.

Proof There are three cases to consider, depending on the rank of $\Gamma_*(f_u)$, the component of $\Gamma(f_u)$ containing the ray labeled R_D. Let x be the initial vertex of D. Recall that $\mathcal{F}^- = \mathcal{F}_{k-1}$ and $\mathcal{F}^+ = \mathcal{F}_k$.

In the first case, $\operatorname{rank}(\Gamma_*(f_u)) = 0$ and we will show that μ is trivial. By Remark 17.6, there exist edges $E_1, \dots, E_m \in \mathcal{E}_f$ with $m \geq 2$ such that $\Gamma_*(f_u)$ is obtained from $R_D \sqcup R_{E_1} \sqcup \dots \sqcup R_{D_m}$ by identifying the initial endpoints of all of these rays. By construction, $\mathsf{L}_{\mathfrak{e}}(\phi) = \{L_1, \dots, L_m\}$, where $L_i = R_{E_i}^{-1} R_D$. The description of $\Gamma_*(g_{u'})$ and $\mathsf{L}_{\mathfrak{e}}(\psi)$ is similar with R_D replaced by $R'_{D'}$. There is a permutation π of $\{1, \dots, m\}$ such that $\mathsf{H}(L_i) = \mathsf{H}(L'_{\pi(i)})$. Writing $R_{E_i} = E_i S_i$ for some ray S_i with height less than that of E_i, we have

$$h_\#(L_i) = [S_i^{-1} \bar{E}_i \bar{\mu} D' \nu S_{D'}] = [S_i^{-1} \bar{E}_i \bar{\mu}] D' [\nu S_{D'}],$$

where $[\,\cdot\,]$ is the tightening operation. On the other hand, letting $l = \pi(i)$, we have $h_\#(L_i) = \alpha^{-1} \bar{E}_l D' \beta$, where α (resp. β) has height less than that of E_l (resp. D') and so

$$[\mu E_i S_i] = E_l \alpha.$$

Note that S_i has a subray that is disjoint from E_l. Since $S_i = u_i \cdot f_\#(u_i) \cdot f_\#^2(u_i) \cdot \dots$ is a coarsening of the complete splitting of S_i, it follows that S_i is disjoint form E_l; see Remark 13.1. If $i \neq l$ then E_l is the first edge of μ. If $i = l$ then μ is either trivial or has the form $E_i \sigma \bar{E}_i$. In either case, μ is either trivial or begins with E_l. As this is true for all $1 \leq i \leq m$, we conclude that μ is trivial.

If $\operatorname{rank}(\Gamma_*(f_u)) = 1$ then either x is contained in a circle component B of the core filtration element realizing $\mathcal{F}_0$, or there exists $j < k$ such that $\mathcal{F}_{j-1} \sqsubset \mathcal{F}_j$ is an LH extension realized by adding a linear edge C_j and a higher-order edge D_j with "new" initial endpoint x; in the former case, we say that D is type (i) and in the latter case we say that D is type (ii). If D is type (i) then B is a single edge e by the (Periodic Edges) property of a CT, and $\mathsf{L}_{\mathfrak{e}}(\phi) = \{e^\infty D S_D, e^{-\infty} D S_D\}$ and likewise $\mathsf{L}_{\mathfrak{e}}(\psi) = \{e^\infty D' S_{D'}, e^{-\infty} D' S_{D'}\}$. Lemma 17.12 implies that $h_\#(e^\infty D S_D) = e^{\pm\infty} D' \beta'$ for some ray $\beta' \subset G(r_D)$. We also have $h_\#(e^\infty D S_D) = [e^\infty \bar{\mu}] D' [\nu S_D]$. We conclude that $\mu = e^m$ for some m and in particular μ is a Nielsen path. If D is type (ii) then $\mathsf{L}_{\mathfrak{e}}(\phi) = \{w_j^{-\infty} \bar{C}_j D S_D, w_j^\infty \bar{C}_j D S_D\}$ and similarly for $\mathsf{L}_{\mathfrak{e}}(\psi)$. As in the previous case, $h_\#(w_j^\infty \bar{C}_j D S_D)$ is equal to both $w_j^{\pm\infty} \bar{C}_j D' \beta'$ and $[w_j^\infty \bar{C}_j \bar{\mu}] D' [\nu S_D]$.

It follows that $\mu C_j w_j^{\infty} = C_j w_j^{\pm\infty}$, which implies that $\mu = C_j w_j^m \overline{C}_j$. This completes the proof if $\operatorname{rank}(\Gamma_*(f_u)) = 1$.

For the final case, assume that $\operatorname{rank}(\Gamma_*(f_u)) \geq 2$ and hence that $x \in G_s$. Given a lift $\tilde{x} \in \widetilde{G}$ of the initial endpoint x of D, we set notation as follows: $\tilde{f}\colon \widetilde{G} \to \widetilde{G}$ is the principal lift that fixes $\tilde{x}$; $\Phi \in \mathcal{P}(\phi)$ is the principal automorphism satisfying $\Phi|\partial F_n = \tilde{f}|\partial F_n$; $\widetilde{D}$ is the lift of D with initial endpoint $\tilde{x}$; $\widetilde{R}_{\widetilde{D}}$ is the lift of R_D whose first edge is $\widetilde{D}$; $\tilde{r}_{\widetilde{D}}$ is the terminal endpoint of $\widetilde{R}_{\widetilde{D}}$; and $\widetilde{N}$ is the set of lines $(\mathsf{Fix}(\Phi), F(\tilde{r}_D))$ and so is a lift of $\mathsf{H}_{\mathfrak{e}}(\phi) = \{[\mathsf{Fix}(\Phi), F(\tilde{r}_D)]\}$. Similarly, given a lift $\tilde{y} \in \widetilde{G}'$ of the initial endpoint y of D', we have: $\tilde{g}\colon \widetilde{G}' \to \widetilde{G}'$, Ψ, $\widetilde{D}'$, $\widetilde{R}_{\widetilde{D}'}$, $\tilde{r}'_{\widetilde{D}'}$ and $\widetilde{N}'$. By Lemma 13.12, $\mathsf{H}_{\mathfrak{e}}(\phi) = \mathsf{H}_{\mathfrak{e}}(\psi)$ so we may choose $\tilde{y}$ so that $\mathsf{Fix}(\Phi) = \mathsf{Fix}(\Psi)$ and $F(\tilde{r}_{\widetilde{D}}) = F(\tilde{r}'_{\widetilde{D}'})$. In particular, $\widetilde{N} = \widetilde{N}'$.

Let C_s be the component of G_s that contains both x and y — which is possible because they are the endpoints of $\mu \subset G_s$ — and let $\widetilde{C}_s \subset \widetilde{G}$ be the lift that contains $\tilde{x}$. Then $\widetilde{C}_s$ is $\tilde{f}$-invariant and $\mathsf{Fix}(\tilde{f}) \subset \widetilde{C}_s$ because G_s contains all Nielsen paths in G with an endpoint at x. There is a free factor F representing $[C_s]$ such that $\partial F = \partial \widetilde{C}_s$. Since $\partial\,\mathsf{Fix}(\Phi)$ is contained in the closure of $\mathsf{Fix}(\tilde{f})$, we have $\partial\,\mathsf{Fix}(\Phi) \subset \partial \widetilde{C}_s = \partial F$. Letting $\widetilde{C}'_s \subset \widetilde{G}'$ be the lift of C_s that contains $\tilde{y}$ and F' the free factor satisfying $\partial F' = \partial \widetilde{C}'_s$, the same argument shows that $\partial\,\mathsf{Fix}(\Psi) \subset \partial F'$. Since $\mathsf{Fix}(\Phi) = \mathsf{Fix}(\Psi)$ is nontrivial, $F = F'$. Since $\phi|F = \psi|F$ and $\mathsf{Fix}(\Phi) = \mathsf{Fix}(\Psi)$ has rank at least two, $\Phi|F = \Psi|F$.

Let $\tilde{h}\colon \widetilde{G} \to \widetilde{G}'$ be the lift of $h\colon G \to G'$ that acts as the identity on ∂F_n and let $p\colon \widetilde{G} \to G$ and $p'\colon \widetilde{G}' \to G'$ be the covering projections. Then $\tilde{h}|\widetilde{C}_s\colon \widetilde{C}_s \to \widetilde{C}'_s$ is a homeomorphism satisfying $p'\tilde{h}(\tilde{z}) = p(\tilde{z})$ for all $\tilde{z} \in \widetilde{C}_s$ where, as usual, we are viewing G_s as a subgraph of both G and G'. Moreover, $\tilde{h}\tilde{f}\tilde{h}^{-1}|\widetilde{C}'_s = \tilde{g}|\widetilde{C}'_s$ because they both project to $g|C'_s$ and induce $\Psi|F'$. In particular $\tilde{g}$ fixes $\tilde{h}(\tilde{x})$. Choose $\widetilde{L} \in \widetilde{N}$ that decomposes as $\widetilde{L} = \tilde{\alpha}\widetilde{R}_{\widetilde{D}}$. Then $h_\#(L) = [h(\alpha)\mu^{-1}]D'\tau'$ for some ray τ' with height lower than that of $\mathcal{D}'$. Since $\tilde{h}_\#(\widetilde{L}) \in \widetilde{N}'$ we have $\tilde{h}_\#(\widetilde{L}) = [\tilde{h}(\tilde{\alpha})\tilde{\mu}^{-1}]\widetilde{D}'\tilde{\tau}'$ for some lift $\tilde{\mu} \subset C'_s$ and some lift $\tilde{\tau}'$ with height lower than that of $\widetilde{D}'$. In particular, $\tilde{\mu}$ connects $\tilde{y}$ to $\tilde{h}(\tilde{x})$ and so is a Nielsen path for $\tilde{g}$. Thus μ is a Nielsen path for $g|G_s = f|G_s$. $\square$

We will complete the proof by constructing a homotopy equivalence $d\colon G' \to G'$ such that the outer automorphism θ^+ determined by $dh\colon G \to G'$ is an element of $\mathcal{X}^+$. By construction, d will be the identity on the complement of E' and satisfy $d(E') = \overline{\mu}'E'\nu'$, where $\mu', \nu' \subset G_s$ are closed paths. In the cases H and LH, μ' will be trivial.

Definition 17.18 We define ν', which always corresponds to D', as follows. By (6), there is a finite path $\kappa_D \subset G_p$ such that $S'_{D'}$ is obtained from $\kappa_D S_D$ by tightening. By construction, $h_\#(ES_D)$ is obtained from $\overline{\mu}E'\nu S_D$ by tightening. Letting

$$\nu' = [\kappa_D \overline{\nu}],$$

it follows that $(dh)_\#(ES_D) = [\overline{\mu}\overline{\mu}']E'[\kappa_D\overline{\nu}\nu S_D] = [\overline{\mu}\overline{\mu}']E'S'_{D'}$. Thus, in the cases HH and LH we have

$$(dh)_\#(R_D) = R'_{D'},$$

and in the case H we have

$$(dh)_\#(R_D) = [\bar{\mu}\bar{\mu}']R'_{D'} = \bar{\mu}R'_{D'},$$

where the second equality comes from the fact that μ' is trivial (see below) in the case H. Since $\nu, \kappa_D \subset G_p$, we have:

- (control of ν') $\nu' \subset G_p$.

In the cases LH and H, μ' is defined to be trivial. In the case HH we choose μ' as we did ν' replacing D with C. The result is that in the case HH,

$$(dh)_\#(R_C) = R'_{C'},$$

and so

$$(dh)_\#(R_C^{-1} R_D) = {R'_{C'}}^{-1} R'_{D'}.$$

Also,

- (control of μ') In the case HH, $\mu' \subset G_q$.

This completes the definition of d.

Lemma 17.19 *The outer automorphism δ represented by $d: G' \to G'$ is an element of* $\mathsf{Ker}(Q) \subset \mathcal{X}$.

Proof We use the following properties of $d: G' \to G'$ to prove that $\delta \in \mathcal{X}$:

(a) The map d preserves every component of every filtration element of G'. In particular, δ preserves $\mathfrak{c}$ and every $[F] \in \mathcal{F} \in \mathfrak{c}$.

(b) If e' is not a higher-order edge in E', then $d(e') = e'$.

These are both obvious from the definition.

(c) The map $d_\#$ fixes each Nielsen path of g.

This follows from (b) and the fact that Nielsen paths do not cross higher-order edges.

- If e' is a higher-order edge and $G'_{p'}$ is the filtration element that realizes $F(r'_{e'})$, then the set of rays of the form $e'\beta'$ with $\beta' \subset G'_{p'}$ is mapped into itself by $d_\#$.

If e' is neither C nor D then this follows from (a) and (b). Otherwise it follows from (a), (control of ν') and (control of μ').

Item (c) implies that δ fixes each component of $\mathsf{Fix}(\phi)$ and every element of $\mathcal{A}(\phi) = \mathcal{A}(\psi)$. Item (b) implies that $d|G'_s$ =identity and hence that δ fixes each element of $\mathsf{SA}(\phi|\mathcal{F}_0)$ and so satisfies defining property (6) of $\mathcal{X}$. Suppose that either L' is an element of $\Omega(\psi)$ or L' is an element of $\mathsf{L}_{\mathfrak{e}'}(\psi)$, where $\mathfrak{e}' \in \mathfrak{c}$ is not large. Then (b), (d) and Lemma 17.12 imply that $\mathsf{H}(L')$ is δ-invariant. Similarly, (c) and (d) imply that if $\mathfrak{e}'$ is large then $\mathsf{H}_{\mathfrak{e}'}(\phi)$ is δ-invariant. We have now shown that $\delta \in \mathcal{X}$. In particular, δ is in the domain of $\bar{Q}$.

To prove that $\delta \in \mathsf{Ker}(Q)$, suppose that $\tilde{b}' = (\tilde{L}'_1, \tilde{L}'_2)$ is a staple pair for ψ with common axis $\tilde{A}'$. By (b), there is a lift $\tilde{d}\colon \tilde{G}' \to \tilde{G}'$ of d that pointwise fixes $\tilde{A}'$. We claim that $\tilde{d}$ preserves both $\mathsf{H}(\tilde{L}'_1)$ and $\mathsf{H}(\tilde{L}'_2)$. The $\tilde{L}'_1$ and $\tilde{L}'_2$ cases are symmetric so we will consider L'_2. If both ends of L'_2 are periodic then $\mathsf{H}(L'_2) = \{L'_2\}$ by Lemma 17.12. Moreover, $\tilde{L}'_2$ does not cross any higher-order edges and so is pointwise fixed by $\tilde{d}$. We may therefore assume that there is a decomposition $\tilde{L}'_2 = \tilde{\alpha}\tilde{R}'_{\tilde{e}'_2}$, where $\tilde{e}'_2$ is a higher-order edge and $\tilde{\alpha}$ does not cross any higher-order edges. It follows that $\tilde{d}$ pointwise fixes $\tilde{\alpha}$ and that $\tilde{e}_2$ is the initial edge of $\tilde{d}_\#(\tilde{R}'_{\tilde{e}'_2})$. Item (d) now implies that $\tilde{d}$ preserves $\mathsf{H}(\tilde{L}'_2)$. This completes the proof of the claim. It now follows from the definitions that $m_{b'}(\delta) = 0$ and hence $Q_{b'}(\delta) = 0$. Since b' is arbitrary, $\delta \in \mathsf{Ker}(Q)$. □

The final step in the algorithm that proves Proposition 16.4 is to return YES and the outer automorphism θ^+ represented by $dh\colon G \to G'$. In conjunction with Lemma 17.4, the following lemma justifies this step.

Lemma 17.20 *The map $dh\colon G \to G'$ represents an element $\theta^+ \in \mathcal{X}^+$.*

Proof Since h represents an element of $\mathsf{Ker}(\bar{Q})$, Lemma 17.19 implies that $\theta^+ \in \mathsf{Ker}(\bar{Q}) \subset \mathcal{X}$. We are therefore reduced to showing that $\theta^+|\mathcal{F}^+$ conjugates $\phi|\mathcal{F}^+$ to $\psi|\mathcal{F}^+$. By Lemma 17.8 and (4), it suffices to prove that:

(b1) A line $L \subset G$ lifts to $\Gamma(f_u)$ if and only if $\theta^+(L) \subset G'$ lifts to $\Gamma(g_{u'})$.

Recall that $\Gamma(f_u)$ is obtained from $\Gamma(f_s)$ by either adding a single new component (the HH and LH cases) or by adding a ray to one of the components $\Gamma_*(f_s)$ of $\Gamma(f_s)$ (the H case). The same is true for $\Gamma(g_{u'})$. The component of $\Gamma(f_u)$ that is not a component of $\Gamma(f_s)$ is denoted by $\Gamma_*(f_u)$; it is the unique component that contains a ray labeled R_D. Likewise, the "new" component $\Gamma_*(g_{u'})$ of $\Gamma(g_{u'})$ is the one that contains a ray labeled $R'_{D'}$. Recall also that $f_s = g_s$, that $\Gamma(f_s) = \Gamma(g_s)$, and that $\Gamma_*(f_s) = \Gamma_*(g_s)$.

Item (b1) is obvious if L lifts into a component of $\Gamma(f_s)$ so we may assume, after reversing orientation on L if necessary, that R_D is a terminal ray of L.

In the case HH, $L = R_C^{-1}R_D$ so (b1) follows from $(dh)_\#(R_C^{-1}R_D) = {R'_{C'}}^{-1}R'_{D'}$; see Definition 17.18.

In the case H, $(dh)_\#(R_D) = \bar{\mu}R'_{D'}$ by Definition 17.18. Let $\hat{x}$ be the initial endpoint of the lift of R_D into $\Gamma(f_u)$ and let x be its projection into G_s. Define $\hat{y}$ and y similarly using $R'_{D'}$ in place of R_D. If R_D is also a terminal ray of L^{-1}, then $L = R_D^{-1}\xi R_D$ for some Nielsen path ξ and $(dh)_\#(L) = R'^{-1}_{D'}[\mu\xi\bar{\mu}]R'_{D'}$, which lifts into $\Gamma(g_{u'})$ because $[\mu\xi\bar{\mu}]$ is a Nielsen path by Lemma 17.17. The remaining case is that $L = \beta^{-1}R_D$ for some ray $\beta \subset G_s$ that lifts to a ray in $\Gamma(f_s)$ based at $\hat{x}$. In this case, $(dh)_\#(L) = \beta^{-1}\bar{\mu}R'_{D'}$, which lifts into $\Gamma^0(g_{u'})$ because μ is a Nielsen path that lifts to a path in $\Gamma(g_s)$ connecting $\hat{y}$ to $\hat{x}$.

In the case LH, $\mathsf{L}_\mathfrak{c}(\phi) = \{L_+, L_-\}$ and $\mathsf{L}_\mathfrak{c}(\psi) = \{L'_+, L'_-\}$, where $L_\pm = w^{\pm\infty}\bar{C}R_D$ and $L'_\pm = w^{\pm\infty}\bar{C}'R'_{D'}$. Since $\mathsf{I}_\mathfrak{c}(\phi) = \mathsf{I}_\mathfrak{c}(\psi)$ and $\theta^+ \in \mathcal{X}$, we have that $\theta^+(L_+)$ is contained in either $\mathsf{H}(L'_+)$ or $\mathsf{H}(L'_-)$. By construction, $\theta^+(L_+) = (dh)_\#(L_+) = [w^\infty\bar{\mu}]\bar{C}'R'_{D'}$. Thus, $\theta^+(L_+) \in \mathsf{H}(L'_+)$ and $[w^\infty\bar{\mu}] = w^\infty$. It follows that $\mu = w^p$ for some $p \in \mathbb{Z}$ and hence that $\theta^+(L_\pm) = L'_\pm$ and $\theta^p(R_D^{-1}w^kR_D) = R'^{-1}_{D'}w^kR'_{D'}$ for all k. This completes the proof of (b1) and hence the proof of the lemma. □

Appendix More on $\mathsf{Ker}\,\bar{Q}$

The main results of this appendix are that $m_b(\theta)$ can be computed for $\theta \in \mathcal{X}$, that $\mathsf{Ker}\,\bar{Q}$ is of type VF (Definition 14.4), and that a finite presentation for $\mathsf{Ker}\,\bar{Q}$ can be computed. This section is needed for future work and is not used in the proof of the main theorem of this paper.

A.1 A Stallings graph for $\mathsf{H}_{\phi,\mathfrak{c}}(\tilde{b})$

We will need the following remark.

Remark A.1 Suppose that G is a marked graph and that for $i = 1, 2$, $\tilde{A}_i$ is the axis of a covering translation $T_i : \tilde{G} \to \tilde{G}$ and that the number of edges in a fundamental domain for $\tilde{A}_i$ is s_i. If $\tilde{A}_1 \cap \tilde{A}_2$ contains at least $s_1 + s_2 + 1$ edges then $\tilde{A}_1 = \tilde{A}_2$. To see this, decompose $\tilde{A}_1 \cap \tilde{A}_2 = e_1 e_2 \dots$ into edges and note that $T_1 T_2(e_1) = e_{s_1+s_2+1} = T_2 T_1(e_1)$. Since $T_1 T_2$ and $T_2 T_1$ agree on an edge they are equal, and so T_1 and T_2 have the same axes.

Notation A.2 Let $f : G \to G$ be a CT for ϕ with $\tilde{f} : \tilde{G} \to \tilde{G}$ a lift to the universal cover. Assume notation as in the definition of m_b (Definition 15.36). In particular, $b = (L_1, L_2) \in \mathcal{S}_2(\phi)$, $\tilde{b} = (\tilde{L}_1, \tilde{L}_2)$ are lifts of (L_1, L_2) such that $\tilde{L}_1^+, \tilde{L}_2^- \in \{\tilde{A}^-, \tilde{A}^+\}$, where $\tilde{A}$ is the common axis of $\tilde{b}$ and $a \in F_n$ is a root-free element with axis $\tilde{A}$ and orientation chosen as in the definition. For $\theta \in \mathcal{X}$, $\Theta_i \in \theta$ is defined uniquely by $\Theta_i(\mathsf{H}_{\phi,\mathfrak{c}}(\tilde{L}_i)) = \mathsf{H}_{\phi,\mathfrak{c}}(\tilde{L}_i)$ and $m_b(\theta)$ is defined so that $\Theta_1 = i_a^{m_b(\theta)} \Theta_2$.

If $\tilde{L}_2^+ = \tilde{r}_2$ for some $r_2 \in \mathcal{R}(\phi)$, then define $\mathsf{H}^2_{\phi,\mathfrak{c}}(\tilde{b}) = F_{\mathfrak{c}}(\tilde{r}_2)$. Otherwise, $\tilde{L}_2^+ \in \{c_2^-, c_2^+\}$ for some root-free $c_2 \in F_n$ representing an element of $\mathcal{A}(\phi)$ and $\mathsf{H}^2_{\phi,\mathfrak{c}}(\tilde{b}) := \langle c_2 \rangle$. Define $\mathsf{H}^1_{\phi,\mathfrak{c}}(\tilde{b})$ similarly using $\tilde{L}_1^-$ in place of $\tilde{L}_2^+$. Finally, define

$$\mathsf{H}_{\phi,\mathfrak{c}}(\tilde{b}) = \langle \mathsf{H}^1_{\phi,\mathfrak{c}}(\tilde{b}), \mathsf{H}^2_{\phi,\mathfrak{c}}(\tilde{b}) \rangle.$$

The covering transformation corresponding to a is denoted by τ. Additionally, $H^i := \mathsf{H}^i_{\phi,\mathfrak{c}}(\tilde{b})$, T^i denotes the minimal subtree for H^i, Γ^i denotes the Stallings graph for H^i, and $\tilde{L}(k)$ denotes the line $[\tilde{L}_1^-, \tau^k(\tilde{L}_2^+)]$.

Remark A.3 Comparing definitions of $\mathsf{H}_{\phi,\mathfrak{c}}(\tilde{L}_i)$ and H^i, Θ_i is the unique $\Theta \in \theta$ fixing $\langle a \rangle$ and H^i.

Lemma A.4 *There is a $k \ge 0$ such that*

- $T^1 \cap \tau^k(T^2) = \varnothing$, *and*
- *the arc $\tilde{\mu}$ spanning between T^1 and $\tau^k(T^2)$ contains more than two fundamental domains of $\tilde{A}$ with orientation agreeing with that of $\tilde{\mu}$.*

Proof The ends $\tilde{A}$ are not ends of T^i. Indeed, if $\tilde{r}_i$ is ray, then the associated higher-order edge separates T^i and the ends of $\tilde{A}$. If not, then T^i is the axis corresponding to the end of $\tilde{L}_i$ that is not an end of $\tilde{A}$. Hence there is a neighborhood of $\tilde{A}^+$ that is disjoint from T_i. Therefore, the conclusion holds for all large k. □

Corollary A.5 *We may compute*:

(1) *For all $l \in \mathbb{Z}$, the Stallings graph for $\langle H^1, i_a^l(H^2)\rangle$.*

(2) *An integer $k \geq 0$ as in Lemma A.4.*

(3) *For all $\theta \in \mathcal{X}$, $m_b(\theta)$.*

Proof (1) By Bass–Serre theory, for k as in Lemma A.4, the Stallings graph for $\langle H^1, i_a^k(H^2)\rangle$ is obtained by attaching at its endpoints a copy of the arc spanning between T^1 and $\tau^k(T^2)$ to the Stallings graphs for $[H^1]$ and $[i_a^k H^2] = [H^2]$.

These latter graphs can be computed. Indeed, if $\widetilde{L}_1^-$ is an eigenray, then, by definition, $[H^1]$ has as its Stallings graph a component of a stratum of G, and otherwise is a circle representing $\langle c_1\rangle$. There is a symmetric argument for $[H^2]$. $\widetilde{L}(k)$ spans between T^1 and $\tau^k(T^2)$. Hence the desired Stallings graph, for large k, is the result of immersing the ends of $\widetilde{L}(k)$ into the Stallings graphs and then performing any folding. By Lemma A.4, folding stops when the copy of $\widetilde{\mu}$ is the spanning arc.

We see that, for large k, $\langle H^1, i_a^k(H^2)\rangle$ is an internal free product. By Remark A.3, $\Phi_1^s(\langle H^1, i_a^k(H^2)\rangle) = \langle H^1, i_a^{sm_b(\phi)+k}(H^2)\rangle$ is also a free product and its Stallings graph can be computed as above (but perhaps the spanning arc is folded away). Recall (Lemma 15.39) that $m_b(\phi) \neq 0$. We note in passing that therefore $[H^1, i_a^l(H^2)]$ is good (Definition 7.13).

(2) For $l = 0, 1, 2, \ldots$, iteratively start computing Stallings graphs for $\langle H^1, i_a^l(H^2)\rangle$. When, after folding in $\widetilde{L}(l)$, more than two correctly oriented fundamental domains of $\widetilde{A} \cap \widetilde{L}(l)$ remain in the spanning arc, stop and set $k = l$.

(3) Let $\theta \in \mathcal{X}$ and $m := m_b(\theta)$. If $\Theta_1(H^1) = H^1$, then by definition,

$$\Theta_1(H^1, i_a^k(H^2)) = (H^1, i_a^{k+m}(H^2)).$$

Hence, m can be read off by comparing the Stallings graphs for $[\langle H^1, i_a^k(H^2)\rangle]$ and $[\langle H^1, i_a^{k+m}(H^2)\rangle]$ for large enough k. The latter, being the Stallings graph for $\theta[\langle H^1, i_a^k(H^2)\rangle]$, can be computed by representing θ as a topological representative $g\colon G \to G$, applying g to the Stallings graph for $[\langle H^1, i_a^k(H^2)\rangle]$, tightening, and taking the core. □

Corollary A.6 (1) *$[H_1, i_a^l H_2]$ is good for all $l \in \mathbb{Z}$.*

(2) *No nontrivial power of a is in $\langle H_1, i_a^l H_2\rangle$.*

Proof (1) This was noted during the proof of Corollary A.5.

(2) Since $\Theta_1(a) = a$, for the second item it is enough to show that

$$a \notin \Theta_1^l(\langle H^1, i_a^l(H^2)\rangle) = \langle H^1, i_a^{l+m}(H^2)\rangle$$

for large $l + m$. So assume that $k > 0$ is as in Lemma A.4. If $a \in \langle H^1, i_a^k(H^2)\rangle$, then there is an immersion of $\widetilde{A}$ with image a closed loop into the Stallings graph for $\langle H^1, i_a^k(H^2)\rangle$ and that overlaps the copy of $\tilde{\mu}$ in its intersection with $\tilde{A}$. The immersion crosses this spanning arc at most once. Indeed, otherwise there would be a covering translation of $\tilde{G}$ taking $\tilde{A}$ to itself reversing orientation, but a and a^{-1} are not conjugate; see Remark A.1. Hence the image of $\tilde{A}$ is not a closed loop. □

Corollary A.7 *For $\theta \in \mathcal{X}$, $m_b(\theta) = 0 \iff [\mathsf{H}_{\phi,c}(\tilde{b})]$ is θ-invariant.*

Proof ($\Longrightarrow$) Suppose $m := m_b(\theta) = 0$. Using Remark A.3 we have

$$\Theta_1(\mathsf{H}_{\phi,c}(\tilde{b})) = \Theta_1(\langle H^1, H^2\rangle) = \langle \Theta_1(H^1), \Theta_1(H^2)\rangle = \langle H^1, i_a^m(H^2)\rangle = \langle H^1, H^2\rangle.$$

($\Longleftarrow$) Suppose $m > 0$, the case for $m < 0$ being similar. Choose l so that $k := ml$ is as in Lemma A.4. We saw in Corollary A.5 that the Stallings graph for $[\langle H^1, i_a^{ml}(H^2)\rangle]$ has an arc spanning between Stallings graphs for $[H^1]$ and $[H^2]$ and that the Stallings graph for $[\Theta_1^{l+1}(\langle H^1, H^2\rangle)]$ is obtained by inserting m copies of a fundamental domain for $\tilde{A}$ into this spanning arc. Since these Stallings graphs are not equal, $\theta([\langle H^1, H^2\rangle]) \neq [\langle H^1, H^2\rangle]$. □

A.2 Ker Q

Recall (Definition 16.2) the homomorphism $Q = Q^\phi \colon \mathcal{X} \to \mathbb{Q}^{\mathcal{S}_2(\phi)}$ given by setting the b^{th} coordinate of $Q(\theta)$ equal to $m_b(\theta)/m_b(\phi)$.

Proposition A.8 *A finite presentation for* Ker Q *can be computed.* Ker Q *is of type* VF.

Proof We use the notation as in Section A.1. For each $b \in \mathcal{S}_2(\phi)$, choose $\tilde{b}$ and replace it with $\Phi_1^k(\tilde{b})$ and compute $[\mathsf{H}_{\phi,c}(\tilde{b})]$, where k is as in Lemma A.4. By Corollary A.7, $\theta \in \operatorname{Ker} Q$ if and only if $\theta \in \mathcal{X}$ and $\theta([\mathsf{H}_{\phi,c}(\tilde{b})]) = [\mathsf{H}_{\phi,c}(\tilde{b})]$ for each b. Hence $\theta \in \operatorname{Ker} Q$ if and only if θ fixes the concatenation of the sequences (J) and $([\mathsf{H}_{\phi,c}(\tilde{b}_1)], \dots, [\mathsf{H}_{\phi,c}(\tilde{b}_N)])$, where $(b_1, \dots b_N)$ is an ordering of $\mathcal{S}_2(\phi)$ and J is as in Definition 14.1. This concatenation is an element of $\overline{\mathsf{IS}}(\mathbb{A}_\bullet)$; see Notation 10.13. By Lemma 11.2, we can compute a finite presentation for Ker Q and, by Proposition 14.5, Ker Q is of type VF. □

A.3 Ker $\bar{Q}$

$\bar{Q}$ is defined in Definition 16.3.

Proposition A.9 *A finite presentation for* Ker $\bar{Q}$ *can be computed.* Ker $\bar{Q}$ *is of type* VF.

Proof Let π denote the quotient map $Q(\mathcal{X}) \to \bar{Q}(\mathcal{X})$. Since we have a finite generating set for $\mathcal{X}$, we can compute a finite presentation for the free abelian group Ker π. Using Proposition A.8, Lemma 9.3 and

$$1 \to \operatorname{Ker} Q \to \operatorname{Ker} \bar{Q} \to \operatorname{Ker} \pi \to 1,$$

we can compute a finite presentation for Ker $\bar{Q}$.

The above short exact sequence shows that Ker $\bar{Q}$ is an extension of a group of type VF (Proposition A.8) by a finitely generated abelian group. It follows from the proposition below that Ker $\bar{Q}$ is of type VF. □

The following proposition is from Moritz Rodenhausen's thesis. As far as we know, it is not published and so for the reader's convenience we copy his proof here.

Proposition A.10 [Rodenhausen 2013, Proposition 13.18] *Suppose that in the short exact sequence* $1 \to G' \xrightarrow{i} G \xrightarrow{\pi} G'' \to 1$, *the group* G' *is of type* VF *and* G'' *is finitely generated abelian. Then* G *is of type* VF.

Proof Suppose first that G'' is infinite cyclic. Let H' be a subgroup of some finite index d in G' such that H' is of type F. The intersection K' of all (finitely many) index d subgroups of G' also is of type F. Furthermore, the group K' is a characteristic subgroup of G'. Let now $t \in G$ be an element such that $\pi(t)$ generates G''. We denote by K the subgroup of G generated t and $i(K')$. It has finite index in G and fits into a short exact sequence

$$1 \to K' \to K \to G'' \to 1.$$

We see that K is an extension of groups of type F and so has type F; see [Geoghegan 2008, Theorem 7.3.4]. Hence G is of type VF.

The case where G'' is isomorphic to $\mathbb{Z}^m$ is proved by induction on m. Let A'' be an infinite cyclic summand of G'', so $G''/A'' \cong \mathbb{Z}^{m-1}$. The short exact sequences

$$1 \to G' \to \pi^{-1}(A'') \to A'' \to 1 \quad \text{and} \quad 1 \to \pi^{-1}(A'') \to G \to G''/A'' \to 1$$

together with the induction hypothesis completes the proof in the case $G'' \cong \mathbb{Z}^m$.

For the general case, let H'' be a free abelian subgroup of G'' of finite index and $H = \pi^{-1}(H'') \subset G$. We obtain a short exact sequence

$$1 \to G' \to H \to H'' \to 1.$$

Hence H, and so also G, is of type VF. □

References

[Bestvina and Brady 1997] **M Bestvina**, **N Brady**, *Morse theory and finiteness properties of groups*, Invent. Math. 129 (1997) 445–470 MR Zbl

[Bestvina and Handel 1992] **M Bestvina**, **M Handel**, *Train tracks and automorphisms of free groups*, Ann. of Math. 135 (1992) 1–51 MR Zbl

[Bestvina et al. 1997] **M Bestvina**, **M Feighn**, **M Handel**, *Laminations, trees, and irreducible automorphisms of free groups*, Geom. Funct. Anal. 7 (1997) 215–244 MR Zbl

[Bestvina et al. 2000] **M Bestvina**, **M Feighn**, **M Handel**, *The Tits alternative for* Out(F_n)*, I: Dynamics of exponentially-growing automorphisms*, Ann. of Math. 151 (2000) 517–623 MR Zbl

[Bestvina et al. 2004] **M Bestvina**, **M Feighn**, **M Handel**, *Solvable subgroups of* Out(F_n) *are virtually abelian*, Geom. Dedicata 104 (2004) 71–96 MR Zbl

[Bestvina et al. 2005] **M Bestvina**, **M Feighn**, **M Handel**, *The Tits alternative for* $\mathrm{Out}(F_n)$*, II: A Kolchin type theorem*, Ann. of Math. 161 (2005) 1–59 MR Zbl

[Bestvina et al. 2023] **M Bestvina**, **M Feighn**, **M Handel**, *A McCool Whitehead type theorem for finitely generated subgroups of* $\mathrm{Out}(F_n)$, Ann. H. Lebesgue 6 (2023) 65–94 MR Zbl

[Bridson and Wilton 2011] **M R Bridson**, **H Wilton**, *On the difficulty of presenting finitely presentable groups*, Groups Geom. Dyn. 5 (2011) 301–325 MR Zbl

[Cohen and Lustig 1999] **M M Cohen**, **M Lustig**, *The conjugacy problem for Dehn twist automorphisms of free groups*, Comment. Math. Helv. 74 (1999) 179–200 MR Zbl

[Cooper 1987] **D Cooper**, *Automorphisms of free groups have finitely generated fixed point sets*, J. Algebra 111 (1987) 453–456 MR Zbl

[Dahmani 2016] **F Dahmani**, *On suspensions and conjugacy of hyperbolic automorphisms*, Trans. Amer. Math. Soc. 368 (2016) 5565–5577 MR Zbl

[Dahmani 2017] **F Dahmani**, *On suspensions, and conjugacy of a few more automorphisms of free groups*, from "Hyperbolic geometry and geometric group theory", Adv. Stud. Pure Math. 73, Math. Soc. Japan, Tokyo (2017) 135–158 MR Zbl

[Dahmani and Touikan 2021] **F Dahmani**, **N Touikan**, *Reducing the conjugacy problem for relatively hyperbolic automorphisms to peripheral components*, preprint (2021) arXiv 2103.16602

[Dahmani and Touikan 2023] **F Dahmani**, **N Touikan**, *Unipotent linear suspensions of free groups*, preprint (2023) arXiv 2305.11274

[Dahmani et al. 2025] **F Dahmani**, **S Francaviglia**, **A Martino**, **N Touikan**, *The conjugacy problem for* $\mathrm{Out}(F_3)$, Forum Math. Sigma 13 (2025) art. id. e41 MR Zbl

[Feighn and Handel 2011] **M Feighn**, **M Handel**, *The recognition theorem for* $\mathrm{Out}(F_n)$, Groups Geom. Dyn. 5 (2011) 39–106 MR Zbl

[Feighn and Handel 2018] **M Feighn**, **M Handel**, *Algorithmic constructions of relative train track maps and CTs*, Groups Geom. Dyn. 12 (2018) 1159–1238 MR Zbl

[Gaboriau et al. 1998] **D Gaboriau**, **A Jaeger**, **G Levitt**, **M Lustig**, *An index for counting fixed points of automorphisms of free groups*, Duke Math. J. 93 (1998) 425–452 MR Zbl

[Geoghegan 2008] **R Geoghegan**, *Topological methods in group theory*, Grad. Texts in Math. 243, Springer (2008) MR Zbl

[Gersten 1984] **S M Gersten**, *On Whitehead's algorithm*, Bull. Amer. Math. Soc. 10 (1984) 281–284 MR Zbl

[Gersten 1987] **S M Gersten**, *Fixed points of automorphisms of free groups*, Adv. Math. 64 (1987) 51–85 MR Zbl

[Handel and Mosher 2011] **M Handel**, **L Mosher**, *Axes in outer space*, Mem. Amer. Math. Soc. 1004, Amer. Math. Soc., Providence, RI (2011) MR Zbl

[Handel and Mosher 2020] **M Handel**, **L Mosher**, *Subgroup decomposition in* $\mathrm{Out}(F_n)$, Mem. Amer. Math. Soc. 1280, Amer. Math. Soc., Providence, RI (2020) MR Zbl

[Kalajdževski 1992] **S Kalajdževski**, *Automorphism group of a free group: centralizers and stabilizers*, J. Algebra 150 (1992) 435–502 MR Zbl

[Kapovich and Benakli 2002] **I Kapovich**, **N Benakli**, *Boundaries of hyperbolic groups*, from "Combinatorial and geometric group theory", Contemp. Math. 296, Amer. Math. Soc., Providence, RI (2002) 39–93 MR Zbl

[Kapovich and Short 1996] **I Kapovich**, **H Short**, *Greenberg's theorem for quasiconvex subgroups of word hyperbolic groups*, Canad. J. Math. 48 (1996) 1224–1244 MR Zbl

[Krstić et al. 2001] **S Krstić**, **M Lustig**, **K Vogtmann**, *An equivariant Whitehead algorithm and conjugacy for roots of Dehn twist automorphisms*, Proc. Edinb. Math. Soc. 44 (2001) 117–141 MR Zbl

[Los 1996] **J E Los**, *On the conjugacy problem for automorphisms of free groups*, Topology 35 (1996) 779–808 MR Zbl

[Lustig 2000] **M Lustig**, *Structure and conjugacy for automorphisms of free groups, I*, preprint 2000-130, Max Planck Inst. (2000) Available at `https://www.mpim-bonn.mpg.de/preblob/834`

[Lustig 2001] **M Lustig**, *Structure and conjugacy for automorphisms of free groups, II*, preprint 2001-4, Max Planck Inst. (2001) Available at `https://www.mpim-bonn.mpg.de/preblob/854`

[Lustig 2007] **M Lustig**, *Conjugacy and centralizers for iwip automorphisms of free groups*, from "Geometric group theory", Birkhäuser, Basel (2007) 197–224 MR Zbl

[Lyndon and Schupp 1977] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer (1977) MR Zbl

[Magnus et al. 1966] **W Magnus**, **A Karrass**, **D Solitar**, *Combinatorial group theory*: *presentations of groups in terms of generators and relations*, Interscience, New York (1966) MR Zbl

[McCool 1975] **J McCool**, *Some finitely presented subgroups of the automorphism group of a free group*, J. Algebra 35 (1975) 205–213 MR Zbl

[Rodenhausen 2013] **M Rodenhausen**, *Centralisers of polynomially growing automorphisms of free groups*, PhD thesis, Universität Bonn (2013) Available at `https://hdl.handle.net/20.500.11811/5725`

[Sela 1995] **Z Sela**, *The isomorphism problem for hyperbolic groups, I*, Ann. of Math. 141 (1995) 217–283 MR Zbl

[Serre 1980] **J-P Serre**, *Trees*, Springer (1980) MR Zbl

[Stallings 1983] **J R Stallings**, *Topology of finite graphs*, Invent. Math. 71 (1983) 551–565 MR Zbl

[Veblen and Franklin 1921] **O Veblen**, **P Franklin**, *On matrices whose elements are integers*, Ann. of Math. 23 (1921) 1–15 MR Zbl

[Whitehead 1936a] **J H C Whitehead**, *On certain sets of elements in a free group*, Proc. Lond. Math. Soc. 41 (1936) 48–56 MR Zbl

[Whitehead 1936b] **J H C Whitehead**, *On equivalent sets of elements in a free group*, Ann. of Math. 37 (1936) 782–800 MR Zbl

Department of Mathematics and Computer Science, Rutgers University
Newark, NJ, United States

Mathematics and Computer Science Department, Herbert H Lehman College (CUNY)
Bronx, NY, United States

`feighn@rutgers.edu`, `michaelxhandel@gmail.com`

Proposed: Martin R Bridson
Seconded: David Fisher, Bruce Kleiner

Received: 2 July 2019
Revised: 7 March 2022

msp

Geometry & Topology **29**:4 (2025) 1819–1849
DOI: 10.2140/gt.2025.29.1819
Published: 27 June 2025

The systole of large genus minimal surfaces in positive Ricci curvature

HENRIK MATTHIESEN
ANNA SIFFERT

We use Colding–Minicozzi lamination theory to show that the systole, and more generally any homology systole, of a sequence of embedded minimal surfaces in an ambient three-manifold of positive Ricci curvature tends to zero as the genus becomes unbounded.

53A10

1 Introduction

In this paper we study properties of compact, embedded minimal surfaces in a closed (ie compact without boundary) ambient three-manifold M of positive Ricci curvature as their genus becomes unbounded. This complements the celebrated theorem of Choi and Schoen [1985]. Recall that this states that for a three-manifold M with positive Ricci curvature, the space of compact, embedded minimal surfaces in M with bounded genus is compact in the C^{ℓ}-topology for any $\ell \geq 2$.

Our main result shows that the systole of such a sequence of minimal surfaces tends to zero. Recall that the *systole* of a closed surface $\Sigma \subset M$ is defined to be

$$\operatorname{sys}(\Sigma) := \inf\{\operatorname{length}(c) \mid c\colon S^1 \to \Sigma \text{ noncontractible}\}.$$

Note that this takes into account all curves that do not bound a disk in Σ. Similarly, the *homology systole* is given by

$$\operatorname{sys}^h(\Sigma) := \inf\{\operatorname{length}(c) \mid 0 \neq [c] \in H_1(\Sigma; \mathbb{Z}/2\mathbb{Z})\},$$

taking into account only curves that are not a boundary in Σ. Clearly, we have

$$\operatorname{sys}(\Sigma) \leq \operatorname{sys}^h(\Sigma).$$

More generally, for $k \in \mathbb{N}^*$, let us define the k^{th} *homology systole* by

$$\operatorname{sys}^h_k(\Sigma) := \inf\Big\{ \max_{i=1,\dots,k} \operatorname{length}(c_i) \mid \operatorname{rank}(\langle c_1, \dots c_k\rangle) = k \Big\},$$

where the span $\langle c_1, \dots, c_k\rangle$ is taken in $H_1(\Sigma; \mathbb{Z}/2\mathbb{Z})$.

We use $\mathbb{Z}/2\mathbb{Z}$-coefficients here to deal with orientable and nonorientable surfaces simultaneously. Of course, for orientable surfaces we can equivalently use $\mathbb{Z}$-coefficients.

We can now state our main result.

Theorem 1.1 *Assume that (M, g) is a three-manifold with positive Ricci curvature. Let $k \in \mathbb{N}^*$ and consider a sequence $(\Sigma_j)_{j\in\mathbb{N}}$ of closed, embedded minimal surfaces in M with $\chi(\Sigma_j) \to -\infty$ as $j \to \infty$. Then the k^{th} homology systole satisfies*

$$\mathrm{sys}_k^h(\Sigma_j) \to 0 \quad \text{as } j \to \infty.$$

Before putting this result into context we briefly discuss the different assumptions that we make.

The reader might wonder if the assumption of the surfaces being minimal is really required. This is because we need to squish the large genus surfaces Σ_j into the compact manifold M, which may force the systole to tend to zero anyway.

Remark 1.2 There is a sequence $(S_j)_{j\in\mathbb{N}^*}$ of embedded (and unknotted) surfaces in S^3 such that $\mathrm{genus}(S_j) \to \infty$ and $\mathrm{sys}(S_j) \geq c_0 > 0$ for some constant c_0.

This can be constructed as follows: Take a surface R_j of genus j with systole $\mathrm{sys}(R_j) \geq 2c_0 > 0$. By the Nash–Kuiper theorem, there is a $C^{1,\alpha}$-isometric embedding of R_j in an arbitrarily small ball $B_\delta \subset \mathbb{R}^3$, where here and below we denote by B_r the Euclidean ball of radius $r > 0$ centered at the origin. After smoothing this and applying stereographic projection, we get a surface $S_j \subset S^3$ of genus j that is closed, unknotted, embedded and has $\mathrm{sys}(S_j) \geq c_0$.

Also the assumption on the surfaces being embedded is crucial, since high-degree covers of a given minimal surface of positive genus provide trivial counterexamples to the immersed version.

Finally, the following example shows that Theorem 1.1 does not hold without any assumptions on the ambient geometry.

Example 1.3 Denote by Σ_γ a closed surface of genus γ for $\gamma \geq 2$. It is shown in [Tollefson 1969] (see also [Neumann 1976] for a generalization) that the three-manifold $M = S^1 \times \Sigma_\gamma$ admits fiber bundles

$$\Sigma_\delta \to M \to S^1 \tag{1.4}$$

for $\delta = \gamma + n(\gamma - 1)$ and $n \in \mathbb{N}$. Since $\pi_2(S^1) = 0$, the long exact sequence for homotopy groups associated to these fibrations implies that $\Sigma_\delta \to M$ is incompressible, ie the induced map $\pi_1(\Sigma_\delta) \to \pi_1(M)$ is injective. It follows from [Schoen and Yau 1979, Theorem 3.1] that there are immersed minimal surfaces S_δ in M which are diffeomorphic to Σ_δ and the induced map on π_1 is given by the inclusion of the fibers from (1.4). Moreover, [Freedman et al. 1983, Theorem 5.1] implies that these are not only immersions but even embeddings. Since $\pi_1(S_\delta) \to \pi_1(M)$ is injective, we have in particular that

$$\mathrm{sys}(S_\delta) \geq \mathrm{sys}(M) > 0.$$

This shows that Theorem 1.1 cannot hold for M.

Remark 1.5 It follows from [Schoen and Yau 1979, Theorem 5.2] that M does not admit any metric of positive scalar curvature, which leaves open the possibility to replace Ricci by scalar curvature in Theorem 1.1.

Let us now put Theorem 1.1 into some more context.

Balacheff, Parlier and Sabourau [2012] provided general results on systoles of surfaces; see Section 2.1 for more details. As a consequence thereof, for any sequence (Σ_j) of surfaces with $\chi(\Sigma_j) \to -\infty$ and area growth in the genus γ bounded by $|\gamma|^\alpha$ for some $\alpha < 1$, and for any $k \in \mathbb{N}$, the k^{th} homology systole tends to zero, ie $\operatorname{sys}^h_k(\Sigma_j) \to 0$ as $j \to \infty$. This motivates the following discussion about area bounds in the genus.

Thanks to the recent work of Chodosh and Mantoulidis [2020] on the Allen–Cahn equation, any closed three-manifold with positive Ricci curvature contains a sequence of embedded minimal surfaces with unbounded genus; see also the related earlier work by Marques and Neves [2017] and Aiex [2018]. Their construction gives a sequence of minimal surfaces $(\Sigma_p)_{p\in\mathbb{N}}$ with

$$\operatorname{area}(\Sigma_p) \sim \operatorname{genus}(\Sigma_p)^{1/3} \sim p^{1/3},$$

ie area growing sublinearly in the genus. In fact, the same result has now also been established using Almgren–Pitts min-max theory through the works of Marques and Neves [2021] and Zhu [2020]. At this point we also would like to point out Song's work [2023] settling the general case of Yau's conjecture and the papers by Irie, Marques and Neves [Irie et al. 2018], Marques, Neves and Song [Marques et al. 2019], and Liokumovich, Marques and Neves [Liokumovich et al. 2018] giving information on the distribution of min-max minimal hypersurfaces in the ambient manifold for generic metrics. In a similar direction Theorem 1.1 provides information on embedded minimal surfaces of high complexity. Because of the sublinear growth of the area, Theorem 1.1 is automatically true for min-max minimal surfaces. However, Theorem 1.1 applies to any family of minimal surfaces, not only those arising from min-max methods.

The best known bound for the area of embedded minimal surfaces in an ambient three manifold of positive Ricci curvature is linear in the genus [Choi and Wang 1983]. More precisely, we have that

$$\operatorname{area}(\Sigma) \le C(\operatorname{genus}(\Sigma) + 1)$$

for a constant C depending on the topology of M and the lower bound on the Ricci curvature. It is by no means clear if this bound is sharp and Theorem 1.1 could be considered as some hint towards the nonsharpness of the linear bound. It appears to be an interesting question to understand the maximal possible area growth of a sequence of embedded, minimal surfaces with genus tending to infinity. To the best knowledge of the authors, among all known families of minimal surfaces in $\mathbb{S}^3$ the Lawson surfaces $\xi_{m,m}$, see [Lawson 1970], exhibit the fastest area growth in terms of the genus. More precisely, $\operatorname{genus}(\xi_{m,m}) = m^2$ while $\operatorname{area}(\xi_{m,m}) \sim m$.

Main problems and strategy

Let us for simplicity focus on the case of M being simply connected, $k = 1$ and the systole instead of the homology systole.

We want to argue by contradiction and consider a sequence of minimal surfaces $\Sigma_j \subset M$ with $\mathrm{sys}(\Sigma_j) \geq l_0 > 0$ and $\mathrm{genus}(\Sigma_j) \to \infty$. In general, we would like to pass to a limit $\Sigma_j \to \mathscr{L}$ in the class of minimal laminations (see eg Definition 3.1) and argue that $\mathscr{L}$ has a stable leaf, which would easily lead to a contradiction since M has positive Ricci curvature.

The problem about this is that we can only do this outside the closed set at which $|A^{\Sigma_j}|^2$ blows-up, where A^{Σ_j} denotes the second fundamental form of Σ_j. A priori, the blow-up set could even be all of M. Work of Colding and Minicozzi gives strong structural information about the blow-up set if the surfaces in question have bounded genus. The main step of our proof is to show that the sequence Σ_j as above can locally be dealt with in this framework.

The reason why this is not obvious is that we do not have $-\Delta_{\Sigma_j} d^2(x,\cdot) \leq 0$ globally (as it is the case for minimal surfaces in $\mathbb{R}^3$). Therefore, the assumption on $\mathrm{sys}(\Sigma_j)$ does not directly imply that there is $R_0 = R_0(l_0)$ such that the intrinsic balls $B^{\Sigma_j}(x, R_0)$ are contained in (intrinsic) *disks* in the intersection $B(x, R_0) \cap \Sigma_j$ with an extrinsic ball. Instead, $B^{\Sigma_j}(x, R_0)$ is contained in some disk $D_x^j \subset \Sigma_j$ but D_x^j could leave any mean-convex ball $B(x,r)$ centered at x. The main step is to show that this is impossible after going to a (potentially much) smaller scale. The proof of this is of global nature and also relies on the positivity of the Ricci curvature of M. It also proceeds by contradiction and follows broadly the same strategy. Given a contradicting sequence we try to find a stable minimal surface in M. The key step to achieve this is to show that Σ_j serves as a good barrier for a minimization problem in M. This in turn is shown by promoting information about singularities of Σ_j for $j \to \infty$ across scales using the maximum principle.

The general case of the theorem follows similar steps but is technically more involved. This requires for instances a more careful blow-up argument in the case $k \geq 2$ and also makes use of some additional elementary topological arguments.

Organization

In Section 2 we give some rather elementary preliminaries on surfaces and topology and recall a fundamental result from systolic geometry which are needed in our arguments. In Section 3 we provide necessary background from [Colding and Minicozzi 2015] on Colding–Minicozzi lamination theory of minimal surfaces with some control on the topology. Section 4 contains two weak chord-arc properties for minimal surfaces contained in small extrinsic balls of an ambient three-manifold. Our main result, Theorem 1.1, is proved first in the case $k = 1$ in Section 5, and then in Section 6 in the general case.

Acknowledgements Both authors would like to thank the Max Planck Institute for Mathematics in Bonn for support and excellent working conditions. We would also like to thank Fabian Henneke for explanations related to Example 1.3 and André Neves for asking us whether it is possible to find more than a single short curve. We are grateful to the referee for many detailed suggestions that greatly benefited the exposition and clarity of the paper.

2 Some preliminaries on surfaces and topology

In this section we recall some elementary and well-known facts about the topology of surfaces. We also recall some results from systolic geometry.

2.1 A result from systolic geometry

We will use the following result from systolic geometry, that relates the area and the k^{th} homology systole.

Theorem 2.1 ([Balacheff et al. 2012, Theorem 1.2]; see also [Gromov 1996]) *Let $\eta\colon \mathbb{N} \to \mathbb{N}$ be a function such that*

$$\lambda := \sup_{\gamma} \frac{\eta(\gamma)}{\gamma} < 1.$$

Then there exists a constant C_λ such that for every closed, orientable Riemannian surface Σ of genus γ, we have

$$\mathrm{sys}^h_{\eta(\gamma)}(\Sigma) \le C_\lambda \frac{\log(\gamma+1)}{\sqrt{\gamma}} \sqrt{\mathrm{area}(\Sigma)}.$$

Recall that a nonorientable surface Σ can be written as a connected sum $\Sigma = \Sigma_1 \# \Sigma_2$, with Σ_1 closed, orientable and Σ_2 diffeomorphic to $\mathbb{RP}^2$ or $\mathbb{RP}^2 \# \mathbb{RP}^2$. If we replace Σ_2 by a disk, Theorem 2.1 easily implies the following for nonorientable surfaces.

Corollary 2.2 *Let η and λ be as above. Then there is a constant C_λ such that for every closed, nonorientable surface of nonorientable genus δ, we have*

$$\mathrm{sys}^h_{\eta(\gamma_\delta)}(\Sigma) \le C_\lambda \frac{\log(\gamma_\delta+1)}{\sqrt{\gamma_\delta}} \sqrt{\mathrm{area}(\Sigma)},$$

where $\gamma_\delta = \lfloor (\delta-1)/2 \rfloor$.

We will only use the following consequence of these results.

Corollary 2.3 *Let (Σ_j) be a sequence of surfaces with $-\chi(\Sigma_j) \to \infty$. If $\mathrm{area}(\Sigma_j) = O((-\chi(\Sigma_j))^\alpha)$ for some $0 \le \alpha < 1$, then, for any $k \in \mathbb{N}$,*

$$\mathrm{sys}^h_k(\Sigma_j) \to 0 \quad \text{as } j \to \infty.$$

To put this into context, notice that the Choi–Wang bound [Choi and Wang 1983] implies, for a closed, embedded, orientable, minimal surface Σ, that

$$\operatorname{area}(\Sigma) \leq C(\operatorname{genus}(\Sigma)+1),$$

where $C = C(k)$, if $\operatorname{Ric}(M) \geq k > 0$.

2.2 Some elementary facts about the topology of surfaces

Lemma 2.4 *Let Σ be a closed surface and $c \subset \Sigma$ a simple closed curve. Then $[c] \neq 0 \in H_1(\Sigma; \mathbb{Z}/2\mathbb{Z})$ if and only if c is nonseparating.*

Proof If c is separating, then $[c] = 0$ in $H_1(\Sigma, \mathbb{Z}/2\mathbb{Z})$. On the other hand, if c is nonseparating, there is a curve d such that $|c \cap d| = 1$. In particular, from the intersection pairing, $[c] \neq 0 \in H_1(\Sigma, \mathbb{Z}/2\mathbb{Z})$. □

Lemma 2.5 *Let Σ be a closed surface and let $c_1, \dots, c_k \subset \Sigma$ be simple closed curves. Assume that $c_i \subset \bigcup_{i=1}^k B_i$ for pairwise disjoint balls B_i. Then, for a simple closed curve $d \subset \Sigma$ such that $0 \neq [d] \in \langle [c_1], \dots, [c_k] \rangle \subseteq H_1(\Sigma, \mathbb{Z}/2\mathbb{Z})$ and $d \subset M \setminus \bigcup_{i=1}^k B_i$, we have that d is separating in $\Sigma \cap \bigl(M \setminus \bigcup_{i=1}^k B_i\bigr)$.*

Remark 2.6 In the case $k = 1$, this states that if c is nonseparating in Σ and $c \subset B$, then any curve in $M \setminus B$ that is homologous to c separates in $M \setminus B$.

Proof Write $B = \bigcup_{i=1}^k B_i$. If d is nonseparating in $\Sigma \setminus B$ we can find a closed curve $e \subset \Sigma \setminus B$ that intersects d exactly once. On the other hand, $c_i \cap e = \varnothing$ for any e, but this is impossible since d is in the span of the c_i. □

Definition 2.7 Let $\Sigma \subset M$ be an embedded surface, $x \in M$ and $r > 0$. We say that $c \colon S^1 \to \Sigma$ is *contractible on scale r at x* if there is a disk $\phi \colon D \to B(x,r) \cap \Sigma$ with $\phi|_{S^1} = c$. If there is some $x \in M$ such that c is contractible on scale r at x, we say that c *is contractible on scale r*.

At this point it is worth recalling the following version of the maximum principle for minimal surfaces.

Theorem 2.8 *Let N be a compact manifold with mean-convex boundary and $\Sigma \subset N$ be a minimal surface (possibly with boundary). Then $(\Sigma \setminus \partial\Sigma) \cap \partial N = \varnothing$ or $\Sigma \subseteq \partial N$.*

In the context of Definition 2.7 this has the following consequence, that we will make use of frequently.

Lemma 2.9 *Choose $r > 0$ such that any ball $B(x,s) \subset M$ with $s \leq r$ and $x \in M$ is mean convex. If $\Sigma \subset M$ is a complete minimal surface and $c \subset \Sigma$ is a simple closed curve that is contractible on scale r (at x), then c is contractible on scale t (at x) for any $t \leq r$ such that $c \subset B(x,t)$.*

Lemma 2.10 *Let $\Sigma \subset M$ be a surface, $x \in \Sigma$ and $R > 0$. If any curve $d \colon S^1 \to B^{\Sigma}(x,R)$ with $\operatorname{length}(d) \leq 3R$ is contractible on scale r at x, then any curve $c \colon S^1 \to B^{\Sigma}(x,R)$ is contractible on scale r at x.*

Proof Let $c\colon S^1 \to B^\Sigma(x,R)$ be a loop. Choose a subdivision

$$0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$$

of [0, 1] such that

$$\operatorname{length}(c_{|[t_i,t_{i+1}]}) \le R.$$

Fix curves $d_i\colon I \to B^\Sigma(x,R)$ with $d_i(0) = x$ and $d_i(1) = c(t_i)$ and such that

$$\operatorname{length}(d_i) \le R.$$

We can then write

$$c = (c_{|[t_{k-1},t_k]} * d_{k-1}) * (\bar{d}_{k-1} * c_{|[t_{k-2},t_{k-1}]} * d_{k-2}) * \cdots * (\bar{d}_2 * c_{|[t_1,t_2]} * d_1) * (\bar{d}_1 * c_{|[t_0,t_1]}),$$

which implies the assertion. □

Since the Hurewicz homomorphism $\pi_1(B^\Sigma(x,R),x) \to H_1(B^\Sigma(x,R);\mathbb{Z})$ as well as the map

$$H_1(B^\Sigma(x,R);\mathbb{Z}) \to H_1(B^\Sigma(x,R);\mathbb{Z}/2\mathbb{Z})$$

are surjective, we immediately get the following corollary.

Corollary 2.11 *Let Σ be a surface, $x \in \Sigma$ and $R > 0$. Then the group $H_1(B^\Sigma(x,R);\mathbb{Z}/2\mathbb{Z})$ is generated by curves of length at most $3R$.*

Lemma 2.12 *Let Σ be a closed surface and $\pi\colon \hat{\Sigma} \to \Sigma$ a covering. Consider a simple closed curve $c \subset \Sigma$ and its preimage $\hat{c} = \pi^{-1}(c) \subset \hat{\Sigma}$. If c is separating, then also $\hat{c}$ is separating.*

Proof If c is separating, we can write $\Sigma \setminus c = \Sigma_+ \cup \Sigma_-$ with connected surfaces $\Sigma_\pm$. Moreover, there is a function $f\colon \Sigma \to [-1,1]$ such that $\{f = 0\} = c$ and $\Sigma_\pm = \{f \gtrless 0\}$. We can then consider the lifted function $\hat{f} = f \circ \pi$, which clearly satisfies $\{\hat{f} = 0\} = \hat{c}$. Therefore, $\hat{c}$ separates $\hat{\Sigma}$ into $\hat{\Sigma}_- = \{\hat{f} < 0\}$ and $\hat{\Sigma}_+ = \{\hat{f} > 0\}$. □

It will be important to keep in mind that the domains $\hat{\Sigma}_\pm$ might be disconnected and $\hat{c}$ is potentially not the boundary of a compact subsurface.

3 Background on Colding–Minicozzi lamination theory

Colding and Minicozzi developed a theory that describes how minimal surfaces of uniformly bounded genus in an ambient three-manifold can degenerate in the absence of curvature bounds. We use this section to provide a very brief introduction to those parts of their theory that will be relevant in the present paper. We will focus here on the case of planar domains, since this is sufficient for our purposes.

We start by recalling the definition of a lamination.

Definition 3.1 [Colding and Minicozzi 2004e, Appendix B] (1) A codimension-one lamination on a three-manifold M is a collection $\mathscr{L}$ of smooth disjoint surfaces $\Gamma \subset M$, the so-called leaves, such that $\bigcup_{\Gamma \in \mathscr{L}} \Gamma$ is closed. Furthermore, for each point $x \in M$, there exists an open neighborhood U of x and a coordinate chart (U, Φ) with $\Phi(U) \subset \mathbb{R}^3$ so that in these coordinates the leaves in $\mathscr{L}$ pass through $\Phi(U)$ in slices of the form $(\mathbb{R}^2 \times \{t\}) \cap \Phi(U)$.

(2) A foliation is a lamination for which one has $M = \bigcup_{\Gamma \in \mathscr{L}} \Gamma$, ie the union of the leaves is all of M.

(3) A minimal lamination is a lamination whose leaves are minimal.

(4) A Lipschitz lamination is a lamination for which the chart maps Φ are Lipschitz.

Given any sequence of minimal surfaces $\Sigma_j \subset M$, we consider the *singular* or *blow-up set*

$$\mathscr{S} = \left\{ z \in M \;\middle|\; \inf_{\delta > 0} \sup_j \sup_{B(z,\delta)} |A^{\Sigma_j}| = \infty \right\},$$

ie the points z where the curvature blows up. Up to taking a subsequence one can always pass to a limit

$$\Sigma_j \to \mathscr{L} \text{ in } M \setminus \mathscr{S},$$

where the convergence is in $C^{0,\alpha}$ and the limit lamination is a minimal Lipschitz lamination.

In the case of minimal surfaces $\Sigma_j \subset B(0, R_j) \subset \mathbb{R}^3$ with bounded genus and $\partial \Sigma_j \subset \partial B(0, R_j)$ one can always extract a subsequence such that either $R_j \to \infty$ or with R_j bounded. In the former case one can reach much stronger conclusions on the structure of the limit lamination, see eg the example in [Colding and Minicozzi 2004a]. Since we only deal with the local case, ie $R_j = R$ is fixed, which in general only allows us to draw significantly weaker conclusions about the structure of the limit lamination, we do not discuss stronger conclusion valid in the global case.

We first consider the case when the Σ_j are disks. Colding and Minicozzi [2004b; 2004c; 2004d; 2004e] proved that every embedded minimal disk is either a graph of a function or is a double spiral staircase where each staircase is a multivalued graph. More precisely, they show that if the curvature blows up at some point (and thus the surface is not a graph), then the surface is a double spiral staircase like the helicoid; see also [Colding and Minicozzi 2004e, Theorem 0.1].

Below we also want to deal with the case where Σ_j are more general domains than disks, namely, so-called uniformly locally simply connected (in short: ULSC) domains.

A sequence of minimal surfaces $\Sigma_j \subset M$ is called *uniformly locally simply connected*[1] if given any compact $K \subset M$ there is some $r > 0$ such that

$$\Sigma_j \cap B(x, r) \text{ consists of disks for any } x \in K.$$

[1] We remark that this is stronger than the definition of Colding–Minicozzi in the case of nonplanar domains.

Moreover, we define

$$\mathcal{S}_{\mathrm{ulsc}} := \{z \in \mathcal{S} \mid \Sigma_j \text{ is ULSC near } z\}.$$

The main local structural result we need for (not necessarily globally planar or bounded genus) ULSC sequences concerns so-called collapsed leaves, whose existence is described in the next lemma. We assume that $\Sigma_j \to \mathcal{L}'$ in $M \setminus \mathcal{S}$, where Σ_j is a ULSC sequence.

Lemma 3.2 [Colding and Minicozzi 2015, Lemma II.2.3] *Given a point $x \in \mathcal{S}_{\mathrm{ulsc}}$, there exists $r_0 > 0$ such that $B(x, r_0) \cap \mathcal{L}'$ has a component Γ_x whose closure $\overline{\Gamma}_x$ is a smooth minimal graph containing x and with boundary in $\partial B(x, r_0)$ (so x is a removable singularity for Γ_x).*

We want to emphasize that while [Colding and Minicozzi 2015] starting at the end of Section II.1 makes the general assumption to be in the global case $R_j \to \infty$ this does not apply to everything contained in the following sections. In particular a look at the proof of Lemma II.2.3 show that this does not make use of this assumption. Similarly, an inspection of the arguments shows the statements from Proposition 3.3 below are valid without this assumption.

The leaves of the limit lamination $\mathcal{L}'$ may not be complete. A special type of incomplete leaves are collapsed leaves. A leaf Γ of $\mathcal{L}'$ is *collapsed* if there exists some $x \in \mathcal{S}_{\mathrm{ulsc}}$ so that Γ contains the local leaf Γ_x given by Lemma 3.2; see Definition II.2.9 in [Colding and Minicozzi 2015].

Until the end of the section, we assume that the ambient manifold is given as $N = M \setminus \{x_1, \dots, x_k\}$, where M is complete and $x_i \in M$. In order to state the key structural results on collapsed leave we need to introduce some notation. Given a leaf $\Gamma \subset \mathcal{L}'$ we fix a point $x \in \Gamma$ and write

$$\Gamma_{\mathrm{clos}} = \bigcup_{R>0} \overline{B^{\Sigma}(x, R)},$$

where the closure is taken in N.

Proposition 3.3 [Colding and Minicozzi 2015, Section II.3] *Each collapsed leaf Γ of $\mathcal{L}'$ has the following properties*:

(1) *Given any $y \in \Gamma_{\mathrm{clos}} \cap \mathcal{S}_{\mathrm{ulsc}}$, there exists $r_0 > 0$ such that the closure in M of each component of $\Gamma \cap B(y, r_0)$ is a compact embedded disk with boundary in $\partial B(y, r_0)$. Furthermore, $\Gamma \cap B(y, r_0)$ must contain the component Γ_y given by Lemma 3.2, and Γ_y is the only component of $\Gamma \cap B(y, r_0)$ with y in its closure.*

(2) *Γ is a limit leaf.*

(3) *Γ extends to a complete minimal surface away from $\{x_1, \dots, x_k\}$.*[2]

[2] In other words, there is a Γ' containing Γ such that if a geodesic in Γ' cannot be extended, it limits to some x_i.

The sequences Σ_j appearing in this manuscript will essentially all be ULSC. This is equivalent to the fact that the singular set $\mathcal{S}$ is given by $\mathcal{S}_{\text{ulsc}}$, ie $\mathcal{S} = \mathcal{S}_{\text{ulsc}}$. Although we will not directly apply the results for non-ULSC surfaces here, some of our arguments (in particular the proof of Lemma 5.12) are inspired by those in [Colding and Minicozzi 2015] for this case.

4 Chord arc properties

We need two weak chord-arc properties for minimal surfaces contained in small extrinsic balls of an ambient three-manifold. Given $x \in M$ and $r > 0$, we write $B(x, r)$ for the metric ball in (M, g). If $z \in \Sigma$ and $r > 0$, we denote by $B^{\Sigma}(z, r)$ the metric ball of radius r in Σ with respect to the induced Riemannian metric.

Let (M, g) be a closed Riemannian three-manifold. Let $R_0 > 0$ be small, so that the metric in all balls of radius R_0 in M is sufficiently close to the Euclidean metric after rescaling to unit size. We will indicate in the proof when making specific assumptions on how small R_0 has to be but point out that all of this will be only dependent on the geometry of M. A first assumption on R_0 is that all balls in M of radius $r \leq R_0$ are mean convex so that Lemma 2.9 will be useful.

We consider minimal embedded disks Σ in $B(x_0, R_0)$ for some $x_0 \in M$. We write

$$\Sigma_{x_0,r} \subseteq \Sigma \cap B(x_0, r)$$

for the connected component of $\Sigma \cap B(x_0, r)$ that contains x_0.

Theorem 4.1 *There are $R_0 > 0$ sufficiently small and $\alpha > 0$ (both depending only on M) such that for any embedded minimal disk $\Sigma \subset B(x_0, R_0) \subset M$ with $x_0 \in \Sigma$, the following holds. For any $R > 0$ with $B^{\Sigma}(x_0, R) \subset \Sigma \setminus \partial\Sigma$, we have $\Sigma_{x_0,\alpha R} \subset B^{\Sigma}(x_0, R/2)$.*

Remark 4.2 (1) This result is proven in [Colding and Minicozzi 2008, Proposition 1.1] for minimal disks in $\mathbb{R}^3$. The proof applies here as well with one minor modification in [loc. cit., Proposition 3.4] that we explain below.

(2) The following property of minimal surfaces in $\mathbb{R}^3$ was used in [loc. cit., Proposition 3.1]. Since minimal surfaces in $\mathbb{R}^3$ have nonpositive curvature it follows that any intrinsic ball B in a minimal disk $D \subset \mathbb{R}^3$ with $B \cap \partial D = \varnothing$ is itself a topological disk. This may not apply in our setting since a minimal surface $\Sigma \subset M$ can have points of positive curvature provided M has such points.

Proof The proof of [Colding and Minicozzi 2008, Proposition 1.1] applies to this setting as well with a few minor modifications. We will provide an outline of the overall proof of [loc. cit., Proposition 1.1] and indicate necessary alterations for the proof of Theorem 4.1.

The proof of [loc. cit., Proposition 1.1] consists of the following two main steps.

Step 1 Colding and Minicozzi first provide this result under the additional assumptions that Σ is compact and that $\partial\Sigma$ is contained in the boundary of an extrinsic ball:

Proposition 4.3 [Colding and Minicozzi 2008, Proposition 2.1] *Let $\Sigma \subset \mathbb{R}^3$ be a compact embedded minimal disk. There exists a constant $\delta_2 > 0$ independent of Σ such that if $x \in \Sigma$ and $\Sigma \subset B_R(x)$ with $\partial\Sigma \subset \partial B_R(x)$, then the component $\Sigma_{x,\delta_2 R}$ of $B_{\delta_2 R}(x) \cap \Sigma$ containing x satisfies*

$$\Sigma_{x,\delta_2 R} \subset B^\Sigma\big(x, \tfrac{1}{2}R\big).$$

The benefit of the above-mentioned additional assumptions is that the authors can directly apply previous results which were provided by Colding and Minicozzi [2004b; 2004c; 2004d; 2004e]. Since this step applies to our setting as well for R_0 chosen sufficiently small depending only on the geometry of M, we do not provide more details, but refer the interested reader to Chapter 2 of [Colding and Minicozzi 2008], which consists of the proof of Proposition 2.1.

Step 2 Colding and Minicozzi [2008, Chapter 3] remove the additional assumptions from Step 1, ie that Σ is compact and that $\partial\Sigma$ is contained in the boundary of an extrinsic ball. In order to formulate the key ingredient for Step 2, the authors define a weak chord arc property for intrinsic balls:

Definition 4.4 An intrinsic ball $B^\Sigma(x,s) \subset \Sigma \setminus \partial\Sigma$ is said to be δ-weakly chord arc for some $\delta > 0$ if we have $\Sigma_{x,\delta s} \subset B^\Sigma(x, s/2)$.

Furthermore, they need the following result.

Lemma 4.5 [Colding and Minicozzi 2008, Lemma 3.6] *There exists $C_0 > 1$ such that for every $C_a > 0$, there exists $\tau > 0$ such that if $B^\Sigma(x_1, C_0)$ and $B^\Sigma(x_2, C_0)$ are disjoint intrinsic balls in $\Sigma \setminus \partial\Sigma$ with*

$$\sup_{B^\Sigma(x_1,C_0)\cup B^\Sigma(x_2,C_0)} |A|^2 \le C_a \quad \text{and} \quad |x_1 - x_2| < \tau,$$

then for $i = 1, 2$ we have

$$B_{10}(x_i) \cap \partial B^\Sigma(x_i, 11) = \varnothing.$$

The key result is as follows, where δ_2 is the constant given in [loc. cit., Proposition 2.1], see Proposition 4.3 above:

Proposition 4.6 [Colding and Minicozzi 2008, Proposition 3.4] *Assume that g is a metric that is sufficiently close (depending only on M) to the Euclidean metric on B_2 and let $\Sigma \subset (B_1(0), g)$ be an embedded minimal disk. There exists a constant $C_b > 1$ independent of Σ such that if $B^\Sigma(y, C_b\, R_0) \subset \Sigma \setminus \partial\Sigma$ is an intrinsic ball and every intrinsic subball $B^\Sigma(z, R_0) \subset B^\Sigma(y, C_b\, R_0)$ is δ_2-weakly chord arc, then, for every $s \le 5\, R_0$, the intrinsic ball $B^\Sigma(y, s)$ is δ_2-weakly chord arc.*

Proof The detailed proof can be found on pages 229–231 of [loc. cit.]. For convenience of the reader we outline the main steps and emphasize where attention is required to apply the arguments in our setting.

After rescaling and translating Σ, we can assume that $R_0 = 1$ and $y = 0$. It suffices to prove the following claim, since applying [loc. cit., Proposition 2.1] to $\Sigma_{0,5}$ then establishes [loc. cit., Proposition 3.4].

Claim *There exists n such that*

$$\Sigma_{0,5} \subset B^{\Sigma}(0, (6n+3)\, C_0), \tag{4.7}$$

where $C_0 > 1$ is given by Lemma 4.5.

Colding and Minicozzi prove the claim, ie (4.7), by arguing by contradiction. So suppose that (4.7) fails for some large n. Consequently, there exists a curve

$$\sigma \subset \Sigma_{0,5} \subset B_5 \tag{4.8}$$

from 0 to a point in $\partial B^{\Sigma}(0, (6n+3)\, C_0)$. For $i = 1, \dots, n$, fix points

$$z_i \in \partial B^{\Sigma}(0, 6i\, C_0) \cap \sigma.$$

It follows that the intrinsic balls $B^{\Sigma}(z_i, 3\, C_0)$

- are disjoint, and
- have centers in $B_5 \subset \mathbb{R}^3$.

Note, however, that these are not guaranteed to be topological disks in the presence of some positive ambient curvature. We will return to this issue in a moment.

Since the n points $\{z_i\}$ are all in the Euclidean ball $B_5 \subset \mathbb{R}^3$, there exist integers i_1 and i_2 with

$$0 < |z_{i_1} - z_{i_2}| < C'\, n^{-1/3}. \tag{4.9}$$

Now we use that each intrinsic ball of radius one about any z_i is δ-weakly chord arc by the assumption that every intrinsic subball $B^{\Sigma}(z, R_0) \subset B^{\Sigma}(y, C_b\, R_0)$ is δ_2-weakly chord arc. Recall that this means that

$$\Sigma_{z_{i_j},\delta} \subseteq B^{\Sigma}\big(z_{i_j}, \tfrac{1}{2}\big) \quad \text{for } j = 1, 2. \tag{4.10}$$

By construction, the two intrinsic balls $B^{\Sigma}\big(z_{i_j}, \frac{1}{2}\big)$ for $j = 1, 2$ are disjoint, which implies that also the surfaces $\Sigma_{z_{i_j},\delta}$ for $j = 1, 2$ need to be disjoint.

We now return to the issue of intrinsic balls not necessarily being disks. Let Σ' be a component of $B(z_{i_j}, \delta) \cap \Sigma$ with $\Sigma' \cap \partial\Sigma = \varnothing$. Consider any simple closed curve $c \subset \Sigma'$. Since Σ is a disk, we know that c is contractible within Σ. But in this scenario, the maximum principle, cf Lemma 2.9, implies that c is contractible on scale δ. This in turn implies that c is contractible within Σ'. Since this applies to any simple closed curve in Σ', we find that Σ' must be a disk.

Thanks to (4.10) this applies to the components $\Sigma_{z_{i_j},\delta}$, proving that these are in fact topological disks with

$$\partial\Sigma_{z_{i_j},\delta} \subset \partial B_\delta(z_{i_j}).$$

Consequently, for n large enough, (4.9) implies that the components Σ_1 and Σ_2 of

$$B_{\delta/2}(z_{i_1}) \cap \Sigma$$

containing z_{i_1} and z_{i_2}, respectively, are compact and have

$$\partial\Sigma_i \subset \partial B_{\delta/2}(z_{i_1}). \tag{4.11}$$

The rest of the proof will remain unchanged in our setting. Therefore we just outline the strategy; for details we refer the reader to [Colding and Minicozzi 2008].

From (4.11) Colding and Minicozzi deduce a curvature bound on the intrinsic balls $B^{\Sigma}(z_{i_j}, \delta(2c))$; namely they have previously shown that if two disjoint embedded minimal disks with boundary in the boundary of a ball both come close to the center, then each has an interior curvature estimate. This curvature bound in turn implies that there exists a constant $r' = r'(\delta, c)$ such that for n sufficiently large, the intrinsic ball $B^{\Sigma}(z_{i_2}, 3\,r')$ can be written as a normal exponential graph of a function u over a domain Ω. Applying the Harnack inequality to u we obtain

$$\sup_{B^{\Sigma}(z_{i_1}, 3\,r')} u \le \widetilde{C}' n^{-1/3}.$$

For n large enough, this inequality guarantees that we can repeat the previous argument with z_{i_1} substituted by a point in $\partial B^{\Sigma}(z_{i_1}, r')$. Thus, by repeatedly combining the above curvature bound and the Harnack inequality, one can extend the curvature bound to larger intrinsic balls. Applying Lemma 4.5 then yields a contradiction. □

Thus the argument from (4.8) till (4.11) also works in our framework, whence the proof. □

We also need a related chord arc property for uniformly locally simply connected surfaces.

Theorem 4.12 *Let $\Sigma \subset B(x_0, R) \subset M$ be a minimal surface with $x_0 \in \Sigma$. Assume that there is an $r > 0$ such that $\Sigma \cap B(y, r)$ consists only of proper disks for any $y \in B(x_0, R - r)$. Then, given $k \in \mathbb{N}$ such that $k\,r \le R$, there is a $\beta_k > 0$ depending only on M such that if $B^{\Sigma}(x_0, \beta_k r) \cap \partial\Sigma = \varnothing$, then $\partial(\Sigma_{x_0,kr}) \subseteq \partial B(x_0, k\,r)$.*

Remark 4.13 (1) This result is stated in [Colding and Minicozzi 2015, Appendix B.1] with the uniformly locally simply connected assumption for intrinsic rather than extrinsic balls, ie it is assumed that all intrinsic balls of a fixed radius are disks.

(2) As already mentioned in Remark 4.2, in our setting intrinsic balls that are contained in a disk may not be disks themselves, which is why we use extrinsic balls in the uniformly locally simply connected assumption.

Proof The argument is analogous to the proof of [Colding and Minicozzi 2008, Proposition 3.4] with some changes that we now explain; compare also the proof of Theorem 4.1.

For simplicity we scale everything so that $r = 1$.

We can follow the argument in [loc. cit., Proposition 3.4] up to (3.20) and consider two disjoint intrinsic balls $B^\Sigma(z_{i_j}, 3C_0) \subset \Sigma \setminus \partial\Sigma$. We first consider the surfaces $\Sigma_{z_{i_j},1}$, which are disks by assumption. Note that clearly $B^\Sigma\big(z_{i_j}, \frac{1}{2}\big) \cap \partial\Sigma_{z_{i_j},1} = \varnothing$. Now we apply Theorem 4.1 with $R = \frac{1}{2}$. This gives that the surfaces

$$\Sigma_{z_{i_j},\alpha/2} \subset B^\Sigma\big(z_{i_j}, \tfrac{3}{2}C_0\big)$$

are disjoint, proper disks, where α is given by Theorem 4.1.

From here on we can again follow the argument in [loc. cit., Proposition 3.4]. □

5 Existence of one short curve

Throughout this section let (M, g) be a closed three-manifold with positive Ricci curvature. In order to prove Theorem 1.1, we want to argue by contradiction. Therefore, we study properties of a sequence $\Sigma_j \subset (M, g)$ of closed, embedded minimal surfaces with $\mathrm{sys}^h_k(\Sigma_j) \geq l_0 > 0$. More precisely, we will be concerned with a limit lamination

$$\Sigma_j \to \mathscr{L} \quad \text{in } M \setminus \mathscr{S}$$

of such a sequence. For the sake of clarity, and since we need the corresponding arguments anyways, we will focus first on the case $k = 1$, ie the first homology systole, and explain the necessary extensions to handle the general case afterwards.

5.1 The singular set is nonempty

We start with a simple observation concerning the maximum of the curvature of a sequence of minimal surfaces in M with unbounded genus. It says, that for a sequence of minimal surfaces of unbounded genus $\Sigma_j \subset M$, we necessarily have $\mathscr{S} \neq \varnothing$. This works without any assumption on the systole.

Lemma 5.1 *Let $\Sigma_j \subset (M, g)$ be a sequence of closed, embedded minimal surfaces with $\chi(\Sigma_j) \to -\infty$. Then there is a sequence of points $z_j \in \Sigma_j$ such that $|A^{\Sigma_j}|^2(z_j) \to \infty$.*

Proof Assume that there is a constant $C > 0$, such that

$$\sup_j \sup_{\Sigma_j} |A^{\Sigma_j}|^2 \leq C. \tag{5.2}$$

By scaling we may for simplicity assume that the sectional curvature satisfies $|\sec(M)| \leq 1$. Thus, by minimality and the theorem of Gauss–Bonnet, the total curvature satisfies

$$\int_{\Sigma_j} |A^{\Sigma_j}|^2 \, d\mu_{\Sigma_j} = -2\int_{\Sigma_j} (K^{\Sigma_j} - \sec(T_x\Sigma_j))\, d\mu_{\Sigma_j}(x) \geq 4\pi|\chi(\Sigma_j)| - 2\,\mathrm{area}(\Sigma_j). \tag{5.3}$$

On the other hand we have

$$\int_{\Sigma_j} |A^{\Sigma_j}|^2 d\mu_{\Sigma_j} \le C\,\mathrm{area}(\Sigma_j) \tag{5.4}$$

by assumption. Combining (5.3) and (5.4), we obtain

$$4\pi|\chi(\Sigma_j)| \le (C+2)\,\mathrm{area}(\Sigma_j).$$

By assumption the left-hand side tends to infinity, therefore we find that

$$\mathrm{area}(\Sigma_j) \to \infty \quad \text{as } \; j \to \infty.$$

We consider the universal covering $\pi: \tilde{M} \to M$, where $\tilde{M}$ is compact by the Bonnet–Myers theorem. Clearly, the minimal surfaces

$$\hat{\Sigma}_j := \pi^{-1}(\Sigma_j)$$

also satisfy the pointwise curvature bound (5.2) and have diverging area,

$$\mathrm{area}(\hat{\Sigma}_j) \to \infty. \tag{5.5}$$

The pointwise curvature bound (5.2) allows us to pass to a subsequence (not relabeled) such that

$$\hat{\Sigma}_j \to \mathscr{L} \text{ in } C^{0,\alpha}(\tilde{M}),$$

where $\mathscr{L}$ is a Lipschitz lamination, whose leaves are smooth, complete minimal surfaces. Moreover, since $\mathrm{area}(\Sigma_j) \to \infty$, we can conclude that there needs to be at least one leaf Γ with stable universal cover, which also implies that Γ is compact, hence diffeomorphic to S^2 thanks to [Fischer-Colbrie and Schoen 1980] and [Schoen and Yau 1983]. For the convenience of the reader we include the argument here following the proof of [Chodosh et al. 2017, Theorem 1.3].

By passing to another subsequence and using (5.5) we find that there has to be a point $p \in \Gamma$ such that

$$\liminf_{j\to\infty} \mathrm{area}(\hat{\Sigma}_j \cap B(p,r)) \to \infty \quad \text{for any } \; r > 0.$$

Since the $\hat{\Sigma}_j$ are embedded and by the curvature bound (5.2), this implies that for j sufficiently large and $r > 0$ sufficiently small (but only depending on the ambient geometry and the curvature bound), $\hat{\Sigma}_j \cap B(p,r)$ is given as the union of $n(j) \to +\infty$ graphical components over $\Gamma \cap B(p,r)$. Let $U \subset \tilde{\Gamma}$ be a bounded and simply connected subset of the universal cover $\tilde{\Gamma}$ of Γ with $\tilde{p} \in U$, where $\tilde{p}$ projects to p. Using the curvature bound, a covering argument and the standard elliptic theory we find that for j sufficiently large we can find at least two functions $v_{1,j}, v_{2,j}$ (out of the lifts of the $n(j)$ components above) defined on U such that the graphs define disjoint minimal surfaces over U and $\inf|v_{2,j} - v_{1,j}| \to 0$. Using the Harnack inequality we find that $w_j = \inf|v_{2,j} - v_{1,j}|^{-1}(v_{2,j} - v_{1,j})$ converges to a nontrivial signed solution of the Jacobi equation, hence U is stable. Since this applies to any such U it follows that $\tilde{\Gamma}$ is stable.

It follows from [Fischer-Colbrie and Schoen 1980] and [Schoen and Yau 1983] that for any disk $D \subset \tilde{\Gamma}$ and any $z \in D$ we have that

$$d(x, \partial D) \leq \frac{2\pi\sqrt{2}}{\sqrt{3\kappa_0}},$$

where $\mathrm{Scal} \geq \kappa_0 > 0$ on D. Since this applies to any such disk it follows that $\tilde{\Gamma}$ is compact, hence a sphere. Since $\tilde{M}$ is simply connected, it does not contain any embedded real projective plane. Therefore, we need to have $\tilde{\Gamma} = \Gamma$. In particular, Γ is a closed, two-sided, stable minimal surface in $\tilde{M}$, which gives the desired contradiction. □

Remark 5.6 Under the additional assumption that $\mathrm{sys}^h(\Sigma_j) \geq l_0 > 0$, we could have used Corollary 2.3 instead of the theorem of Gauss–Bonnet to obtain that $\mathrm{area}(\Sigma_j)$ has to be unbounded. However, this relies on the assumption on the systole and is less elementary. We will exploit such an argument below, in the proof of the existence of multiple pinching curves.

5.2 Localized systole and contractibility radius I

We now start to aim for Theorem 1.1 for $k = 1$, ie we show that there is at least one homologically nontrivial curve that becomes arbitrarily short. By Lemma 5.1, in order to prove Theorem 1.1 using a contradiction argument invoking a limit lamination, we are forced to study the structure of a limit lamination of $(\Sigma_j)_{j\in\mathbb{N}}$ in the presence of a nonempty singular set. In this subsection we use the global positivity of the Ricci curvature to rule out rather general neck-pinch singularities under appropriate assumptions.

We now fix $r_0 > 0$ sufficiently small such that, firstly, the results from Section 4 apply in any ball $B(x, r_0)$ and, secondly, all balls $B(x, r) \subset M$ with $r \leq r_0$ have strictly mean-convex boundary.

For an embedded closed surface $\Sigma \subset M$ and a point $x \in M$ we write

$$C(x, r) = C^{\Sigma}(x, r) = \{c : S^1 \to \Sigma \cap B(x, r) \mid 0 \neq [c] \in \pi_1(\Sigma \cap B(x, r))\}.$$

Note that $c \in C(x, r)$ could still be globally contractible in Σ. We also write

$$C(x) = C^{\Sigma}(x) = C(x, r_0).$$

At this point, recall that the maximum principle Lemma 2.9 says that if $C(x, r) = \varnothing$ for some $r \leq r_0$, then $C(x, s) = \varnothing$ for any $s \leq r$.

Definition 5.7 We call

$$c(\Sigma) = \inf_{x\in M} \sup\{r > 0 \mid \pi_1(\Sigma \cap B(x, r), x) = 0\}$$

the *contractibility radius* of Σ, and

$$\mathrm{sys}_r(\Sigma) = \inf_{x\in M} \inf_{c\in C(x,r)} \mathrm{length}(c)$$

the *r-local systole* of Σ.

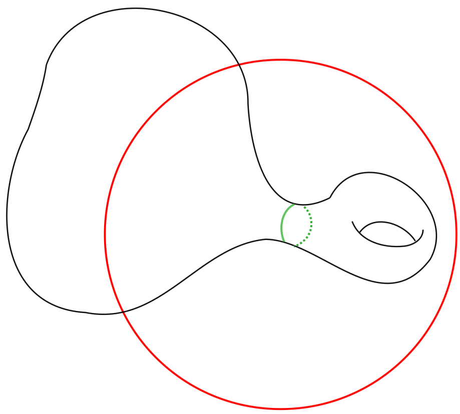


Figure 1: A surface with small r-local systole but not too small systole as the curve γ is globally contractible.

While the two definitions seem closely related, there is an important difference to be pointed out. Note that both of these are defined by looking at the intersection of Σ with extrinsic balls. However, the r-local systole still refers to intrinsic distances, ie we only localize extrinsically. Also observe that we have

$$\operatorname{sys}_r(\Sigma) = \inf_{x\in\Sigma} \operatorname{sys}(\Sigma \cap B(x,r)).$$

Note that both the contractibility radius and the r-local systole refer to extrinsic balls of radius r. One of the main challenges of our arguments is that in general there might be no strong connection between the systole (which is intrinsic) and the extrinsic r-local systole as indicated in Figure 1.

The goal of this section is to show that if $\Sigma \subset M$ is a minimal surface with homology systole bounded below by some constant l_0, then Σ is uniformly contractible on some potentially much smaller scale r_1 that depends on the ambient geometry and l_0 but not Σ otherwise; cf Proposition 5.21. Note that if $C(x) = \varnothing$ for any $x \in \Sigma$, then we automatically have that $c(\Sigma) \geq r_0$, where we recall that r_0 only depends on the ambient geometry. Similarly, in this case we have that $\operatorname{sys}_r(\Sigma) = \infty$ for any $r < r_0$. We are therefore mainly concerned with the case $C(x) \neq \varnothing$ for some $x \in M$.

The next lemma is our key scale-breaking argument indicated in Figure 2. Via the maximum principle we transfer some connectedness properties from the scale of certain singularities of a limit lamination to a definite scale. Given a very short and separating curve we show that both connected components of the complement have to extend a definite amount away.

Lemma 5.8 *Let $\Sigma \subset M$ be a closed minimal surface such that $\operatorname{sys}^h(\Sigma) \geq l_0$. There is an $l_1 = l_1(M, l_0) \leq \min(r_0/4, l_0/4)$ with the following property. Suppose that $\operatorname{sys}_{r_0}(\Sigma) \leq l_1$ and that $c \in C(x)$ is a simple closed curve for some $x \in M$ such that*

$$\operatorname{length}(c) \leq 2\operatorname{sys}_{r_0}(\Sigma).$$

Then $\Sigma \setminus c$ has two connected components Σ_1 and Σ_2, and these satisfy

$$\Sigma_i \cap \partial B(x, r_0) \neq \varnothing \quad \text{for } i = 1, 2.$$

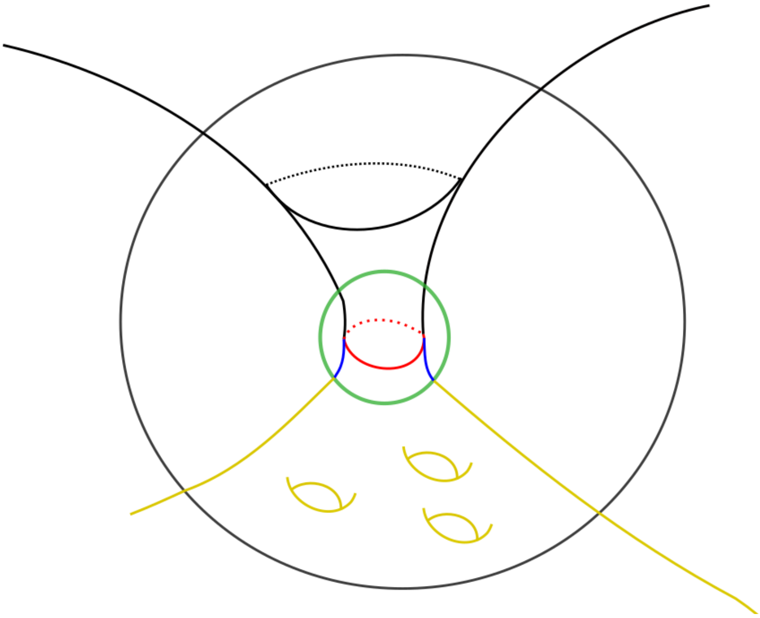


Figure 2: The proof of Lemma 5.8: Because $B(x,r)$ is mean convex for any $r \le r_0$ at least one component of $\Sigma \setminus c$ has to leave $B(x,r_0)$, say Σ_1. Once we can show that Σ_2 is forced to leave $B(x,9R_0)$ (still on the scale of c), the maximum principle gets us all the way to $\partial B(x,r_0)$.

Proof Write $R_0 = \operatorname{length}(c)/8$ and also note that our assumptions on c, l_0 and l_1 imply that

$$\operatorname{length}(c) \le 2l_1 \le \min\big(\tfrac{1}{2}r_0, \tfrac{1}{2}l_0\big)$$

thanks to our assumptions. Note the following two consequences of these choices.

Firstly, since the length of c is strictly below l_0, we find that c is homologically trivial. This means that $\Sigma \setminus c$ has two connected components, denoted by Σ_1 and Σ_2 with $\partial\Sigma_i = c$.

Secondly, note that since $x \in c$ we have that

(5.9) $$\partial\Sigma_i \subset B(x,4R_0) \subset B\big(x,\tfrac{1}{4}r_0\big).$$

We first show that these choices imply that there is no nontrivial topology on intrinsic scales below R_0. More precisely, we let $y \in \Sigma \cap B(x,r_0/2)$ and claim that there is a unique disk $D_y \subset \Sigma \cap B(x,r_0)$ with

$$B^{\Sigma}(y,R_0) \subset D_y \quad \text{and} \quad \partial D_y \subset \partial B^{\Sigma}(y,R_0).$$

This can be seen as follows. By Lemma 2.10, if there were a curve $\sigma \subset B^{\Sigma}(y,R_0)$ that is noncontractible on scale r_0 at x, we could find a simple closed curve $\sigma' \subset B^{\Sigma}(y,R_0)$ also noncontractible on scale r_0 at x, with

$$\operatorname{length}(\sigma') \le 3R_0 < \tfrac{1}{2}\operatorname{length}(c) \le \operatorname{sys}_{r_0}(\Sigma)$$

by our choice of the curve c, but this is impossible by the definition of the r_0-local systole. We conclude that any simple closed curve contained in $B^{\Sigma}(y,R_0)$ admits a filling disk contained in $\Sigma \cap B(x,r_0)$, from which the existence of D_y follows. If Σ is not a sphere, it follows immediately that such a disk is unique. In the case of Σ being a sphere there are two such disks in Σ. However, by the choice of r_0 and the maximum principle, these disks cannot both be entirely contained in $B(x, r_0)$.

It follows from Theorem 4.1 and the convex hull property, that we can find some small $\alpha > 0$ such that

$$\Sigma \cap B(y, \alpha R_0) \text{ consists of disks for any } y \in B\big(x, \tfrac{1}{2} r_0\big).$$

Now choose $k \in \mathbb{N}$ such that $k\alpha \geq 9$, and let $\beta_k > 1$ be given by Theorem 4.12. First assume that we can find $z \in \Sigma_i$ such that

$$B^{\Sigma}(z, \beta_k \alpha R_0) \cap \partial \Sigma_i = \varnothing. \tag{5.10}$$

Also assume that

$$B^{\Sigma}(z, \beta_k \alpha R_0) \subset B\big(x, \tfrac{1}{2} r_0\big), \tag{5.11}$$

since the conclusion otherwise follows from the maximum principle thanks to (5.9). Under these assumptions it follows from Theorem 4.12 that

$$\partial((B^{\Sigma}(z, \beta_k \alpha R_0))_{z, 9R_0}) \subset \partial B(z, 9R_0),$$

which clearly implies that

$$\Sigma_i \cap \partial B(x, 9R_0) \neq \varnothing,$$

since Σ_i is connected. Since on the other hand $\partial \Sigma_i \subset B(x, 4R_0)$, we then find from the maximum principle that

$$\Sigma_i \cap \partial B(x, r_0) \neq \varnothing.$$

Note that to go from one scale to the other scale the maximum principle is applied on all balls of radii between the two scales.

We still need to justify why we can assume (5.10). Take l_1 such that $\beta_k \alpha R_0 \leq 16 \beta_k \alpha l_1 \leq \frac{1}{12} r_0$. If with these choices (5.10) fails for any $z \in \Sigma_i$, we then have that

$$\operatorname{diam}(\Sigma_i) \leq 32 \beta_k \alpha l_1 \leq \tfrac{1}{6} r_0.$$

Suppose first that Σ_i is a disk. In this case the diameter estimate implies that c is contractible on scale r_0, contradicting our choice of c.

If Σ_i is not a disk it contains at least one nonseparating curve d, since $\partial \Sigma_i$ is connected. Thanks to the diameter estimate, Corollary 2.11 then implies that we can find a nonseparating curve d' having

$$\operatorname{length}(d') \leq 48 \beta_k \alpha R_0 < l_0,$$

contradicting the assumptions. □

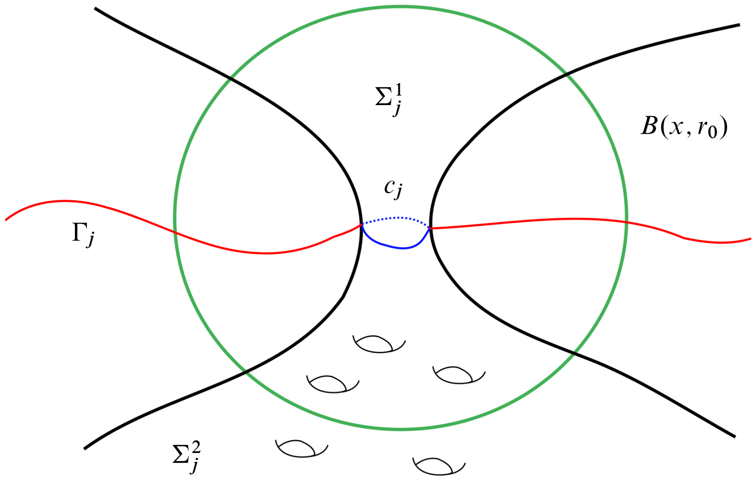


Figure 3: The construction in the proof of Lemma 5.12. The surface Σ_j is a good barrier for the Plateau problem: Both components of $\Sigma_j \setminus c_j$ extend out of $B(x, r_0)$ by Lemma 5.8.

Below, we solve a Plateau problem in $M \setminus \Sigma$ with boundary given by a curve c as above. In this situation, Lemma 5.8 implies that Σ is a useful barrier.

Lemma 5.12 *Given $l_0 > 0$ there is an $l_2 = l_2(M, l_0) > 0$ with the following property. Let $\Sigma \subset M$ be a closed minimal surface with $\mathrm{sys}^h(\Sigma) \geq l_0$. Then the r_0-local systole satisfies*

$$\mathrm{sys}_{r_0}(\Sigma) \geq l_2.$$

Note that this achieves two things simultaneously. Firstly, it shows that the systole is bounded away from zero if the homology systole is. Secondly, we also find that curves of controlled length which are very short, but potentially on a much smaller scale than the systole, can be contracted in an extrinsically controlled neighborhood.

This corresponds to the fact, already present in the argument for Lemma 5.8, that the proof handles two types of curves. On the one hand, applied to homologically trivial noncontractible curves, this implies that the homology systole of a sequence Σ_j tends to 0 if we can show that the systole does so. On the other hand, we will apply it to short curves bounding (large) disks in Σ_j in order to understand the convergence of Σ_j to a limit lamination.

Proof Let us first consider the case of M being simply connected. Afterwards we reduce the general case to this special case. We argue by contradiction and assume that we can find a sequence of minimal surfaces $(\Sigma_j)_{j \in \mathbb{N}}$ such that:

(1) All nonseparating curves in Σ_j have length at least l_0, ie $\mathrm{sys}^h(\Sigma_j) \geq l_0$.

(2) We have $\mathrm{sys}_{r_0}(\Sigma_j) \to 0$.

Up to taking a subsequence we then find $x \in M$, radii $r_j \to 0$, and simple closed curves $c_j \in C^{\Sigma_j}(x, r_j)$ such that

$$\operatorname{length}(c_j) \le 2\,\operatorname{sys}_{r_0}(\Sigma_j) \to 0.$$

Since M is simply connected, Σ_j separates M into two mean-convex connected components

$$M \setminus \Sigma_j = M_j^1 \cup M_j^2.$$

Clearly, once j is large enough such that $4\,\operatorname{sys}_{r_0}(\Sigma_j) \le l_0$, we have that c_j is null-homologous in the closure of both of these components.

In addition, we claim that at least one of M_j^1 and M_j^2 has the following property: If $\operatorname{length}(c_j) \le l_1$ from Lemma 5.8, then any minimal surface $S \subset M_j^i$ with $\partial S = c_j$ satisfies

$$S \cap \partial B(x, r_0) \neq \varnothing. \tag{5.13}$$

If this was not the case, we would find $S_j^1 \subset M_j^1 \cap B(x, r_0)$ and $S_j^2 \subset M_j^2 \cap B(x, r_0)$ such that $\partial S_j^i = c_j$. The surface $S_j = S_j^1 \cup S_j^2 \subset B(x, r_0)$ is a closed surface and separates $B(x, r_0)$ into two connected components. Moreover, (5.13) does not hold for S, so that one of these components is contained in $B(x, r_0 - \delta)$ for some small $\delta > 0$. By construction, this component contains a component of $\Sigma_j \setminus c_j$ contradicting Lemma 5.8.

Let M_j^1 be the component having property (5.13). By [Hardt and Simon 1979] we can find a stable minimal surface $\Gamma_j \subset M_j^1$ with $\partial \Gamma_j = c_j$ which minimizes area among all surfaces in M_j^1 which have boundary c_j. It follows from (5.13) that

$$\Gamma_j \cap \partial B(x, r_0) \neq \varnothing \tag{5.14}$$

for j sufficiently large. Moreover, by the curvature estimates [Schoen 1983], there is a constant C such that

$$\sup_j \sup_{\Gamma_j \cap (M \setminus B(x,r))} (r - r_j)^2 |A^{\Gamma_j}|^2 \le C$$

for any $r > r_j$, where r_j was defined such that $c_j \in C^{\Sigma_j}(x, r_j)$. In particular, we can pass to a subsequence such that

$$\Gamma_j \to \mathscr{L}$$

in $C^{0,\alpha}_{\text{loc}}(M \setminus \{x\})$, where $\mathscr{L}$ is a minimal Lipschitz lamination. Since Γ_j is stable, the same argument as in [Chodosh et al. 2017, Lemma 4.1] implies[3] that the lamination $\mathscr{L}$ extends to a lamination $\widetilde{\mathscr{L}}$ across $\{x\}$.

We claim that also $\widetilde{\mathscr{L}}$ has stable leaves thanks to a standard argument using the log cut-off trick. The details are as follows. Let Γ be a leaf of $\widetilde{\mathscr{L}}$ passing through x. Let $r > 0$ such that on $B(x, r)$ we find

[3] If all leaves are two-sided this follows immediately from [Chodosh et al. 2017, Proposition D.3]. The argument in [Chodosh et al. 2017, Lemma 4.1] explains how this can be assumed by passing to the double cover.

Lipschitz coordinates for the lamination $\widetilde{\mathscr{L}}$. We write Γ_x for the component of $\Gamma \cap B(x,r)$ passing through x. Fix some $0 < R < \min(r,1)$ and denote by $\eta_R\colon \Gamma \to [0,1]$ the log cut-off function given by

$$\eta_R(y) = \begin{cases} 0 & \text{for } d(x,y) \le R^2 \text{ and } y \in \Gamma_x, \\ 2 - \dfrac{\log(d(x,y))}{\log(R)} & \text{for } R^2 < d(x,y) \le R \text{ and } y \in \Gamma_x, \\ 1 & \text{else,} \end{cases}$$

where d denotes the distance function in M. Moreover, for a smooth vector field X in the normal bundle of Γ, we write

$$L_\Gamma X = \Delta_\Gamma^N X + F(X)$$

for the stability operator of Γ. Here, F is a linear operator of order zero with smooth coefficients. The precise form of F is irrelevant to the argument.

Let X be a smooth vector field in the normal bundle of Γ with compact support and such that $|X|, |\nabla_\Gamma X| \in L^\infty(\Gamma)$. Note that Γ_x has bounded area by construction and that $\eta_R = 1$ on $\Gamma \setminus (\Gamma_x \cap B(x,R))$. Since X has compact support and the coefficients of F are locally bounded, it thus it follows from the dominated convergence theorem that

$$\lim_{R\to 0} \int_\Gamma \langle F(\eta_R X), \eta_R X\rangle = \int_\Gamma \langle F(X), X\rangle. \tag{5.15}$$

We now turn to the gradient term. First, note that

$$\int_\Gamma |\nabla_\Gamma X|^2 - \int_\Gamma |\nabla_\Gamma(\eta_R X)|^2 = \int_{\Gamma_x \cap B(x,R)} |\nabla_\Gamma X|^2 - \int_{\Gamma_x \cap B(x,R)} |\nabla_\Gamma(\eta_R X)|^2. \tag{5.16}$$

Since

$$\lim_{R\to 0} \int_{\Gamma_x \cap B(x,R^2)} |\nabla_\Gamma X|^2 = 0 = \int_{\Gamma_x \cap B(x,R^2)} |\nabla_\Gamma(\eta_R X)|^2,$$

thanks to dominated convergence once again, we are left with considering the contribution on the annulus $(B(x,R) \setminus B(x,R^2)) \cap \Gamma_x$.

From the dominated convergence theorem we have that

$$\lim_{R\to 0} \int_{B(x,R)\cap \Gamma_x} |\nabla_\Gamma X^2| = 0,$$

which when combined with $|\eta_R| \le 1$ and $|\nabla_\Gamma(\eta_R X)|^2 \le 2\eta_R^2 |\nabla_\Gamma X|^2 + 2|\nabla_\Gamma \eta_R|^2 |X|^2$ implies that we only have to estimate

$$\int_{(B(x,R)\setminus B(x,R^2))\cap\Gamma_x} |X|^2 |\nabla_\Gamma \eta_R|^2.$$

To this end, we decompose into dyadic annuli via

$$\big(B(x,R) \setminus B(x,R^2)\big) \cap \Gamma_x = \bigcup_{i=1}^{\lceil \log(1/R)\rceil} A_i, \quad \text{where } A_i \subseteq \big(\bar{B}(x, e^i R^2) \setminus B(x, e^{i-1} R^2)\big) \cap \Gamma_x.$$

Note that since Γ_x has bounded area, we find from the monotonicity formula that

$$\text{area}(A_i) \le \text{area}(B(x, e^i R^2) \cap \Gamma_x) \le C e^{2i} R^4 \,\text{area}(\Gamma_x), \tag{5.17}$$

where C is a constant that depends on M and r. Also note that

$$|\nabla_\Gamma \eta_R| \le \frac{1}{e^{i-1} R^2 \log(1/R)} \quad \text{on } A_i. \tag{5.18}$$

We thus find from (5.17) and (5.18), also using that X is bounded, that

$$\int_{(\bar{B}(x,e^i R^2)\setminus B(x,e^{i-1}R^2))\cap\Gamma_x} |X^2||\nabla_\Gamma \eta_R|^2 \le \frac{C}{\log(1/R)^2} \sum_{i=1}^{\lceil \log(1/R)\rceil} \frac{e^{2i} R^4}{e^{2i-2} R^4} \le \frac{C}{\log(1/R)}.$$

We have thus shown that

$$\lim_{R\to 0} \int_\Gamma |\nabla_\Gamma(\eta_R X)|^2 = \int_\Gamma |\nabla_\Gamma X|^2. \tag{5.19}$$

Combining (5.15) and (5.19) and integrating by parts gives that

$$-\int_\Gamma \langle L_\Gamma X, X\rangle = \int_\Gamma (|\nabla_\Gamma X|^2 - \langle F(X), X\rangle) = \lim_{R\to 0} \int_\Gamma \big(|\nabla_\Gamma(\eta_R X)|^2 - \langle F(\eta_R X), \eta_R X\rangle\big) \ge 0,$$

since $\eta_R X$ has compact support in $\Gamma \setminus \{x\}$ which is stable. This proves that Γ is stable.

From (5.14), we find that there is a leaf $\bar{\Gamma} \subset \widetilde{\mathscr{L}}$ with

$$\bar{\Gamma} \cap \partial B(x, r_0) \ne \varnothing.$$

In particular, $\bar{\Gamma}$ is nonempty. Moreover, invoking [Fischer-Colbrie and Schoen 1980] and [Schoen and Yau 1983] once again, $\bar{\Gamma}$ is closed. Thus, since M is simply connected, we find that $\bar{\Gamma}$ is two-sided. Since M has positive Ricci curvature, this is a contradiction since $\bar{\Gamma}$ is a nonempty, two-sided, closed, stable minimal surface in M.

We now consider the general case in which we can assume that M is not simply connected. We can pass to the universal covering $\pi : \widetilde{M} \to M$, which is compact by the Bonnet–Myers theorem. In particular, there is a finite group G acting freely on M such that $M = \widetilde{M}/G$. We obtain minimal surfaces

$$\hat{\Sigma}_j = \pi^{-1}(\Sigma_j) \subset \widetilde{M}.$$

Since M has positive Ricci curvature, by the Frankel property, the surfaces $\hat{\Sigma}_j$ are connected.

We may assume that r_0 is chosen sufficiently small that

$$g(B(x, r_0)) \cap B(x, r_0) = \varnothing$$

for any $g \in G \setminus \{e\}$. If there is a noncontractible curve $c_j \subset \Sigma_j \cap B(x, r_0)$, with

$$\text{length}(c_j) \le l_0,$$

we may again assume that c_j is chosen to have properties (1) and (2) from above. It follows from our assumption that c_j is separating. Therefore, by Lemma 2.12, also $\hat{c}_j := \pi^{-1}(c_j)$ is separating. Moreover, by the choice of r_0, and recalling $l_0 \le r_0$, we see that $\hat{c}_j$ consists of $|G|$ disjoint, closed curves. We can now argue exactly as above and minimize area in the correct component of $\widetilde{M} \setminus \hat{\Sigma}_j$ relative to the boundary $\hat{c}_j$. Finally, by Lemma 5.8,[4] the limit lamination will be nonempty and we can conclude as in the first case. □

Remark 5.20 For curves that are noncontractible in $\Sigma \cap B(x,r)$ but contractible in Σ, it should be possible to extend Lemma 5.8 to bumpy metrics of positive scalar curvature. In this situation one component of $\Sigma_j \setminus c_j$ is a planar domain and one can write large parts of this component as a graph over Γ_j. This can then be used to construct a nontrivial Jacobi field on Γ.

Proposition 5.21 *For any $l_0 > 0$ there is an $r_1 > 0$ such that for any closed minimal surface $\Sigma \subset M$ with $\mathrm{sys}^h(\Sigma) \ge l_0$, we have for the contractibility radius that $c(\Sigma) \ge r_1$.*

Proof If we apply Lemma 5.12 to Σ we get some $l_2 > 0$ such that all curves in Σ of length at most l_2 are contractible in the intersection of Σ with some mean-convex ball $B(x, r_0)$. In particular, it follows from Lemma 2.10 that any intrinsic ball $B^{\Sigma}(z, l_2/3)$ is contained in some disk D_z with

$$B^{\Sigma}\left(z, \tfrac{1}{3}l_2\right) \subset D_z \subset \Sigma_j \cap B(z, r_0).$$

The claim now follows with $r_1 = \frac{1}{3}\alpha l_2$ from Theorem 4.1, where also $\alpha > 0$ is from Theorem 4.1. □

5.3 The first homology systole

At this stage we are in a position to prove the special case $k = 1$ of our main result.

Proof of Theorem 1.1 for $k = 1$ We argue by contradiction and assume that we have a sequence of minimal surfaces $\Sigma_j \subset M$ with $-\chi(\Sigma_j) \to \infty$ and

$$\mathrm{sys}^h(\Sigma_j) \ge l_0 > 0$$

for some positive constant l_0. Thanks to Proposition 5.21 we find that the sequence (Σ_j) is ULSC, ie

$$\Sigma_j \cap B(x, r_1) \text{ consists of disks for any } x \in M. \tag{5.22}$$

Clearly, after potentially decreasing r_1, property (5.22) holds for the surfaces $\hat{\Sigma}_j \subset \widetilde{M}$ as well. Therefore, it suffices to derive a contradiction from (5.22) if M is simply connected.

Thanks to (5.22) and [White 2015] (see also [Colding and Minicozzi 2015] which gives Lipschitz curves), we can pass to a subsequence such that

$$\Sigma_j \to \mathscr{L} \quad \text{in } M \setminus \mathscr{S}$$

outside the singular set $\mathscr{S}$ which is contained in a union of C^1-curves. It follows from Lemma 5.1, that $\mathscr{S} \neq \varnothing$. In particular, we can pick $x \in \mathscr{S}$ and the associated collapsed leaf Γ_x. Moreover, since Γ_x is

[4] We apply this to Σ_j and observe that this trivially implies (5.14) for $\hat{\Sigma}_j$.

a limit leaf of $\mathscr{L}$ it is stable by [Meeks et al. 2010]. It follows from Proposition 3.3 that Γ_x extends to a complete minimal surface $\overline{\Gamma}$ in M and that $\mathscr{S}\cap\overline{\Gamma}$ is discrete. In particular, $\overline{\Gamma}$ is also stable and by [Fischer-Colbrie and Schoen 1980] and [Schoen and Yau 1983], its universal cover is diffeomorphic to S^2. Since M is simply connected, it does not contain any one-sided surfaces and we conclude that $\overline{\Gamma}$ is a two-sided, closed, stable minimal surface in M. This is clearly a contradiction, since M has positive Ricci curvature. □

6 Existence of multiple short curves

We now proceed to the proof of the general case of Theorem 1.1.

Recall that we assume M to be a closed three-manifold with positive Ricci curvature. Assume we have a sequence of minimal surfaces $(\Sigma_j)_{j\in\mathbb{N}}$ in M with the following properties. There is a natural number $k\geq 2$ and for each $j\in\mathbb{N}$ a set $\{c_j^1,\dots,c_j^{k-1}\}$ of simple closed curves in Σ_j such that

(1) $\operatorname{length}(c_j^i)\to 0$ for $i=1,\dots,k-1$ as $j\to\infty$,

(2) $\operatorname{rank}\langle[c_j^1],\dots,[c_j^{k-1}]\rangle=k-1$ in $H_1(\Sigma_j;\mathbb{Z}/2\mathbb{Z})$,

(3) there is an $l_0>0$ such that if a closed curve $d_j\subset\Sigma_j$ has $\operatorname{length}(d_j)\leq l_0$, then $[d_j]\in\langle[c_j^1],\dots,[c_j^{k-1}]\rangle$.

Note that (3) allows for $[d_j]=0$.

By taking a subsequence we may assume that $c_j^i\subset B(x_i,s_j)$ for a sequence of radii $s_j\to 0$.

We now follow the same steps that we used for the case of the first homology systole, but have to deal with several new difficulties.

6.1 Additional points in the singular set

In a first step we show that the singular points arising from the curves c_j^i do not comprise the entire singular set. This is the analogue of Lemma 5.1. In contrast to Lemma 5.1 the argument in this case relies on the assumption on the homology systole.

Lemma 6.1 *We have* $\mathscr{S}\cap M\setminus\bigcup_{i=1}^{k-1}B(x_i,r_3)\neq\varnothing$ *for some* $r_3>0$.

Proof Assume that $\mathscr{S}\subset\{x_1,\dots,x_{k-1}\}$. By Corollary 2.3, we can assume that $\operatorname{area}(\Sigma_j)$ is unbounded. For simplicity, let us scale M to have $|\sec|\leq 1$, and write $B_s=\bigcup_{i=1}^{k-1}B(x_i,s)$. The monotonicity formula then implies

$$\begin{aligned}\operatorname{area}(\Sigma_j\cap(B_{2r_3}\setminus B_{r_3}))&=\operatorname{area}(\Sigma_j\cap B_{2r_3})-\operatorname{area}(\Sigma_j\cap B_{r_3})\\&\geq\left(\frac{4}{e^{2r_3}}-1\right)\operatorname{area}(\Sigma_j\cap B_{r_3})\geq\operatorname{area}(\Sigma_j\cap B_{r_3})\end{aligned}$$

if $r_3\leq\log(2)/2$, which in turn implies

(6.2) $$2\operatorname{area}(\Sigma_j\setminus B_{r_3})\geq\operatorname{area}(\Sigma_j\setminus B_{r_3})+\operatorname{area}(\Sigma_j\cap(B_{2r_3}\setminus B_{r_3}))\geq\operatorname{area}(\Sigma_j)\to\infty.$$

Now we can argue exactly as in Lemma 5.1 and obtain a limit lamination $\mathscr{L} \subset M \setminus \mathscr{S}$. Thanks to (6.2) we can conclude that $\mathscr{L}$ has a leaf with stable universal cover. We then use stability to extend it across the isolated singularities $\mathscr{S}$ and eventually use the log cut-off trick to conclude that this is still stable, which gives the desired contradiction. □

6.2 Localized systole and contractibility radius II

In a next step we prove that Σ_j is ULSC off the set $\{x_1, \dots x_{k-1}\}$.

Proposition 6.3 *Assume $(\Sigma_j)_{j\in\mathbb{N}}$ is as above. Given $r > 0$ there is an $r_2 = r_2(M, g, l_0, r)$ such that the contractibility radius satisfies $c\big(\Sigma_j \cap \big(M \setminus \bigcup_{i=1}^{k-1} B(x_i, 4r)\big)\big) \geq r_2$ for j sufficiently large.*

We want to follow the same strategy that we used to obtain Proposition 5.21, for which in turn Lemma 5.8 was the key input. Because of the short curves c_j^i, we need to be more careful in how we select the scale on which we work. Recall that in the case of Lemma 5.8 this was the smallest intrinsic scale of nontrivial topology. It turns out that there are two cases to consider in the more general case, depending on whether a potentially contradicting curve is separating or not. The instance of separating curves is more delicate, and we introduce some notation here related to this case. In order to find the correct scale, we define functions $l_j, f_j \colon \Sigma_j \to [0,\infty)$ as follows. For $x \in \Sigma_j$, we consider the set C_j' of curves in Σ_j given by

$$C_j'(x) := \{c \colon S_1 \to \Sigma_j \mid 0 \neq [c] \in \pi_1(\Sigma_j \cap B(x, r_0), x), 0 = [c] \in H_1(\Sigma_j; \mathbb{Z}/2\mathbb{Z})\}.$$

Note that we only take into account separating curves here. Then the first function is defined via

$$l_j(x) := \min\big\{1, \inf\{\text{length}(c) \mid c \in C_j'(x)\}\big\},$$

and f_j is a scale-invariant version of (the inverse of) this, incorporating the distance to the short curves c_j^i, given by

$$f_j(x) = l_j(x)^{-1} \operatorname{dist}(x, c_j^1 \cup \cdots \cup c_j^{k-1}).$$

Proof of Proposition 6.3 We argue by contradiction and assume that we can find a simple closed curve $d_j \subset \Sigma_j$ such that

(6.4) $$\text{length}(d_j) \to 0$$

and

(6.5) $$d_j \subset M \setminus \bigcup_{i=1}^{k-1} B(x_i, 2r),$$

but

(6.6) $$d_j \text{ is noncontractible on scale } r_0.$$

If we cannot find such a curve, the assertion follows from Theorem 4.1 combined with Lemma 2.10 and the convex hull property exactly as in the proof of Proposition 5.21.

Up to taking a subsequence, by (6.4) and (6.5) we can assume that

$$d_j \to y \in M \setminus \bigcup_{i=1}^{k-1} B(x_i, 2r).$$

Observe that (6.4) combined with the assumption (3) implies that $[d_j] \in \langle [c_j^1], \dots, [c_j^{k-1}] \rangle$ for j sufficiently large, which we simply assume to be the case from here on.

We have to distinguish the following two cases:

(a) The curve d_j is nonseparating.

(b) The curve d_j is separating, ie $[d_j] = 0$.

We start with case (a). In this case it follows from Lemma 2.5 that

$$\Sigma_j \cap \left(M \setminus \bigcup_{i=1}^{k-1} B(x_i, s_j) \right) = \Sigma_j^1 \cup \Sigma_j^2,$$

where now Σ_j^i are connected, disjoint minimal surfaces with

$$\partial \Sigma_j^i \subset d_j \cup \bigcup_{i=1}^{k-1} \partial B(x_i, s_j).$$

Since d_j is nonseparating in Σ_j, it follows immediately that

$$\Sigma_j^i \cap \partial B(y, r_0) \neq \varnothing \tag{6.7}$$

holds for $i = 1, 2$ and for j sufficiently large. By the same arguments as in the proof of Lemma 5.12 we may assume that M is simply connected. We now want to minimize area with boundary d_j in $M \setminus \bigcup_{i=1}^{k-1} B(x_i, s_j)$ instead of all of M. In order to do so we first slightly modify the metric near $\bigcup_{i=1}^{k-1} \partial B(x_i, s_j)$ to obtain a mean-convex domain. Using a partition of unity we may simply choose a metric g_j on $M \setminus B(x_i, s_j)$ that agrees with the original metric outside of $\bigcup_{i=1}^{k-1} B(x_i, 2s_j)$ and has mean-convex boundary. We can now solve the Plateau problem as before in $\left(M \setminus \bigcup_{i=1}^{k-1} B(x_i, s_j), g_j\right)$ with prescribed boundary d_j. After passing to a subsequential limit we find a nonempty (thanks to (6.7)) limit lamination in $(M \setminus \{x_1, \dots, x_{k-1}\}, g)$. By stability, the limit lamination extends also across the set $\{x_1, \dots, x_{k-1}\}$ and we can argue as in the proof of Lemma 5.12.

For the remaining case (b), we prove the stronger assertion that f_j is uniformly bounded. This handles case (b) as follows. If $f_j \leq C$, then for $x \in M \setminus \bigcup_{i=1}^{k-1} B(x_i, 2r)$, we find that

$$l_j(x) \geq C^{-1} \operatorname{dist}(x, c_j^1 \cup \dots \cup c_j^{k-1}) \geq C^{-1} r$$

for j sufficiently large, which contradicts (6.4)–(6.6).

In order to show that f_j is uniformly bounded, we argue by contradiction and assume that

$$\liminf_{j \to \infty} \sup_{\Sigma_j} f_j \to \infty, \tag{6.8}$$

which we simply assume to be a full limit after taking another subsequence. Note that $f_j \le C_j$ for some constant $C_j > 0$, since Σ_j is a smooth and closed surface; therefore, we can pick[5] $z_j \in \Sigma_j$ such that

$$2 f_j(z_j) \ge \sup_{\Sigma_j} f_j .$$

The assumption (6.8) implies that there is a loop $e_j \in C_j'(z_j)$ based at z_j that is noncontractible on scale r_0 such that

$$\text{length}(e_j) \le o(\text{dist}(z_j, c_j^1 \cup \cdots \cup c_j^{k-1})). \tag{6.9}$$

We can assume that any other loop $e_j' \in C_j'(z_j)$ has

$$\text{length}(e_j) \le 2\,\text{length}(e_j'), \tag{6.10}$$

since we could otherwise replace our original loop e_j by an even shorter one satisfying all other assumptions.

For $z \in \Sigma_j \cap B(z_j, 2\,\text{length}(e_j)))$, we find from (6.9) that

$$\text{dist}(z, c_j^1 \cup \cdots \cup c_j^{k-1}) \ge \text{dist}(z_j, c_j^1 \cup \cdots \cup c_j^{k-1}) - 2\,\text{length}(e_j) \ge \tfrac{1}{2}\,\text{dist}(z_j, c_j^1 \cup \cdots \cup c_j^{k-1})$$

for j sufficiently large. Therefore, by the choice of z_j, we have

$$4 l_j(z) \ge l_j(z_j) \tag{6.11}$$

for any $z \in \Sigma_j \cap B(z_j, 2\,\text{length}(e_j)))$.

Since e_j is separating (recall that all curves in C_j' are separating by definition) we can write

$$\Sigma \setminus e_j = \Sigma_j^1 \cup \Sigma_j^2$$

for connected minimal surfaces Σ_j^i with boundary e_j. We claim that

$$\Sigma_j^i \cap \partial B(z_j, r_0) \neq \varnothing \quad \text{for } i = 1, 2. \tag{6.12}$$

For ease of notation, we prove (6.12) for Σ_j^1; the argument for Σ_j^2 is analogous.

We again distinguish two cases. In the first case we assume that there is a simple closed curve $g_j \subset \Sigma_j^1$ with $0 \neq [g_j] \in \langle [c_j^1], \dots, [c_j^{k-1}] \rangle$. This case follows for homological reasons: We can pick a closed curve $h_j \subset \Sigma_j$ that intersects g_j exactly once. Since e_j is separating and $g_j \subset \Sigma_j^1$, we can even choose h_j with $h_j \subset \Sigma_j^1$. But then h_j has to intersect at least one of the curves c_j^l, which implies that

$$\Sigma_j^1 \cap B(x_l, s_j) \neq \varnothing.$$

Thanks to (6.9) this implies that

$$\Sigma_j^1 \cap \partial B(z_j, \text{length}(e_j)) \neq \varnothing.$$

Since $\partial \Sigma_j^1 = e_j \subset B(z_j, \text{length}(e_j)/2)$, we obtain (6.12) from the convex hull property applied to Σ_j^1.

[5]This is the standard selection procedure for such scales adapted to our situation.

In the remaining case we have that if $g_j \subset \Sigma_j^1$ is a simple closed curve with $\operatorname{length}(g_j) \le l_0$, then $[g_j] = 0$. Moreover, the bound (6.11) combined with the choice (6.10) then implies that any simple closed curve $g_j \subset \Sigma_j^1 \cap B(z_j, 2\operatorname{length}(e_j))$ with $\operatorname{length}(g_j) \le \operatorname{length}(e_j)/4$ is contractible on scale $2\operatorname{length}(e_j)$ at z_j. At this point we are (for $\Sigma_j^1 \cap B(z_j, 2\operatorname{length}(e_j))$) in the setting of Lemma 5.8, and can simply repeat the same argument to obtain

$$\Sigma_j^1 \cap \partial B(z_j, 2\operatorname{length}(e_j)) \neq \varnothing,$$

which in turn implies (6.12) by the convex hull property using that $\partial \Sigma_j^1 \subset B(z_j, \operatorname{length}(e_j))$.

We can now once again argue as in the proof of Lemma 5.12 and conclude the proposition. □

6.3 Proof of the main result

We now give the proof for the general case of our main result.

Proof of Theorem 1.1 For $\pi : \tilde{M} \to M$ the universal covering, consider the surfaces $\hat{\Sigma}_j = \pi^{-1}(\Sigma_j)$ and write $\mathscr{X} = \pi^{-1}(\{x_1, \dots, x_{k-1}\})$. We can pass to a subsequential limit

$$\hat{\Sigma}_j \to \mathscr{L} \quad \text{in } C^{0,\alpha}_{\mathrm{loc}}(\tilde{M} \setminus \mathscr{S}),$$

where clearly $\mathscr{X} \subset \mathscr{S}$. It follows from Proposition 6.3 that the surfaces are ULSC away from $\mathscr{X}$. Moreover, thanks to Lemma 6.1, we can find a collapsed leaf $\Gamma \subset \mathscr{L}$, which extends across $\mathscr{S} \setminus \mathscr{X}$ by Proposition 3.3. Moreover, since this is stable, it also extends across the isolated points $\mathscr{X}$ to a complete, stable minimal surface, which implies a contradiction as in the case of the first homology systole. □

References

[Aiex 2018] **N S Aiex**, *Non-compactness of the space of minimal hypersurfaces*, Math. Ann. 370 (2018) 191–208 MR Zbl

[Balacheff et al. 2012] **F Balacheff**, **H Parlier**, **S Sabourau**, *Short loop decompositions of surfaces and the geometry of Jacobians*, Geom. Funct. Anal. 22 (2012) 37–73 MR Zbl

[Chodosh and Mantoulidis 2020] **O Chodosh**, **C Mantoulidis**, *Minimal surfaces and the Allen–Cahn equation on 3-manifolds: index, multiplicity, and curvature estimates*, Ann. of Math. 191 (2020) 213–328 MR Zbl

[Chodosh et al. 2017] **O Chodosh**, **D Ketover**, **D Maximo**, *Minimal hypersurfaces with bounded index*, Invent. Math. 209 (2017) 617–664 MR Zbl

[Choi and Schoen 1985] **H I Choi**, **R Schoen**, *The space of minimal embeddings of a surface into a three-dimensional manifold of positive Ricci curvature*, Invent. Math. 81 (1985) 387–394 MR Zbl

[Choi and Wang 1983] **H I Choi**, **A N Wang**, *A first eigenvalue estimate for minimal hypersurfaces*, J. Differential Geom. 18 (1983) 559–562 MR Zbl

[Colding and Minicozzi 2004a] **T H Colding**, **W P Minicozzi, II**, *Embedded minimal disks: proper versus nonproper, global versus local*, Trans. Amer. Math. Soc. 356 (2004) 283–289 MR Zbl

[Colding and Minicozzi 2004b] **T H Colding**, **W P Minicozzi, II**, *The space of embedded minimal surfaces of fixed genus in a* 3*-manifold, I: Estimates off the axis for disks*, Ann. of Math. 160 (2004) 27–68 MR Zbl

[Colding and Minicozzi 2004c] **T H Colding**, **W P Minicozzi, II**, *The space of embedded minimal surfaces of fixed genus in a* 3*-manifold, II: Multi-valued graphs in disks*, Ann. of Math. 160 (2004) 69–92 MR Zbl

[Colding and Minicozzi 2004d] **T H Colding**, **W P Minicozzi, II**, *The space of embedded minimal surfaces of fixed genus in a* 3*-manifold, III: Planar domains*, Ann. of Math. 160 (2004) 523–572 MR Zbl

[Colding and Minicozzi 2004e] **T H Colding**, **W P Minicozzi, II**, *The space of embedded minimal surfaces of fixed genus in a* 3*-manifold, IV: Locally simply connected*, Ann. of Math. 160 (2004) 573–615 MR Zbl

[Colding and Minicozzi 2008] **T H Colding**, **W P Minicozzi, II**, *The Calabi–Yau conjectures for embedded surfaces*, Ann. of Math. 167 (2008) 211–243 MR Zbl

[Colding and Minicozzi 2015] **T H Colding**, **W P Minicozzi, II**, *The space of embedded minimal surfaces of fixed genus in a* 3*-manifold, V: Fixed genus*, Ann. of Math. 181 (2015) 1–153 MR Zbl

[Fischer-Colbrie and Schoen 1980] **D Fischer-Colbrie**, **R Schoen**, *The structure of complete stable minimal surfaces in* 3*-manifolds of nonnegative scalar curvature*, Comm. Pure Appl. Math. 33 (1980) 199–211 MR Zbl

[Freedman et al. 1983] **M Freedman**, **J Hass**, **P Scott**, *Least area incompressible surfaces in* 3*-manifolds*, Invent. Math. 71 (1983) 609–642 MR Zbl

[Gromov 1996] **M Gromov**, *Systoles and intersystolic inequalities*, from "Actes de la Table Ronde de Géométrie Différentielle" (A L Besse, editor), Sémin. Congr. 1, Soc. Math. France, Paris (1996) 291–362 MR Zbl

[Hardt and Simon 1979] **R Hardt**, **L Simon**, *Boundary regularity and embedded solutions for the oriented Plateau problem*, Ann. of Math. 110 (1979) 439–486 MR Zbl

[Irie et al. 2018] **K Irie**, **F C Marques**, **A Neves**, *Density of minimal hypersurfaces for generic metrics*, Ann. of Math. 187 (2018) 963–972 MR Zbl

[Lawson 1970] **H B Lawson, Jr**, *Complete minimal surfaces in* S^3, Ann. of Math. 92 (1970) 335–374 MR Zbl

[Liokumovich et al. 2018] **Y Liokumovich**, **F C Marques**, **A Neves**, *Weyl law for the volume spectrum*, Ann. of Math. 187 (2018) 933–961 MR Zbl

[Marques and Neves 2017] **F C Marques**, **A Neves**, *Existence of infinitely many minimal hypersurfaces in positive Ricci curvature*, Invent. Math. 209 (2017) 577–616 MR Zbl

[Marques and Neves 2021] **F C Marques**, **A Neves**, *Morse index of multiplicity one min-max minimal hypersurfaces*, Adv. Math. 378 (2021) art. id. 107527 MR Zbl

[Marques et al. 2019] **F C Marques**, **A Neves**, **A Song**, *Equidistribution of minimal hypersurfaces for generic metrics*, Invent. Math. 216 (2019) 421–443 MR Zbl

[Meeks et al. 2010] **W H Meeks, III**, **J Pérez**, **A Ros**, *Limit leaves of an* H *lamination are stable*, J. Differential Geom. 84 (2010) 179–189 MR Zbl

[Neumann 1976] **D A Neumann**, 3*-manifolds fibering over* S^1, Proc. Amer. Math. Soc. 58 (1976) 353–356 MR Zbl

[Schoen 1983] **R Schoen**, *Estimates for stable minimal surfaces in three-dimensional manifolds*, from "Seminar on minimal submanifolds" (E Bombieri, editor), Ann. of Math. Stud. 103, Princeton Univ. Press (1983) 111–126 MR Zbl

[Schoen and Yau 1979] **R Schoen**, **S T Yau**, *Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature*, Ann. of Math. 110 (1979) 127–142 MR Zbl

[Schoen and Yau 1983] **R Schoen**, **S T Yau**, *The existence of a black hole due to condensation of matter*, Comm. Math. Phys. 90 (1983) 575–579 MR Zbl

[Song 2023] **A Song**, *Existence of infinitely many minimal hypersurfaces in closed manifolds*, Ann. of Math. 197 (2023) 859–895 MR Zbl

[Tollefson 1969] **J L Tollefson**, *3-manifolds fibering over S^1 with nonunique connected fiber*, Proc. Amer. Math. Soc. 21 (1969) 79–80 MR Zbl

[White 2015] **B White**, *Curvatures of embedded minimal disks blow up on subsets of C^1 curves*, J. Differential Geom. 100 (2015) 389–394 MR Zbl

[Zhou 2020] **X Zhou**, *On the multiplicity one conjecture in min-max theory*, Ann. of Math. 192 (2020) 767–820 MR Zbl

HM: *Max Planck Institute for Mathematics*
Bonn, Germany

Current address: *Department of Mathematics, University of Chicago*
Chicago, IL, United States

AS: *Mathematisches Institut, University of Münster*
Münster, Germany

hmatthiesen@math.uchicago.edu, asiffert@uni-muenster.de

Proposed: Tobias H Colding
Seconded: Gang Tian, David Fisher

Received: 2 October 2020
Revised: 16 August 2024

Geometry & Topology 29:4 (2025) 1851–1907
DOI: 10.2140/gt.2025.29.1851
Published: 27 June 2025

The Manhattan curve, ergodic theory of topological flows and rigidity

STEPHEN CANTRELL
RYOKICHI TANAKA

For every nonelementary hyperbolic group, we introduce the Manhattan curve associated to each pair of left-invariant hyperbolic metrics which are quasi-isometric to a word metric. It is convex; we show that it is continuously differentiable and moreover is a straight line if and only if the corresponding two metrics are roughly similar, ie they are within bounded distance after multiplying by a positive constant. Further, we prove that the Manhattan curve associated to two strongly hyperbolic metrics is twice continuously differentiable. The proof is based on the ergodic theory of topological flows associated to general hyperbolic groups and analyzing the multifractal structure of Patterson–Sullivan measures. We exhibit some explicit examples including a hyperbolic triangle group and compute the exact value of the mean distortion for pairs of word metrics.

20F67; 37D35, 37D40

1 Introduction

Let Γ be a nonelementary hyperbolic group. Given a pair of hyperbolic metrics d and d_* which are left-invariant and quasi-isometric to a word metric on Γ (hence they are quasi-isometric each other), we determine exactly when they are *roughly similar*, ie d and d_* are within bounded distance after rescaling by a positive constant, in terms of the *Manhattan curve* for the pair of metrics.

For d (resp. d_*), let us define the stable translation length by

$$\ell[x] := \lim_{n\to\infty} \frac{1}{n} d(o, x^n) \quad \text{for } x \in \Gamma,$$

(resp. $\ell_*[x]$), where the limit exists since the function $n \mapsto d(o, x^n)$ is subadditive, and o denotes the identity element in Γ. Note that the stable translation length for x depends only on the conjugacy class of x, and thus defines the function on the set of conjugacy classes **conj** in Γ. Let us consider the series with two parameters

$$\mathcal{Q}(a,b) = \sum_{[x]\in\mathbf{conj}} \exp(-a\ell_*[x] - b\ell[x]) \quad \text{for } a, b \in \mathbb{R}.$$

The Manhattan curve $\mathscr{C}_M$ for a pair (d, d_*) is defined by the boundary of the following convex set

$$\{(a,b) \in \mathbb{R}^2 : \mathcal{Q}(a,b) < \infty\}.$$

This curve was introduced by Burger [1993] for a class of groups acting on a noncompact symmetric space of rank one.

Let $\mathscr{D}_\Gamma$ be the set of hyperbolic metrics which are left-invariant and quasi-isometric to some (equivalently, every) word metric in Γ. Recall that for $d, d_* \in \mathscr{D}_\Gamma$, we say that d and d_* are roughly similar if there exist constants $\tau > 0$ and $C \geq 0$ such that

$$|d_*(x,y) - \tau d(x,y)| \leq C \quad \text{for all } x, y \in \Gamma.$$

Theorem 1.1 *Let Γ be a nonelementary hyperbolic group. For every pair (d, d_*) in $\mathscr{D}_\Gamma$, the Manhattan curve $\mathscr{C}_M$ for (d, d_*) is continuously differentiable, and it is a straight line if and only if d and d_* are roughly similar.*

Note that, if v (resp. v_*) is the abscissa of convergence for $\mathscr{Q}(0,b)$ in b (resp. $\mathscr{Q}(a,0)$ in a), then

$$v := \lim_{r\to\infty} \frac{1}{r} \log \# B(o,r), \quad \text{where } B(o,r) := \{x \in \Gamma : d(o,x) \leq r\},$$

(similarly v_* for d_*) and $\# B$ denotes the cardinality of B. In particular, $(0, v)$ and $(v_*, 0)$ lie on $\mathscr{C}_M$; see Section 3.

Theorem 1.2 *Let Γ be a nonelementary hyperbolic group. For every pair (d, d_*) in $\mathscr{D}_\Gamma$, the limit*

$$\tau(d_*/d) := \lim_{r\to\infty} \frac{1}{\# B(o,r)} \sum_{x \in B(o,r)} \frac{d_*(o,x)}{r} \tag{1-1}$$

exists, where the balls $B(o,r)$ are defined for d, and we have

$$\tau(d_*/d) \geq \frac{v}{v_*}. \tag{1-2}$$

Moreover, the following are equivalent:

1. *The equality $\tau(d_*/d) = v/v_*$ holds.*
2. *There exists a constant $c > 0$ such that $\ell_*[x] = c\ell[x]$ for all $[x] \in$ **conj**.*
3. *The metrics d and d_* are roughly similar.*

Let us call $\tau(d_*/d)$ defined by (1-1) the *mean distortion* of d_* over d, and the inequality (1-2) the *distortion inequality*. The proof of Theorem 1.2 is based on Theorem 1.1 and relies on the following property of the Manhattan curve $\mathscr{C}_M$: the slope of $\mathscr{C}_M$ at $(0, v)$ is $-\tau(d_*/d)$ (Theorem 3.12).

If both metrics d and d_* are strongly hyperbolic (eg induced by an isometric cocompact action on a CAT(-1)-space; see Definition 2.2), then we have the following.

Theorem 1.3 *Let Γ be a nonelementary hyperbolic group. If d and d_* are strongly hyperbolic metrics in $\mathscr{D}_\Gamma$, then the Manhattan curve $\mathscr{C}_M$ for (d, d_*) is twice continuously differentiable.*

We show this in Theorem 4.16 and prove the analogous result for pairs of word metrics in Theorem 4.14.

The various methods we use throughout our work allow us to connect the geometrical features of $\mathscr{C}_M$ with properties of the corresponding metrics. For example, by comparing our methods to those of the first author in [Cantrell 2021], we connect the second differential of $\mathscr{C}_M$ at $(0, v)$ with the variance of a central limit theorem for uniform counting measures on spheres; see Theorem 4.17 and the remark thereafter. We also connect the asymptotic gradients of $\mathscr{C}_M$ with the dilation constants

$$\mathrm{Dil}_- := \inf_{[x]\in\mathbf{conj}_{>0}} \frac{\ell_*[x]}{\ell[x]} \quad \text{and} \quad \mathrm{Dil}_+ := \sup_{[x]\in\mathbf{conj}_{>0}} \frac{\ell_*[x]}{\ell[x]},$$

where $\mathbf{conj}_{>0}$ is the set of $[x] \in \mathbf{conj}$ such that $\ell[x]$ (and hence $\ell_*[x]$) is nonzero. We also show that, for every pair of word metrics, Dil_- and Dil_+ are rational (Proposition 4.22). Note that $\mathrm{Dil}_- = \mathrm{Dil}_+$ if and only if $\tau(d_*/d) = v/v_*$ by Theorem 1.2.

1.1 Historical backgrounds

Burger [1993] introduced the Manhattan curve associated to a finitely generated, nonelementary group Γ which acts on a rank one symmetric space X properly discontinuously, convex cocompactly and without fixed points. For each convex cocompact realization of Γ into the isometry group of X there is a natural length function defined on the conjugacy classes of Γ: one can assign to each conjugacy class in Γ the geometric length of the corresponding closed geodesic in the quotient. Burger's Manhattan curve is defined using two of these length functions. He showed that the curve is continuously differentiable and it is a straight line if and only if the isomorphism of lattices associated to the two corresponding realizations extends to an isomorphism of the ambient Lie groups; see [Burger 1993, Theorem 1]. An important special case includes two isomorphic copies of torsion-free cocompact Fuchsian groups acting on the hyperbolic plane. In this case, Sharp [1998] has shown that the associated Manhattan curve is real analytic by employing thermodynamic formalism for geodesic flows. Recently, Kao [2020] has shown that the Manhattan curve is real analytic for a class of noncompact hyperbolic surfaces. A similar rigidity problem related to the Manhattan curve is discussed for cusped Hitchin representations in [Bray et al. 2022].

Kaimanovich, Kapovich and Schupp [Kaimanovich et al. 2007] have extensively studied similar problems for a free group F of rank at least 2. They compared the pair of word metrics for the generating sets S and $\phi(S)$, where S is the free generating set and ϕ is an automorphism of F. They introduced the *generic stretching factor* $\lambda(\phi)$, which is defined as the average or typical growth rate of $|\phi(x_n)|/n$ when x_n is chosen uniformly at random from the words of length n in S. Using our terminology, it amounts to considering $\tau(S/\phi(S)) = \lambda(\phi)$. An automorphism ϕ of F is called simple if it is a composition of an inner automorphism and a permutation of S. It has been shown [Kaimanovich et al. 2007, Theorem F] that ϕ is simple if and only if $\lambda(\phi) = 1$ (in which case ϕ gives rise to a rough similarity on the Cayley graph of (F, S)). Sharp [2010] has pointed out its connection to the corresponding Manhattan curve: he has identified $\lambda(\phi)$ with the slope of the normal line at the point $(\log(2k-1), 0)$, where k is the rank of F.

The mean distortion is a generalization of the generic stretching factor for hyperbolic groups and has appeared in the work of Calegari and Fujiwara [2010] (though the identification is not immediately clear at first sight). They have shown that for every pair of word metrics the distortion inequality holds [Calegari and Fujiwara 2010, Remark 4.28], and also that the mean distortion is an algebraic integer [Calegari and Fujiwara 2010, Corollary 4.27]. Furthermore, a (possibly degenerate) central limit theorem (CLT) has been shown [Calegari and Fujiwara 2010, Theorem 4.25] (see also [Calegari 2013, Section 3.6]), and their result has been generalized in [Cantrell 2021, Theorem 1.2] and [Gekhtman et al. 2022, Theorem 1.1]. Moreover, the variance of CLT is zero if and only if two metrics are roughly similar; see [Cantrell 2021, Lemma 5.1] and [Gekhtman et al. 2022, Theorem 1.1].

Furman [2002] has proposed a general framework which can be used to compare metrics belonging to $\mathscr{D}_\Gamma$ when Γ is a torsion-free hyperbolic group. In his work he introduced an abstract geodesic flow in a measurable category for a general hyperbolic group, and showed that two metrics are roughly similar if and only if the associated Bowen–Margulis currents are not mutually singular on the boundary square $\partial^2\Gamma$ (which is the set of two distinct ordered pairs of points in the Gromov boundary $\partial\Gamma$); see [Furman 2002, Theorem 2]. He also claims that two metrics in $\mathscr{D}_\Gamma$ are roughly similar if and only if their associated translation length functions are proportional. A main motivation of the present paper is to incorporate the properties of the Manhattan curve into rigidity statements that characterize rough similarity. As a consequence we strengthen and generalize Furman's result for all nonelementary hyperbolic groups.

1.2 Outline of proofs

Let us sketch the proof of Theorem 1.1. First we consider the following series in $a, b \in \mathbb{R}$,

$$\mathscr{P}(a,b) = \sum_{x\in\Gamma} \exp(-ad_*(o,x) - bd(o,x)),$$

and identify the Manhattan curve $\mathscr{C}_M$ with the graph of $b = \theta(a)$, where $\theta(a)$ is the abscissa of convergence in b for each fixed a; see Proposition 3.1. In what follows, we also call the function θ which parametrizes $\mathscr{C}_M$ the Manhattan curve. Next we perform the Patterson–Sullivan construction for $ad_* + bd$ for (a,b) with $b = \theta(a)$ and construct a one-parameter family of measures $\mu_{a,b}$ on $\partial\Gamma$ for $(a,b) \in \mathscr{C}_M$ (Corollary 2.10). A key step in the proof of Theorem 1.1 is understanding the variation of $\mu_{a,b}$ in a. Every measure $\mu_{a,b}$ (which is not necessarily unique) is ergodic with respect to the Γ-action on $\partial\Gamma$ for each fixed $(a,b) \in \mathscr{C}_M$ and the proof of this is adapted from classical arguments in [Coornaert 1993]. Moreover, $\mu_{a,b}$ is doubly ergodic, ie $\mu_{a,b} \otimes \mu_{a,b}$ is ergodic with respect to the diagonal action of Γ on $\partial^2\Gamma$. Proving this amounts to showing that the geodesic flow is ergodic if it is properly defined eg in the case of manifolds (where we owe this idea to the work of Kaimanovich [1990]). Furman [2002] has constructed a framework where machinery concerning geodesic flows works for general hyperbolic groups (see also [Bader and Furman 2017]), but the space in his setup has only a measurable structure and so difficulties arise when we discuss a family of flow-invariant measures (on the same space) and carry out limiting arguments with those measures. We employ a compact model of geodesic flow defined by Mineyev [2005], and show that there exist associated flow-invariant probability measures $m_{a,b}$ (which is unique) for each $(a,b) \in \mathscr{C}_M$ and they are continuous in $a \in \mathbb{R}$ in the weak-star topology (Section 2.7).

Finally we define the *local intersection number* $\tau(\xi)$ at $\xi \in \partial\Gamma$ for a pair (d, d_*) as the limit of $d_*(\gamma_\xi(0), \gamma_\xi(t))/d(\gamma_\xi(0), \gamma_\xi(t))$ as $t \to \infty$, where γ_ξ is a quasigeodesic such that $\gamma_\xi(t) \to \xi$ as $t \to \infty$, if the limit exists. We show that the limit exists and is equal to a constant $\tau_{a,b}$ for $\mu_{a,b}$-almost every $\xi \in \partial\Gamma$. Furthermore, $\tau_{a,b}$ is continuous in $(a,b) \in \mathscr{C}_M$ (Section 3.2). In fact, $\theta'(a) = -\tau_{a,b}$ if θ is differentiable at $a \in \mathbb{R}$. Note that θ is convex (as seen from the definition) and thus is continuous everywhere and differentiable at all but at most countably many points. Since we have shown that $\tau_{a,b}$ is continuous in a, the function θ is C^1. In the last step, where we show that $\tau_{a,b}$ coincides with $-\theta'(a)$ (if it exists), we prove that $\mu_{a,b}$ assigns a full measure to the set where the local intersection number $\tau(\xi)$ is defined and is equal to $-\theta'(a)$. This discussion naturally leads us to study the multifractal spectrum of $\tau(\xi)$, ie to determine all the possible values α of $\tau(\xi)$ and the size of the level sets for which $\tau(\xi) = \alpha$. This spectrum is actually the multifractal profile of a Patterson–Sullivan measure for d_*, where the Hausdorff dimension is defined by a quasimetric associated with d (Theorem 3.8). Furthermore, the profile function is the Legendre transform of the Manhattan curve and is defined on the interval $(\mathrm{Dil}_-, \mathrm{Dil}_+)$. (It would be interesting if it is defined on $[\mathrm{Dil}_-, \mathrm{Dil}_+]$, including the two extrema, which is indeed the case for some special case, eg word metrics. See Remark 3.9 and our investigation of this point in Section 4.6.) Based on this discussion, we show that the Manhattan curve $\mathscr{C}_M$ for a pair (d, d_*) is a straight line if and only if d and d_* are roughly similar (Theorem 3.10) and thus conclude Theorem 1.1. The discussion also yields Theorem 1.2 by identifying $-\theta'(0)$ with the mean distortion $\tau(d_*/d)$; see Theorem 3.12.

Let us briefly describe the proof of Theorem 1.3, as our methods are quite different to those used in the proof of Theorem 1.1. We introduce a subshift of finite type coming from the coding built upon Cannon's automatic structure. This allows us to employ techniques from thermodynamic formalism, which we apply within the strengthened thermodynamic framework of Gouëzel [2014]; see Section 4. This enables us to relate $\mathscr{C}_M$ to a family of real analytic (pressure) functions associated to suitable potentials on the subshift. Unfortunately, in general, we do not know whether the subshift is topologically transitive, ie whether Cannon's automatic structure has a single recurrent component which is dominant. This introduces additional difficulties. Since the automatic structure may contain various large components, our pressure functions of interest are not coming from a single component. In particular, we must compare the corresponding dominating eigenvalues of a collection of transfer operators, each of which depend on one of these components. To overcome these difficulties we use the following ideas. First, we introduce a multiparameter family of Patterson–Sullivan measures developed in Section 2.6. Second, we compare these measures to a collection of pressure functions and show that the first and second-order partial derivatives of these functions coincide at certain points; our proof of this latter part is developed upon an argument of Calegari and Fujiwara [2010, Section 4.5]. This allows us to show that a function $\widetilde{\theta}(a,b)$, which we obtain from gluing together our analytic pressure functions, is twice continuously differentiable.

Finally, we realize $\mathscr{C}_M$ as the solution to $\widetilde{\theta}(a,b) = 0$ and then apply the implicit function theorem to conclude the proof.

After showing Theorem 1.3 (Theorem 4.16), we investigate further properties of the Manhattan curve for strongly hyperbolic and word metrics. For a pair of word metrics, we obtain a finer version of Theorem 1.1 and show that two word metrics are not roughly similar if and only if the corresponding Manhattan curve is globally strictly convex (Theorem 4.17). Furthermore we show that there are two lines, the slopes of which we can express explicitly, which are within bounded distance of the Manhattan curve at $\pm\infty$ (Proposition 4.22). As an application we obtain a precise large deviation principle for $d_*(o, x_n)/n$ when x_n uniformly distributes on the sphere $d(o,x) = n$, for word metrics d and d_*. We identify the effective domain (where the rate function is finite) with $[\mathrm{Dil}_-, \mathrm{Dil}_+]$ (Theorem 4.23 and Corollary 4.25).

1.3 Organization of the paper

In Section 2.1, we review basic theory on hyperbolic groups, a classical Patterson–Sullivan construction in a generalized form and a topological flow. In Section 3, we show Theorems 1.1 and 1.2. In Section 4, we discuss thermodynamic formalism. We show Theorem 1.3 in Theorem 4.16, the analogous result for word metrics in Theorem 4.14, a stronger version of Theorem 1.1 for word metrics in Theorem 4.17, and that Dil_- and Dil_+ are rational for word metrics in Proposition 4.22. The proof of this proposition is based on a finer analysis of a transfer operator in Proposition 4.20. We exhibit an application to a large deviation principle in Theorem 4.23 and Corollary 4.25. In Section 5, we compute explicit examples: the free group of rank 2 and the $(3,3,4)$-triangle group. In particular, in the case of the latter group, we find a pair of word metrics for which the mean distortion is algebraic irrational. In the appendix, we show Lemma 3.11, which we use in the proof of main rigidity result Theorem 3.10 (the second part of Theorem 1.1).

Notation Throughout the article, we denote by $C, C', C'', \dots$ constants whose explicit values may change from line to line, and by $C_R, C'_R, C''_R, \dots$ constants with subscript R to indicate their dependency on a parameter R. For real-valued functions $f(t)$ and $g(t)$ in $t \in \mathbb{R}$, we write $f(t) \asymp g(t)$ if there exist constants $C_1, C_2 > 0$ independent of t such that $C_1 g(t) \le f(t) \le C_2 g(t)$, and $f(t) \asymp_R g(t)$ if those constants C_1 and C_2 depend only on R. Further we use the big-O and small-o notation: $f(t) = O(g(t))$ if there exist constants $C > 0$ and $T > 0$ such that $|f(t)| \le C|g(t)|$ for all $t \ge T$, while $f(t) = O_R(g(t))$ if the implied constant is C_R, and $f(t) = g(t) + o(t)$ as $t \searrow 0$ if $|f(t) - g(t)|/t \to 0$ as $t \to 0$ for $t > 0$. We say that two measures μ_1 and μ_2 defined on the common measurable spaces are *comparable* if there exists constant $C > 0$ such that $C^{-1}\mu_1 \le \mu_2 \le C\mu_1$. We use the notation $\#A$ which stands for the cardinality of a set A.

Acknowledgements We thank Professor Richard Sharp for his suggestion in Remark 4.21, and the referee for useful comments. Tanaka is partially supported by JSPS Grant-in-Aid for Scientific Research JP20K03602 and JST, ACT-X grant JPMJAX190J, Japan.

2 Preliminaries

2.1 Hyperbolic groups

We briefly review some fundamental material concerning hyperbolic groups. See the original work by Gromov [1987] and Ghys and de la Harpe [1990] for background. Let (X, d) be a metric space. The Gromov product is defined by

$$(x|y)_w := \frac{d(w,x) + d(w,y) - d(x,y)}{2} \quad \text{for } x, y, w \in X.$$

For $\delta \geq 0$, a metric space (X, d) is called *δ-hyperbolic* if

$$(x|y)_w \geq \min\{(x|z)_w, (z|y)_w\} - \delta \quad \text{for all } x, y, z, w \in X.$$

We say that a metric space is *hyperbolic* if it is δ-hyperbolic for some $\delta \geq 0$.

Let Γ be a finitely generated group. We call a finite set of generators S of Γ *symmetric* if $s^{-1} \in S$ whenever $s \in S$. The word metric associated to a symmetric finite set of generators S is defined by

$$d_S(x, y) := |x^{-1}y|_S \quad \text{for } x, y \in \Gamma,$$

where $|x|_S := \min\{k \geq 0 : x = s_1 \cdots s_k,\ s_i \in S\}$ and 0 for the identity element. We say that Γ is a *hyperbolic* group if the pair (Γ, d_S) is hyperbolic for some word metric d_S. If (Γ, d_S) is δ-hyperbolic, then for every finite, symmetric set of generators S', the pair $(\Gamma, d_{S'})$ is δ'-hyperbolic for some δ'. A hyperbolic group is called *nonelementary* if it is nonamenable, and *elementary* otherwise. Elementary hyperbolic groups are either finite groups or contain $\mathbb{Z}$ as a finite-index subgroup.

We say that two metrics d and d_* on Γ are *quasi-isometric* if there exist constants $L > 0$ and $C \geq 0$ such that

$$L^{-1}d(x, y) - C \leq d_*(x, y) \leq Ld(x, y) + C \quad \text{for all } x, y \in \Gamma,$$

and *roughly similar* if there exist constants $\lambda > 0$ and $C \geq 0$ such that

$$\lambda d(x, y) - C \leq d_*(x, y) \leq \lambda d(x, y) + C \quad \text{for all } x, y \in \Gamma.$$

Suppose that (Γ, d) is δ-hyperbolic. If d_* is roughly similar to d, then (Γ, d_*) is δ'-hyperbolic for some (possibly different) δ'. However, if d_* is just quasi-isometric to d, then (Γ, d_*) is not necessarily hyperbolic. We will discuss a category of metrics which are hyperbolic and quasi-isometric to some hyperbolic metric in Γ.

Let Γ be a nonelementary hyperbolic group. We define $\mathcal{D}_\Gamma$ to be the set of metrics on Γ that are left-invariant, ie $d(gx, gy) = d(x, y)$ for all x, y and $g \in \Gamma$, hyperbolic, and quasi-isometric to some (equivalently, every) word metric.

Example 2.1 Let Γ be the fundamental group of a compact negatively curved manifold (M, d_M). The group Γ acts on the universal cover $(\widetilde{M}, d_{\widetilde{M}})$ isometrically and freely. For each point p in $\widetilde{M}$, if we define $d(x,y) := d_{\widetilde{M}}(xp, yp)$ for $x, y \in \Gamma$, then d yields a metric, which is left-invariant, hyperbolic and quasi-isometric to a word metric by the Milnor–Švarc lemma. Such d therefore belongs to $\mathcal{D}_\Gamma$. More generally, if Γ is a nonelementary hyperbolic group and acts on a CAT(-1)-space isometrically and freely with a precompact fundamental domain, then as in the same way above the metric of the CAT(-1)-space yields a metric on Γ in $\mathcal{D}_\Gamma$.

A particular subclass of metrics that we will be interested in are strongly hyperbolic metrics.

Definition 2.2 A hyperbolic metric d on Γ is called *strongly hyperbolic* if there exist $L \geq 0$, $c > 0$ and $R_0 \geq 0$ such that for all $x, x', y, y' \in \Gamma$ and all $R \geq R_0$, the condition

$$d(x,y) - d(x,x') - d(y,y') + d(x',y') \geq R$$

implies that

$$|d(x,y) - d(x',y) - d(x,y') + d(x',y')| \leq Le^{-cR}.$$

Every hyperbolic group Γ admits a *strongly hyperbolic metric* in $\mathcal{D}_\Gamma$. This was shown by Mineyev [2005, Theorem 32]; see also Nica and Špakula [2016]. We use the existence of such a metric in the course of our proofs.

Let us consider a metric d on Γ. We say that for an interval I in $\mathbb{R}$, a map $\gamma\colon I \to (\Gamma, d)$ is an (L,C)-*quasigeodesic* for constants $L > 0$ and $C \geq 0$ if it holds that

$$L^{-1}|s-t| - C \leq d(\gamma(s), \gamma(t)) \leq L|s-t| + C \quad \text{for all } s,t \in I,$$

and a C-*rough geodesic* for $C \geq 0$ if it holds that

$$|s-t| - C \leq d(\gamma(s), \gamma(t)) \leq |s-t| + C \quad \text{for all } s,t \in I.$$

A *geodesic* is a C-rough geodesic with $C = 0$. A metric d is called C-roughly geodesic if for all $x, y \in \Gamma$ there exists a C-rough geodesic $\gamma\colon [a,b] \to \Gamma$ such that $\gamma(a) = x$ and $\gamma(b) = y$, and called *roughly geodesic* if it is C-roughly geodesic for some $C \geq 0$. If $d \in \mathcal{D}_\Gamma$, then (Γ, d) is not necessarily a geodesic metric space, but it is a roughly geodesic space [Bonk and Schramm 2000, Proposition 5.6]. In many places, we use the following fact which we refer to as the *Morse lemma*: if d is a proper (ie all balls of finite radius consist of finitely many points) C_0-roughly geodesic hyperbolic metric in Γ, then every (L,C)-quasigeodesic γ in (Γ, d), there exists a C_0-rough geodesic γ' such that γ and γ' are within Hausdorff distance D where D depends only on C_0, L, C and the hyperbolic constant of d (cf [Ghys and de la Harpe 1990, Théorèmes 21 et 25, Chapitre 5] and [Bonk and Schramm 2000, the proof of Proposition 5.6]).

2.2 Boundary at infinity

Let us define the (geometric) boundary of Γ. Let o be the identity element in Γ. Fix $d \in \mathscr{D}_\Gamma$ and consider the corresponding Gromov product in Γ. We say that a sequence $\{x_n\}_{n=0}^\infty$ is *divergent* if $(x_n|x_m)_o \to \infty$ as $n, m \to \infty$, and define an equivalence relation in the set of divergent sequences by

$$\{x_n\}_{n=0}^\infty \sim \{x'_n\}_{n=0}^\infty \iff (x_n|x'_m)_o \to \infty \quad \text{as } n, m \to \infty.$$

Let us define $\partial\Gamma$ the set of equivalence classes of divergent sequences in Γ and call it the *boundary* of (Γ, d). For $\xi \in \partial\Gamma$, if $\{x_n\}_{n=0}^\infty \in \xi$, then we write $x_n \to \xi$ as $n \to \infty$. We extend the Gromov product to $\Gamma \cup \partial\Gamma$ by setting

$$(\xi|\eta)_o := \sup\Bigl\{\liminf_{n\to\infty}(x_n|y_n)_o : \{x_n\}_{n=0}^\infty \in \xi, \{y_n\}_{n=0}^\infty \in \eta\Bigr\},$$

where if ξ or η is in Γ, then $(\xi|\eta)_o$ is defined by taking the constant sequences $\xi_n = \xi$ or $\eta_n = \eta$. Note that if divergent sequences $\{x_n\}_{n=0}^\infty$ and $\{y_n\}_{n=0}^\infty$ are equivalent to $\{x'_n\}_{n=0}^\infty$ and $\{y'_n\}_{n=0}^\infty$, respectively, then

$$\liminf_{n\to\infty}(x'_n|y'_n)_o \ge \limsup_{n\to\infty}(x_n|y_n)_o - 2\delta.$$

This implies that for all $\xi, \eta, \zeta \in \Gamma \cup \partial\Gamma$,

$$(\xi|\eta)_o \ge \min\{(\xi|\zeta)_o, (\zeta|\eta)_o\} - 3\delta.$$

Let us define a *quasimetric* by

$$\rho(\xi, \eta) := e^{-(\xi|\eta)_o} \quad \text{for } \xi, \eta \in \partial\Gamma.$$

In general, ρ is not a metric in $\partial\Gamma$, but it satisfies that $\rho(\xi, \eta) = 0$ if and only if $\xi = \eta$, $\rho(\xi, \eta) = \rho(\eta, \xi)$ for all $\xi, \eta \in \partial\Gamma$, and there exists a constant $C > 0$ such that

$$\rho(\xi, \eta) \le C \max\{\rho(\xi, \zeta), \rho(\zeta, \eta)\} \quad \text{for all } \xi, \eta, \zeta \in \partial\Gamma.$$

The quasimetric ρ associated to $d \in \mathscr{D}_\Gamma$ defines a topology on $\partial\Gamma$ that is compact, separable and metrizable. In fact, for arbitrary two metrics $d, d_* \in \mathscr{D}_\Gamma$ the corresponding boundaries with the topologies constructed above are homeomorphic. We refer to $\partial\Gamma$ the underlying topological space.

2.3 Shadows

For all $R \ge 0$ and $x \in \Gamma$, we define the *shadow* by

$$O(x, R) := \{\xi \in \partial\Gamma : (\xi|x)_o \ge d(o, x) - R\}.$$

Let us denote by $B(\xi, r)$ the ball of radius $r \ge 0$ centered at ξ in $\partial\Gamma$ relative to the quasimetric $\rho(\xi, \eta) = e^{-(\xi|\eta)_o}$. The δ-hyperbolic inequality yields the following comparison between balls and shadows.

Lemma 2.3 *Let (Γ, d) be δ-hyperbolic for $\delta \geq 0$. For each $\tau \geq 0$, if $R \geq \tau + 3\delta$, then for all $\xi \in \partial\Gamma$ and all $x \in \Gamma$ such that $(o|\xi)_x \leq \tau$, we have*

$$B(\xi, e^{-3\delta+R-d(o,x)}) \subset O(x, R) \subset B(\xi, e^{3\delta+R-d(o,x)}).$$

Proof See eg [Blachère et al. 2011, Proposition 2.1]; we omit the details. □

Note that if d and d_* are in $\mathcal{D}_\Gamma$ and (L, C)-quasi-isometric, then for all $R \geq 0$ there exists $R' \geq 0$ depending on L, C and their hyperbolicity constants such that

$$O(x, R) \subset O'(x, R') \quad \text{for all } x \in \Gamma,$$

where $O(x, R)$ (resp. $O'(x, R')$) are the shadows defined by d (resp. d_*). This follows from the stability of rough geodesics and the fact that every pair of points in $\Gamma \cup \partial\Gamma$ are connected by a C-rough geodesic in (Γ, d) for some C. Therefore omitting the dependency on d in the shadow $O(x, R)$ will not cause any confusion, up to changing the thickness parameter R.

2.4 Hausdorff dimension

For every $d \in \mathcal{D}_\Gamma$, let $\rho(\xi, \eta) = \exp(-(\xi|\eta)_o)$ be the corresponding quasimetric in $\partial\Gamma$. Although it is not a metric in general, we may define the Hausdorff dimension of sets and measures in $\partial\Gamma$ relative to ρ as in the case of metrics. It is known that there exists a constant $\varepsilon > 0$ such that ρ^ε is bi-Lipschitz to a genuine metric d_ε (eg [Heinonen 2001, Proposition 14.5]), in which case the Hausdorff dimension relative to d_ε will be $1/\varepsilon$ times the Hausdorff dimension relative to ρ.

For every subset E in $\partial\Gamma$, let us denote by $\rho(E) := \sup\{\rho(\xi, \eta) : \xi, \eta \in E\}$. For all $s \geq 0$ and $\Delta > 0$, we define

$$\mathcal{H}^s_\Delta(E, \rho) := \inf\left\{ \sum_{i=0}^{\infty} \rho(E_i)^s : E \subset \bigcup_{i=0}^{\infty} E_i \text{ and } \rho(E_i) \leq \Delta \right\},$$

$$\mathcal{H}^s(E, \rho) := \sup_{\Delta > 0} \mathcal{H}^s_\Delta(E, \rho) = \lim_{\Delta \to 0} \mathcal{H}^s_\Delta(E, \rho).$$

The *Hausdorff dimension* of a set E in $(\partial\Gamma, \rho)$ is defined by

$$\dim_{\mathrm{H}}(E, \rho) := \inf\{s \geq 0 : \mathcal{H}^s(E, \rho) = 0\} = \sup\{s \geq 0 : \mathcal{H}^s(E, \rho) > 0\}.$$

For every Borel measure ν on $\partial\Gamma$, the (upper) *Hausdorff dimension* of ν is defined by

$$\dim_{\mathrm{H}}(\nu, \rho) := \inf\{\dim_{\mathrm{H}}(E, \rho) : \nu(\partial\Gamma \setminus E) = 0 \text{ and } E \text{ is Borel}\}.$$

Lemma 2.4 (the Frostman-type lemma) *Let ν be a Borel probability measure on $(\partial\Gamma, \rho)$. For $s_1, s_2 \geq 0$, let*

$$E(s_1, s_2) := \left\{ \xi \in \partial\Gamma : s_1 \leq \liminf_{r \to 0} \frac{\log \nu(B(\xi, r))}{\log r} \leq s_2 \right\},$$

where $B(\xi, r) = \{\eta \in \partial\Gamma : \rho(\xi, \eta) \leq r\}$. If $\nu(E(s_1, s_2)) = 1$, then

$$s_1 \leq \dim_{\mathrm{H}}(E(s_1, s_2), \rho) \leq s_2 \quad \text{and} \quad s_1 \leq \dim_{\mathrm{H}}(\nu, \rho) \leq s_2.$$

Proof It suffices to show that $\dim_H(E(s_1, s_2), \rho) \le s_2$ and $s_1 \le \dim_H(\nu, \rho)$. These follow as in the case when ρ is a metric; see eg [Heinonen 2001, Section 8.7]. □

2.5 Distance (Busemann) quasicocycles

For $d \in \mathcal{D}_\Gamma$, let us define

$$\beta_w(x, \xi) := \sup\Bigl\{\limsup_{n\to\infty}(d(x, \xi_n) - d(w, \xi_n)) : \{\xi_n\}_{n=0}^{\infty} \in \xi\Bigr\}$$

for $w, x \in \Gamma$ and for $\xi \in \partial\Gamma$, and call $\beta_w : \Gamma \times \partial\Gamma \to \mathbb{R}$ the *Busemann function* based at w. We note that

$$d(x, z) - d(o, z) = d(o, x) - 2(x|z)_o \quad \text{for } x, z \in \Gamma,$$

and thus the δ-hyperbolicity implies that

$$|\beta_o(x, \xi) - (d(o, x) - 2(x|\xi)_o)| \le 2\delta \quad \text{for } (x, \xi) \in \Gamma \times \partial\Gamma.$$

The Busemann function β_o satisfies the following cocycle identity with an additive error:

$$|\beta_o(xy, \xi) - (\beta_o(y, x^{-1}\xi) + \beta_o(x, \xi))c| \le 4\delta \quad \text{for } x, y \in \Gamma \text{ and } \xi \in \partial\Gamma.$$

Let us consider a strongly hyperbolic metric $\hat{d}$ in $\mathcal{D}_\Gamma$ (Definition 2.2) and denote by $\langle x|y\rangle_o$ the corresponding Gromov product. Then there exists a constant $\varepsilon > 0$ such that

$$\exp(-\varepsilon\langle x|y\rangle_w) \le \exp(-\varepsilon\langle x|z\rangle_w) + \exp(-\varepsilon\langle z|y\rangle_w) \quad \text{for all } x, y, z, w \in \Gamma,$$

by [Nica and Špakula 2016, Lemma 6.2, Definition 4.1] (in fact, this property characterizes the strong hyperbolicity). This shows that the Gromov product based at o for a strongly hyperbolic metric extends to $\Gamma \cup \partial\Gamma$ as genuine limits. This also shows that the corresponding Busemann function $\hat{\beta}_o$ is defined as limits and satisfies the cocycle identity,

$$\hat{\beta}_o(xy, \xi) = \hat{\beta}_o(y, x^{-1}\xi) + \hat{\beta}_o(x, \xi) \quad \text{for } x, y \in \Gamma \text{ and } \xi \in \partial\Gamma.$$

We use a strongly hyperbolic metric to construct an analogue of geodesic flow in Section 2.7.

2.6 Patterson–Sullivan construction

For $d \in \mathcal{D}_\Gamma$, let us denote the ball of radius r centered at x relative to d by

$$B(x, r) := \{y \in \Gamma : d(x, y) \le r\} \quad \text{for } x \in \Gamma \text{ and } r \ge 0.$$

We define the *exponential volume growth rate* relative to d as

$$v := \limsup_{r\to\infty} \frac{1}{r} \log \#B(o, r).$$

Since Γ is nonamenable, v is finite and nonzero.

We recall the classical construction of Patterson–Sullivan measures for $d \in \mathcal{D}_\Gamma$. Consider the Dirichlet series

$$\mathcal{P}(s) := \sum_{x \in \Gamma} e^{-sd(o,x)},$$

which has the divergence exponent v. Suppose for a moment that the series diverges at $s = v$. Then the sequence of probability measures on Γ,

$$\mu_s := \frac{1}{\mathcal{P}(s)} \sum_{x \in \Gamma} e^{-sd(o,x)} \delta_x,$$

where δ_x is the Dirac measure at x, considered as measures on the compactified space $\Gamma \cup \partial\Gamma$, has a convergent subsequence as $s \searrow v$. A limit point μ is a probability measure supported on $\partial\Gamma$, and there exists a constant $C_\delta > 0$ such that for $x \in \Gamma$ and for $\xi \in \partial\Gamma$,

$$C_\delta^{-1} e^{-v\beta_o(x,\xi)} \le \frac{dx_*\mu}{d\mu}(\xi) \le C_\delta e^{-v\beta_o(x,\xi)}. \tag{2-1}$$

All limit points satisfy the above estimates (2-1). If the series $\mathcal{P}(s)$ does not diverge at $s = v$, then a slight modification yields a measure satisfying (2-1). We call a probability measure satisfying (2-1) a *Patterson–Sullivan measure* for $d \in \mathcal{D}_\Gamma$. For details, see [Coornaert 1993, Théorème 5.4].

The above construction applies to the following setting where the distance is replaced by a more general function. Let us consider a function $\psi \colon \Gamma \times \Gamma \to \mathbb{R}$ and define

$$\psi(x|y)_z := \frac{\psi(x,z) + \psi(z,y) - \psi(x,y)}{2} \quad \text{for } x, y, z \in \Gamma,$$

as a generalization of the Gromov product. Note that the order of x, y, z matters since ψ may not satisfy $\psi(x,y) = \psi(y,x)$. We assume that $\psi(\cdot|\cdot)_o$ admits a "quasiextension" to $\Gamma \times (\Gamma \cup \partial\Gamma)$, ie there exist a function $\psi(\cdot|\cdot)_o \colon \Gamma \times (\Gamma \cup \partial\Gamma) \to \mathbb{R}$ and a constant $C \ge 0$ such that

$$\limsup_{n\to\infty} \psi(x|\xi_n)_o - C \le \psi(x|\xi)_o \le \liminf_{n\to\infty} \psi(x|\xi'_n)_o + C \tag{QE}$$

for all $(x,\xi) \in \Gamma \times (\Gamma \cup \partial\Gamma)$ and for all $\{\xi_n\}_{n=0}^\infty, \{\xi'_n\}_{n=0}^\infty \in \xi$. This allows us to define the following function analogous to the Busemann function for $(x,\xi) \in \Gamma \times \partial\Gamma$,

$$\beta_o^\psi(x,\xi) := \sup\Bigl\{\limsup_{n\to\infty}(\psi(x,\xi_n) - \psi(o,\xi_n)) : \{\xi_n\}_{n=0}^\infty \in \xi\Bigr\}.$$

Furthermore, if ψ is Γ-invariant, ie $\psi(gx,gy) = \psi(x,y)$ for all $g, x, y \in \Gamma$, then β_o^ψ satisfies the quasicocycle relation:

$$|\beta_o^\psi(xy,\xi) - (\beta_o^\psi(y,x^{-1}\xi) + \beta_o^\psi(x,\xi))| \le 4C.$$

Recall that if $d \in \mathcal{D}_\Gamma$, then (Γ, d) is a C-rough geodesic metric space for some $C \ge 0$. Let us consider the following "rough geodesic" condition: for all large enough $C, R \ge 0$, there exists $C_0 \ge 0$ such that for all C-rough geodesics γ between x and y, and for all z in the R-neighborhood of γ,

$$|\psi(x,y) - (\psi(x,z) + \psi(z,y))| \le C_0. \tag{RG}$$

If ψ satisfies (RG) relative to $d \in \mathscr{D}_\Gamma$, then there exists a constant C' such that for a large enough R and for all $x \in \Gamma$,

$$|\beta_o^\psi(x,\xi) + \psi(o,x)| \le C' \quad \text{for all } \xi \in O(x,R). \tag{2-2}$$

Definition 2.5 We say that a function $\psi\colon \Gamma \times \Gamma \to \mathbb{R}$ is a *tempered potential* relative to $d \in \mathscr{D}_\Gamma$ if ψ satisfies (QE) and (RG) relative to d.

Example 2.6 For all $d, d_* \in \mathscr{D}_\Gamma$, by the Morse lemma, d_* satisfies (RG) relative to d. This implies that for every $a \in \mathbb{R}$, the function $\psi_a = ad_*$ satisfies (RG) relative to d. Moreover ψ_a satisfies (QE) and is Γ-invariant. Therefore $\psi_a = ad_*$ is a Γ-invariant tempered potential relative to d for every $a \in \mathbb{R}$. The same argument applies to an arbitrary triple $d, d_*, d_{**} \in \mathscr{D}_\Gamma$ and every linear combination

$$\psi_{a,b} := ad_* + bd \quad \text{for } a,b \in \mathbb{R}.$$

For every $a,b \in \mathbb{R}$, the function $\psi_{a,b}$ is a Γ-invariant tempered potential relative to d_{**}. The functions ψ_a and $\psi_{a,b}$ are the main tools in Sections 3.2 and 4.4, respectively.

For $d \in \mathscr{D}_\Gamma$, let ψ be a Γ-invariant tempered potential relative to d. We say that a probability measure μ on $\partial\Gamma$ satisfies the "quasiconformal" property with exponent $\theta \in \mathbb{R}$ relative to (ψ, d) if there exists a constant C depending only on ψ and d such that

$$C^{-1} \le \exp(\beta_o^\psi(x,\xi) + \theta\beta_o(x,\xi)) \cdot \frac{dx_*\mu}{d\mu}(\xi) \le C \tag{QC}$$

for all $x \in \Gamma$ and μ-almost every ξ in $\partial\Gamma$, where β_o is the Busemann function associated to d. We simply say that μ satisfies (QC) if θ and (ψ, d) are fixed and apparent from the context.

Proposition 2.7 *For $d \in \mathscr{D}_\Gamma$, let ψ be a Γ-invariant tempered potential relative to d. Then the abscissa of convergence θ of the series in s,*

$$\sum_{x\in\Gamma} \exp(-\psi(o,x) - sd(o,x)), \tag{2-3}$$

is finite and there exists a probability measure μ_ψ on $\partial\Gamma$ satisfying (QC) with exponent θ relative to (ψ, d). Moreover, every finite Borel measure μ satisfying (QC) has the property

$$C'^{-1}\exp(-\psi(o,x) - \theta d(o,x)) \le \mu(O(x,R)) \le C'\exp(-\psi(o,x) - \theta d(o,x)) \tag{2-4}$$

for all $x \in \Gamma$, where C' is a constant depending on C, C_0 and R.

Proof Note that θ is given by

$$\limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in S(n,R_0)} e^{-\psi(o,x)}, \quad \text{where } S(n,R_0) = \{x \in \Gamma : |d(o,x) - n| \le R_0\}.$$

Since ψ satisfies (RG) relative to d and is Γ-invariant, then for all $n, m \geq 0$,

$$\sum_{x \in S(n+m, R_0)} e^{-\psi(o,x)} \leq e^C \sum_{x \in S(n, R_0)} e^{-\psi(o,x)} \cdot \sum_{x \in S(m, R_0)} e^{-\psi(o,x)},$$

which implies that θ is finite. Let us define the family of probability measures for $s > \theta$ by

$$\mu_{\psi,s} := \frac{\sum_{x \in \Gamma} \exp(-\psi(o,x) - s d(o,x)) \delta_x}{\sum_{x \in \Gamma} \exp(-\psi(o,x) - s d(o,x))}.$$

If the series (2-3) diverges at θ, then letting $s \searrow \theta$ yields a weak limit μ_ψ after passing to a subsequence. The measure μ_ψ is supported on $\partial\Gamma$. For $x, y \in \Gamma$, we have that

$$x_* \mu_{\psi,s}(y) = \frac{\exp(-\psi(o, x^{-1}y) - s d(o, x^{-1}y))}{\sum_{z \in \Gamma} \exp(-\psi(o,z) - s d(o,z))} = \frac{\exp(-\psi(o, x^{-1}y))}{\exp(-\psi(o,y))} e^{-s(d(o,x^{-1}y) - d(o,y))} \mu_{\psi,s}(y).$$

By the assumption, $\psi(x,y) - \psi(o,y)$ is $\beta_o^\psi(x,\xi)$ up to a uniform additive constant as y tends to ξ. Further, $d(x,y) - d(o,y)$ coincides with $\beta_o(x,\xi)$ up to a constant depending only on the hyperbolicity constant of d uniformly on a neighborhood of ξ in $\Gamma \cup \partial\Gamma$. This yields (QC). If the series (2-3) does not diverge at θ, then the argument as in the classical setting provides (QC); cf [Coornaert 1993, Théorème 5.4] and [Tanaka 2017, Theorem 3.3] for a special case.

Further, since ψ satisfies (RG) relative to d, we have (2-2). Suppose that a finite measure μ satisfies (QC), and μ is a probability measure without loss of generality. Then for all $x \in \Gamma$,

$$x_* \mu(O(x,R)) \asymp_{C,\theta} \exp(\psi(o,x) + \theta d(o,x)) \mu(O(x,R)).$$

For all small enough $0 < \varepsilon_0 < 1$, there exists a large enough R such that

$$\mu(x^{-1} O(x,R)) \geq 1 - \varepsilon_0 \quad \text{for all } x \in \Gamma$$

(cf [Coornaert 1993, Proposition 6.1]), and thereby we obtain (2-4). □

Lemma 2.8 *For $d \in \mathscr{D}_\Gamma$, if ψ is a Γ-invariant tempered potential relative to d, then there exist constants $\theta \in \mathbb{R}$ and $C, R_0 > 0$ such that for all $n \geq 0$,*

$$C^{-1} e^{\theta n} \leq \sum_{x \in S(n, R_0)} e^{-\psi(o,x)} \leq C e^{\theta n},$$

where $S(n, R_0) := \{x \in \Gamma : |d(o,x) - n| \leq R_0\}$.

We say that θ is the *exponent* of (ψ, d) abusing the notation; the proof actually shows that if there is a finite Borel measure μ satisfying (QC) relative to (ψ, d) with some exponent, then that exponent has to be θ.

Proof For (Γ, d), fix large enough constants $R_0, R > 0$ so that for every $n \geq 0$ the shadows $O(x,R)$ for $x \in S(n, R_0)$ cover $\partial\Gamma$. Since (Γ, d) is hyperbolic, there exists a constant M such that for every n, each

$\xi \in \partial\Gamma$ is included in at most M shadows $O(x, R)$ with $x \in S(n, R_0)$. By Proposition 2.7, there exists a probability measure μ_ψ which satisfies (2-4). The first inequality in (2-4) shows that for all $n \geq 0$,

$$e^{-\theta(n+R_0)} \sum_{x \in S(n,R_0)} e^{-\psi(o,x)} \leq C \sum_{x \in S(n,R_0)} \mu_\psi(O(x, R)) \leq CM.$$

The second inequality in (2-4) shows that for all $n \geq 0$,

$$1 = \mu_\psi(\partial\Gamma) \leq \sum_{x \in S(n,R_0)} \mu_\psi(O(x, R)) \leq Ce^{-\theta(n-R_0)} \sum_{x \in S(n,R_0)} e^{-\psi(o,x)},$$

hence we obtain the claim. □

We say that a (finite) Borel measure μ on $\partial\Gamma$ is *doubling* relative to a quasimetric ρ if $\mu(B(\xi, r)) > 0$ for all $\xi \in \partial\Gamma$ and for all $r > 0$, and there exists a C such that for every $r \geq 0$ and $\xi \in \partial\Gamma$,

$$\mu(B(\xi, 2r)) \leq C\mu(B(\xi, r)),$$

where $B(\xi, r)$ is the ball defined by ρ in $\partial\Gamma$.

Lemma 2.9 *For $d \in \mathscr{D}_\Gamma$, if ψ is a Γ-invariant tempered potential relative to d, then every finite Borel measure μ on $\partial\Gamma$ satisfying (QC) is doubling relative to a quasimetric ρ. Moreover, an arbitrary pair of finite Borel measures on $\partial\Gamma$ satisfying (QC) with the same exponent and (ψ, d) are mutually absolutely continuous and their densities are uniformly bounded from above and below. In particular, every measure μ_ψ is ergodic with respect to the Γ-action on $\partial\Gamma$, ie every Γ-invariant Borel set A in $\partial\Gamma$ satisfies that either $\mu_\psi(A) = 0$ or $\mu_\psi(\partial\Gamma \setminus A) = 0$.*

Proof First, by Proposition 2.7, every finite Borel measure μ with (QC) satisfies (2-4), which shows that $\mu(O(x, R)) > 0$ for all $x \in \Gamma$ and

$$\mu(O(x, 2R)) \asymp_R \mu(O(x, R)) \quad \text{for all } x \in \Gamma,$$

where R is a large enough fixed constant. Applying this estimate finitely many times if necessary, by Lemma 2.3 we find that μ is doubling relative to ρ.

Next, (2-4) implies that for arbitrary two finite Borel measures μ, μ' satisfying (QC) with common exponent and (ψ, d), the ratio of the measures of balls relative to μ and μ' are uniformly bounded from above and below. Since both measures are doubling relative to ρ, the Vitali covering theorem [Heinonen 2001, Theorem 1.6], adapted to a quasimetric ρ, shows that μ and μ' are mutually absolutely continuous and their densities are uniformly bounded from above and below.

Finally, for μ satisfying (QC), if A is an arbitrary Γ-invariant Borel set in $\partial\Gamma$ such that $\mu(A) > 0$, then the restriction $\mu|_A$ also satisfies (QC) with the same exponent and (ψ, d). Therefore what we have shown implies that $\mu|_A \asymp \mu$ and thus $\mu(\partial\Gamma \setminus A) = 0$. This in particular applies to μ_ψ. □

A central example of the construction is a family of measures $\mu_{a,b}$ for $(a,b) \in \mathscr{C}_M$ associated to a pair (d, d_*). Let us single out the following corollary, which we use in Section 2.7.

Corollary 2.10 *Let us consider a pair $d, d_* \in \mathscr{D}_\Gamma$.*

(1) *For each $(a,b) \in \mathscr{C}_M$, there exists a probability measure $\mu_{a,b}$ on $\partial\Gamma$ such that for all $x \in \Gamma$,*

$$C_{a,b}^{-1} e^{-a\beta_{*o}(x,\xi)-b\beta_o(x,\xi)} \le \frac{dx_*\mu_{a,b}}{d\mu_{a,b}}(\xi) \le C_{a,b} e^{-a\beta_{*o}(x,\xi)-b\beta_o(x,\xi)},$$

where β_{*o} and β_o are Busemann functions for d_* and d, respectively, and $C_{a,b}$ is a constant of the form $C_{a,b} = C_{d_*}^{|a|} C_d^{|b|}$. Moreover, we have that

$$C'^{-1} \exp(-ad_*(o,x) - bd(o,x)) \le \mu_{a,b}(O(x,R)) \le C' \exp(-ad_*(o,x) - bd(o,x))$$

for all $x \in \Gamma$, where C' is a constant depending on $C_{a,b}$ and R.

(2) *For every $a \in \mathbb{R}$,*

$$\theta(a) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{x \in S(n,R_0)} e^{-ad_*(o,x)},$$

where $S(n,R_0) := \{x \in \Gamma : |d(o,x) - n| \le R_0\}$ for some constant R_0, and the function θ is convex and continuous on $\mathbb{R}$.

(3) *For each $(a,b) \in \mathscr{C}_M$, every probability measure $\mu_{a,b}$ is ergodic with respect to the Γ-action on $\partial\Gamma$.*

Proof For each $a \in \mathbb{R}$, if we let $\psi(x,y) = ad_*(x,y)$, then ψ is a Γ-invariant tempered potential (Example 2.6) and $\theta = b$ for $(a,b) \in \mathscr{C}_M$. Therefore Proposition 2.7 implies (1), where the constant $C_{a,b} = C_{d_*}^{|a|} C_d^{|b|}$ is obtained from the proof of Proposition 2.7. Lemma 2.8 and the Hölder inequality imply that $\theta(a)$ is finite and convex in $a \in \mathbb{R}$, hence θ is continuous on $\mathbb{R}$, showing (2), and Lemma 2.9 shows (3). □

Note that letting v and v_* be the exponential volume growth rates for d and d_* respectively, we have that $(0,v), (v_*,0) \in \mathscr{C}_M$, and $\mu_{0,v}$ and $\mu_{v_*,0}$ are (classical) Patterson–Sullivan measures for d and d_*, respectively.

2.7 Topological flow

In this section, we follow the discussion in [Tanaka 2021, Section 3]. Let $\partial^2\Gamma := (\partial\Gamma)^2 \setminus \{\text{diagonal}\}$, where Γ acts on $\partial^2\Gamma$ by $x \cdot (\xi,\eta) := (x\xi, x\eta)$ for $x \in \Gamma$ and $(\xi,\eta) \in \partial^2\Gamma$. Consider the space $\partial^2\Gamma \times \mathbb{R}$ and fix a strongly hyperbolic metric $\hat{d} \in \mathscr{D}_\Gamma$. There exists a constant $C \ge 0$ such that for each $(\xi,\eta) \in \partial^2\Gamma$ there is a C-rough geodesic $\gamma_{\xi,\eta} : \mathbb{R} \to (\Gamma, \hat{d})$ satisfying that $\gamma_{\xi,\eta}(-t) \to \xi$ and $\gamma_{\xi,\eta}(t) \to \eta$ as $t \to \infty$, respectively [Bonk and Schramm 2000, Proposition 5.2(3)]. Shifting the parameter $t \mapsto t + T$ by some T if necessary, we parametrize $\gamma_{\xi,\eta}$ in such a way that

$$\hat{d}(\gamma_{\xi,\eta}(0), o) = \min_{t\in\mathbb{R}} \hat{d}(\gamma_{\xi,\eta}(t), o).$$

We define

$$\mathbf{ev}\colon \partial^2\Gamma\times\mathbb{R}\to\Gamma \quad\text{by}\quad \mathbf{ev}(\xi,\eta,t):=\gamma_{\xi,\eta}(t).$$

Note that the map **ev** depends on the choice of C-rough geodesics; however, every other choice yields the map whose image lies in a uniformly bounded distance: if $\gamma_{\xi,\eta}$ and $\gamma'_{\xi,\eta}$ are two C-rough geodesics with the same pair of extreme points, then

$$\max_{t\in\mathbb{R}}\hat{d}(\gamma_{\xi,\eta}(t),\gamma'_{\xi,\eta}(t))<C'$$

for some positive constant C' depending only on the metric. Let us endow the space of C-rough geodesics on $(\Gamma,\hat{d}\,)$ with the pointwise convergence topology. We define $\mathbf{ev}\colon\partial^2\Gamma\times\mathbb{R}\to\Gamma$ as a measurable map by assigning $\gamma_{\xi,\eta}$ to $(\xi,\eta)\in\partial^2\Gamma$ in a Borel measurable way: first fix a set of generators S in Γ and an order on it; second, consider C-rough geodesics evaluated on the set of integers as sequences of group elements and choose lexicographically minimal ones $\gamma^0_{\xi,\eta}$ for each $(\xi,\eta)\in\partial^2\Gamma$; and finally, define $\gamma_{\xi,\eta}(t):=\gamma^0_{\xi,\eta}(\lfloor t\rfloor)$ for $t\in\mathbb{R}$, where $\lfloor t\rfloor$ stands for the largest integer at most t.

Letting $\hat{\beta}_o\colon\Gamma\times\partial\Gamma\to\mathbb{R}$ be the Busemann function based at o associated with $\hat{d}$, we define the cocycle

$$\kappa\colon\Gamma\times\partial^2\Gamma\to\mathbb{R},\quad \kappa(x,\xi,\eta):=\tfrac{1}{2}(\hat{\beta}_o(x^{-1},\xi)-\hat{\beta}_o(x^{-1},\eta)),$$

where the cocycle identity for κ follows from that of $\hat{\beta}_o$ (Section 2.5). Then, Γ acts on $\partial^2\Gamma\times\mathbb{R}$ through κ by

$$x\cdot(\xi,\eta,t):=(x\xi,x\eta,t-\kappa(\xi,\eta,t)).$$

Let us call this Γ-action the (Γ,κ)-*action* on $\partial^2\Gamma\times\mathbb{R}$. It is shown that the (Γ,κ)-action on $\partial^2\Gamma\times\mathbb{R}$ is properly discontinuous and cocompact, namely, the quotient topological space $\Gamma\backslash(\partial^2\Gamma\times\mathbb{R})$ is compact [Tanaka 2021, Lemma 3.2]. Let

$$\mathscr{F}_\kappa:=\Gamma\backslash(\partial^2\Gamma\times\mathbb{R}),$$

where we define a continuous $\mathbb{R}$-action as in the following. The $\mathbb{R}$-action $\tilde{\Phi}$ on $\partial^2\Gamma\times\mathbb{R}$ is defined by the translation in the $\mathbb{R}$-component,

$$\tilde{\Phi}_s(\xi,\eta,t):=(\xi,\eta,t+s).$$

This action and the (Γ,κ)-action commute, and thus the $\mathbb{R}$-action $\tilde{\Phi}$ descends to the quotient

$$\Phi_s[\xi,\eta,t]:=[\xi,\eta,t+s]\quad\text{for }[\xi,\eta,t]\in\mathscr{F}_\kappa.$$

Then $\mathbb{R}$ acts on $\mathscr{F}_\kappa$ via Φ continuously. We call the $\mathbb{R}$-action Φ on $\mathscr{F}_\kappa$ the *topological flow* (or, simply, the *flow*) on $\mathscr{F}_\kappa$.

Let us consider finite measures invariant under the flow on $\mathscr{F}_\kappa$. Let Λ be a Γ-invariant Radon measure on $\partial^2\Gamma$, ie $x_*\Lambda=\Lambda$ for all $x\in\Gamma$ and Λ is Borel regular and finite on every compact set. Then every measure of the form $\Lambda\otimes dt$, where dt is the (normalized) Lebesgue measure on $\mathbb{R}$, yields a flow-invariant finite measure on $\mathcal{F}_\kappa$. Namely, for every Γ-invariant Radon measure Λ on $\partial^2\Gamma$, there exists a unique finite Radon measure m invariant under the flow on $\mathcal{F}_\kappa$ such that

$$\int_{\partial^2\Gamma\times\mathbb{R}} f\,d\Lambda\otimes dt = \int_{\mathcal{F}_\kappa} \bar{f}\,dm \tag{2-5}$$

for all compactly supported continuous functions f on $\partial^2\Gamma\times\mathbb{R}$, where $\bar{f}$ is the Γ-invariant function

$$\bar{f}(\xi,\eta,t) := \sum_{x\in\Gamma} f(x\cdot(\xi,\eta,t)),$$

considered as a function on $\mathcal{F}_\kappa$ [Tanaka 2021, Lemma 3.4] (and we further note that every continuous function φ on $\mathcal{F}_\kappa$ is of the form $\varphi = \bar{f}$ by invoking Urysohn's lemma). If we take a Borel fundamental domain D in $\partial^2\Gamma\times\mathbb{R}$ with respect to the (Γ,κ)-action and a measurable section $\iota\colon \mathcal{F}_\kappa\to D$, then

$$\Lambda\otimes dt = \sum_{x\in\Gamma} x_*(\iota_* m).$$

Note that it is not necessarily the case that the restriction $\Lambda\otimes dt|_D$ coincides with $\iota_* m$ unless the (Γ,κ)-action is free. We always normalize Λ in such a way that the corresponding flow-invariant measure m has total measure 1 (and so is a probability measure on $\mathcal{F}_\kappa$).

For all $d\in\mathcal{D}_\Gamma$, an associated Patterson–Sullivan measure μ on $\partial\Gamma$ yields a Γ-invariant Radon measure Λ_d on $\partial^2\Gamma$ equivalent to

$$\exp(2v(\xi|\eta)_o)\mu\otimes\mu,$$

with the Radon–Nikodym density uniformly bounded from above and from below by positive constants, and the corresponding flow-invariant probability measure m_d on $\mathcal{F}_\kappa$ is ergodic with respect to the flow, ie for every Borel set A such that $\Phi_{-t}(A) = A$ for all $t\in\mathbb{R}$, either $m_d(A) = 0$ or 1; see [Tanaka 2021, Proposition 2.11 and Theorem 3.6]. The same construction applies to measures $\mu_{a,b}$ for all $(a,b)\in\mathscr{C}_M$.

Proposition 2.11 *For each $(a,b)\in\mathscr{C}_M$, there exists a Γ-invariant Radon measure $\Lambda_{a,b}$ on $\partial^2\Gamma$ which is equivalent to*

$$\exp(2a(\xi|\eta)_{*o} + 2b(\xi|\eta)_o)\mu_{a,b}\otimes\mu_{a,b}, \tag{2-6}$$

with the Radon–Nikodym density uniformly bounded from above and below by positive constants of the form $C_{d_*}^{|a|}C_d^{|b|}$. Moreover, $\Lambda_{a,b}$ is ergodic with respect to the Γ-action on $\partial^2\Gamma$, ie for every Γ-invariant Borel set A in $\partial^2\Gamma$, either the set A or the complement has zero $\Lambda_{a,b}$-measure, and the corresponding flow-invariant probability measure $m_{a,b}$ is ergodic with respect to the flow on $\mathcal{F}_\kappa$.

Proof If we denote the measure (2-6) by ν, then we have that

$$C^{-1} \le \frac{dx_*\nu}{d\nu}(\xi,\eta) \le C$$

for all $x\in\Gamma$ and for ν-almost all $(\xi,\eta)\in\partial^2\Gamma$, where C is a positive constant of the form $C_{d_*}^{|a|}C_d^{|b|}$. If we define

$$\varphi(\xi,\eta) := \sup_{x\in\Gamma}\frac{dx_*\nu}{d\nu}(\xi,\eta),$$

then $\Lambda_{a,b} := \varphi(\xi,\eta)\nu$ is a Γ-invariant Radon measure, as desired. The details follow as in Proposition 2.11, Theorem 3.6 and Corollary 3.7 in [Tanaka 2021]. □

Lemma 2.12 *If Λ and Λ' are Γ-invariant ergodic Radon measures on $\partial^2\Gamma$, then either Λ and Λ' are mutually singular, or there exists a positive constant $c > 0$ such that $\Lambda = c\Lambda'$.*

Proof Let us decompose Λ as a sum of two measures $\Lambda = \Lambda_{\mathrm{ac}} + \Lambda_{\mathrm{sing}}$, where Λ_{ac} (resp. Λ_{sing}) is the absolutely continuous (resp. singular) part with respect to Λ'. Note that Λ_{ac} and Λ_{sing} are Γ-invariant Radon measures since Λ and Λ' are so. Suppose that $\Lambda_{\mathrm{ac}} \neq 0$. Then the Radon–Nikodym density $d\Lambda_{\mathrm{ac}}/d\Lambda'$ is locally integrable and Γ-invariant, and thus constant since Λ' is ergodic with respect to the Γ-action. Hence there exists a positive constant $c > 0$ such that $\Lambda_{\mathrm{ac}} = c\Lambda'$, and since Λ is ergodic with respect to the Γ-action, $\Lambda_{\mathrm{sing}} = 0$ and $\Lambda = c\Lambda'$, as desired. □

Corollary 2.13 *For each $(a,b) \in \mathscr{C}_M$, let $m_{a,b}$ be the flow-invariant probability measure on $\mathscr{F}_\kappa$ corresponding to the (normalized) Γ-invariant Radon measure $\Lambda_{a,b}$ on $\partial^2\Gamma$. If $(a,b) \to (a_0,b_0)$ in $\mathscr{C}_M$, then $m_{a,b}$ weakly converges to m_{a_0,b_0}.*

Proof If $(a,b) \to (a_0,b_0)$, then up to taking a subsequence, there exists a normalized Γ-invariant Radon measure Λ_* on $\partial^2\Gamma$ such that $\int_{\partial^2\Gamma} f\,d\Lambda_{a,b}$ converges to $\int_{\partial^2\Gamma} f\,d\Lambda_*$ for each compactly supported continuous function f on $\partial^2\Gamma$ (where we use the fact that $\partial^2\Gamma$ is σ-compact). Let Λ_* be an arbitrary such limit point. Taking a further subsequence, we have that $\mu_{a,b}$ weakly converges to some probability measure μ_*, which is comparable to μ_{a_0,b_0} by Proposition 2.7 (in the form of Corollary 2.10) and Lemma 2.9. This together with Proposition 2.11 shows that Λ_* is equivalent to Λ_{a_0,b_0}. Lemma 2.12 implies that Λ_* coincides with Λ_{a_0,b_0} up to a multiplicative constant, and if they are normalized, then $\Lambda_* = \Lambda_{a_0,b_0}$. Therefore by (2-5) for every limit point m_* of $m_{a,b}$ as $(a,b) \to (a_0,b_0)$, we have that $m_* = m_{a_0,b_0}$, hence $m_{a,b}$ weakly converges to m_{a_0,b_0}. □

3 The Manhattan curve for general hyperbolic metrics

3.1 Fundamental properties of the Manhattan curve

For $d \in \mathscr{D}_\Gamma$, we recall that the stable translation length of $x \in \Gamma$ with respect to d is given by $\ell[x] = \lim_{n\to\infty} d(o,x^n)/n$, where ℓ defines a function on the set of conjugacy classes **conj** and $[x]$ denotes the conjugacy class of $x \in \Gamma$. For $d_* \in \mathscr{D}_\Gamma$, we denote the corresponding function by ℓ_*. For $a,b \in \mathbb{R}$, let

$$\mathscr{Q}(a,b) := \sum_{[x]\in\mathbf{conj}} \exp(-a\ell_*[x] - b\ell[x]),$$

and for each fixed $a \in \mathbb{R}$, we define $\Theta(a)$ as the abscissa of convergence of $\mathscr{Q}(a,b)$ in b. Recall that for $a \in \mathbb{R}$, we have defined $\theta(a)$ as the abscissa of convergence of $\mathscr{P}(a,b)$ in b, where

$$\mathscr{P}(a,b) = \sum_{x\in\Gamma} \exp(-ad_*(o,x) - bd(o,x)).$$

Proposition 3.1 *For all $a \in \mathbb{R}$, we have $\theta(a) = \Theta(a)$.*

We will also call the functions θ as well as Θ the Manhattan curve for the pair (d, d_*). The proof follows the ideas from [Coornaert and Knieper 2002, Section 5] and [Knieper 1983, Section II] (the latter is indicated in [Burger 1993, Section 4.1]); we provide the main argument adapted to our setting for the sake of completeness. We use the following lemma in the proof.

Lemma 3.2 *For $d \in \mathfrak{D}_\Gamma$, there exists a constant C_0 such that for all $x \in \Gamma$, if $d(o,x) - 2(x|x^{-1})_o > C_0$, then*

$$|\ell[x] - (d(o,x) - 2(x|x^{-1})_o)| \le C_0,$$

and there exists $p \in \Gamma$ such that $|\ell[x] - d(p, xp)| \le C_0$.

Sketch of proof Recall that if $d \in \mathfrak{D}_\Gamma$, then (Γ, d) is a C-rough geodesic metric space, ie for all $x, y \in \Gamma$, there exists a C-rough geodesic $\gamma : [a,b] \to \Gamma$ such that $\gamma(a) = x$ and $\gamma(b) = y$. We provide an outline of the proof when d is geodesic for the sake of convenience (a detailed proof is found in [Maher and Tiozzo 2018, Proposition 5.8]); the same argument applies to C-rough geodesic metrics with slight modifications. For all $x, y \in \Gamma$, let us denote by $[x, y]$ the image of a geodesic between x and y. On the one hand, for each $x \in \Gamma$, let us consider a geodesic triangle on o, x and x^2, and take p as a midpoint of $[o, x]$. If $d(o,x) - 2(x|x^{-1})_o$ is large enough, then $d(p, x) > (x|x^{-1})_o$, and thus

$$d(p, xp) \le d(o,x) - 2(x|x^{-1})_o + \delta. \tag{3-1}$$

On the other hand, for an arbitrary positive integer $n > 0$, let us consider a geodesic $\gamma := [o, x^n]$. It holds that if $d(o,x) - 2(x|x^{-1})_o$ is large enough, then

$$\max_{0 \le k \le n} d(x^k, \gamma) \le (x|x^{-1})_o + 3\delta. \tag{3-2}$$

Indeed, let us write $x_k := x^k$ for $0 \le k \le n$ and use p_k to denote a nearest point from x_k on γ. Suppose that x_k is one of the furthest points among $x_0, \dots, x_n$ from γ. Consider a geodesic quadrangle on x_{k-1}, p_{k-1}, p_{k+1} and x_{k+1}, in this order. Let q be a nearest point from x_k on $[x_{k-1}, x_{k+1}]$. By δ-hyperbolicity there is a point r with $d(q,r) \le 2\delta$ on $[x_{k-1}, p_{k-1}] \cup [p_{k-1}, p_{k+1}] \cup [p_{k+1}, x_{k+1}]$, and we see that r is in fact on $[p_{k-1}, p_{k+1}]$; this shows (3-2).

Finally, (3-2) together with the triangle inequality implies that for all $n > 0$,

$$d(o, x_n) \ge d(o, x_{n-1}) + d(o,x) - 2(x|x^{-1})_o - 6\delta,$$

which yields $\ell[x] \ge d(o,x) - 2(x|x^{-1})_o - 6\delta$. Combining this with (3-1) shows the claim, since $\ell[x] \le d(p, xp)$. □

Proof of Proposition 3.1 To simplify the notation we write $|x| := d(o,x)$ and $|x|_* := d_*(o,x)$. Note that for all large enough $L \ge 0$ and for each $a \in \mathbb{R}$,

$$\theta(a) = \limsup_{n\to\infty} \frac{1}{n} \log \sum_{||x|-n| \le L} e^{-a|x|_*} \quad \text{and} \quad \Theta(a) = \limsup_{n\to\infty} \frac{1}{n} \log \sum_{[x] : |\ell[x]-n| \le L} e^{-a\ell_*[x]}.$$

Also note that for each $x \in \Gamma$, there exists $p \in \Gamma$ such that

$$|\ell_*[x] - d_*(p, xp)| \le C_1 \quad \text{and} \quad |\ell[x] - d(p, xp)| \le C_2,$$

by the proof of Lemma 3.2, where C_1 and C_2 are constants depending only on the hyperbolicity constants of d_* and d. This yields $\theta(a) \ge \Theta(a)$ for each $a \in \mathbb{R}$.

For a large enough $R > 0$ and all $z, w \in \Gamma$, let

$$O(z, w, R) := \{g \in \Gamma \cup \partial\Gamma : (z|g)_w \le R\}.$$

Let us take a pair of hyperbolic elements x, y such that $n \mapsto x^n$ and $n \mapsto y^n$ for $n \in \mathbb{Z}$ yield quasigeodesics and their extremes points are distinct; there exists such a pair since Γ is nonelementary (cf [Ghys and de la Harpe 1990, 37.-Théorème in Section 3, Chapitre 8]). Taking large enough powers of x if necessary, we define for a large enough $R > 0$,

$$U := O(o, x^{-1}, R), \quad V := O(o, x, R), \quad \widetilde{V} := O(x^{-3}, x^{-2}, R) \quad \text{and} \quad \widetilde{U} := O(x^3, x^2, R),$$

such that

$$U \cap V = \varnothing, \quad (\Gamma \cup \partial\Gamma) \setminus U \subset \widetilde{V} \quad \text{and} \quad (\Gamma \cup \partial\Gamma) \setminus V \subset \widetilde{U}.$$

Further, taking large enough powers of y if necessary, we assume that $U' := yU$, $V' := yV$, U and V are disjoint. For a fixed positive constant $L > 0$ and every positive integer n, let

$$S_{n,L} := \{z \in \Gamma : ||z| - n| \le L\} \quad \text{and} \quad S_{n,L}(U, V) := \{z \in S_{n,L} : U \cap zV = \varnothing\},$$

and similarly, $S_{n,L}(V, U)$ and $S_{n,L}(U', V')$.

First we note that if $z \in S_{n,L}(U, V)$, then

$$(x^3 z x^3)\widetilde{V} \subset V \quad \text{and} \quad (x^3 z x^3)^{-1}\widetilde{U} \subset U,$$

since $(x^3 z x^3)\widetilde{V} = (x^3 z)V \subset x^3(\Gamma \cup \partial\Gamma \setminus U) \subset x^3\widetilde{V} = V$, and the latter is analogous. Therefore if we define

$$U_{n,L}(x, x^{-1}, R) := \{z \in S_{n,L} : z^{-1} \in O(o, x^{-1}, R),\ z \in O(o, x, R)\},$$

then

$$x^3 S_{n,L}(U, V) x^3 \subset U_{n,L+6|x|}(x, x^{-1}, R), \tag{3-3}$$

since $o \in \widetilde{U}$ and $o \in \widetilde{V}$. Moreover, we have that

$$x^3 S_{n,L}(V, U)^{-1} x^3 \subset U_{n,L+6|x|}(x, x^{-1}, R) \quad \text{and} \quad y^{-1} S_{n,L}(U', V') y \subset S_{n,L+2|y|}(U, V), \tag{3-4}$$

where the former follows from $S_{n,L}(V, U)^{-1} = S_{n,L}(U, V)$ and (3-3), and the latter holds by the definition of U' and V'. The map $z \mapsto x^3 z x^3$ yields an injection from $S_{n,L}(U, V)$ into $U_{n,L+6|x|}(x, x^{-1}, R)$. Similarly, the map $z \mapsto x^3 z^{-1} x^3$ yields an injection from $S_{n,L}(V, U)$ into $U_{n,L+6|x|}(x, x^{-1}, R)$, and the map $z \mapsto y^{-1} z y$ yields an injection from $S_{n,L}(U', V')$ into $U_{n,L+6|x|+2|y|}(x, x^{-1}, R)$.

Second, let us show that there exists a finite set $F_{U,V}$ in Γ independent of n such that

$$S_{n,L} \setminus F_{U,V} \subset S_{n,L}(U,V) \cup S_{n,L}(V,U) \cup S_{n,L}(U',V'). \tag{3-5}$$

Indeed, if $z \in S_{n,L}$ and z is not included in any one of $S_{n,L}(U,V)$, $S_{n,L}(V,U)$ or $S_{n,L}(U',V')$, then one has

$$U \cap zV \neq \varnothing, \quad V \cap zU \neq \varnothing \quad \text{and} \quad U' \cap zV' \neq \varnothing.$$

Note that those elements z for which $U \times V \times U'$ and $z(V \times U \times V')$ intersect are finite; this follows since $U \times V \times U'$ and $V \times U \times V'$ are in $(\Gamma \cup \partial\Gamma)^{(3)}$, where

$$(\Gamma \cup \partial\Gamma)^{(3)} := \{(\xi,\eta,\zeta) \in (\Gamma \cup \partial\Gamma)^3 : \xi, \eta \text{ and } \zeta \text{ are distinct}\},$$

and the diagonal action of Γ on $(\Gamma \cup \partial\Gamma)^{(3)}$ is properly discontinuous. (Note that it is more standard to state that the diagonal action of Γ on the space of distinct ordered triples of points in the boundary $\partial\Gamma$ is properly discontinuous; the same proof works for the case of $(\Gamma \cup \partial\Gamma)^{(3)}$ where we endow $\Gamma \cup \partial\Gamma$ with the compactified topology, cf [Gromov 1987, 8.2.M] and [Bowditch 1999, Lemma 1.2 and Proposition 1.12].) Hence (3-5) holds for some finite set $F_{U,V}$ in Γ independent of n.

Finally, if $z \in U_{n,L}(x,x^{-1},R)$, then since $z \in V = O(o,x,R)$ and $z^{-1} \in U = O(o,x^{-1},R)$, we have that by the δ-hyperbolicity,

$$|(z|z^{-1})_o - (x|x^{-1})_o| \leq C_{R,\delta}.$$

Lemma 3.2 implies that for all such z,

$$\ell[z] = |z| - 2(z|z^{-1})_o + O_{R,\delta}(1) = |z| + O_{x,R,\delta}(1).$$

The analogous relations hold for $\ell_*[z]$. Given $z \in U_{n,L}(x,x^{-1},R)$, let us count the number of elements in the set

$$C_n(z;x,L,R) := \{g\langle z\rangle \in \Gamma/\langle z\rangle : gzg^{-1} \in U_{n,L}(x,x^{-1},R)\},$$

ie the number of $g \in \Gamma$ modulo powers of z such that $gzg^{-1} \in U_{n,L}(x,x^{-1},R)$; see Figure 1. Let $[o,z]$ denote a C-rough geodesic segment between o and z (with respect to d), and define $\gamma(z) := \bigcup_{k\in\mathbb{Z}} z^k[o,z]$, which is (the image of) a (A,B)-quasigeodesic line invariant under z for some $A, B > 0$. Similarly, $\gamma(gzg^{-1})$ is a (A,B)-quasigeodesic line invariant under gzg^{-1}. Note that $g\gamma(z)$ is also a (A,B)-quasigeodesic line invariant under gzg^{-1}, and thus $g\gamma(z)$ and $\gamma(gzg^{-1})$ lie within a bounded Hausdorff distance. This shows that if $g\langle z\rangle \in C_n(z;x,L,R)$, then $g\gamma(z)$ passes through near o within a bounded distance $C_{A,B,\delta}$ of o. Crucially, A and B depend only on the hyperbolicity constant and the x, R used in $C_n(z;x,L,R)$. Since for all such g the inverse g^{-1} lies in a neighborhood of $[o,z]$ up to translation by z, and $\gamma(z)$ is z-invariant, counting such g^{-1} modulo the $\langle z\rangle$-action on $\gamma(z)$ yields $\#C_n(z;x,L,R) \leq C'_{A,B,\delta}n$. Now we obtain

$$\sum_{z\in U_{n,L}(x,x^{-1},R)} e^{-a|z|_*} \leq C'n \sum_{[z]:|\ell[z]-n|\leq L'} e^{-a\ell_*[z]},$$

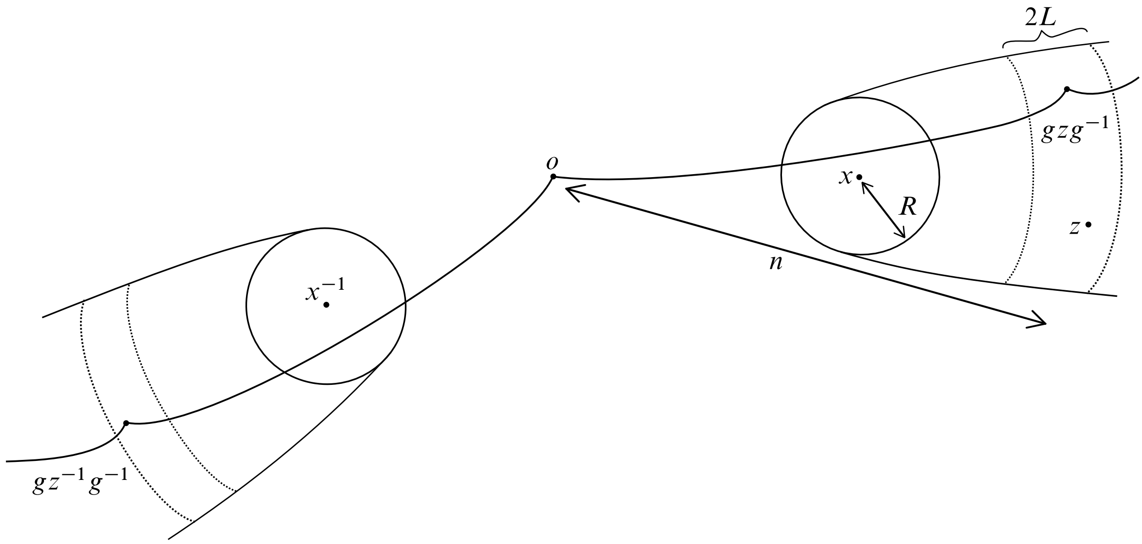


Figure 1

where C' and L' are constants depending only on a, x, R, L and the hyperbolicity constants of d and d_*. Therefore, noting that

$$\big||x^3zx^3|_* - |z|_*\big| \le 6|x|_*, \quad \big||x^3z^{-1}x^3|_* - |z|_*\big| \le 6|x|_* \quad \text{and} \quad \big||y^{-1}zy|_* - |z|_*\big| \le 2|y|_*,$$

we obtain by (3-3) and (3-4) together with (3-5), for all large enough n,

$$\sum_{z \in S_{n,L} \setminus F_{U,V}} e^{-a|z|_*} \le Cn \sum_{[z]:|\ell[z]-n| \le L'} e^{-a\ell_*[z]},$$

where L and L' are large enough fixed constants depending only on x and y. Since $F_{U,V}$ is a finite set of elements independent of n, we have that $\theta(a) \le \Theta(a)$, as required. □

Let

$$\alpha_{\min} := -\lim_{a \to \infty} \frac{\theta(a)}{a} \quad \text{and} \quad \alpha_{\max} := -\lim_{a \to -\infty} \frac{\theta(a)}{a},$$

where $\alpha_{\min}$ and $\alpha_{\max}$ are positive and finite since d and d_* are quasi-isometric. Recall that

$$\mathrm{Dil}_- := \inf_{[x] \in \mathbf{conj}_{>0}} \frac{\ell_*[x]}{\ell[x]} \quad \text{and} \quad \mathrm{Dil}_+ := \sup_{[x] \in \mathbf{conj}_{>0}} \frac{\ell_*[x]}{\ell[x]}$$

where $\mathbf{conj}_{>0}$ is the set of elements $[x] \in \mathbf{conj}$ such that $\ell[x]$ (and hence $\ell_*[x]$) is nonzero.

Corollary 3.3 *For all $d, d_* \in \mathscr{D}_\Gamma$, we have*

$$\mathrm{Dil}_- = \alpha_{\min} \quad \textit{and} \quad \mathrm{Dil}_+ = \alpha_{\max}.$$

Proof Fix a large enough $L > 0$. For all $a > 0$ and all integers $n \ge 0$, we have that

$$\sum_{[x]:\ |\ell[x]-n| \le L} e^{-a\ell_*[x]} \le \sum_{[x]:\ |\ell[x]-n| \le L} e^{-a\mathrm{Dil}_-\ell[x]},$$

which together with Proposition 3.1 implies that $\theta(a) \le -a\mathrm{Dil}_- + O(1)$ and thus $\alpha_{\min} \ge \mathrm{Dil}_-$. Further for all $\varepsilon > 0$, there exist infinitely many $[x] \in \mathbf{conj}_{>0}$ such that

$$\frac{\ell_*[x]}{\ell[x]} \le \mathrm{Dil}_- + \varepsilon.$$

Hence for all $a > 0$ and infinitely many integers $n \ge 0$,

$$\sum_{[x]:\ |\ell[x]-n|\le L} e^{-a\ell_*[x]} \ge c e^{-a(\mathrm{Dil}_- + \varepsilon)n},$$

where c is a positive constant independent of n, and thus by Proposition 3.1,

$$\theta(a) \ge -a(\mathrm{Dil}_- + \varepsilon).$$

Therefore $\alpha_{\min} \le \mathrm{Dil}_-$. We obtain $\alpha_{\min} = \mathrm{Dil}_-$. Showing that $\alpha_{\max} = \mathrm{Dil}_+$ is analogous. □

Lemma 3.4 *Let Γ be a nonelementary hyperbolic group and $d, d_* \in \mathscr{D}_\Gamma$. The following are equivalent*:

(1) *The Manhattan curve $\mathscr{C}_M$ for the pair d and d_* is a straight line.*

(2) *There exists a constant $c > 0$ such that $\ell[x] = c\ell_*[x]$ for all $[x] \in \mathbf{conj}$.*

Proof If the Manhattan curve $\mathscr{C}_M$ is a straight line, then $\mathrm{Dil}_- = \mathrm{Dil}_+$. Furthermore, $-\mathrm{Dil}_-$ and $-\mathrm{Dil}_+$ are equal to the gradient of the Manhattan curve. Since $\mathscr{C}_M$ goes through the points $(0, v)$ and $(v_*, 0)$, this gradient is $-v/v_*$, where $v, v_* > 0$ if Γ is nonelementary. Hence for all $[x] \in \mathbf{conj}_{>0}$,

$$\frac{v}{v_*} = \mathrm{Dil}_- = \inf_{[g]\in\mathbf{conj}_{>0}} \frac{\ell_*[g]}{\ell[g]} \le \frac{\ell_*[x]}{\ell[x]} \le \sup_{[g]\in\mathbf{conj}_{>0}} \frac{\ell_*[g]}{\ell[g]} = \mathrm{Dil}_+ = \frac{v}{v_*},$$

implying that $v\ell[x] = v_*\ell_*[x]$ for all $[x] \in \mathbf{conj}$. The converse follows from the definition of the Manhattan curve. □

3.2 Proof of the C^1-regularity

Fix a pair of metrics $d, d_* \in \mathscr{D}_\Gamma$. For each $\xi \in \partial\Gamma$ and quasigeodesic $\gamma_\xi : [0, \infty) \to (\Gamma, d)$ with $\gamma_\xi(t) \to \xi$ as $t \to \infty$, we define

$$\tau_{\inf}(\xi) := \liminf_{t\to\infty} \frac{d_*(\gamma_\xi(0), \gamma_\xi(t))}{d(\gamma_\xi(0), \gamma_\xi(t))} \quad \text{and} \quad \tau_{\sup}(\xi) := \limsup_{t\to\infty} \frac{d_*(\gamma_\xi(0), \gamma_\xi(t))}{d(\gamma_\xi(0), \gamma_\xi(t))}.$$

The Morse lemma (applied to (Γ, d)) implies that $\tau_{\inf}(\xi)$ and $\tau_{\sup}(\xi)$ are independent of the choice of quasigeodesics converging to ξ, or of their starting points, respectively. If $\tau_{\inf}(\xi) = \tau_{\sup}(\xi)$, then we denote the common value by $\tau(\xi)$ and call it the *local intersection number* at ξ for the pair (d, d_*).

Lemma 3.5 *For each $(a, b) \in \mathscr{C}_M$, we have that $\tau_{\inf}(\xi) = \tau_{\sup}(\xi)$ for $\mu_{a,b}$-almost every $\xi \in \partial\Gamma$, and further, there exists a constant $\tau_{a,b}$ such that*

$$\tau(\xi) = \tau_{a,b} \quad \text{for } \mu_{a,b}\text{-almost every } \xi \in \partial\Gamma.$$

Moreover, $\tau_{a,b}$ is continuous in $(a, b) \in \mathscr{C}_M$.

Proof Let $\widehat{d}$ be a strongly hyperbolic metric in $\mathscr{D}_\Gamma$. We consider the map $\mathbf{ev}\colon \partial^2\Gamma \times \mathbb{R} \to (\Gamma, \widehat{d}\,)$ and the flow space $\mathscr{F}_\kappa$, where κ is the cocycle associated with $\widehat{d}$ defined in Section 2.7. We write $w_t := \Phi_t(w)$ for $w \in \mathscr{F}_\kappa$. Taking a measurable section $\iota\colon \mathscr{F}_\kappa \to D$ for a Borel fundamental domain D in $\partial^2\Gamma \times \mathbb{R}$ for the (Γ, κ)-action, we define $\widetilde{w} := \iota(w)$ and set $\widetilde{w}_t := \widetilde{\Phi}_t(\widetilde{w})$. Let

$$c_*(w_s, w_t) := d_*(\mathbf{ev}(\widetilde{w}_s), \mathbf{ev}(\widetilde{w}_t)).$$

Then $c_*(w_s, w_t)$ is subadditive, ie for all $t, s \in [0, \infty)$,

$$c_*(w_0, w_{s+t}) \le c_*(w_0, w_s) + c_*(w_s, w_{s+t}), \tag{3-6}$$

and superadditive up to an additive constant, ie there exists a $C \ge 0$ such that for all $t, s \in [0, \infty)$,

$$c_*(w_0, w_{s+t}) \ge c_*(w_0, w_s) + c_*(w_s, w_{s+t}) - C, \tag{3-7}$$

by the Morse lemma on (Γ, d_*). Since $m_{a,b}$ is ergodic with respect to the flow on $\mathscr{F}_\kappa$, the Kingman subadditive ergodic theorem implies that there exists a constant $\chi_*(a,b)$ such that

$$\lim_{t\to\infty} \frac{1}{t} c_*(w_0, w_t) = \chi_*(a,b) \quad \text{for } m_{a,b}\text{-almost every } w \text{ in } \mathscr{F}_\kappa,$$
$$\lim_{t\to\infty} \frac{1}{t} \int_{\mathscr{F}_\kappa} c_*(w_0, w_t)\, dm_{a,b} = \chi_*(a,b).$$

Let us show that $\chi_*(a,b)$ is continuous in $(a,b) \in \mathscr{C}_M$. For each $(a_0, b_0) \in \mathscr{C}_M$, if $(a,b) \to (a_0, b_0)$, then $m_{a,b}$ weakly converges to $m_0 := m_{a_0,b_0}$ by Corollary 2.13. Then the subadditivity (3-6) yields for all $(a,b) \in \mathscr{C}_M$ and all $t > 0$,

$$\frac{1}{t} \int_{\mathscr{F}_\kappa} c_*(w_0, w_t)\, dm_{a,b} \ge \inf_{t>0} \frac{1}{t} \int_{\mathscr{F}_\kappa} c_*(w_0, w_t)\, dm_{a,b} = \chi_*(a,b),$$

so we have that for each $t \ge 0$,

$$\frac{1}{t} \int_{\mathscr{F}_\kappa} c_*(w_0, w_t)\, dm_0 \ge \limsup_{a\to a_0} \chi_*(a,b),$$

and similarly, (3-7) implies that for each $t > 0$,

$$\frac{1}{t} \int_{\mathscr{F}_\kappa} (c_*(w_0, w_t) - C)\, dm_0 \le \liminf_{a\to a_0} \chi_*(a,b).$$

Letting $t \to \infty$, we obtain $\lim_{a\to a_0} \chi_*(a,b) = \chi_*(a_0, b_0)$, ie $\chi_*(a,b)$ is continuous in $(a,b) \in \mathscr{C}_M$.

We apply the same discussion to d: letting

$$c(w_s, w_t) := d(\mathbf{ev}(\widetilde{w}_s), \mathbf{ev}(\widetilde{w}_t)),$$

we have that there exists a constant $\chi(a,b)$ such that

$$\lim_{t\to\infty} \frac{1}{t} c(w_0, w_t) = \chi(a,b) \quad \text{for } m_{a,b}\text{-almost every } w \text{ in } \mathscr{F}_\kappa,$$

and $\chi(a,b)$ is continuous in $(a,b) \in \mathscr{C}_M$.

Therefore for $m_{a,b}$-almost every $w = [\xi_-, \xi_+, t_0] \in \mathscr{F}_\kappa$,

$$\lim_{t\to\infty} \frac{c_*(w_0, w_t)}{c(w_0, w_t)} = \frac{\chi_*(a,b)}{\chi(a,b)}.$$

Recall that $\Lambda_{a,b} \otimes dt = \sum_{x\in\Gamma} x_*(\iota_* m_{a,b})$, $\Lambda_{a,b}$ is equivalent to $\mu_{a,b} \otimes \mu_{a,b}$ and that d_* and d are left-invariant. Hence if we define $\tau_{a,b} := \chi_*(a,b)/\chi(a,b)$, then $\tau_{\inf}(\xi) = \tau_{\sup}(\xi)$ for $\mu_{a,b}$-almost every $\xi \in \partial\Gamma$, and

$$\tau(\xi) = \tau_{a,b} \quad \text{for } \mu_{a,b}\text{-almost every } \xi \in \partial\Gamma.$$

Since $\chi(a,b)$ and $\chi_*(a,b)$ are positive and continuous, $\tau_{a,b}$ is continuous in $(a,b) \in \mathscr{C}_M$, as required. □

For every real value $r \in \mathbb{R}$, let

$$E_r := \{\xi \in \partial\Gamma : \tau_{\inf}(\xi) = \tau_{\sup}(\xi) = r\}.$$

The set E_r is possibly empty for some r. Note that a point ξ is in E_r if and only if for some (equivalently, every) quasigeodesic γ_ξ converging to ξ, we have

$$\lim_{t\to\infty} \frac{d_*(\gamma_\xi(0), \gamma_\xi(t))}{d(\gamma_\xi(0), \gamma_\xi(t))} = r.$$

Recall that the Manhattan curve $\mathscr{C}_M$ is the graph of the function θ, ie $(a,b) \in \mathscr{C}_M$ if and only if $b = \theta(a)$, and since θ is convex, it is differentiable except for at most countably many points.

Lemma 3.6 *Fix a pair $d, d_* \in \mathscr{D}_\Gamma$. For each $(a,b) \in \mathscr{C}_M$, if θ is differentiable at a and $r = -\theta'(a)$, then $\mu_{a,b}(E_r) = 1$.*

Proof Fix a large enough constant $C \geq 0$. Let us endow the space of C-rough geodesic rays from o in (Γ, d) with the pointwise convergence topology. For each $\xi \in \partial\Gamma$, we associate a C-rough geodesic γ_ξ from o to ξ, and define this correspondence in a Borel measurable way as in Section 2.7, where we have done the same but for rough geodesics. For each nonnegative integer n and $\xi \in \partial\Gamma$, we abbreviate notation by writing $\xi_n := \gamma_\xi(n)$ and $|\xi_n| := d(o, \xi_n)$ (resp. $|\xi_n|_* := d_*(o, \xi_n)$). We use $O(x, R)$ to denote the shadows associated to the metric d, for a large enough thickness parameter R.

For $(a,b) \in \mathscr{C}_M$, let us suppose that $r = -\theta'(a)$. For a Patterson–Sullivan measure μ_* for d_*, we show that

$$\liminf_{n\to\infty} -\frac{1}{v_*|\xi_n|} \log \mu_*(O(\xi_n, R)) \geq r \quad \text{for } \mu_{a,b}\text{-almost every } \xi \in \partial\Gamma. \tag{3-8}$$

For every $\varepsilon > 0$, the Markov inequality shows the following: for every $s > 0$, integrating

$$\mathbf{1}_{\{\xi\in\partial\Gamma:\mu_*(O(\xi_n,R))\geq e^{-(r-\varepsilon)v_*|\xi_n|}\}} \leq \mu_*(O(\xi_n, R))^s \cdot e^{s(r-\varepsilon)v_*|\xi_n|}$$

over $\xi \in \partial\Gamma$ with respect to $\mu_{a,b}$ yields

$$\mu_{a,b}(\{\xi \in \partial\Gamma : \mu_*(O(\xi_n, R)) \geq e^{-(r-\varepsilon)v_*|\xi_n|}\}) \leq \int_{\partial\Gamma} \mu_*(O(\xi_n, R))^s e^{s(r-\varepsilon)v_*|\xi_n|} \, d\mu_{a,b}(\xi).$$

Since

$$\mu_*(O(\xi_n,R)) \asymp_R \exp(-v_*|\xi_n|_*) \quad \text{and} \quad \mu_{a,b}(O(\xi_n,R)) \asymp_R \exp(-a|\xi_n|_* - b|\xi_n|),$$

the integral on the right-hand side is at most

$$C_R \sum_{x\in S(n,R)} \exp(-sv_*|x|_* + s(r-\varepsilon)v_*|x| - a|x|_* - b|x|), \tag{3-9}$$

where we recall that $S(n,R)=\{x\in\Gamma : |d(o,x)-n|\le R\}$, up to a multiplicative constant depending only on R and the δ-hyperbolicity constant of d. Moreover, (3-9) is at most

$$C'_R \exp(s(r-\varepsilon)v_*n - bn) \sum_{x\in S(n,R)} \exp(-sv_*|x|_* - a|x|_*) \le C''_R \exp(s(r-\varepsilon)v_*n - bn + \theta(sv_*+a)n),$$

where we have used Lemma 2.8:

$$\sum_{x\in S(n,R)} \exp(-sv_*|x|_* - a|x|_*) \asymp \exp(\theta(sv_*+a)n).$$

Since $b=\theta(a)$ and $r=-\theta'(a)$,

$$\theta(sv_*+a)-\theta(a) = -rsv_* + o(s) \quad \text{as } s\searrow 0,$$

and we obtain

$$\begin{aligned}\mu_{a,b}(\{\xi\in\partial\Gamma : \mu_*(O(\xi_n,R)) \ge e^{-(r-\varepsilon)v_*|\xi_n|}\}) &\le C''_R \exp(s(r-\varepsilon)v_*n + (\theta(sv_*+a)-\theta(a))n)\\ &\le C''_R \exp(-s\varepsilon v_*n + o(s)n) \le C''_R \exp(-c(\varepsilon,s)n)\end{aligned}$$

for some constant $c(\varepsilon,s)>0$ for all $n\ge 0$. Hence the Borel–Cantelli lemma shows that

$$\liminf_{n\to\infty} -\frac{1}{v_*|\xi_n|}\log \mu_*(O(\xi_n,R)) \ge r-\varepsilon \quad \text{for } \mu_{a,b}\text{-almost every } \xi\in\partial\Gamma,$$

and since this holds for every $\varepsilon>0$, we obtain (3-8).

Similarly, it holds that

$$\limsup_{n\to\infty} -\frac{1}{v_*|\xi_n|}\log \mu_*(O(\xi_n,R)) \le r \quad \text{for } \mu_{a,b}\text{-almost every } \xi\in\partial\Gamma. \tag{3-10}$$

Indeed, for every $\varepsilon>0$ and every $s>0$, we have that

$$\mu_{a,b}(\{\xi\in\partial\Gamma : \mu_*(O(\xi_n,R)) \le e^{-(r+\varepsilon)v_*|\xi_n|}\}) \le \int_{\partial\Gamma} \mu_*(O(\xi_n,R))^{-s} e^{-s(r+\varepsilon)v_*|\xi_n|}\, d\mu_{a,b}(\xi),$$

and the rest follows as in the same way above; we omit the details.

Combining (3-8) and (3-10), we obtain

$$\lim_{n\to\infty} \frac{|\xi_n|_*}{|\xi_n|} = \lim_{n\to\infty} -\frac{1}{v_*|\xi_n|}\log \mu_*(O(\xi_n,R)) = r$$

for $\mu_{a,b}$-almost every $\xi\in\partial\Gamma$. Therefore we have that $\tau_{\inf}(\xi)=\tau_{\sup}(\xi)$ for $\mu_{a,b}$-almost every $\xi\in\partial\Gamma$, and $\mu_{a,b}(E_r)=1$ if $b=\theta(a)$ and $r=-\theta'(a)$, as required. □

Theorem 3.7 *For every pair $d, d_* \in \mathscr{D}_\Gamma$, the Manhattan curve $\mathscr{C}_M$ is C^1, ie the function θ is continuously differentiable on $\mathbb{R}$. Moreover, $\theta'(a) = -\tau_{a,b}$ for all $(a,b) \in \mathscr{C}_M$.*

Proof Recall that since θ is convex, θ is differentiable except for at most countably many points. For each $(a,b) \in \mathscr{C}_M$, if $r = -\theta'(a)$, then Lemma 3.6 implies that $\tau(\xi) = r$ for $\mu_{a,b}$-almost every $\xi \in \partial\Gamma$; on the other hand, Lemma 3.5 implies that $\tau(\xi) = \tau_{a,b}$ for $\mu_{a,b}$-almost every $\xi \in \partial\Gamma$. Therefore if $b = \theta(a)$ and $r = -\theta'(a)$, then

$$\theta'(a) = -\tau_{a,b}.$$

Since this holds for Lebesgue almost every a in $\mathbb{R}$ and $\tau_{a,b}$ is continuous in $(a,b) \in \mathscr{C}_M$ by Lemma 3.5, θ is differentiable everywhere and the derivative coincides with $-\tau_{a,b}$, which is continuous. □

The above proof yields the multifractal spectrum of every Patterson–Sullivan measure μ_* with respect to $\rho(\xi,\eta) = \exp(-(\xi|\eta)_o)$ in $\partial\Gamma$, and the profile is the Legendre transform of the Manhattan curve.

Theorem 3.8 (the multifractal spectrum) *For every pair $d, d_* \in \mathscr{D}_\Gamma$, let μ_* be an arbitrary Patterson–Sullivan measure relative to d_* and $\rho(\xi,\eta) = \exp(-(\xi|\eta)_o)$ be the quasimetric relative to d on $\partial\Gamma$. For $\alpha \in \mathbb{R}$ we define*

$$E(\alpha) := \left\{\xi \in \partial\Gamma : \lim_{r\to 0} \frac{\log \mu_*(B(\xi,r))}{\log r} = \alpha\right\},$$

where $B(\xi,r) = \{\eta \in \partial\Gamma : \rho(\xi,\eta) < r\}$. Then we have

$$\dim_{\mathrm{H}}(E(v_*\alpha),\rho) = \inf_{a\in\mathbb{R}}\{a\alpha + \theta(a)\} \quad \text{for } \alpha \in (\alpha_{\min},\alpha_{\max}), \tag{3-11}$$

where

$$\alpha_{\min} = -\lim_{a\to\infty}\frac{\theta(a)}{a} \quad \text{and} \quad \alpha_{\max} = -\lim_{a\to-\infty}\frac{\theta(a)}{a}.$$

Proof Note that the function θ is C^1 and $\theta'(a) = -\tau_{a,b}$ for all $(a,b) \in \mathscr{C}_M$ by Theorem 3.7. Hence Lemmas 3.6 and 3.5 together with Lemma 2.3 imply that for all $(a,b) \in \mathscr{C}_M$,

$$\lim_{r\to 0}\frac{\log \mu_*(B(\xi,r))}{\log r} = v_*\tau_{a,b} \qquad \text{for } \mu_{a,b}\text{-almost every } \xi \in \partial\Gamma,$$

$$\lim_{r\to 0}\frac{\log \mu_{a,b}(B(\xi,r))}{\log r} = a\tau_{a,b} + b \quad \text{for } \mu_{a,b}\text{-almost every } \xi \in \partial\Gamma,$$

where we have used $\mu_{a,b}(O(x,R)) \asymp_R \exp(-ad_*(o,x) - bd(o,x))$ for $x \in \Gamma$. The Frostman-type lemma (Lemma 2.4) shows that

$$\dim_{\mathrm{H}}(E(v_*\tau_{a,b}),\rho) = a\tau_{a,b} + b.$$

Since θ is continuously differentiable and convex, for each $\alpha \in (\alpha_{\min},\alpha_{\max})$ there exists $a \in \mathbb{R}$ such that $\alpha = -\theta'(a)$, and

$$\dim_{\mathrm{H}}(E(v_*\alpha),\rho) = -a\theta'(a) + \theta(a),$$

where the right-hand side is the Legendre transform of θ. Therefore we conclude the claim. □

Remark 3.9 If we have

$$\theta(a) = -a\alpha_{\min} + O(1) \quad \text{as } a \to \infty \qquad \text{and} \qquad \theta(a) = -a\alpha_{\max} + O(1) \quad \text{as } a \to -\infty,$$

then the formula (3-11) is valid for all $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ including two extreme points $\alpha_{\min}$ and $\alpha_{\max}$. This is the case for example when both d_* and d are word metrics; see Proposition 4.22 in the following Section 4.

3.3 Rough similarity rigidity

In this section we prove Theorem 1.2 and the rigidity statement in Theorem 1.1. We begin with the following.

Theorem 3.10 *For every pair $d, d_* \in \mathcal{D}_\Gamma$, the following are equivalent*:

(i) *The Manhattan curve $\mathcal{C}_M$ is a straight line between $(0, v)$ and $(v_*, 0)$, where v and v_* are the exponential volume growth rates of (Γ, d) and (Γ, d_*), respectively.*

(ii) *The metrics d_* and d are roughly similar.*

We use the following lemma in the proof. Recall that $d \in \mathcal{D}_\Gamma$ is a roughly geodesic metric and there exists a constant $C \geq 0$ such that for all $\xi \in \partial\Gamma$, one may take a C-rough geodesic ray γ_ξ from o converging to ξ on (Γ, d).

Lemma 3.11 *Let ν be a finite Borel regular measure on $\partial\Gamma$ and μ be a doubling measure on $\partial\Gamma$ relative to a quasimetric ρ for $d \in \mathcal{D}_\Gamma$. If we decompose $\nu = \nu_{ac} + \nu_{sing}$, where ν_{ac} is the absolutely continuous part of ν and ν_{sing} is the singular part of ν relative to μ, then for a large enough $R > 0$,*

$$\limsup_{n\to\infty} \frac{\nu_{ac}(O(\xi_n, R))}{\mu(O(\xi_n, R))} < \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{\nu_{sing}(O(\xi_n, R))}{\mu(O(\xi_n, R))} = 0$$

for μ-almost every $\xi \in \partial\Gamma$, where $\xi_n := \gamma_\xi(n)$ for $n \geq 1$.

The proof of Lemma 3.11 follows from the classical Lebesgue differentiation theorem and the weak maximal inequality — we include a proof for the sake of completeness in the appendix.

Proof of Theorem 3.10 If (ii) holds, then the Manhattan curve $\mathcal{C}_M$ is actually a straight line on $\mathbb{R}$ since $\tau_{a,b} = \tau$ for a constant $\tau > 0$ for all $(a, b) \in \mathcal{C}_M$ by Lemma 3.5 and $\theta'(a) = -\tau$ for all $a \in \mathbb{R}$ by Theorem 3.7 (or by Lemma 3.4).

Suppose that (i) holds. Then $(a, b) := (v_*/2, v/2) \in \mathcal{C}_M$. By Corollary 2.10(1), we have that for all $x \in \Gamma$,

$$\mu_{a,b}(O(x, R)) \asymp \exp\left(-\frac{v_*}{2}|x|_* - \frac{v}{2}|x|\right),$$

$$\mu_*(O(x, R)) \asymp \exp(-v_*|x|_*) \quad \text{and} \quad \mu(O(x, R)) \asymp \exp(-v|x|). \tag{3-12}$$

This implies that

$$\frac{\mu_*(O(x,R))}{\mu_{a,b}(O(x,R))}\cdot\frac{\mu(O(x,R))}{\mu_{a,b}(O(x,R))}\asymp 1 \quad \text{for all } x\in\Gamma. \tag{3-13}$$

Fix a large enough $R>0$. Letting $\xi_n := \gamma_\xi(n)$ for integers $n \ge 0$, we have that

$$\limsup_{n\to\infty}\frac{\mu_*(O(\xi_n,R))}{\mu_{a,b}(O(\xi_n,R))}<\infty \quad \text{and} \quad \limsup_{n\to\infty}\frac{\mu(O(\xi_n,R))}{\mu_{a,b}(O(\xi_n,R))}<\infty$$

for $\mu_{a,b}$-almost every $\xi\in\partial\Gamma$ by Lemma 3.11. Here we are using that μ_* and μ are finite Borel regular measures and that $\mu_{a,b}$ is doubling relative to a quasimetric ρ in $\partial\Gamma$. Hence if either μ_* and $\mu_{a,b}$, or μ and $\mu_{a,b}$ are mutually singular, then Lemma 3.11 together with (3-13) leads to a contradiction. Therefore both μ_* and μ have nonzero absolutely continuous parts relative to $\mu_{a,b}$, and thus for the corresponding Γ-invariant Radon measures Λ_*, Λ and $\Lambda_{a,b}$ for μ_*, μ and $\mu_{a,b}$, respectively, both Λ_* and Λ have nonzero absolutely continuous parts relative to $\Lambda_{a,b}$. By Lemma 2.12, there exist positive constants $c,c'>0$ such that $\Lambda_* = c\Lambda_{a,b}$ and $\Lambda = c'\Lambda_{a,b}$. In particular, $\Lambda_* = (c/c')\Lambda$ and this implies that μ_* and μ are mutually absolutely continuous. Letting $\varphi := d\mu_*/d\mu$, we shall show that φ is uniformly bounded away from 0 and from above. We have that

$$\varphi(\xi)\varphi(\eta)e^{2v_*(\xi|\eta)_{*o}}\asymp e^{2v(\xi|\eta)_o} \quad \text{for } (\xi,\eta)\in\partial^2\Gamma.$$

If φ is unbounded on $B(\xi,\varepsilon)$ for all $\varepsilon>0$, then for a fixed $\eta\neq\xi$ such that $\varphi(\eta)>0$, it is possible that for $\xi'\in B(\xi,\varepsilon)$, the value $\varphi(\xi')\varphi(\eta)$ is arbitrarily large; however, $(\xi'|\eta)_{*o}$ and $(\xi'|\eta)_o$ are uniformly bounded, and this is a contradiction. This shows that $\mu_*\asymp\mu$ and thus by the above estimates (3-12), there exists a constant $C\ge 0$ such that

$$|v_*|x|_* - v|x|| \le C \quad \text{for all } x\in\Gamma,$$

ie d_* and d are roughly similar; we conclude the claim. □

We can now conclude the proof of our first main result.

Proof of Theorem 1.1 Combining Theorems 3.7 and 3.10 concludes the proof of Theorem 1.1. □

Let us now move on to the proof of Theorem 1.2. We will break the proof into two parts. For every pair $d,d_*\in\mathscr{D}_\Gamma$ define

$$\tau(d_*/d) := \limsup_{r\to\infty}\frac{1}{\#B(o,r)}\sum_{x\in B(o,r)}\frac{d_*(o,x)}{r},$$

where $B(o,r) := \{x\in\Gamma : d(o,x)\le r\}$ for a real $r>0$. We begin by proving the following.

Theorem 3.12 *For every pair $d,d_*\in\mathscr{D}_\Gamma$, the following limit exists:*

$$\tau(d_*/d) = \lim_{r\to\infty}\frac{1}{\#B(o,r)}\sum_{x\in B(o,r)}\frac{d_*(o,x)}{r},$$

and $\tau_{0,v} = \tau(d_*/d)$*. Moreover, we have that*

$$\tau(d_*/d) \geq \frac{v}{v_*},$$

where v *and* v_* *are the exponential volume growth rates of* (Γ, d) *and* (Γ, d_*)*, respectively.*

Proof Fix a pair $d, d_* \in \mathscr{D}_\Gamma$ and consider the point $(0, v)$ on the associated Manhattan curve $\mathscr{C}_M$. By Lemma 3.5, there exists a constant $\tau_{0,v}$ such that

$$\tau(\xi) = \tau_{0,v} \quad \text{for } \mu_{0,v}\text{-almost every } \xi \in \partial\Gamma,$$

where we note that $\mu_{0,v}$ is a Patterson–Sullivan measure for the metric d. In particular, for $\mu_{0,v}$-almost every $\xi \in \partial\Gamma$, $d_*(o, \gamma_\xi(n))/n \to \tau_{0,v}$ as $n \to \infty$, where γ_ξ is an arbitrary rough geodesic ray (with respect to d) starting from o. Let us define

$$A_{n,\varepsilon} := \left\{ x \in \Gamma : \frac{|d_*(o,x) - n\tau_{0,v}|}{n} > \varepsilon \right\} \quad \text{for } n \geq 0 \text{ and } \varepsilon > 0.$$

Consider $S(n, R) := \{x \in \Gamma : |d(o,x) - n| \leq R\}$ and fix a sufficiently large $R_0 > 0$. Since the shadows $O(x, R_0)$ for $x \in S(n, R)$ cover the boundary $\partial\Gamma$ with a bounded multiplicity, we have

$$\frac{\#(A_{n,\varepsilon} \cap S(n,R))}{\#S(n,R)} \leq C \sum_{x \in A_{n,\varepsilon} \cap S(n,R)} \mu_{0,v}(O(x, R_0)) \leq C' \mu_{0,v}\Bigg(\bigcup_{x \in A_{n,\varepsilon} \cap S(n,R)} O(x, R_0) \Bigg).$$

Note that the last term tends to 0 as $n \to \infty$ since if ξ belongs to $O(x, R_0)$ for some $x \in A_{n,\varepsilon} \cap S(n, R)$, then $|d_*(o, \gamma_\xi(n)) - n\tau_{0,v}| \geq \varepsilon n - R_0 L_*$, where

$$L_* := \sup\{d_*(o,x) : d(o,x) \leq R_0\}.$$

This shows that if x is sampled uniformly at random from $S(n, R)$, then for all $\varepsilon > 0$ and for all large enough n, we have $|d_*(o,x) - n\tau_{0,v}| \leq \varepsilon n$ with probability at least $1 - \varepsilon$, implying that

$$\tau_{0,v} = \lim_{n \to \infty} \frac{1}{\#S(n,R)} \sum_{x \in S(n,R)} \frac{d_*(o,x)}{n}$$

for all large enough R. For all real $r > 0$, let us take $n := \lfloor r \rfloor$ the largest integer at most r. Note that if x_n is sampled uniformly at random from the ball $B(o, r)$, then we have $x_n \in A_{n,\varepsilon}$ with probability at most $O(e^{-vR})$ for all large enough n, since the probability that x is not in $S(n, R)$ is at most $O(e^{-vR})$ (following from Lemma 2.8: $\#S(n, R) \asymp_R e^{vn}$). Therefore first letting $r \to \infty$ and then $R \to \infty$, we obtain

$$\tau_{0,v} = \lim_{r \to \infty} \frac{1}{\#B(o,r)} \sum_{x \in B(o,r)} \frac{d_*(o,x)}{r},$$

and thus $\tau(d_*/d) = \tau_{0,v}$. Furthermore, this reasoning shows that for each fixed, sufficiently large R,

$$\#B_*(o, (\tau_{0,v} + \varepsilon)r) \geq (1 - O(e^{-vR})) \cdot \#B(o,r) \quad \text{as } r \to \infty,$$

where $B_*(o, R)$ stands for the ball of radius R centered at o with respect to d_*. Therefore $\tau(d_*/d) \geq v/v_*$, where v and v_* are exponential volume growth rate relative to d and d_*, respectively. □

We can now conclude the proof of Theorem 1.2.

Proof of Theorem 1.2 We have already proven the first part of the theorem in Theorem 3.12. Let us show the equivalence of statements (1), (2) and (3). Note that the equivalence of (2) and (3) is a consequence of Lemma 3.4 and Theorem 3.10. (If two metrics d, d_* are roughly similar, then the corresponding Manhattan curve is a straight line on the entire part, not just on the part connecting $(0, v)$ and $(v_*, 0)$.) We therefore just need to prove the equivalence of (1) and (3), which we prove below:

Consider the Manhattan curve $\mathscr{C}_M$ and the function $\theta(a)$ for the pair (d, d_*) and recall that, by Theorem 3.7, we have that $\theta'(0) = -\tau(d_*/d)$. It follows, since the curve $\mathscr{C}_M$ passes through $(0, v)$ and $(v_*, 0)$, that $\tau(d_*/d) = v/v_*$ if and only if θ is a straight line on $[0, v_*]$. By Theorem 3.10 this is the case if and only if d and d_* are roughly similar. This concludes the proof. □

Let us record the following result on the asymptotics of a typical ratio between two stable translation lengths, as it is of interest in its own right.

Corollary 3.13 *For all $d, d_* \in \mathscr{D}_\Gamma$, we have that*

$$\frac{\ell_*[\gamma_\xi(t)]}{\ell[\gamma_\xi(t)]} \to \tau(d_*/d) \quad \text{as } t \to \infty \text{ for } \mu\text{-almost every } \xi \in \partial\Gamma,$$

where μ is a Patterson–Sullivan measure relative to d and γ_ξ is a quasigeodesic ray γ_ξ converging to ξ.

Proof This follows from Lemmas 3.2 and 3.5 since $\tau_{0,v} = \tau(d_*/d)$ by Theorem 3.12. □

4 The C^2 regularity for strongly hyperbolic metrics

The aim of this section is to deduce better regularity (ie higher-order differentiability) for the Manhattan curve under the additional assumption that d and d_* are strongly hyperbolic metrics; see Definition 2.2 in Section 2.1. The method we use also applies to word metrics, in which case it is (in principle) possible to compute explicit examples; we provide some in the subsequent section (Section 5). We will use automatic structures to introduce a symbolic coding for our group Γ. This will allow us to use techniques from thermodynamic formalism. We begin with some introductory material on these techniques. For the thermodynamic formalism on nontopologically transitive systems, we follow [Gouëzel 2014, Sections 3.2 and 3.3].

4.1 Automatic structures

Fix a finite (symmetric) set of generators S for Γ. An automaton $\mathscr{A} = (\mathscr{G}, \pi, S)$ is a triple consisting of a finite directed graph $\mathscr{G} = (V, E, s_*)$, where s_* is a distinguished vertex called the initial state, a labeling $\pi : E \to S$ on edges by S and a finite (symmetric) set of generators S. Associated to every directed path

$\omega = (e_0, e_1, \dots, e_{n-1})$ in the graph $\mathcal{G}$ where the terminus of e_i is the origin of e_{i+1}, there is a path $\pi(\omega)$ in the Cayley graph $\mathrm{Cay}(\Gamma, S)$ issuing from the identity id, $\pi(e_0)$, $\pi(e_0)\pi(e_1), \dots, \pi(e_0)\cdots\pi(e_{n-1})$. Let us denote by $\pi_*(\omega)$ the terminus of the path $\pi(\omega)$, ie $\pi_*(\omega) := \pi(e_0)\cdots\pi(e_{n-1})$.

Definition 4.1 An automaton $\mathcal{A} = (\mathcal{G}, \pi, S)$, where $\mathcal{G} = (V, E, s_*)$ and $\pi : E \to S$ is a labeling, is called a *strongly Markov automatic structure* if

(1) for every vertex $v \in V$ there is a directed path from the initial state s_* to v,

(2) for every directed path ω in $\mathcal{G}$ the associated path $\pi(\omega)$ is a geodesic in the Cayley graph $\mathrm{Cay}(\Gamma, S)$, and

(3) the map π_* evaluating the terminus of a path yields a bijection from the set of directed paths from s_* in $\mathcal{G}$ to Γ.

We sometimes abuse notation by identifying $\mathcal{A}$ with the underlying finite directed graph $\mathcal{G}$. By a theorem of Cannon [1984] every hyperbolic group admits a strongly Markov automatic structure for every finite symmetric set of generators S; cf [Calegari 2013]. Given an automatic structure $\mathcal{A} = (\mathcal{G}, \pi, S)$ for (Γ, S), we write Σ^* for the set of finite directed paths in $\mathcal{G}$ (not necessarily starting from s_*) and Σ for the set of semi-infinite directed paths $\omega = (\omega_i)_{i=0,1,\dots}$ in $\mathcal{G}$. Let $\bar{\Sigma} := \Sigma^* \cup \Sigma$. The function $\pi_* : \Sigma^* \to \Gamma$ naturally extends to

$$\pi_* : \bar{\Sigma} \to \Gamma \cup \partial\Gamma, \quad \omega \mapsto \pi_*(\omega),$$

by mapping a sequence to the terminus of the geodesic segment or ray $\pi(\omega)$ starting at id in $\mathrm{Cay}(\Gamma, S)$. We define a metric $d_{\bar{\Sigma}}$ on $\bar{\Sigma}$ by $d_{\bar{\Sigma}}(\omega, \omega') = 2^{-n}$ if $\omega \neq \omega'$ and ω and ω' coincide up to the n^{th} entry, and $d_{\bar{\Sigma}}(\omega, \omega') = 0$ if $\omega = \omega'$.

4.2 Thermodynamic formalism

The shift map $\sigma : \bar{\Sigma} \to \bar{\Sigma}$ takes a (possibly finite) sequence $\omega = (\omega_o)_{i=0,1,\dots}$ and maps it to $\sigma(\omega) = (\omega_{i+1})_{i=0,1,\dots}$. To ensure that σ is well-defined, we include the empty path in $\bar{\Sigma}$. For every real-valued Hölder continuous function $\varphi : \bar{\Sigma} \to \mathbb{R}$ (which we call a *potential*), the transfer operator $\mathcal{L}_\varphi$ acting on the space of continuous functions f on $\bar{\Sigma}$ is defined by

$$\mathcal{L}_\varphi f(\omega) = \sum_{\sigma(\omega')=\omega} e^{\varphi(\omega')} f(\omega'),$$

where for the empty path $\omega = \varnothing$ the preimages of σ are defined only by nonempty paths. We say that the directed graph $\mathcal{G}$ is *recurrent* if there is a directed path between arbitrary two vertices. We say that $\mathcal{G}$ is *topologically mixing* if there exists n such that every pair of vertices is connected by a path of length n. If $\mathcal{G}$ is recurrent but not topologically mixing, then there is an integer $p > 1$ such that every loop (ie path starting and ending at the same vertex) in $\mathcal{G}$ has length divisible by p. Furthermore the set of vertices of $\mathcal{G}$ decomposes into p subsets $V = \bigsqcup_{j \in \mathbb{Z}/p\mathbb{Z}} V_j$, where every edge with the origin in V_j has the terminus in V_{j+1}. We call this decomposition a *cyclic decomposition* of V. Restricting σ^p to V_j, we obtain a topological mixing subshift of finite type. If $\mathcal{G}$ is not recurrent, then we decompose $\mathcal{G}$ into components — these are the maximal induced subgraphs which are recurrent. For each component $\mathscr{C}$, we define the transfer operator $\mathscr{L}_{\mathscr{C}}$ by restricting φ to the paths staying in $\mathscr{C}$. The spectral radius of $\mathscr{L}_{\mathscr{C}}$ is given by $e^{\mathrm{Pr}_{\mathscr{C}}(\varphi)}$ for some real value $\mathrm{Pr}_{\mathscr{C}}(\varphi)$. This constant is obtained from the limit

$$\mathrm{Pr}_{\mathscr{C}}(\varphi,\sigma) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{[\omega_0,\dots,\omega_{n-1}]} \exp(S_{[\omega_0,\dots,\omega_{n-1}]}(\varphi)), \tag{4-1}$$

where the summation is taken over all the cylinder sets of length n,

$$S_{[\omega_0,\dots,\omega_{n-1}]}(\varphi) := \sup\{S_n\varphi(\omega) : \omega \in [\omega_0,\dots,\omega_{n-1}]\} \quad \text{and} \quad S_n\varphi := \sum_{i=0}^{n-1} \varphi\circ\sigma^i.$$

(See [Parry and Pollicott 1990, Theorem 2.2]; this follows from the Gibbs property of an eigenmeasure for each component in the cyclic decomposition.)

Let

$$\mathrm{Pr}(\varphi) := \max_{\mathscr{C}} \mathrm{Pr}_{\mathscr{C}}(\varphi),$$

where the maximum is taken over all components $\mathscr{C}$ of $\mathscr{G}$. We call a component $\mathscr{C}$ *maximal* if $\mathrm{Pr}_{\mathscr{C}}(\varphi) = \mathrm{Pr}(\varphi)$. Note that the set of maximal components depends on φ. We are interested in potentials that satisfy the following condition.

Definition 4.2 A potential φ is called *semisimple* if there are no directed paths from any maximal component to any other maximal components.

We denote by $\mathscr{H}$ the space of Hölder continuous functions on $\bar{\Sigma}$ with some fixed exponent, whose explicit value is not used, and by $\|\cdot\|_{\mathscr{H}}$ the corresponding Hölder norm.

Theorem 4.3 [Gouëzel 2014, Theorem 3.8] *Let φ be a semisimple potential and $\mathscr{C}_1,\dots,\mathscr{C}_I$ be the corresponding maximal components, each with period p_i and cyclic decomposition $\mathscr{C}_i = \bigsqcup_{j\in\mathbb{Z}/p_i\mathbb{Z}} \mathscr{C}_{i,j}$. Then there exist Hölder continuous functions $h_{i,j}$ and measures $\lambda_{i,j}$ with $\int_{\bar{\Sigma}} h_{i,j}\,d\lambda_{i,j} = 1$ such that*

$$\left\| \mathscr{L}_\varphi^n f - e^{\mathrm{Pr}(\varphi)n} \sum_{i=1}^{I} \sum_{j\in\mathbb{Z}/p_i\mathbb{Z}} \left(\int_{\bar{\Sigma}} f\,d\lambda_{i,(j-n \bmod p_i)} \right) h_{i,j} \right\|_{\mathscr{H}} \le C\|f\|_{\mathscr{H}} e^{(\mathrm{Pr}(\varphi)-\varepsilon_0)n}$$

for every Hölder continuous function f, where positive constants C and $\varepsilon_0 > 0$ are independent of f, and the probability measures $\mu_i = (1/p_i)\sum_{j\in\mathbb{Z}/p_i\mathbb{Z}} h_{i,j}\lambda_{i,j}$ are invariant under the shift σ. The measures μ_i are also ergodic.

Remark 4.4 In the statement of Theorem 4.3, if we define $\mathscr{C}_{i,j,\to}$ to be the set of edges which can be reached by a path from $\mathscr{C}_{i,j}$ of length divisible by p_i, and $\mathscr{C}_{\to,i,j}$ to be the set of edges which we can reach $\mathscr{C}_{i,j}$ with a path of length divisible by p_i, then the function $h_{i,j}$ is bounded from below on the paths starting with edges in $\mathscr{C}_{i,j,\to}$ and the empty path, and takes 0 elsewhere. Furthermore the measure $\lambda_{i,j}$ is supported on the set of infinite paths starting with edges in $\mathscr{C}_{\to,i,j}$ and eventually staying in $\mathscr{C}_i$.

We will use both of the measures μ_i and $\lambda_i := \sum_{j=0}^{p_i-1} \lambda_{i,j}$ for each $i = 1, \dots, I$. They have different supports on the space of paths $\bar{\Sigma}$, and μ_i is σ-invariant while λ_i is not.

Lemma 4.5 [Gouëzel 2014, Lemma 3.9] *In the notation in Theorem 4.3, let $\lambda_i := \sum_{j=0}^{p_i-1} \lambda_{i,j}$. Then $\sigma_*\lambda_i$ is absolutely continuous with respect to λ_i.*

Proposition 4.6 [Gouëzel 2014, Proposition 3.10] *Suppose that φ is a semisimple potential in $\mathscr{H}$ and $\mathscr{C}_1, \dots, \mathscr{C}_I$ are the corresponding maximal components. Then there exist positive constants $C, \varepsilon_0 > 0$ such that for all small enough $\psi \in \mathscr{H}$, there exist Hölder continuous functions $h_{i,j}^{\psi}$ and measures $\lambda_{i,j}^{\psi}$ with the same support as $h_{i,j}$ and $\lambda_{i,j}$, respectively, such that*

$$\left\| \mathscr{L}_{\varphi+\psi}^n f - \sum_{i=1}^{I} e^{\mathrm{Pr}_{\mathscr{C}_i}(\varphi+\psi)n} \sum_{j\in\mathbb{Z}/p_i\mathbb{Z}} \left(\int_{\bar{\Sigma}} f \, d\lambda_{i,(j-n \bmod p_i)}^{\psi} \right) h_{i,j}^{\psi} \right\|_{\mathscr{H}} \le C \|f\|_{\mathscr{H}} e^{(\mathrm{Pr}(\varphi)-\varepsilon_0)n}$$

for all $f \in \mathscr{H}$. Moreover, the maps $\psi \mapsto \mathrm{Pr}_{\mathscr{C}_i}(\varphi+\psi)$, $\psi \mapsto h_{i,j}^{\psi}$ and $\psi \mapsto \lambda_{i,j}^{\psi}$ (from $\mathscr{H}$ to $\mathbb{R}$, $\mathscr{H}$ and the dual of $\mathscr{H}$, respectively) are each real analytic in a small neighborhood of 0 in $\mathscr{H}$.

Let $[E_*]$ denote the set of paths in $\bar{\Sigma}$ starting at s_*. If $\mathbf{1}_{[E_*]}$ denotes the corresponding indicator function, then

$$\mathscr{L}_{\varphi}^n \mathbf{1}_{[E_*]}(\varnothing) = \sum e^{S_n\varphi(\omega)}, \quad \text{where } S_n\varphi(\omega) = \sum_{k=0}^{n-1} \varphi(\sigma^k(\omega)),$$

and the summation is taken over all paths ω of length n starting from s_*.

Lemma 4.7 *For every Hölder continuous potential φ and for every integer $k \ge 1$, if there exists a path from s_* in $\mathscr{A}$ containing edges successively from k different maximal components for φ, then there exists a constant $C > 0$ such that for all $n \ge 1$,*

$$\mathscr{L}_{\varphi}^n \mathbf{1}_{[E_*]}(\varnothing) \ge C n^{k-1} e^{n\mathrm{Pr}(\varphi)}.$$

On the other hand, if there are L components in $\mathscr{A}$, then there exists a constant $C > 0$ such that for all $n \ge 1$,

$$\mathscr{L}_{\varphi}^n \mathbf{1}_{[E_*]}(\varnothing) \le C n^{L} e^{n\mathrm{Pr}(\varphi)}.$$

Proof This lemma is a special case of [Gouëzel 2014, Lemma 3.7]; the proof of the second part is found in [Tanaka 2017, Lemma 4.7]. □

4.3 Semisimple potentials

In this section we use thermodynamic formalism to link the geometric measures constructed in Section 2.6 to certain measures on $\bar{\Sigma}$. The key result that allows us to do this is the following.

Lemma 4.8 *Let ψ be a Γ-invariant tempered potential relative to d_S on Γ with exponent θ; see Definition 2.5. If for a strongly Markov automatic structure $\mathcal{A} = (\mathcal{G}, \pi, S)$ the corresponding shift space $(\bar{\Sigma}, \sigma)$ admits a Hölder continuous potential Ψ such that*

$$S_n\Psi(\omega) = \sum_{i=0}^{n-1} \Psi(\sigma^i(\omega)) = -\psi(o, \pi_*(\omega)) \quad \text{for all } \omega = (\omega_0, \dots, \omega_{n-1}) \in \Sigma^*, \tag{4-2}$$

then Ψ is semisimple and $\mathrm{Pr}(\Psi) = \theta$. Moreover, for each $a \in \mathbb{R}$, the potential $a\Psi$ is semisimple.

Proof Note that for the potential Ψ, we have that for all n,

$$\mathcal{L}_\Psi^n \mathbf{1}_{[E_*]}(\varnothing) = \sum_{|x|_S = n} e^{-\psi(o,x)}.$$

Hence Lemmas 2.8 and 4.7 show that $\mathrm{Pr}(\Psi) = \theta$. If the potential Ψ is not semisimple, then there is a directed path in the automatic structure $\mathcal{A} = (\mathcal{G}, \pi, S)$ starting from s_* passing through k distinct maximal components for Ψ for $k > 1$. The first part of Lemma 4.7 implies that $\mathcal{L}_\Psi^n \mathbf{1}_{[E_*]}(\varnothing) \geq C n^{k-1} e^{n\mathrm{Pr}(\Psi)}$ for all $n \geq 0$. This however contradicts Lemma 2.8. Therefore Ψ is semisimple. Furthermore, for every $a \in \mathbb{R}$, the same proof applies to $a\psi$, and the potential $a\Psi$ is semisimple. □

For each semisimple potential Ψ on $\bar{\Sigma}$, let $\mathcal{C}_i$ for $i = 1, \dots, I$ be the maximal components for Ψ. Let $\lambda_{i,j}$ for $i = 1, \dots, I$ and $j = 0, \dots, p_i$ be the measures obtained in Theorem 4.3 applied to the potential Ψ. We define $\lambda_i := \sum_{j=0}^{p_i-1} \lambda_{i,j}$ and $\lambda_\Psi := \sum_{i=1}^{I} \lambda_i$. Let us denote by μ_ψ a finite Borel measure on $\partial\Gamma$ satisfying (QC) with exponent θ relative to (ψ, d) (which has been constructed in Proposition 2.7).

Lemma 4.9 *Assume that ψ and Ψ are as in Lemma 4.8. Then the pushforward of $\lambda_\Psi(\cdot \cap [E_*])$ by π_* is comparable to μ_ψ.*

Proof For all n, let $\tilde{m}_n$ be the finite measure on $\bar{\Sigma}$ defined by the positive linear functional $f \mapsto e^{-n\mathrm{Pr}(\Psi)} \cdot \mathcal{L}_\Psi^n f(\varnothing)$. If the maximal components for the potential Ψ have periods p_i for $i = 1, \dots, I$, then let p be the least common multiple of these periods. Theorem 4.3 shows that for every Hölder continuous function f on $\bar{\Sigma}$, we have that for each $q = 0, \dots, p-1$,

$$e^{-(np+q)\mathrm{Pr}(\Psi)} \cdot \mathcal{L}_\Psi^{np+q} f(\varnothing) \to \sum_{i=1}^{I} \sum_{j=0}^{p_i} \int_{\bar{\Sigma}} f \, d\lambda_{i,(j-q \bmod p_i)} h_{i,j}(\varnothing) \quad \text{as } n \to \infty.$$

This convergence holds for all continuous functions f on $\bar{\Sigma}$; indeed, we approximate f by Hölder continuous functions and use $|e^{-n\mathrm{Pr}(\Psi)} \cdot \mathcal{L}_\Psi^n f(\varnothing)| \leq C\|f\|_\infty$ for all n, where $\|\cdot\|_\infty$ stands for the supremum norm. This shows that $\tilde{m}_{np+q}$ weakly converges to a measure $\tilde{m}_q$ for each $q = 0, \dots, p-1$. Since $c_1 \leq h_{i,j}(\varnothing) \leq c_2$ for some $c_1, c_2 > 0$ (see Theorem 4.3 and Remark 4.4), all $\tilde{m}_q$ are comparable with $\sum_{i,j} \lambda_{i,j}$. If we denote by $\tilde{m}_\infty$ the weak limit of $\sum_{k=0}^{n} \tilde{m}_k(\cdot \cap [E_*]) / \sum_{k=0}^{n} \tilde{m}_k([E_*])$, then the measure $\pi_*\widetilde{m}_\infty$ is actually μ_ψ. Indeed, for every continuous function f on $\Gamma \cup \partial\Gamma$, we have

$$\begin{aligned} e^{-n\mathrm{Pr}(\Psi)}\mathscr{L}_\Psi^n(\mathbf{1}_{[E_*]} \cdot f \circ \pi_*)(\varnothing) &= e^{-n\mathrm{Pr}(\Psi)} \sum_{\omega \text{ of length } n \text{ from } s_*} e^{S_n\Psi(\omega)} f(\pi_*(\omega)) \\ &= e^{-n\mathrm{Pr}(\Psi)} \sum_{|x|_S=n} e^{-\psi(o,x)} f(x), \end{aligned}$$

where the last line follows since the map π_* induces a bijection from the set of paths of length n starting at s_* to the set of $x \in \Gamma$ with $|x|_S = n$ and (4-2). Since $\mathrm{Pr}(\Psi) = \theta$ by Lemma 4.8, this shows that the measure $\pi_*\widetilde{m}_\infty$ is comparable with μ_ψ obtained by the Patterson–Sullivan procedure, and for $\lambda_\Psi := \sum_{i,j} \lambda_{i,j}$, the measure $\pi_*\lambda(\,\cdot \cap [E_*])$ is comparable with μ_ψ. □

Example 4.10 For every pair of finite symmetric sets of generators S and S_*, there exist a strongly Markov automatic structure $\mathscr{A} = (\mathscr{G}, \pi, S)$ and a function $d\phi_{S_*} : E(\mathscr{G}) \to \mathbb{Z}$ such that

$$|\pi_*(\omega)|_{S_*} = \sum_{i=0}^{n-1} d\phi_{S_*}(\omega_i) \quad \text{for every path } \omega = (\omega_0, \dots, \omega_{n-1}) \text{ from } s_* \text{ on } \mathscr{G},$$

where $\mathscr{G} = (V(\mathscr{G}), E(\mathscr{G}))$ is the underlying directed graph of $\mathscr{A}$. This is proved in [Calegari and Fujiwara 2010, Lemma 3.8]; see also [Calegari 2009, Theorem 6.39]. Let us define a function $\Psi_{S_*} : \overline{\Sigma} \to \mathbb{R}$ by setting

$$\Psi_{S_*}(\omega) = -d\phi_{S_*}(\omega_0) \quad \text{for } \omega \in \overline{\Sigma}.$$

This function depends only on the first coordinate of ω and is Hölder continuous with respect to the metric $d_{\overline{\Sigma}}$. Further, by construction, for $\omega = (\omega_0, \dots, \omega_{n-1}) \in \Sigma^*$ we have that

$$S_n\Psi_{S_*}(\omega) = -\sum_{i=0}^{n-1} d\phi_{S_*}(\omega_i) = -d_{S_*}(o, \pi_*(\omega)).$$

This shows that, on Σ^*, the Birkhoff sums of Ψ_{S_*} encode information about the metric d_{S_*}.

Example 4.11 Let $d \in \mathscr{D}_\Gamma$ be a strongly hyperbolic metric. For every finite (symmetric) set of generators S, we consider a subshift $\overline{\Sigma}$ arising from a strongly Markov structure $\mathscr{A} = (\mathscr{G}, \pi, S)$. Since the Busemann function for a strongly hyperbolic metric is defined as limits (Section 2.6), if we define

$$\beta_o(x, y) := d(x, y) - d(o, y) \quad \text{for } (x, y) \in \Gamma \times (\Gamma \cup \partial\Gamma),$$

then its restriction on $\Gamma \times \partial\Gamma$ is the original Busemann function (based at o) for d. Let

$$\Psi(\omega) := \beta_o(\pi_*(\omega_0), \pi_*(\omega)) \quad \text{for } \omega \in \overline{\Sigma}.$$

Note that Ψ is Hölder continuous with respect to $d_{\overline{\Sigma}}$ by the definition of strong hyperbolicity; see Section 2.6. Furthermore, for $\omega \in \Sigma^*$, if $n = |\pi_*(\omega)|_S$, then we have that

$$S_n\Psi(\omega) = -d(o, \pi_*(\omega)).$$

Remark 4.12 It is important to note that for *any* strongly Markov structure $\mathcal{A} = (\mathcal{G}, \pi, S)$ and *any* strongly hyperbolic metric $d \in \mathcal{D}_\Gamma$, we can find a function Ψ on $\bar{\Sigma}$ encoding d. It is not clear if we can do the same when d is a word metric; in which case we only know the existence of *some* strongly Markov structure and a function on $\bar{\Sigma}$ encoding d; see Example 4.10. We will exploit this freedom of choice for strongly hyperbolic metrics in our proof of Theorem 1.3.

4.4 Proof of the C^2-regularity

In general, restricting to each component $\mathscr{C}_i$, the pressure function $\Pr_{\mathscr{C}_i}(\Psi)$ is real analytic in Ψ. Furthermore, for every $\Psi_0, \Psi \in \mathscr{H}$,

$$\Pr_{\mathscr{C}_i}(\Psi_0 + s\Psi) = \Pr_{\mathscr{C}_i}(\Psi_0) + s\tau_i + \tfrac{1}{2}s^2\sigma_i^2 + O(s^3) \quad \text{as } s \to 0,$$

where

$$\tau_i := \int_\Sigma \Psi \, d\mu_i \quad \text{and} \quad \sigma_i^2 := \lim_{n\to\infty} \frac{1}{n} \int_\Sigma (S_n\Psi - n\tau_i)^2 \, d\mu_i.$$

We will prove that the τ_i (resp. the σ_i^2) coincide on all maximal components for Ψ_0.

Proposition 4.13 *Let ψ be a Γ-invariant tempered potential relative to d_S on Γ with exponent θ, and*

$$\theta(a) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{|x|_S = n} e^{-a\psi(o,x)} \quad \textit{for } a \in \mathbb{R}.$$

Suppose that the shift space $(\bar{\Sigma}, \sigma)$ corresponding to a strongly Markov automatic structure $\mathcal{A} = (\mathcal{G}, \pi, S)$ admits a Hölder continuous potential Ψ satisfying (4-2). Then the function $\theta(a)$ is twice continuously differentiable in $a \in \mathbb{R}$.

Proof By Lemma 4.8, we have that $\theta(a) = \max_{\mathscr{C}} \Pr_{\mathscr{C}}(a\Psi)$ for every $a \in \mathbb{R}$. Proposition 4.6 shows that for each maximal component $\mathscr{C}_i$ with $i = 1, \dots, I$, for $a\Psi$ we have $\Pr'_{\mathscr{C}_i}(a\Psi) = \int_{\Sigma_i} \Psi \, d\mu_i$, where Σ_i is the set of paths staying in $\mathscr{C}_i$ for all time. Let us prove that $\int_{\Sigma_i} \Psi \, d\mu_i = \int_{\Sigma_j} \Psi \, d\mu_j$ for all $i, j \in \{1, \dots, I\}$. For every $\tau \in \mathbb{R}$, let $A(\tau)$ be the set of boundary points ξ in $\partial\Gamma$ for which there exists a unit-speed geodesic ray ξ_n in $\mathrm{Cay}(\Gamma, S)$ converging to ξ such that

$$\lim_{n\to\infty} \frac{1}{n} \psi(o, \xi_n) = -\tau.$$

Note that if this convergence holds for some geodesic ray toward ξ, then in fact this holds for every geodesic ray toward ξ since an arbitrary pair of geodesic rays converging to the same extreme point are eventually within bounded distance up to shifting the parametrizations (where all geodesic rays are parametrized with unit speed). This shows that the set $A(\tau)$ is Γ-invariant. Let $\tau_i := \int_{\Sigma_i} \Psi \, d\mu_i$, where $\mu_i = (1/p_i) \sum_{j=0}^{p_i - 1} h_{i,j} \lambda_{i,j}$. If we define

$$U_i := \left\{ \omega \in \Sigma_i : \lim_{n\to\infty} \frac{1}{n} S_n\Psi(\omega) = \tau_i \right\},$$

then $\pi_*(U_i) \subset A(\tau_i)$ and the Birkhoff ergodic theorem implies that $\mu_i(U_i) = 1$ since μ_i is ergodic by Theorem 4.3. We shall show that $\tau_i = \tau_j$ for all $i, j \in \{1, \dots, I\}$. Let $U_i^c := \bar{\Sigma} \setminus U_i$. Since λ_i and μ_i are equivalent on Σ_i, we have $\lambda_i(\Sigma_i \cap U_i^c) = 0$. This implies that $\lambda_i(U_i^c) = 0$. Indeed, note that $\sigma^k U_i^c \subset U_i^c$, and $\sigma^{-k}\Sigma_i \cap U_i^c = \sigma^{-k}(\Sigma_i \cap \sigma^k U_i^c) \subset \sigma^{-k}(\Sigma_i \cap U_i^c)$. Since $\lambda_i \circ \sigma^{-1}$ is absolutely continuous with respect to λ_i by Lemma 4.5, we have $\lambda_i(\sigma^{-k}\Sigma_i \cap U_i^c) = 0$, and since $\bar{\Sigma} = \bigcup_{k=0}^{\infty} \sigma^{-k}\Sigma_i$ modulo λ_i-null sets, we obtain $\lambda_i(U_i^c) = 0$.

We then have that $\lambda_i(U_i \cap [E_*]) > 0$ since U_i has full λ_i-measure and λ_i assigns positive measure to $[E_*]$ by Theorem 4.3. Therefore

$$\lambda_i(\pi_*^{-1} A(\tau_i) \cap [E_*]) \geq \lambda_i(U_i \cap [E_*]) > 0$$

and Lemma 4.9 implies that $\mu_\psi(A(\tau_i)) > 0$. As we have noted, $A(\tau_i)$ is Γ-invariant, and since μ_ψ is ergodic with respect to the Γ-action on $\partial\Gamma$ by Lemma 2.9, the set $A(\tau_i)$ has the full μ_ψ measure. Since this is true for all $i = 1, \dots, I$, all τ_i and thus $\mathrm{Pr}'_{\mathscr{C}_i}(a\Psi)$ coincide. This shows that $\theta(a)$ is differentiable at every $a \in \mathbb{R}$.

Let τ be the common value of all of the τ_i at $a \in \mathbb{R}$. For each $i = 1, \dots, I$, we have by [Parry and Pollicott 1990, Proposition 4.11], $\mathrm{Pr}''_{\mathscr{C}_i}(a\Psi) = a^2\sigma_i^2$, where

$$\sigma_i^2 = \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma_i} (S_n\Psi - n\tau)^2 \, d\mu_i.$$

We define the set $B(\sigma_i)$ of points ξ in $\partial\Gamma$ for which there exists a (unit-speed) geodesic ray ξ in $\mathrm{Cay}(\Gamma, S)$ converging to ξ such that the following double limits hold:

$$\sigma_i^2 = \lim_{n\to\infty} \frac{1}{n}\sigma_i^2(n), \quad \text{where } \sigma_i^2(n) := \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m-1} (-\psi(\xi_k, \xi_{n+k}) - n\tau)^2.$$

Note that $B(\sigma_i)$ is Γ-invariant since ψ is Γ-invariant. If we define $\mu := \sum_{i=1}^{I} \mu_i$, then applying the Birkhoff ergodic theorem countably many times on a dense subset of the space of continuous functions on Σ, we have that for μ-almost every $\omega \in \Sigma$, the measures $(1/n)\sum_{k=0}^{n-1} \delta_{\sigma^k\omega}$ weakly converge to a measure μ_ω on Σ, and for μ_i-almost every $\omega \in \Sigma_i$ one has $\mu_\omega = \mu_i$ for each $i = 1, \dots, I$. Note that

$$\mu_\omega \circ \sigma^{-1} = \mu_{\sigma\omega} = \mu_\omega \quad \mu\text{-almost everywhere.}$$

Let us define

$$V_i := \left\{ \omega \in \Sigma_i : \sigma_i^2 = \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} (S_n\Psi - n\tau)^2 \, d\mu_\omega \right\}.$$

We then have that $\pi_*(V_i) \subset B(\sigma_i)$ since ψ is a Γ-invariant tempered potential relative to d_S and $\mu_i(V_i) = 1$. Since V_i is σ-invariant modulo μ_i-null sets, the same argument as above implies that $\mu_\psi(B(\sigma_i)) > 0$ and all σ_i^2 coincide. This shows that $\mathrm{Pr}''_{\mathscr{C}_i}(a\Psi)$ coincide for all $i = 1, \dots, I$. Since each $\mathrm{Pr}_{\mathscr{C}_i}(a)$ is real analytic and $\theta(a)$ coincides with the maximum of finitely many $\mathrm{Pr}_{\mathscr{C}_i}(a)$ on a neighborhood of a by Proposition 4.6, the function $\theta(a)$ is twice continuously differentiable in $a \in \mathbb{R}$. □

Theorem 4.14 *For every pair of finite symmetric sets of generators S and S_*, if*

$$\theta_{S_*/S}(a) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{|x|_S=n} e^{-a|x|_{S_*}} \quad \text{for } a \in \mathbb{R},$$

then $\theta_{S_*/S}$ is twice continuously differentiable in $a \in \mathbb{R}$.

Proof This follows from Proposition 4.13 and Example 4.10. □

We now consider the case when $d, d_* \in \mathcal{D}_\Gamma$ are both strongly hyperbolic and we want to show that the associated Manhattan curve is C^2. In the previous part we exploited the fact that the subshift $\overline{\Sigma}$ encoded one of the metrics d that we were considering. It is not clear how to exploit this fact when both metrics are strongly hyperbolic. To get around this issue we take a word metric d_S on Γ associated to a finite symmetric set of generators S and use this to introduce a subshift $\overline{\Sigma}$ on which we are able to encode information about the metrics d, d_*.

For the rest of this section assume that we have two strongly hyperbolic metrics $d, d_* \in \mathcal{D}_\Gamma$ and that we have arbitrarily chosen a finite symmetric set of generators S for Γ. We begin by constructing a useful two-parameter family of measures on $\partial\Gamma$. For each $(a,b) \in \mathbb{R}^2$, let $\widetilde{\theta}(a,b)$ be the abscissa of convergence of

$$\sum_{x\in\Gamma} \exp(-ad_*(o,x) - bd(o,x) - sd_S(o,x))$$

as s varies.

To understand this summation we use the measures we constructed in Section 2.6. Since $ad_* + bd$ is a Γ-invariant tempered potential relative to d_S for each $(a,b) \in \mathbb{R}^2$ (Example 2.6), Proposition 2.7 implies that for each $(a,b) \in \mathbb{R}^2$, there exists a measure $\mu_{a,b,S}$ on $\partial\Gamma$ such that for $x \in \Gamma$,

$$C_{a,b}^{-1} \le \exp(a\beta_{*o}(x,\xi) + b\beta_o(x,\xi) + \widetilde{\theta}(a,b)\beta_{So}(x,\xi)) \cdot \frac{dx_*\mu_{a,b,S}}{d\mu_{a,b,S}}(\xi) \le C_{a,b},$$

where β_{*o}, β_o and β_{So} are Busemann functions (based at o) for d_*, d and d_S respectively, and $C_{a,b}$ is a positive constant. By Lemma 2.8, we have that

$$\sum_{|x|_S=n} e^{-ad_*(o,x)-bd(o,x)} \asymp_{a,b} \exp(\widetilde{\theta}(a,b)n) \quad \text{for all integers } n \ge 0. \tag{4-3}$$

For each fixed $a \in \mathbb{R}$, the function $b \mapsto \widetilde{\theta}(a,b)$ is continuous by the Hölder inequality. Note that for each fixed $a \in \mathbb{R}$,

$$b > \theta(a) \implies \widetilde{\theta}(a,b) < 0 \quad \text{and} \quad b < \theta(a) \implies \widetilde{\theta}(a,b) > 0.$$

Therefore, combining with the continuity of $\widetilde{\theta}(a,b)$ in b, we have that $\widetilde{\theta}(a,b) = 0$ if and only if $b = \theta(a)$.

Now consider the subshift of finite type $\overline{\Sigma}$ arising from a coding corresponding to S. Let $\Psi, \Psi_* : \Sigma \to \mathbb{R}$ be Hölder continuous potentials that encode d and d_*, respectively, as in Example 4.11. For each $(a,b) \in \mathbb{R}^2$, by Lemma 4.8 (adapted to (4-3)) the potential $\Psi_{a,b} = a\Psi_* + b\Psi$ is semisimple and $\Pr(\Psi_{a,b}) = \widetilde{\theta}(a,b)$.

Consider a point $(a,b) \in \mathbb{R}^2$ and take a maximal component $\mathscr{C}$ for $\Psi_{a,b}$. Let $\mu_{\mathscr{C}}$ denote the measure corresponding to $\Psi_{a,b}$ on $\mathscr{C}$ from applying Theorem 4.3. Since for each $\mathscr{C}$, the function $\Psi_{a,b}$ is real analytic in $(a,b) \in \mathbb{R}^2$, it admits a Taylor expansion with Jacobian $J_{\mathscr{C}}(a,b)$ and a symmetric Hessian $\mathrm{Cov}_{\mathscr{C}}(a,b)$. More precisely, letting $\Psi_1 := \Psi_*$ and $\Psi_2 := \Psi$, we have $J_{\mathscr{C}}(a,b) = (J_1(a,b), J_2(a,b))$, where

$$J_i(a,b) := \int_{\Sigma_{\mathscr{C}}} \Psi_i \, d\mu_{\mathscr{C}} = \frac{\partial}{\partial s_i}\Big|_{(a,b)} \mathrm{Pr}_{\mathscr{C}}(\Psi_{s_1,s_2}) \quad \text{for } i = 1,2,$$

and $\mathrm{Cov}_{\mathscr{C}}(a,b)$ has entries

$$\begin{aligned}\mathrm{Cov}_{\mathscr{C}}(a,b)_{i,j} &= \frac{\partial^2}{\partial s_i \partial s_j}\Big|_{(a,b)} \mathrm{Pr}_{\mathscr{C}}(\Psi_{s_1,s_2}) \\ &= \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma_{\mathscr{C}}} (S_n\Psi_i(\omega) - nJ_i(a,b))(S_n\Psi_j(\omega) - nJ_j(a,b)) \, d\mu_{\mathscr{C}}(\omega)\end{aligned}$$

for $i, j = 1,2$; this follows from the one-parameter case by considering $s \mapsto \Psi_{a+s,b+s}$ and differentiating at $s = 0$ [Parry and Pollicott 1990, Proposition 4.11]. We know that the potentials Ψ_1 and Ψ_2 satisfy a (possibly degenerate) multidimensional central limit theorem with respect to $\mu_{\mathscr{C}}$ on $\Sigma_{\mathscr{C}}$. That is, the distribution of

$$\frac{(S_n\Psi_1(\omega), S_n\Psi_2(\omega)) - nJ_{\mathscr{C}}(a,b)}{\sqrt{n}}$$

under $\mu_{\mathscr{C}}$ weakly converges to a two-dimensional Gaussian distribution with covariance matrix $\mathrm{Cov}_{\mathscr{C}}(a,b)$ as $n \to \infty$. It is useful to keep this in mind throughout the following. Furthermore, we note that Ψ and Ψ_* may vanish at certain points; however, there exists n such that $S_n\Psi$ and $S_n\Psi_*$ are strictly negative functions. Therefore for $\Psi_{a,b}$ and for the corresponding on maximal components $\mathscr{C}$, we have that $J_{\mathscr{C}}(a,b) \neq (0,0)$ whenever $b = \theta(a)$ for all $a \in \mathbb{R}$. This fact is crucial when we appeal to the implicit function theorem later in our discussion.

Proposition 4.15 *For every $(a,b) \in \mathbb{R}^2$, the Jacobian $J_{\mathscr{C}}(a,b)$ and the Hessian matrix $\mathrm{Cov}_{\mathscr{C}}(a,b)$ do not depend on the choice of maximal component $\mathscr{C}$.*

Proof Showing that $J_{\mathscr{C}}(a,b)$ is independent of the choice of maximal component $\mathscr{C}$ for $\Psi_{a,b}$ is analogous to the first part in the proof of Proposition 4.13; we omit the details. We will show that $\mathrm{Cov}_{\mathscr{C}}(a,b)$ is independent of the maximal component $\mathscr{C}$. The proof follows the same lines as in the proof of Proposition 4.13 but we need to adapt the arguments to the multidimensional setting.

To simplify the following exposition, let us fix a and b and suppress their dependence in the notation. We also write $d_1 := d_*$ and $d_2 := d$, and denote the corresponding potentials by $\Psi_1 = \Psi_*$ and $\Psi_2 = \Psi$. For $\mathrm{Cov}_{\mathscr{C}} = (\sigma_{i,j})_{i,j=1,2}$, we define the set $B(\mathrm{Cov}_{\mathscr{C}})$ of points in $\partial\Gamma$ for which there exists a geodesic ray $\xi \in \mathrm{Cay}(\Gamma, S)$ converging to ξ such that for all $i, j = 1,2$, we have

$$\sigma_{i,j} = \lim_{n\to\infty} \frac{1}{n} \sigma_{i,j}(n),$$

where

$$\sigma_{i,j}(n) = \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m-1} (-d_i(\xi_k, \xi_{n+k}) - nJ_i)(-d_j(\xi_k, \xi_{n+k}) - nJ_j).$$

For each $\Psi_{a,b}$ we have a measure $\lambda_{a,b}$ by Lemma 4.9 such that the pushforward of $\lambda_{a,b}$ restricted on $[E_*]$ by π_* is comparable to $\mu_{a,b,S}$, which is ergodic with respect to the Γ-action on $\partial\Gamma$ by Lemma 2.9. By comparing the set $B(\mathrm{Cov}_{\mathscr{C}})$ to the set

$$V_{\mathscr{C}} := \bigcap_{i,j=1,2} \left\{ \omega \in \Sigma_{\mathscr{C}} : \sigma_{i,j} = \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} (S_n\Psi_i - nJ_i)(S_n\Psi_j - nJ_j)\, d\mu_\omega \right\},$$

where μ_ω is as in the proof of Proposition 4.13, we see that the matrix $\mathrm{Cov}_{\mathscr{C}}$ does not depend on the component $\mathscr{C}$. This concludes the proof. □

Theorem 4.16 *For every pair of strongly hyperbolic metrics d and d_* on Γ, the corresponding function $\theta(a)$ is twice continuously differentiable in $a \in \mathbb{R}$.*

Proof For each $(a,b) \in \mathbb{R}^2$, let $\mathscr{C}$ be a maximal component for the potential $\Psi_{a,b}$. Proposition 4.15 shows that $\mathrm{Pr}_{\mathscr{C}}(\Psi_{s_1,s_2})$ admits the Taylor expansion whose terms up to the second order are independent of the choice of $\mathscr{C}$. This implies that since $\mathrm{Pr}(\Psi_{a,b})$ is given by the maximum over finitely many functions $\mathrm{Pr}_{\mathscr{C}}(\Psi_{a,b})$ and $\mathrm{Pr}(\Psi_{a,b}) = \widetilde{\theta}(a,b)$, the function $\widetilde{\theta}(a,b)$ is twice continuously differentiable in $(a,b) \in \mathbb{R}^2$. Note that $\widetilde{\theta}(a,b) = 0$ if and only if $b = \theta(a)$ for all $(a,b) \in \mathbb{R}^2$, and for every $(a,b) \in \mathbb{R}^2$ with $\widetilde{\theta}(a,b) = 0$ and for every maximal component $\mathscr{C}$, we have that $J_{\mathscr{C}}(a,b) \neq (0,0)$; see the discussion just before Proposition 4.15. Therefore the implicit function theorem implies that θ is twice continuously differentiable. □

Now Theorem 1.3 follows from Theorem 4.16.

Note that this result and the arguments we applied to prove it are independent of the choice of S. However in the case when Γ admits a finite symmetric set of generators S such that the underlying directed graph of an automaton has only one recurrent component, then the Manhattan curve associated to two strongly hyperbolic metrics in $\mathscr{D}_\Gamma$ is real analytic. This is because, in this case, $\mathrm{Pr}(\Psi_{a,b})$ is real analytic in $(a,b) \in \mathbb{R}$ as there is only one maximal component which is recurrent and all the others are transient (not recurrent). For example, the fundamental groups of closed orientable surface of genus at least 2 admit such automata with the standard set of generators since they have a single relator. For more general cocompact Fuchsian groups, see [Series 1981].

4.5 Pairs of word metrics

In this section we deduce further rigidity results for word metrics. Recall that the Manhattan curve $\theta_{S_*/S} : \mathbb{R} \to \mathbb{R}$ associated to every pair of word metrics d_S, d_{S_*} is twice continuously differentiable.

Theorem 4.17 *Let Γ be a nonelementary hyperbolic group and d_S and d_{S_*} be word metrics associated to finite symmetric sets of generators S and S_*, respectively. If we denote the Manhattan curve for the pair (d_S, d_{S_*}) by*

$$\mathscr{C}_M = \{(a,b) \in \mathbb{R}^2 : b = \theta_{S_*/S}(a)\},$$

then the following are equivalent:

(1) *The metrics d_S and d_{S_*} are not roughly similar.*

(2) *The Manhattan curve $\mathscr{C}_M$ is strictly convex at 0, ie $\theta''_{S_*/S}(0) > 0$.*

(3) *The Manhattan curve $\mathscr{C}_M$ is strictly convex at every point, ie $\theta''_{S_*/S}(a) > 0$ for every $a \in \mathbb{R}$.*

Remark 4.18 Let $\boldsymbol{S}_n = \{x \in \Gamma : |x|_S = n\}$. If one of the equivalence statements in Theorem 4.17 holds, then the law of

$$\frac{d_{S_*}(o, x_n) - n\tau(S_*/S)}{\sqrt{n}}, \tag{4-4}$$

where x_n is chosen uniformly at random from $\boldsymbol{S}_n$, converges, as $n \to \infty$, to the normal distribution with mean 0 and variance $\theta''_{S_*/S}(0) > 0$. This follows from [Cantrell 2021, Theorems 1.1 and 1.2] and [Gekhtman et al. 2022, Theorem 1.1] (where Gekhtman et al. have established their result in a more general setting). Observe that if d_S and d_{S_*} are roughly similar, then the limiting distribution of (4-4) is the Dirac mass at 0.

Let $\overline{\Sigma}$ denote a subshift of finite type associated to a strongly Markov automatic structure $\mathscr{A} = (\mathscr{G}, \pi, S)$.

Lemma 4.19 *Let $\Psi_* \colon \overline{\Sigma} \to \mathbb{R}$ be a Hölder continuous function such that*

$$S_n\Psi_*(\omega) = -d_{S_*}(o, \pi(\omega_0)\cdots\pi(\omega_{n-1})) \quad \text{for } \omega = (\omega_0, \dots, \omega_{n-1}) \in \Sigma^*,$$

as in Example 4.10. The function $\Psi_* \colon \Sigma_{\mathscr{C}} \to \mathbb{R}$ is cohomologous to a constant on a maximal component $\mathscr{C}$, ie there exist a constant $c_0 \in \mathbb{R}$ and a Hölder continuous function $u \colon \Sigma_{\mathscr{C}} \to \mathbb{R}$ such that

$$\Psi_* = c_0 + u - u \circ \sigma$$

if and only if d_S and d_{S_*} are roughly similar.

Proof For a maximal component $\mathscr{C}$, let $\Gamma_{\mathscr{C}}$ be the set of group elements that are realized as the image of a word corresponding to a finite path in $\mathscr{C}$. Recall that Ψ_* is cohomologous to a constant on $\Sigma_{\mathscr{C}}$ if and only if there exists $\tau \in \mathbb{R}$ such that the set

$$\{S_n\Psi_*(\omega) + n\tau : \omega \in \Sigma_{\mathscr{C}} \text{ and } n \geq 0, n \in \mathbb{Z}\}$$

is a bounded subset of $\mathbb{R}$. By the definition of Ψ_*, this holds if and only if

$$\{d_{S_*}(o,x) - \tau d_S(o,x) : x \in \Gamma_{\mathscr{C}}\}$$

is a bounded subset of $\mathbb{R}$. We show that this is equivalent to the fact that d_S and d_{S_*} are roughly similar.

Let us prove that for each maximal component $\mathscr{C}$ there exists a finite set of group elements $B \subset \Gamma$ such that $B\Gamma_{\mathscr{C}}B = \Gamma$. This claim, which concludes the proof, is essentially observed in Lemma 4.6 of [Gouëzel et al. 2018]; we provide a proof below for the sake of completeness.

For an element $w \in \Gamma$ and a real number $\Delta \geq 0$, we say that an S-geodesic word Δ-*contains* w if it contains a subword h such that $h = h_1 w h_2$ for some $h_1, h_2 \in \Gamma$ with $|h_1|_S, |h_2|_S \leq \Delta$. Let $Y_{w,\Delta}$ be the set of group elements $x \in \Gamma$ such that x is represented by some S-geodesic word which does not Δ-contain w. It is known that there exists $\Delta_0 > 0$ such that for all $w \in \Gamma$,

$$\lim_{n\to\infty} \frac{\#(Y_{w,\Delta_0} \cap \boldsymbol{S}_n)}{\#\boldsymbol{S}_n} = 0,$$

see [Arzhantseva and Lysenok 2002, Theorem 3]. Since $\mathscr{C}$ is a maximal component in the underlying directed graph $\mathscr{G}$ and thus the spectral radius is the exponential volume growth rate relative to S, the upper density of $\Gamma_{\mathscr{C}}$ is strictly positive, ie

$$\limsup_{n\to\infty} \frac{\#(\Gamma_{\mathscr{C}} \cap \boldsymbol{S}_n)}{\#\boldsymbol{S}_n} > 0.$$

Fix $w \in \Gamma$. Since $\Gamma_{\mathscr{C}}$ has positive upper density and Y_{w,Δ_0} has vanishing density, $\Gamma_{\mathscr{C}} \backslash Y_{w,\Delta_0} \neq \varnothing$, ie there is an element x of $\Gamma_{\mathscr{C}}$ whose every S-geodesic representation Δ_0-contains w, in particular there exists $y \in \Gamma_{\mathscr{C}}$ (corresponding to a subword) such that $y = h_1 w h_2$ with $|h_1|_S, |h_2|_S \leq \Delta_0$. Hence if we let B denote the ball of radius Δ_0 with respect to d_S centered at the identity in Γ, then $\Gamma = B\Gamma_{\mathscr{C}}B$. □

Proof of Theorem 4.17 First let us show that (1) $\iff$ (2). By Theorem 4.14 (see the proof of Proposition 4.13), the second derivative $\theta''_{S_*/S}(0)$ coincides with the second derivative at $t = 0$ of the function $\mathrm{Pr}_{\mathscr{C}}(t\Psi_*)$ for each fixed maximal component $\mathscr{C}$ (at $t = 0$). By Proposition 4.12 of [Parry and Pollicott 1990] this second derivative is strictly positive if and only if $\Psi_* \colon \Sigma_{\mathscr{C}} \to \mathbb{R}$ is not cohomologous to a constant. Lemma 4.19 implies that this is true if and only if d_S and d_{S_*} are not roughly similar.

Next let us show that (2) $\iff$ (3). We shall in fact show that if $\theta''_{S_*/S}(t) > 0$ for some $t \in \mathbb{R}$, then $\theta''_{S_*/S}(t) > 0$ for all $t \in \mathbb{R}$. Let us fix $t_0 \in \mathbb{R}$ and let $\mathscr{C}_1, \dots, \mathscr{C}_I$ be the maximal components of $t\Psi_*$ at $t = t_0$. By Proposition 4.6 there exists $\varepsilon > 0$ such that

$$\theta_{S_*/S}(t) = \max_{i=1,\dots,I} \mathrm{Pr}_{\mathscr{C}_i}(t\Psi_*)$$

for all $|t - t_0| < \varepsilon$. Since each $\mathrm{Pr}_{\mathscr{C}_i}(t\Psi_*)$ is real analytic in $t \in \mathbb{R}$, changing $\varepsilon > 0$ if necessary we find at most two components $\mathscr{C}_1$ and $\mathscr{C}_2$ (possibly $\mathscr{C}_1 = \mathscr{C}_2$) such that

$$\theta_{S_*/S}(t) = \begin{cases} \mathrm{Pr}_{\mathscr{C}_1}(t\Psi_*) & \text{for } t_0 \leq t < t_0 + \varepsilon, \\ \mathrm{Pr}_{\mathscr{C}_2}(t\Psi_*) & \text{for } t_0 - \varepsilon < t \leq t_0. \end{cases} \tag{4-5}$$

Moreover, $\mathrm{Pr}_{\mathscr{C}_1}(t\Psi_*)$ (resp. $\mathrm{Pr}_{\mathscr{C}_2}(t\Psi_*)$) is strictly convex at all points if and only if it is strictly convex at some point, since this is equivalent to the fact that Ψ_* is not cohomologous to a constant function on $\Sigma_{\mathscr{C}_1}$ (resp. $\Sigma_{\mathscr{C}_2}$) [Parry and Pollicott 1990, Proposition 4.12]. It follows that if $\theta''_{S_*/S}(t_0) > 0$, then both $\mathrm{Pr}_{\mathscr{C}_1}(t\Psi_*)$ and $\mathrm{Pr}_{\mathscr{C}_2}(t\Psi_*)$ are strictly convex at all points, since

$$\theta''_{S_*/S}(t_0) = \mathrm{Pr}''_{\mathscr{C}_1}(t\Psi_*)|_{t=t_0} = \mathrm{Pr}''_{\mathscr{C}_2}(t\Psi_*)|_{t=t_0} > 0$$

by the proof of Proposition 4.13. Note that since there are only finitely many components in the underlying directed graph, the set of $t_0 \in \mathbb{R}$ where (4-5) holds for two distinct $\mathscr{C}_1$ and $\mathscr{C}_2$ is at most countable and discrete in $\mathbb{R}$. Applying the same argument to such t_0 at most countably many times, we see that if $\theta_{S_*/S}$ is strictly convex at some point, then it is strictly convex at all points, as desired. □

4.6 Tightness of the tangent lines at infinity for the Manhattan curve

In this section we prove an inequality for pressure curves that will be a useful tool in understanding the asymptotic properties of $\mathscr{C}_M$. This inequality will have subsequent applications to a large deviation principle (Theorem 4.23), our results on the multifractal spectrum (Theorem 3.8) and to proving the rationality of the dilation constants associated to word metrics (Proposition 4.22).

Proposition 4.20 *Let (Σ,σ) be a transitive subshift of finite type and $\Psi\colon \Sigma \to \mathbb{R}$ be a Hölder continuous function. If we define*

$$P_\infty(\Psi) := \sup\left\{\exp\left(\frac{S_n\Psi(\omega)}{n}\right) : \sigma^n\omega = \omega \text{ for } \omega \in \Sigma \text{ and } n \ge 1\right\},$$

then we have that

$$\frac{1}{\rho_A} e^{\mathrm{Pr}(t\Psi)} \le P_\infty(t\Psi) \le e^{\mathrm{Pr}(t\Psi)} \quad \text{for all } t \in \mathbb{R},$$

where $\mathrm{Pr}(t\Psi)$ is the pressure for $t\Psi$ and ρ_A is the spectral radius of the adjacency matrix for (Σ,σ).

In particular, it holds that

$$\mathrm{Pr}(t\Psi) = t\log P_\infty(\Psi) + O(1) \quad \text{and} \quad \mathrm{Pr}(-t\Psi) = t\log P_\infty(-\Psi) + O(1) \quad \text{as } t \to \infty.$$

Proof Let E be a finite set of alphabets and $A = (A_{e,e'})_{e,e' \in E}$ be the adjacency matrix which defines the transitive subshift of finite type (Σ,σ) on E. We consider the associated finite directed graph $\mathscr{G}$ whose set of edges is E. Let us denote by Ω_E the set of cycles in $\mathscr{G}$, ie the set of finite paths $w = (\omega_0,\dots,\omega_{n-1})$ whose terminus coincides with the origin, and by $|w| = n$ the length of w. Recall that the transfer operator

$$\mathscr{L}_\Psi f(\omega) = \sum_{\sigma(\omega')=\omega} e^{\Psi(\omega')} f(\omega') \quad \text{for } f\colon \Sigma \to \mathbb{R}$$

has spectral radius $e^{\mathrm{Pr}(\Psi)}$.

First we consider a special case (from which the general case will be reduced); $\Psi\colon \Sigma \to \mathbb{R}$ depends only on the first coordinate, ie $\Psi(\omega) = \psi(\omega_0)$ for some function $\psi\colon E \to \mathbb{R}$. Let

$$P_{\max}(\psi) := \max\left\{\prod_{e\in w} e^{\psi(e)/|w|} : w \in \Omega_E\right\}, \tag{4-6}$$

where we note that the maximum is attained by some simple cycle (ie a cycle consisting of pairwise distinct vertices). Then we have that

$$P_{\max}(\psi) \le e^{\Pr(\Psi)}$$

by the definition of spectral radius, and moreover,

$$e^{\Pr(\Psi)} \le \rho_A \cdot P_{\max}(\psi),$$

where ρ_A stands for the spectral radius of the adjacency matrix A by [Friedland 1986, Theorem 2] — we have applied the nonnegative matrix $(e^{\psi(e)} A_{e,e'})_{e,e' \in E}$; note that we have $\rho_A = e^{\Pr(0)}$. Therefore we obtain

$$P_{\max}(t\psi) \le e^{\Pr(t\Psi)} \le \rho_A \cdot P_{\max}(t\psi) \quad \text{for all } t \in \mathbb{R}. \tag{4-7}$$

Next we consider the general case. Since Ψ is Hölder continuous, for all $\varepsilon > 0$ sufficiently small there exists a function $\Psi_0 \colon \Sigma \to \mathbb{R}$ such that Ψ_0 depends on finitely many coordinates and satisfies $\|\Psi_0 - \Psi\|_\infty \le \varepsilon$. Let us recode the subshift Σ, where an induced potential depends only on the first coordinate: if Ψ_0 depends only on at most the first K coordinates for some $K > 0$, then we replace all K-length strings allowed by A with a new symbol. We thereby obtain a new transitive subshift Σ_B with an adjacency matrix B. Note that the natural bijective map $\Sigma \to \Sigma_B$, $\omega = (\omega_i)_{i=0}^\infty \mapsto \widetilde{\omega} = (\omega_i \omega_{i+1} \cdots \omega_{i+K-1})_{i=0}^\infty$, defined by concatenation of subsequent K-alphabets yields an isomorphism between the shifts. If we define $\Psi_0' \colon \Sigma_B \to \mathbb{R}$ by $\Psi_0'(\widetilde{\omega}) = \Psi_0(\omega)$, then $\Psi_0'(\widetilde{\omega}) = \psi(\widetilde{\omega}_0)$ for some function ψ on Σ_B, and the spectral radii of transfer operators for Ψ_0 and Ψ_0' coincide, therefore so do those of the adjacency matrices A and B. Applying (4-7), we obtain

$$e^{\Pr(t\Psi_0)} \le \rho_B \cdot P_{\max}(t\psi).$$

Further, since $\rho_A = \rho_B$ and the inequality

$$|\Pr(t\Psi) - \Pr(t\Psi_0)| \le \|t\Psi - t\Psi_0\|_\infty \quad \text{for all } t \in \mathbb{R}, \tag{4-8}$$

which follows from (4-1), we have that

$$e^{\Pr(t\Psi)} \le e^{\varepsilon|t|} \rho_A \cdot P_{\max}(t\psi).$$

Combining with

$$P_{\max}(t\psi) \le e^{\varepsilon|t|} \sup\left\{\exp\left(\frac{tS_n\Psi(\omega)}{n}\right) : \sigma^n\omega = \omega \text{ for } \omega \in \Sigma \text{ and } n \ge 1\right\},$$

which follows from the definition (4-6) and (4-8), we obtain $e^{\Pr(t\Psi)} \le e^{2\varepsilon|t|}\rho_A \cdot P_\infty(t\Psi)$. Noting that this estimate is uniform in $t \in \mathbb{R}$ as $\varepsilon \to 0$, we have that $e^{\Pr(t\Psi)} \le \rho_A \cdot P_\infty(t\Psi)$ for all $t \in \mathbb{R}$. Similarly, by using (4-7) we have $P_\infty(t\Psi) \le e^{\Pr(t\Psi)}$ for all $t \in \mathbb{R}$, concluding the first claim. Further noting that $P_\infty(t\Psi) = P_\infty(\Psi)^t$ for all $t \ge 0$, we obtain the second claim. □

Remark 4.21 Richard Sharp has suggested that the above proof can be simplified by appealing to the fact that the set of uniform measures on periodic orbits is dense in the set of all σ-invariant Borel probability measures in the weak topology [Sigmund 1970, Theorem 1]. We have provided a more elementary approach which gives a clearer insight into the following proposition.

Proposition 4.22 *For every pair of word metrics d_S and d_{S_*} associated to finite symmetric sets of generators S and S_*, respectively, the corresponding Manhattan curve satisfies*

$$\theta_{S_*/S}(t) = -\alpha_{\min} t + O(1) \quad \text{as } t \to \infty \qquad \text{and} \qquad \theta_{S_*/S}(t) = -\alpha_{\max} t + O(1) \quad \text{as } t \to -\infty.$$

Moreover, $\alpha_{\min}$ and $\alpha_{\max}$ are rational.

Proof We apply Proposition 4.20 to each transitive component in the subshift in Example 4.10; for each $t \in \mathbb{R}$, we have that

$$\theta_{S_*/S}(t) = \max_{\mathscr{C}} \mathrm{Pr}_{\mathscr{C}}(t\Psi_{S_*}),$$

where $\mathscr{C}$ is a component in the underlying directed graph by Lemma 4.8. Although the components which attain the maximum can depend on the t, since we have $\alpha_{\min} = -\lim_{t\to\infty} \theta_{S_*/S}(t)/t$ and there are only finitely many components, there exists $\mathscr{C}$ such that $\alpha_{\min} = -\log P_\infty(\Psi_{S_*|\mathscr{C}})$, where $\Psi_{S_*|\mathscr{C}}$ is the restriction of Ψ_{S_*} to $\Sigma_{\mathscr{C}}$. Therefore $\theta_{S_*/S}(t) = -\alpha_{\min} t + O(1)$ as $t \to \infty$. Similarly, we have $\alpha_{\max} = \log P_\infty(-\Psi_{S_*|\mathscr{C}})$ for a possibly different $\mathscr{C}$, implying that $\theta_{S_*/S}(t) = -\alpha_{\max} t + O(1)$ as $t \to -\infty$.

Furthermore, since $\Psi_{S_*}(\omega) = -d\phi_{S_*}(\omega_0)$ for $\omega \in \Sigma$ and $d\phi_{S_*}$ takes integer values (Example 4.10), by (4-6) in the proof of Proposition 4.20, we have that

$$\alpha_{\min} = \sum_{e \in w} \frac{d\phi_{S_*}(e)}{|w|} \quad \text{for a cycle } w = (e_0, \dots, e_{k-1}) \text{ and } k = |w|,$$

and $\alpha_{\max}$ has a similar form. Hence $\alpha_{\min}$ and $\alpha_{\max}$ are rational. □

4.7 An application to large deviations

In Section 3 we compared the typical growth rates of two metrics $d, d_* \in \mathscr{D}_\Gamma$ by studying two related limits. In Lemma 3.5 we studied the limiting ratio of the metrics as we travel along "typical" quasigeodesic rays. We then, in Theorem 3.12, considered the limiting average of the ratio of two metrics, where we average over n-balls in one of the metrics. In this section we investigate a finer statistical result that compares a metric $d \in \mathscr{D}_\Gamma$ with a word metric d_S. More precisely, we study the distribution of $d(o,x)/n$ when x is sampled uniformly at random from the set of all words of length n in d_S.

It has been shown that if $d \in \mathscr{D}_\Gamma$ is a word metric or is strongly hyperbolic, then there exists a positive real number τ such that

$$\text{(4-9)} \qquad \frac{1}{\#\boldsymbol{S}_n} \sum_{|x|_S = n} \frac{d(o,x)}{n} \to \tau \quad \text{as } n \to \infty,$$

where $\boldsymbol{S}_n = \{x \in \Gamma : |x|_S = n\}$; see [Cantrell 2021, Theorem 1.1]. (In fact, a stronger result was shown: the left-hand side of (4-9) is $\tau + O(1/n)$ as $n \to \infty$.) Furthermore, the values $d(o,x)/n$ concentrate exponentially near τ, ie for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{\#\boldsymbol{S}_n} \#\left\{x \in \boldsymbol{S}_n : \left|\frac{d(o,x)}{n} - \tau\right| > \varepsilon\right\}\right) < 0.$$

A detailed analysis of the Manhattan curve allows us to establish a precise global large deviation result that is valid for every metric $d \in \mathscr{D}_\Gamma$.

Theorem 4.23 *Let Γ be a nonelementary hyperbolic group equipped with a finite symmetric set of generators S. Let $\boldsymbol{S}_n := \{x \in \Gamma : |x|_S = n\}$ for nonnegative integers $n \geq 0$. If $d \in \mathscr{D}_\Gamma$, then for every open set U and every closed set V in $\mathbb{R}$ such that $U \subset V$, we have that*

$$\begin{aligned} -\inf_{s\in U} I(s) &\leq \liminf_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{\#\boldsymbol{S}_n} \#\left\{x \in \boldsymbol{S}_n : \frac{d(o,x)}{n} \in U\right\}\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{\#\boldsymbol{S}_n} \#\left\{x \in \boldsymbol{S}_n : \frac{d(o,x)}{n} \in V\right\}\right) \leq -\inf_{s\in V} I(s), \end{aligned}$$

where

$$I(s) = \theta_{d/d_S}(0) + \sup_{t\in\mathbb{R}} \{ts - \theta_{d/d_S}(-t)\},$$

and θ_{d/d_S} is the Manhattan curve for the pair of metrics d and d_S. Furthermore, I has a unique zero at the mean distortion $\tau(d/d_S)$, is finite on $(\alpha_{\min}, \alpha_{\max})$, and infinite outside of $[\alpha_{\min}, \alpha_{\max}]$, where

$$\alpha_{\min} = -\lim_{a\to\infty} \frac{\theta_{d/d_S}(a)}{a} \quad \text{and} \quad \alpha_{\max} = -\lim_{a\to-\infty} \frac{\theta_{d/d_S}(a)}{a}.$$

Remark 4.24 In a recent work, Gekhtman, Taylor and Tiozzo [Gekhtman et al. 2022, Theorem 1.1] have established the corresponding central limit theorem in the case when d is a (not necessarily proper) hyperbolic metric; we are not aware of the corresponding large deviation principle in the nonproper setting, see eg [Gekhtman et al. 2018, Theorem 7.3] (for a recent analogous result on random walks, see [Boulanger et al. 2023]).

Proof of Theorem 4.23 For every $t \in \mathbb{R}$, we have that

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{\#\boldsymbol{S}_n} \sum_{|x|_S = n} e^{td(o,x)}\right) = \theta_{d/d_S}(-t) - \theta_{d/d_S}(0)$$

by Lemma 2.8, and furthermore the right-hand side is continuously differentiable by Theorem 3.7. The theorem then follows from the Gärtner–Ellis theorem (eg [Dembo and Zeitouni 1998, Theorem 2.3.6]) and the definitions of I and θ_{d/d_S}. □

Corollary 4.25 *Suppose $d, d_S \in \mathscr{D}_\Gamma$ are as in Theorem 4.23. Further assume that d is a word metric associated to a finite symmetric set of generators S_*, ie $d = d_{S_*}$. Then $I(\alpha_{\min}), I(\alpha_{\max}) < \infty$ and further, $\alpha_{\min}$ and $\alpha_{\max}$ are rational.*

Proof This follows from Proposition 4.22. □

Remark 4.26 If we take a strongly hyperbolic metric d in Theorem 4.23, then $I(\alpha_{\min}), I(\alpha_{\max}) < \infty$ by Proposition 4.20 and by the first part in the proof of Proposition 4.22. We are unsure whether both $I(\alpha_{\min})$ and $I(\alpha_{\max})$ are finite for all $d \in \mathscr{D}_\Gamma$, which we leave open.

5 Examples

In this section we compute the Manhattan curve for two examples, focusing on a pair of word metrics. In the first, we provide an exact formula for a Manhattan curve associated to a free group. In the second, we analyze a hyperbolic triangle group, which is another explicit class of hyperbolic groups other than free groups. We obtain a Manhattan curve as an implicit function for some pair of word metrics in the $(3,3,4)$-triangle group. For both cases, we use the GAP package [GAP Group 2020] to produce explicit forms of automatic structures.

5.1 The free group of rank 2

The following example is considered in [Calegari 2009, Example 6.5.5]. We extend the discussion found in [Calegari 2009] by commenting on the Manhattan curve. Let $F = \langle a, b \mid \rangle$ be the free group of rank 2. We consider the standard set of generators $S = \{a, b, a^{-1}, b^{-1}\}$ and another symmetric set of generators

$$S_* := \{a, b, c, a^{-1}, b^{-1}, c^{-1}\}, \quad \text{where } c := ab.$$

Our aim is to compute the Manhattan curve θ_{S/S_*}. Note that a word on S_* is reduced if and only if it contains no subword of the form $a^{-1}c, cb^{-1}, c^{-1}a, bc^{-1}$, and furthermore each element in F has a unique reduced word representative on S_*. Henceforth we use A, B, C to denote a^{-1}, b^{-1}, c^{-1} respectively. An automatic structure for (F, S) (resp. for (F, S_*)) is given in Figure 2 (resp. Figure 3). (In the figures the arrows are solid if they are in the strongly connected component, and dotted otherwise. The initial state is denoted by "1".)

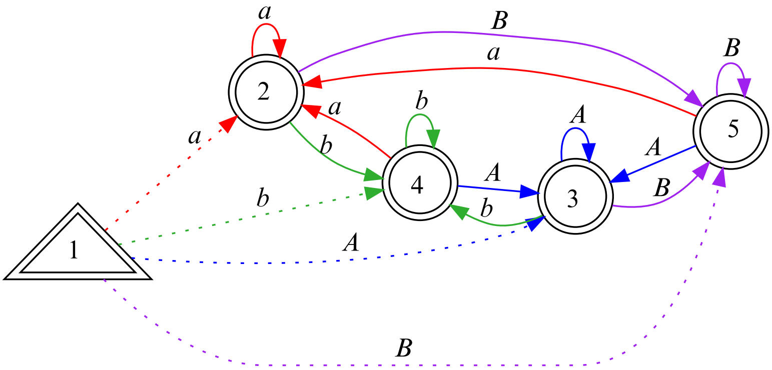


Figure 2: An automatic structure for (Γ, S) in Section 5.1.

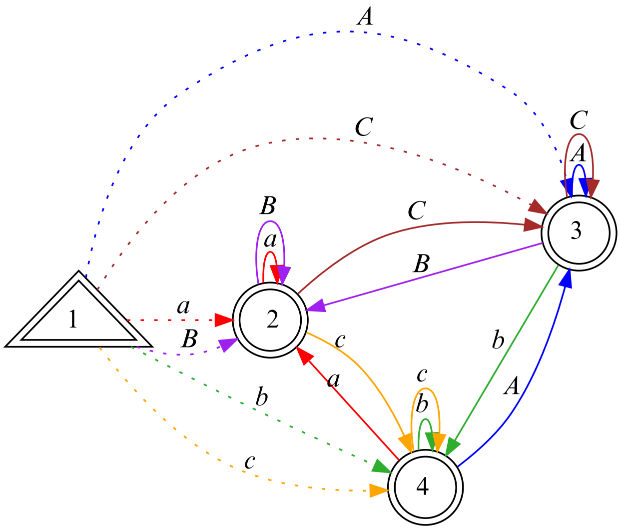


Figure 3: An automatic structure for (Γ, S_*) in Section 5.1.

The word-length with respect to S is computed by setting

$$d\phi_S(e) = \begin{cases} 1 & \text{if } e \text{ has the label } a, b, A, B, \\ 2 & \text{if } e \text{ has the label } c, C, \end{cases}$$

where $d\phi_S$ is defined as in Example 4.10. The adjacency matrix of this directed graph is of size 12 (because we use all the edges in the strongly connected component as indices), but it is enough to deal with a smaller one: we observe that the "flip" of the labels by $c \leftrightarrow C$, $a \leftrightarrow B$ and $b \leftrightarrow A$ keeps the directed graph structure with labeling. This allows us to consider the following matrix of size 6, where indices correspond to the set of labels a, b, c, A, B, C:

$$\begin{array}{c} \\ a \\ b \\ c \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cccccc} a & b & c & A & B & C \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

Then the Manhattan curve is computed as $\theta_{S/S_*}(t) = \log P(e^{-t})$, where $P(s)$ is the spectral radius of the matrix

$$\begin{pmatrix} s & 0 & s^2 & 0 & s & s^2 \\ s & s & s^2 & s & 0 & 0 \\ s & s & s^2 & s & 0 & 0 \\ 0 & s & 0 & s & s & s^2 \\ s & 0 & s^2 & 0 & s & s^2 \\ 0 & s & 0 & s & s & s^2 \end{pmatrix}$$

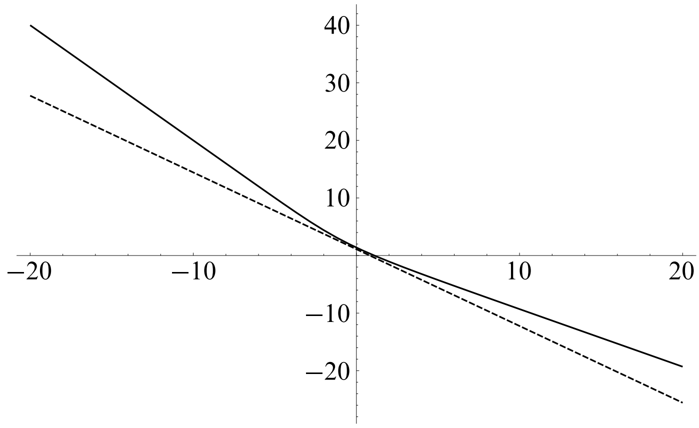


Figure 4: The Manhattan curve (solid) for the example in Section 5.1 and the tangent line at 0 (dotted) for comparison.

(see Lemma 4.8). Hence we obtain

$$\theta_{S/S_*}(t) = \log\bigl(\tfrac{1}{2}e^{-t}(e^{-t} + \sqrt{e^{-t}(e^{-t}+8)} + 4)\bigr),$$
$$\alpha_{\max} = \lim_{t\to-\infty} -\frac{1}{t}\theta_{S/S_*}(t) = 2 \quad \text{and} \quad \alpha_{\min} = \lim_{t\to\infty} -\frac{1}{t}\theta_{S/S_*}(t) = 1,$$

see Figure 4. Moreover, the mean distortion $\tau(S/S_*)$ and v_*/v, where v (resp. v_*) is the exponential volume growth rate for (Γ, S) (resp. (Γ, S_*)), are given by

$$\tau(S/S_*) = -\theta'_{S/S_*}(0) = \frac{4}{3} = 1.33333\ldots \quad \text{and} \quad \frac{v_*}{v} = \frac{\log 4}{\log 3} = 1.26186\ldots.$$

5.2 The (3, 3, 4)-triangle group

We now turn our attention to computing a Manhattan curve for a pair of word metrics in the case of the (3, 3, 4)-triangle group. Let

$$\Gamma := \langle a, b, c \mid a^3, b^3, c^4, abc\rangle,$$

where we denote the standard set of generators by

$$S := \{a, b, c, a^{-1}, b^{-1}, c^{-1}\},$$

and another symmetric set of generators by

$$S_* := \{a, b, c, d, a^{-1}, b^{-1}, c^{-1}, d^{-1}\}, \quad \text{where } d := c^2.$$

Note that Γ has a presentation with respect to S_* given by

$$\langle a, b, c, d \mid a^3, b^3, c^4, abc, d^{-1}c^2\rangle.$$

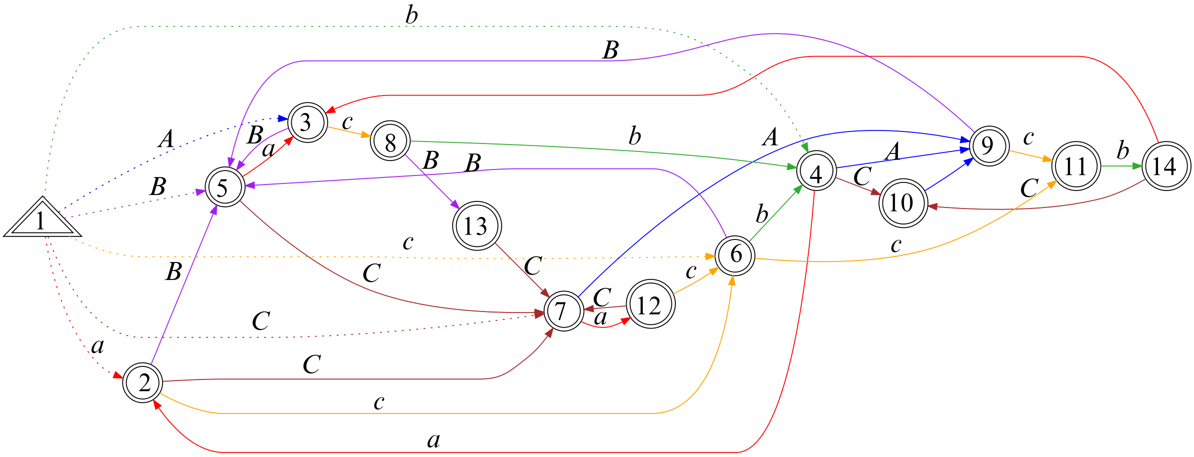


Figure 5: An automatic structure for (Γ, S) in Section 5.2.

Based on these presentations for Γ, we compute the exponential growth rate v (resp. v_*) for (Γ, S) (resp. (Γ, S_*)) to be

$$v = 0.674756\ldots \quad \text{and} \quad v_* = 0.732858\ldots,$$

and we have produced automatic structures for (Γ, S) and for (Γ, S_*) in Figures 5 and 6.

Following the method in Section 5.1, we find a matrix representation of the transfer operator, which is read off from the strongly connected component with appropriately defined weights in Figure 7. The characteristic polynomial of the transfer matrix of size 21 is

$$-e^{-9s}x^{13}(-1+e^s x)(1+e^s x)(1+e^s x+e^{2s}x^2)(1-e^s x-e^{2s}x^2-e^{3s}x^2-e^{4s}x^3+e^{5s}x^4).$$

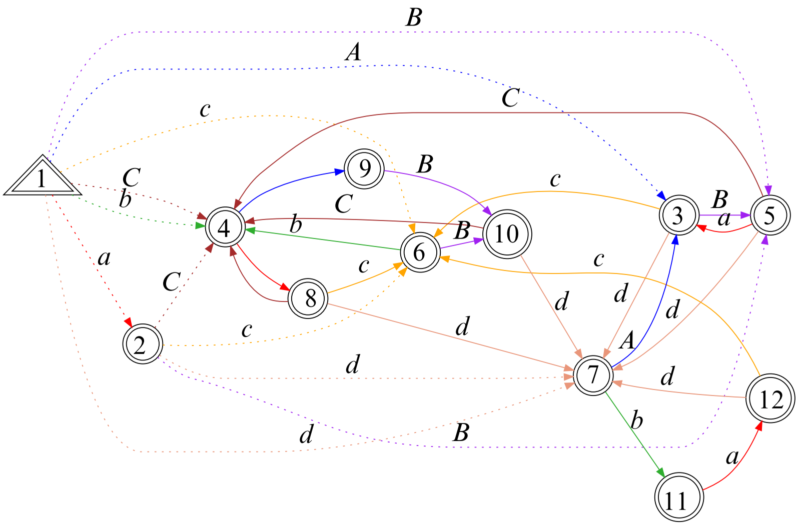


Figure 6: An automatic structure for (Γ, S_*) in Section 5.2.

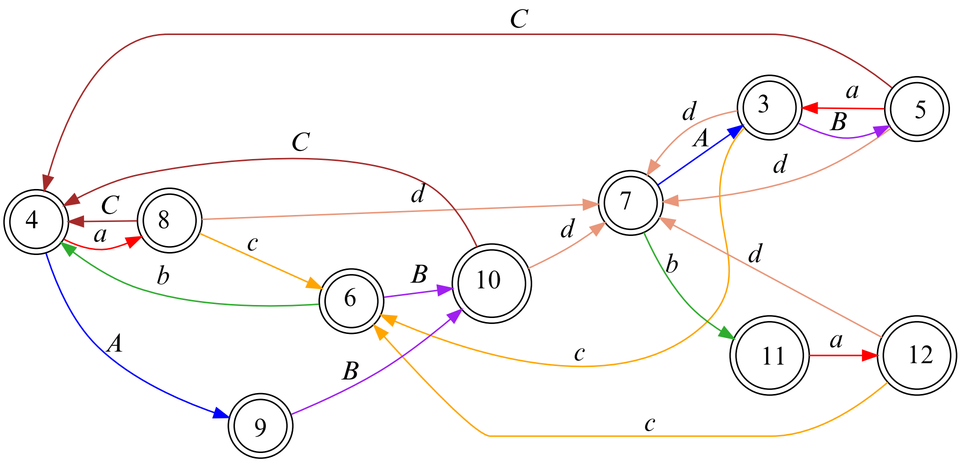


Figure 7: The strongly connected component of the automatic structure for (Γ, S_*) in Figure 6. The weight 1 is assigned on the directed edges with labels a, A, b, B, c, C, and the weight 2 is assigned on the directed edge with label d.

Let $s \mapsto r(s)$ be the branch given as a root of

$$e^{3s} - e^{4s}x - e^{5s}x^2 - x^{6s}x^2 - e^{7s}x^3 + e^{8s}x^4$$

such that $r(s)$ coincides with the spectral radius of the transfer operator at $s = 0$. Then the Manhattan curve is obtained by $\theta_{S/S_*}(s) = \log r(s)$; see Figure 8. We find that $\alpha_{\max} = \frac{3}{2}$ and $\alpha_{\min} = 1$. Moreover, the expansion of $\theta_{S/S_*}(s)$ at 0 has the form

$$\theta_{S/S_*}(s) = 0.732858 \cdots - 1.18937\ldots s + 0.0515301\ldots s^2 + O(s^3),$$

and the mean distortion $\tau(S/S_*)$ and v_*/v are given by

$$\tau(S/S_*) = \tfrac{1}{68}(85 - \sqrt{17}) = 1.18937\ldots \quad \text{and} \quad \frac{v_*}{v} = 1.08611\ldots.$$

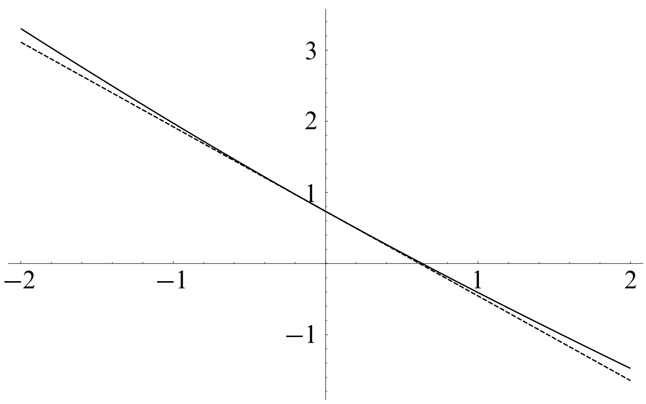


Figure 8: The Manhattan curve (solid) for the example in Section 5.2, and the tangent line at 0 (dotted) for comparison.

Note that $\tau(S/S_*)$ is quadratic irrational although it is a priori a root of a higher-degree polynomial with rational coefficients. We are still far from having a systematic understanding of the exact class of numbers to which $\tau(S/S_*)$ can belong in general. We believe that this deserves further investigation.

Appendix Proof of Lemma 3.11

In this section we will refer to some fundamental facts from [Heinonen 2001]. Let X be a topological space endowed with a quasimetric ρ which is compatible with the topology on X, and let us call (X, ρ) a *quasimetric space*. Let μ be a Borel regular measure on a topological space (X, ρ). For $r > 0$ and $x \in X$, we denote the ball of radius r centered at x relative to the quasimetric ρ by

$$B(x,r) := \{y \in X : \rho(x,y) < r\}.$$

A measure μ is called *doubling* if all balls have *finite* and *positive* μ-measure and there exists a constant $C(\mu) > 0$ such that $\mu(B(x,2r)) \le C(\mu)\mu(B(x,r))$ for every ball $B(x,r)$ in X. We call $C(\mu)$ a *doubling constant* of μ.

Let (X, ρ) be a quasimetric space which admits a doubling measure μ. For every Borel regular finite (nonnegative) measure ν on X, ie $\nu(X) < \infty$, let us decompose

$$\nu = \nu_{\mathrm{ac}} + \nu_{\mathrm{sing}},$$

where ν_{ac} is the absolutely continuous part of ν and ν_{sing} is the singular part of ν relative to μ. Since ν is finite, ν_{ac} is also finite and thus $d\nu_{\mathrm{ac}}/d\mu$ is integrable.

Lemma A.1 *For μ-almost every $x \in X$, we have that*

$$\lim_{r\to 0} \frac{\nu_{\mathrm{ac}}(B(x,r))}{\mu(B(x,r))} = \frac{d\nu_{\mathrm{ac}}}{d\mu}(x) < \infty \quad \textit{and} \quad \limsup_{r\to 0} \frac{\nu_{\mathrm{sing}}(B(x,r))}{\mu(B(x,r))} = 0.$$

Proof By adapting the proof of the Lebesgue differentiation theorem [Heinonen 2001, Theorem 1.8] to a quasimetric, we have

$$\lim_{r\to 0} \frac{\nu_{\mathrm{ac}}(B(x,r))}{\mu(B(x,r))} = \frac{d\nu_{\mathrm{ac}}}{d\mu}(x) \quad \text{for } \mu\text{-almost every } x \in X.$$

Now we show the second claim. Let us write $\nu = \nu_{\mathrm{sing}}$. Since ν and μ are mutually singular, there exists a measurable set N in X such that $\nu(N) = \nu(X)$ and $\mu(N) = 0$. Further, (the singular part) ν is also Borel regular and the inner regularity implies that for all $\varepsilon > 0$ there exists a compact set K_ε in N such that $\nu(N \setminus K_\varepsilon) < \varepsilon$. Let $\nu_\varepsilon := \nu|_{X\setminus K_\varepsilon}$ be the restriction of ν on $X \setminus K_\varepsilon$, for which $\nu_\varepsilon(X) < \varepsilon$. For all $x \in X \setminus K_\varepsilon$, there exists a small enough $r > 0$ such that $B(x,r) \subset X \setminus K_\varepsilon$. If we define

$$L\nu(x) := \limsup_{r\to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))} \quad \text{and} \quad M\nu_\varepsilon(x) := \sup_{r>0} \frac{\nu_\varepsilon(B(x,r))}{\mu(B(x,r))} \quad \text{for } x \in X,$$

then for all $t > 0$,

$$\{L\nu > t\} \subset K_\varepsilon \cup \{M\nu_\varepsilon > t\}.$$

The weak maximal inequality shows that

$$\mu(\{M\nu_\varepsilon > t\}) \leq \frac{C_\mu}{t}\nu_\varepsilon(X)$$

for all $t > 0$ and for a constant C_μ depending only on the doubling constant of μ (where the proof follows exactly as for (2.3) in [Heinonen 2001, Theorem 2.2]), and thus

$$\mu(\{L\nu > t\}) \leq \mu(K_\varepsilon) + \mu(\{M\nu_\varepsilon > t\}) \leq \frac{C_\mu}{t}\nu_\varepsilon(X) < \frac{C_\mu}{t}\varepsilon,$$

where we have used the fact that $\mu(K_\varepsilon) \leq \mu(N) = 0$. Therefore for each $t > 0$, letting $\varepsilon \to 0$, we have that

$$\mu(\{L\nu > t\}) = 0,$$

and this shows that $L\nu(x) = 0$ for μ-almost every $x \in X$, as required. □

Remark A.2 The above fact is standard in metric spaces with doubling measures, and the part adapted to a quasimetric is mainly the place where we use Vitali's covering theorem. We avoid repeating all the details: see [Heinonen 2001, Chapters 1 and 2].

Proof of Lemma 3.11 If we fix a large enough $R > 0$ given C for which $d \in \mathscr{D}_\Gamma$ is a C-rough geodesic metric, then Lemma 2.3 and Lemma A.1 together with the fact that μ is doubling show that

$$\limsup_{n\to\infty} \frac{\nu_{\mathrm{ac}}(O(\xi_n, R))}{\mu(O(\xi_n, R))} \leq C_{\mu,R}\frac{d\nu_{\mathrm{ac}}}{d\mu}(\xi) < \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{\nu_{\mathrm{sing}}(O(\xi_n, R))}{\mu(O(\xi_n, R))} = 0$$

for μ-almost every ξ in $\partial\Gamma$, where $C_{\mu,R,\delta}$ is a positive constant depending only on a doubling constant of μ and R as well as the δ-hyperbolicity constant of d, and we recall that $\xi_n = \gamma_\xi(n)$ for $n \geq 1$ and γ_ξ is a C-rough geodesic ray from o converging to ξ. This concludes the claim. □

References

[Arzhantseva and Lysenok 2002] **G N Arzhantseva**, **I G Lysenok**, *Growth tightness for word hyperbolic groups*, Math. Z. 241 (2002) 597–611 MR Zbl

[Bader and Furman 2017] **U Bader**, **A Furman**, *Some ergodic properties of metrics on hyperbolic groups*, preprint (2017) arXiv 1707.02020

[Blachère et al. 2011] **S Blachère**, **P Haïssinsky**, **P Mathieu**, *Harmonic measures versus quasiconformal measures for hyperbolic groups*, Ann. Sci. École Norm. Sup. 44 (2011) 683–721 MR Zbl

[Bonk and Schramm 2000] **M Bonk**, **O Schramm**, *Embeddings of Gromov hyperbolic spaces*, Geom. Funct. Anal. 10 (2000) 266–306 MR Zbl

[Boulanger et al. 2023] **A Boulanger**, **P Mathieu**, **C Sert**, **A Sisto**, *Large deviations for random walks on Gromov-hyperbolic spaces*, Ann. Sci. École Norm. Sup. 56 (2023) 885–944 MR Zbl

[Bowditch 1999] **B H Bowditch**, *Convergence groups and configuration spaces*, from "Geometric group theory down under", de Gruyter, Berlin (1999) 23–54 MR Zbl

[Bray et al. 2022] **H Bray**, **R Canary**, **L-Y Kao**, **G Martone**, *Counting, equidistribution and entropy gaps at infinity with applications to cusped Hitchin representations*, J. Reine Angew. Math. 791 (2022) 1–51 MR Zbl

[Burger 1993] **M Burger**, *Intersection, the Manhattan curve, and Patterson–Sullivan theory in rank* 2, Int. Math. Res. Not. 1993 (1993) 217–225 MR Zbl

[Calegari 2009] **D Calegari**, *scl*, MSJ Memoirs 20, Math. Soc. Japan, Tokyo (2009) MR Zbl

[Calegari 2013] **D Calegari**, *The ergodic theory of hyperbolic groups*, from "Geometry and topology down under", Contemp. Math. 597, Amer. Math. Soc., Providence, RI (2013) 15–52 MR Zbl

[Calegari and Fujiwara 2010] **D Calegari**, **K Fujiwara**, *Combable functions, quasimorphisms, and the central limit theorem*, Ergodic Theory Dynam. Systems 30 (2010) 1343–1369 MR Zbl

[Cannon 1984] **J W Cannon**, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata 16 (1984) 123–148 MR Zbl

[Cantrell 2021] **S Cantrell**, *Statistical limit laws for hyperbolic groups*, Trans. Amer. Math. Soc. 374 (2021) 2687–2732 MR Zbl

[Coornaert 1993] **M Coornaert**, *Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific J. Math. 159 (1993) 241–270 MR Zbl

[Coornaert and Knieper 2002] **M Coornaert**, **G Knieper**, *Growth of conjugacy classes in Gromov hyperbolic groups*, Geom. Funct. Anal. 12 (2002) 464–478 MR Zbl

[Dembo and Zeitouni 1998] **A Dembo**, **O Zeitouni**, *Large deviations techniques and applications*, 2nd edition, Appl. Math. (NY) 38, Springer (1998) MR Zbl

[Friedland 1986] **S Friedland**, *Limit eigenvalues of nonnegative matrices*, Linear Algebra Appl. 74 (1986) 173–178 MR Zbl

[Furman 2002] **A Furman**, *Coarse-geometric perspective on negatively curved manifolds and groups*, from "Rigidity in dynamics and geometry", Springer (2002) 149–166 MR Zbl

[GAP Group 2020] **GAP Group**, *GAP*: *groups*, *algorithms*, *and programming* (2020) Version 4.11.0 Available at `http://www.gap-system.org`

[Gekhtman et al. 2018] **I Gekhtman**, **S J Taylor**, **G Tiozzo**, *Counting loxodromics for hyperbolic actions*, J. Topol. 11 (2018) 379–419 MR Zbl

[Gekhtman et al. 2022] **I Gekhtman**, **S J Taylor**, **G Tiozzo**, *Central limit theorems for counting measures in coarse negative curvature*, Compos. Math. 158 (2022) 1980–2013 MR Zbl

[Ghys and de la Harpe 1990] **É Ghys**, **P de la Harpe** (editors), *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progr. Math. 83, Birkhäuser, Boston, MA (1990) MR Zbl

[Gouëzel 2014] **S Gouëzel**, *Local limit theorem for symmetric random walks in Gromov-hyperbolic groups*, J. Amer. Math. Soc. 27 (2014) 893–928 MR Zbl

[Gouëzel et al. 2018] **S Gouëzel**, **F Mathéus**, **F Maucourant**, *Entropy and drift in word hyperbolic groups*, Invent. Math. 211 (2018) 1201–1255 MR Zbl

[Gromov 1987] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR Zbl

[Heinonen 2001] **J Heinonen**, *Lectures on analysis on metric spaces*, Springer (2001) MR Zbl

[Kaimanovich 1990] **V A Kaimanovich**, *Invariant measures of the geodesic flow and measures at infinity on negatively curved manifolds*, Ann. Inst. H. Poincaré Phys. Théor. 53 (1990) 361–393 MR Zbl

[Kaimanovich et al. 2007] **V Kaimanovich**, **I Kapovich**, **P Schupp**, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel J. Math. 157 (2007) 1–46 MR Zbl

[Kao 2020] **L-Y Kao**, *Manhattan curves for hyperbolic surfaces with cusps*, Ergodic Theory Dynam. Systems 40 (2020) 1843–1874 MR Zbl

[Knieper 1983] **G Knieper**, *Das Wachstum der Äquivalenzklassen geschlossener Geodätischer in kompakten Mannigfaltigkeiten*, Arch. Math. (Basel) 40 (1983) 559–568 MR Zbl

[Maher and Tiozzo 2018] **J Maher**, **G Tiozzo**, *Random walks on weakly hyperbolic groups*, J. Reine Angew. Math. 742 (2018) 187–239 MR Zbl

[Mineyev 2005] **I Mineyev**, *Flows and joins of metric spaces*, Geom. Topol. 9 (2005) 403–482 MR Zbl

[Nica and Špakula 2016] **B Nica**, **J Špakula**, *Strong hyperbolicity*, Groups Geom. Dyn. 10 (2016) 951–964 MR Zbl

[Parry and Pollicott 1990] **W Parry**, **M Pollicott**, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque 187-188, Soc. Math. France, Paris (1990) MR Zbl

[Series 1981] **C Series**, *The infinite word problem and limit sets in Fuchsian groups*, Ergodic Theory Dynam. Systems 1 (1981) 337–360 MR Zbl

[Sharp 1998] **R Sharp**, *The Manhattan curve and the correlation of length spectra on hyperbolic surfaces*, Math. Z. 228 (1998) 745–750 MR Zbl

[Sharp 2010] **R Sharp**, *Distortion and entropy for automorphisms of free groups*, Discrete Contin. Dyn. Syst. 26 (2010) 347–363 MR Zbl

[Sigmund 1970] **K Sigmund**, *Generic properties of invariant measures for Axiom A diffeomorphisms*, Invent. Math. 11 (1970) 99–109 MR Zbl

[Tanaka 2017] **R Tanaka**, *Hausdorff spectrum of harmonic measure*, Ergodic Theory Dynam. Systems 37 (2017) 277–307 MR Zbl

[Tanaka 2021] **R Tanaka**, *Topological flows for hyperbolic groups*, Ergodic Theory Dynam. Systems 41 (2021) 3474–3520 MR Zbl

Department of Mathematics, University of Warwick
Coventry, United Kingdom

Department of Mathematics, Kyoto University
Kyoto, Japan

stephen.cantrell@warwick.ac.uk, rtanaka@math.kyoto-u.ac.jp

Proposed: Anna Wienhard — Received: 9 November 2021
Seconded: Ian Agol, Dmitri Burago — Revised: 23 September 2023

Geometry & Topology 29:4 (2025) 1909–1973
DOI: 10.2140/gt.2025.29.1909
Published: 27 June 2025

Realizability in tropical geometry and unobstructedness of Lagrangian submanifolds

JEFFREY HICKS

We say that a tropical subvariety $V \subset \mathbb{R}^n$ is B-realizable if it can be lifted to an analytic subset of $(\Lambda^*)^n$. When V is a smooth curve or hypersurface, there always exists a Lagrangian submanifold lift $L_V \subset (\mathbb{C}^*)^n$. We prove that whenever L_V has well-defined Floer cohomology, we can find for each point of V a Lagrangian torus brane whose Lagrangian intersection Floer cohomology with L_V is nonvanishing. Assuming an appropriate homological mirror symmetry result holds for toric varieties, it follows that whenever L_V is a Lagrangian submanifold that can be made unobstructed by a bounding cochain, the tropical subvariety V is B-realizable.

As an application, we show that the Lagrangian lift of a genus-0 tropical curve is unobstructed, thereby giving a purely symplectic argument for Nishinou and Siebert's proof that genus-0 tropical curves are B-realizable. We also prove that tropical curves inside tropical abelian surfaces are B-realizable.

14T20, 53D37

1 Introduction

Mirror symmetry is a collection of equivalences between symplectic geometry (A-model) and algebraic geometry (B-model) on a pair of mirror spaces. A general proposal for constructing mirror pairs of a symplectic space X_A and algebraic space X_B comes from Strominger, Yau, and Zaslow [56], who conjectured that mirror pairs can be presented as dual torus fibrations over an integral affine manifold Q. One relation between these spaces arises in the form of Kontsevich's homological mirror symmetry (HMS) conjecture [36], which predicts an equivalence between the Fukaya category of X_A and the category of coherent sheaves on a mirror manifold X_B. Roughly, the objects of the Fukaya category of X_A are Lagrangian submanifolds $L \subset X_A$. A blueprint for mirror symmetry is that Lagrangian submanifolds of X_A relate to sheaves supported on a subvariety of X_B via mutual comparison to tropical subvarieties on the base Q.

We consider the relatively well-understood example of $X_A = T^*\mathbb{R}^n / T^*_{\mathbb{Z}}\mathbb{R}^n$, $X_B = (\Lambda^*)^n$, and $Q = \mathbb{R}^n$. On the A side, it will be convenient for us to identify X_A with $(\mathbb{C}^*)^n$, which has holomorphic coordinates

$x_i = e^{q_i + i\theta_i}$ and standard symplectic form $\sum_{i=1}^n dq_i \wedge d\theta_i$. Note that X_A does not naturally come with a complex structure. On the B-side, we take Λ to be the Novikov field

$$\Lambda := \left\{ \sum_{i=0}^{\infty} a_i T^{\lambda_i} \;\middle|\; a_i \in \mathbb{C}, \lambda \in \mathbb{R}, \lim_{i\to\infty} \lambda_i = \infty \right\},$$

whose valuation map $\mathrm{val}\colon \Lambda \to \mathbb{R} \cup \{\infty\}$ is the smallest exponent appearing the expansion of $\sum_{i=0}^{\infty} a_i T^{\lambda_i}$. On the A-side, the torus fibration is given by

$$\pi_A\colon X_A \to Q, \quad (x_1, \dots, x_n) \mapsto (\log|x_1|, \dots, \log|x_n|) = (q_1, \dots, q_n),$$

whose fibers are Lagrangian tori. The dual fibration $\mathrm{TropB}\colon X_B \to Q$ is given by taking coordinatewise valuation

$$\mathrm{TropB}\colon X_B \to Q, \quad (z_1, \dots, z_n) \mapsto (\mathrm{val}(z_1), \dots, \mathrm{val}(z_n)).$$

Instead of using tropical geometry as an intuition for HMS, we use HMS and our understanding of the tropical-to-A correspondence to study the tropical-to-B correspondence. We now review these correspondences before stating our results.

Tropical-to-*B* correspondence The tropical-to-complex correspondence and its applications to enumerative geometry have been a particularly rich field of study since the pioneering work of Mikhalkin [41], which related counts of tropical curves in $\mathbb{R}^2$ to counts of curves in the *complex* algebraic torus (as opposed to the Λ analytic torus we study). This relation consists of two parts: *tropicalization*, which associates to a holomorphic curve in $(\mathbb{C}^*)^2$ a tropical curve in $\mathbb{R}^2$, and *realization*, which lifts every tropical curve $V \subset \mathbb{R}^2$ to a holomorphic curve in $(\mathbb{C}^*)^2$. Both of these constructions have been extended to greater generality; we provide a coarse overview of the constructions here:

- ***B*-Tropicalization** The tropicalization map associates to a closed analytic subset $Y \subset X_B$ its tropicalization $\mathrm{TropB}(Y) \subset Q$. The expectation (which holds for algebraic subvarieties by work of Bieri and Groves [11]) is that the tropicalization is a *tropical subvariety* (Definition 2.2.1).
- ***B*-Realization** Starting with $V \subset Q$ a tropical subvariety, we say that V is *B-realizable* if there exists a closed analytic subset $Y \subset X_B$ with $\mathrm{TropB}(Y) = V$.

One goal of tropical geometry is to determine which tropical subvarieties $V \subset Q$ are B-realizable. Examples such as that of Mikhalkin [40, Example 5.12] show that there exist tropical curves $V \subset \mathbb{R}^n$ for $n > 2$ that are nonrealizable. In some cases, there are criteria determining if a tropical subvariety is B-realizable.

For example, if $V \subset Q$ is a tropical hypersurface, then there exists a tropical polynomial (piecewise integral affine convex function) $f\colon Q \to \mathbb{R}$ such that V is the locus of points where f is nonaffine. The function f is called a tropical polynomial as it can be written using the tropical sum and product operations:

$$\oplus\colon (\mathbb{R} \cup \{\infty\}) \times (\mathbb{R} \cup \{\infty\}) \to (\mathbb{R} \cup \{\infty\}), \quad q_1 \oplus q_2 = \min(q_1, q_2),$$

$$\odot\colon (\mathbb{R} \cup \{\infty\}) \times (\mathbb{R} \cup \{\infty\}) \to (\mathbb{R} \cup \{\infty\}), \quad q_1 \odot q_2 = q_1 + q_2.$$

Let $f = \bigoplus_{\alpha\in\mathbb{N}^n} a_\alpha \odot q^{\odot\alpha}$ be a tropical polynomial whose nonaffine locus is V. Let Λ be a complete non-Archimedean valued field, and let $X_B = (\Lambda^*)^n$ be the algebraic torus. For each a_α, select a coefficient $c_\alpha \in \Lambda$ whose valuation is $\operatorname{val}(c_\alpha) = a_\alpha$. Then the zero set of the polynomial $\sum_{\alpha\in\mathbb{N}^n} c_\alpha z^\alpha$ defines a subvariety of X_B which is the B-realization of V. Note that this construction does not produce a unique lift.

The other examples where we have B-realization criteria are tropical curves. In [45], Nishinou and Siebert showed that if $V \subset Q$ is a trivalent tropical curve of genus 0, then V is realizable. This was extended to all balanced maps from trees by Ranganathan [50]. In higher genus, the space of deformations of a tropical curve may have higher dimension than the expected dimension of a possible B-realization. In this case, we say that the tropical curve is superabundant; see Mikhalkin [41]. We expect that a generically chosen superabundant curve is not B-realizable. It is known that all 3-valent nonsuperabundant curves are realizable; see Cheung, Fantini, Park, and Ulirsch [14]. In the superabundant setting, Speyer [55, Theorem 3.4] established that if $V \subset Q$ is a tropical curve of genus 1 and satisfies a condition called *well-spacedness*, then V is realizable.

Tropical-to-A correspondence The tropical-to-Lagrangian correspondence is a more recent construction, independently arrived at by Hicks [30], Mak and Ruddat [38], Matessi [39], and Mikhalkin [42]. Each of the papers associates to a (certain type of) tropical subvariety $V \subset Q$ a Lagrangian submanifold $L_V^\varepsilon \subset X_A$ whose projection to the base of the Lagrangian torus fibration $\pi_A(L_V^\varepsilon)$ is contained within an ϵ-neighborhood of the tropical subvariety V. We call this a geometric Lagrangian lift of V. When V is a hypersurface, [30] proves that under homological mirror symmetry $L_V^\varepsilon \subset (\mathbb{C}^*)^n$ is identified with a sheaf whose support is a hypersurface $Y \subset (\Lambda^*)^n$.

In contrast to B-realization, the constructions in [30; 38; 39; 42] can construct a geometric Lagrangian lift L_V of *any* smooth tropical curve $V \subset Q$. This difference occurs because the map $\pi_A\colon X_A \to Q$ does not provide a good tropicalization map. For example, for any subset $U \subset Q$ and $\varepsilon > 0$ there exists a Lagrangian submanifold $L \subset X_A$ with the property that the Hausdorff distance between $\pi_A(L)$ and U is less than ε. Additionally, it would be desirable to have a tropicalization map that only depends on the Hamiltonian isotopy class of the Lagrangian submanifold — and $\pi_A(L)$ can change substantially when we apply a Hamiltonian isotopy to L.

To obtain a correspondence from Lagrangian submanifolds to tropical subsets of Q, and justify why the Lagrangian L_V is the "correct" A-model realization of a tropical curve V, one needs to employ techniques from Floer theory. Not all Lagrangian submanifolds are amenable to such analysis. We call a Lagrangian submanifold unobstructed if its filtered A_∞ algebra $\mathrm{CF}^\bullet(L)$ admits a bounding cochain. The pair (L, b) of Lagrangian submanifold equipped with a bounding cochain is called a Lagrangian brane. Examples of unobstructed Lagrangian submanifolds include those which bound no pseudoholomorphic disks for a given choice of almost complex structure. In particular, if L is exact, it is unobstructed.

If (L_1, b_1) and (L_2, b_2) are Lagrangian branes then there exists a cochain complex $\mathrm{CF}^\bullet((L_1, b_1), (L_2, b_2))$ generated by the intersections of L_1 and L_2, and whose cohomology groups $\mathrm{HF}^\bullet((L_1, b_1), (L_2, b_2))$ are invariant under Hamiltonian isotopies of either L_1 or L_2. We can use this to define A-tropicalization and A-realization.

- ***A*-Tropicalization** Starting with the fibration $X_A \to Q$ and a Lagrangian brane $(L, b) \subset X_A$, we define the A-tropicalization

$$\mathrm{TropA}(L, b) := \{q \in Q \mid \exists (F_q, \nabla) \text{ with } \mathrm{HF}^\bullet((L, b), (F_q, \nabla)) \neq 0\},$$

where $F_q = \pi_A^{-1}(q)$ is equipped with a unitary local system ∇, and $\mathrm{HF}^\bullet((L, b), (F_q, \nabla))$ is the Lagrangian intersection Floer cohomology of (L, b) with F_q deformed by the local system ∇. An advantage of $\mathrm{TropA}(L, b)$ over $\pi_A(L)$ is that the former depends only on the Hamiltonian isotopy class of L.

- ***A*-Realizability** In light of the definition of A-tropicalization, we say that $V \subset Q$ is A-realizable if there exists a Lagrangian brane $(L, b) \subset X_A$ with $\mathrm{TropA}(L) = V$.

The Lagrangian submanifold L_V^ε associated to V provides a geometric candidate for an A-realization of V. However, to verify A-realizability, one still needs to check that L_V^ε is unobstructed with bounding cochain b and that $\mathrm{TropA}(L_V, b) = V$. We call this last condition *faithfulness*.

1.1 Results

The three components (geometric realizability, unobstructedness, and faithfulness) of the A-realizability problem and its implications for the B-realizability problem in $Q = \mathbb{R}^n$ are summarized in Figure 1.

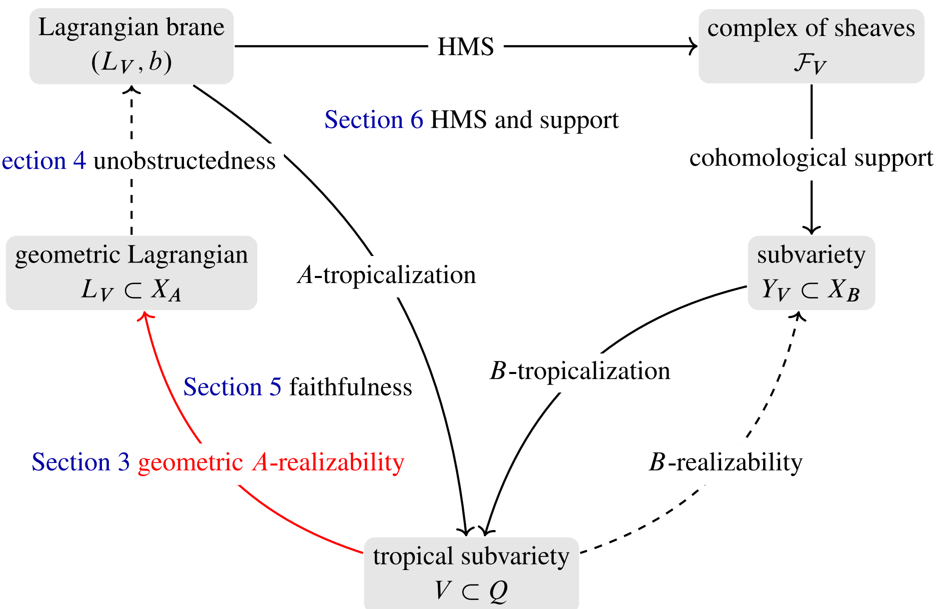


Figure 1

The correspondences given by solid black lines always exist. Geometric A-realizability (the solid red arrow) is only known to exist for certain examples of tropical subvarieties of Q. For the applications we consider (smooth tropical hypersurfaces and curves) we always have geometric A-realizability. We conjecture that every tropical subvariety of Q is geometrically A-realizable. For any given tropical subvariety $V \subset Q$, there is no reason for either of the dashed arrows to hold. However, the following conjecture seems natural:

Conjecture 1.1.1 *Let $V \subset \mathbb{R}^n$ be a tropical curve. Then a geometric Lagrangian lift L_V is unobstructed if and only if V is B-realizable.*

The main step in proving the forward direction of the conjecture is to establish the faithfulness of the Lagrangian brane lift, that is showing that $\mathrm{TropA}((L_V, b)) = V$. Our primary result is to prove faithfulness (for *all* tropical subvarieties admitting unobstructed Lagrangian lifts).

Theorem A (restatement of Lemma 5.2.2) *Let $V \subset Q$ be a tropical subvariety. Let (L^ε_V, b) be a Lagrangian brane lift of V. Then $\mathrm{TropA}(L^\varepsilon_V, b) = V$.*

When we can apply homological mirror symmetry, we obtain the forward direction of Conjecture 1.1.1. Depending on the affine manifold Q and Lagrangian L_V, we may require Assumption 6.1.2, which states that the family Floer construction of Abouzaid [6] extends to the noncompact and unobstructed setting.

Theorem B (restatement of Corollary 6.2.1) *Suppose Assumption 6.1.2. Let $V \subset \mathbb{R}^n$ be a tropical subvariety. Suppose there exists $(L_V, b) \subset (\mathbb{C}^*)^n$ a Lagrangian brane lift of V. Then V is B-realizable.*

Our second goal is to show that this can be used to produce realizability criteria. We first recover a theorem of Nishinou and Siebert [45]:

Corollary C (restatement of Corollary 4.3.3) *Suppose Assumption 6.1.2. Every smooth genus-0 tropical curve $V \subset \mathbb{R}^n$ has a Lagrangian brane lift (L_V, b) and is, therefore, B-realizable.*

The results of Nishinou [44] give necessary and sufficient conditions for when a tropical curve can be realized by a family of algebraic curves in a degenerating family of complex tori. In contrast to those results, our results show that every 3-valent tropical curve can be realized by a closed analytic subset. The following result *does not* assume Assumption 6.1.2:

Corollary D (restatement of Corollary 6.2.4) *Let $Q = T^2$ be a tropical abelian surface. Let $V \subset Q$ be a 3-valent tropical curve. V has a Lagrangian brane lift $(L_V, 0)$, and is, therefore, B-realizable.*

In summary, we can recover B-realizability results using unobstructedness for the first five cases in Table 1. We also provide some insight into the existence of holomorphic curves with boundary on tropical Lagrangian submanifolds.

V and Q	A-model (unobstructedness)	B-model (realizability)	HMS status
curves in abelian surfaces	Corollary 6.2.4	[44][1]	✓
curves in $\mathbb{R}^2$	[31]	[41]	$(*)$
hypersurfaces of $\mathbb{R}^n$	[31]	folklore	$(*)+(**)$
hypersurfaces in abelian varieties	Corollary 6.2.3	—	$(**)$
genus-0 curves in $\mathbb{R}^n$	Theorem C	[45]	$(*)+(**)$
compact genus-0 curves in $\dim(Q)=3$	[38]	[45]	—
well-spaced genus-1 curves	spec. in Section 6.4	[55]	$(*)+(**)$

Table 1: Relating A-unobstructedness to B-realizability. Here $(*)$ and $(**)$ refer to the needed extensions of family Floer cohomology (Assumption 6.1.2) to the noncompact and nontautologically unobstructed settings.

Example E (restatement of Example 6.3.2) Let $V_c \subset \mathbb{R}^3$ be a generic tropical line. The Lagrangian L_{V_c} is unobstructed, but not tautologically unobstructed.

Outline In Section 2, we give a toy computation that explores the entire roadmap above for a simple example, $V_{\text{pants}} \subset \mathbb{R}^2$, the tropical pair of pants. In addition to providing context for the remainder of the paper, the computation reviews some background for tropical geometry and symplectic geometry. We also use this section to fix notation. It is our hope that this section will be accessible to both tropical and symplectic geometers.

Section 3 discusses the geometric lifting problem. Definition 3.0.2 specifies when a family of Lagrangian submanifolds L_V^ε is a geometric Lagrangian lift of a tropical subvariety V. We show that Definition 3.0.2 distinguishes tropical subvarieties among all polyhedral complexes as the ones which permit geometric Lagrangian lifts. Definition 3.0.2 requires that geometric Lagrangian lifts are monomially admissible, graded, and spin. In Sections 3.1–3.3, we show that known constructions of geometric Lagrangian lifts of tropical subvarieties of Hicks [30], Matessi [39], and Mikhalkin [42] satisfy these conditions. We also prove Lemma 3.3.1, which shows that for smooth genus-0 tropical curves, the map $H^2(L_V) \to H^2(\partial L_V)$ is an injection.

Section 4 investigates Lagrangian submanifolds which can be unobstructed by a bounding cochain (Lagrangian branes). We provide a brief overview of the pearly model for Lagrangian submanifolds in Section 4.1. This is followed by examples of unobstructed geometric Lagrangian lifts (Lagrangian brane lifts) of tropical subvarieties in Section 4.2 (summarized in Table 1). Section 4.3 gives a new method for checking unobstructedness of Lagrangian submanifolds inside noncompact symplectic spaces which have a potential function $W: X_A \to \mathbb{C}$ (see Definition A.0.1).

[1]The realization result of Nishinou and Siebert [45] considers B-tropicalizations coming from degenerating families of abelian surfaces so that a tropical curve is realized by a parametrized algebraic curve. The B-realization we take is by closed analytic subsets. In the setting of genus-0 stable tropical curves in toric varieties, these tropicalizations can be related by Ranganathan [50].

Theorem F (restatement of Theorem 4.3.1) *Let $W\colon X_A \to \mathbb{C}$ be a symplectic fibration outside of a compact set of $\mathbb{C}$. Let $L \subset X_A$ be a W-admissible Lagrangian submanifold with boundary $M \subset W^{-1}(t)$ for $t \in \mathbb{R}_{\gg 0}$. Suppose M is a tautologically unobstructed Lagrangian submanifold of $W^{-1}(t)$, and the connecting map $H^1(M) \to H^2(L, M)$ is surjective. Then there exists a bounding cochain b such that (L, b) is a Lagrangian brane.*

The proof uses a lemma on filtered A_∞ algebras (Lemma B.2.8). Since we have previously proven in Lemma 3.3.1 that the geometric Lagrangian lifts L_V of smooth genus-0 tropical curves satisfy the criterion of Theorem 4.3.1, we obtain that such L_V are unobstructed (Corollary 4.3.3).

In Section 5, we prove faithfulness (Lemma 5.2.2), which shows that the A-tropicalization (Floer-theoretic support) of a Lagrangian brane lift L_V is V. The proof uses that the Lagrangian intersection Floer cohomology between (L_V, b) and F_q is a deformation of the cohomology of a subtorus of F_q. An application of Lemma B.3.1 shows that this can be "undeformed" by a bounding cochain, so that $\mathrm{HF}^0((L, \nabla_0, b_0), (F_q, \nabla, b)) = \Lambda$.

Section 6 applies the previous constructions to address questions of realizability for tropical subvarieties. Abouzaid [5, Remark 1.1] states that we expect that the family Floer functor can be adapted to include unobstructed Lagrangians. We instead use Assumption 6.1.2 — the weaker assumption that the family Floer construction of Abouzaid [6] can be employed for unobstructed Lagrangian submanifolds in the Lagrangian torus fibration $(\mathbb{C}^*)^n \to \mathbb{R}^n$ to construct a sheaf on the mirror space. We give a brief outline of the modifications to [6] which would be required to prove Assumption 6.1.2. With this assumption, we prove the forward direction of Conjecture 1.1.1 in Corollary 6.2.1. We also discuss the first five cases in Table 1.

Finally, we discuss evidence towards the reverse direction of Conjecture 1.1.1. This requires us to understand some of the holomorphic disks which appear on Lagrangian lifts of tropical subvarieties. In Example 6.3.2, we show that the lift of the tropical line in $\mathbb{R}^3$ bounds a holomorphic disk whose symplectic area is dictated by the internal edge length on the tropical line. We also discuss applications of B nonrealizability to obstructedness in Section 6.4. We consider the superabundant tropical elliptic curve $V \subset \mathbb{R}^3$ of Mikhalkin [40, Example 5.12] and provide a sketch for how Speyer's well-spacedness criterion might be recovered from holomorphic disk counts on L_V. Section 6.5 looks at how to relate tropical line bundles on tropical curves to Lagrangian isotopies of their geometric Lagrangian lifts. We conjecture a relation between superabundance of a tropical curve V and the relative ranks of $\mathrm{HF}(L_V, b)$ and $H(L_V)$ (wide versus nonwide).

We provide some auxiliary results in the appendices. Appendix A discusses how to adapt the pearly model of Floer cohomology of Charest and Woodward [12] to the setting of noncompact spaces equipped with a potential $W\colon X \to \mathbb{C}$. In Appendix B, we prove the results on filtered A_∞ algebras used in this paper.

Acknowledgments I'd particularly like to thank Ailsa Keating, Denis Auroux, Mark Gross, Grigory Mikhalkin, Nick Sheridan, and Ivan Smith for the many mathematical conversations which directly and indirectly contributed to this project. Additionally, I thank Andrew Hanlon and Umut Varolgunes for providing me with some thoughtful feedback and suggestions for this article. I'm especially indebted to Dhruv Ranganathan, who also answered my many questions about tropical geometry, and Jean-Philippe Chassé and Yoon Jae Nick Nho, who corrected an error in the reverse isoperimetric inequality for Lagrangian intersection Floer cohomology. This article benefited from the thoughtful comments and feedback of the referee. Finally, the computation of holomorphic strips in Section 2.5 stems from a discussion with Diego Matessi. A portion of this work is adapted from Section 4.3 of my PhD thesis at UC Berkeley. This paper was supported by EPSRC grant N03189X/1 at the University of Cambridge and ERC grant 850713 (Homological mirror symmetry, Hodge theory, and symplectic topology) at the University of Edinburgh.

2 A guided calculation to the support of the Lagrangian pair of pants

This section contains an expository computation that is designed to frame the main ideas of the paper, provide background, and fix notation. The exposition here is not intended to be comprehensive, although we hope that through explicit examples, direct computations, and additional references, we've made this section accessible to both the tropical and symplectic geometry communities. As a result, the materials outside of Sections 2.5 and 2.5 are expository. As we will frequently use notation from Examples 2.4.3 and 2.4.4, we suggest that the readers take a look at these computations of Lagrangian intersection Floer cohomology for conormal bundles in the cotangent bundle of the torus.

2.1 A-model, B-model, and Lagrangian torus fibrations

We provide a high-level overview of the viewpoint of [26; 56] on mirror symmetry. Let Q be an integral affine manifold, that is, a manifold equipped with a choice of integrable full-rank lattice $T_{\mathbb{Z}}Q \subset TQ$. This identifies a dual lattice $T^*_{\mathbb{Z}}Q \subset T^*Q$, and also a flat connection on TQ. There are three kinds of geometries that we may associate with Q: symplectic geometry, complex geometry, and tropical geometry.

A-model A symplectic manifold is a $2n$-manifold X_A with a choice of 2-form $\omega \in \Omega^2(X_A)$ which is closed ($d\omega = 0$) and nondegenerate ($\omega^n \neq 0$). The submanifolds of interest for us in X_A are Lagrangian submanifolds $L \subset X_A$, which are n-dimensional submanifolds on which the symplectic form vanishes ($\omega|_L = 0$). For any manifold Q, the cotangent bundle T^*Q (whose local coordinates are (q, p)) carries a canonical symplectic form $\sum_{i=1}^n dq_i \wedge dp_i$. This descends to a symplectic form on the quotient $X_A := T^*Q/T^*_{\mathbb{Z}}Q$.

Given an integral affine submanifold $\underline{V} \subset Q$ such that $T_{\mathbb{Z}}\underline{V} \subset T_{\mathbb{Z}}Q$, the periodized conormal bundle $L_{\underline{V}} := N^*\underline{V}/N^*_{\mathbb{Z}}\underline{V} \subset X_A$ is an example of a Lagrangian submanifold. The simplest example is when we pick a point $q \in Q$ such that

$$(1) \qquad L_{\underline{q}} = N^*q/N^*_{\mathbb{Z}}q = T^*_qQ/T^*_{q,\mathbb{Z}}Q$$

is a Lagrangian torus of X_A. We will call this Lagrangian torus $F_q \subset X_A$. For this reason, we call the projection $\pi_A\colon X_A \to Q$ a Lagrangian torus fibration.

***B*-model** We can also build an almost-complex manifold from the data of Q. An almost complex structure on $X_B^{\mathbb{C}}$ is an endomorphism $J\colon TX_B^{\mathbb{C}} \to TX_B^{\mathbb{C}}$ which squares to $-\mathrm{id}$. The submanifolds of interest in the B-model are the *almost-complex submanifolds* $Y^{\mathbb{C}} \subset X_B^{\mathbb{C}}$ whose tangent spaces are fixed under the almost complex structure, so that $J(T_y Y^{\mathbb{C}}) = T_y Y^{\mathbb{C}}$.

As Q is integral affine, there exists a connection on TQ whose flat sections are locally constant sections of $T_{\mathbb{Z}} Q$. This provides a splitting $T(TQ) = T_q Q \oplus \ker(\pi)$. We define an almost complex structure on TQ which interchanges the components of this splitting with a sign:

$$J := \begin{pmatrix} 0 & -\mathrm{id} \\ \mathrm{id} & 0 \end{pmatrix}.$$

The almost complex structure on TQ descends to an almost complex structure on $X_B^{\mathbb{C}} := T^*Q/T^*_{\mathbb{Z}}Q$; the fibers of $\pi_B\colon X_B^{\mathbb{C}} \to Q$ are real tori.

Given an integral affine submanifold $\underline{V} \subset Q$, the periodized tangent bundle

$$Y_{\underline{V}}^{\mathbb{C}} := T\underline{V}/T_{\mathbb{Z}}\underline{V} \subset X_B^{\mathbb{C}}$$

is an example of an almost-complex submanifold. If we start with $q \subset Q$ a point, we see that $Y_q \subset X_B^{\mathbb{C}}$ is a point of $X_B^{\mathbb{C}}$.

Mirror symmetry from Lagrangian torus fibrations We now describe in more detail the relationship between the Lagrangian tori of X_A and the points of $X_B^{\mathbb{C}}$. First, we note that for fixed $q \in Q$, there are a torus worth of points z in $X_B^{\mathbb{C}}$ with the property that $\pi_B(z) = q$.

In contrast to the complex lift, there is only one Lagrangian torus $F_q \subset X_A$ with $\pi_A(F_q) = \{q\}$. To get a matching family of Lagrangian lifts to our complex lift, we consider Lagrangian tori equipped with the additional data of a local system. Let (F_q, ∇) be a pair consisting of a Lagrangian torus F_q and a choice of $U(1)$ local system on F_q. Then there is a bijection between pairs $(F_q, \nabla) \subset X_A$ and points $z \in X_B^{\mathbb{C}}$. A similar story holds for the Lagrangian and complex lifts of integral affine subspace $\underline{V} \subset Q$.

To generalize beyond the submanifolds $\underline{V}$, $L_{\underline{V}}$, and $Y_{\underline{V}}$ discussed above, we need to look at tropical geometry.

Notation 2.1.1 Unless otherwise stated, we only consider $Q = \mathbb{R}^n$, so that $X_A = (\mathbb{C}^*)^n$ and $X_B^{\mathbb{C}} = (\mathbb{C}^*)^n$.

2.2 A quick introduction to tropical geometry and *B*-tropicalization

A convex polyhedral domain is the intersection of finitely many closed half-spaces in $\mathbb{R}^n$,

$$\underline{V} = \{q \in Q \mid \langle q, \vec{v}_i \rangle \geq \lambda_i\},$$

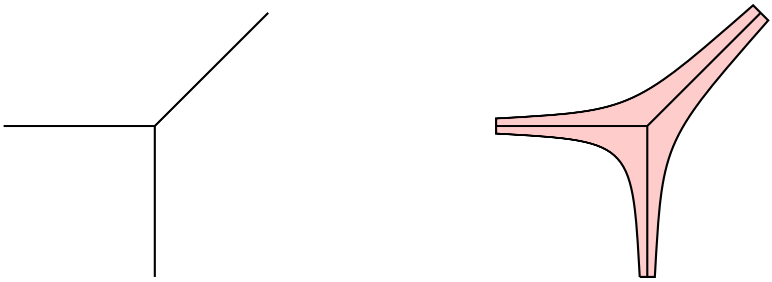


Figure 2: The tropical pair of pants (left) approximates the Amoeba of a curve (sketched on the right).

where $\vec{v}_i$ is a collection of vectors in $\mathbb{R}^n$, and λ_i is some set of constants in $\mathbb{R}$. We say that this is a rational convex domain if $\vec{v}_i \in \mathbb{Z}^n$ for all i, equivalently if there is a full lattice $T_{\mathbb{Z}}\underline{V} \subset T\underline{V}$ which is a sublattice of $T_{\mathbb{Z}}Q$. A tropical subvariety is built out of these pieces.

Definition 2.2.1 A k-dimensional tropical subvariety $V \subset Q$ is a collection of k-dimensional rational convex polyhedral domains $\{\underline{V}_s \subset Q\}$ and weights $\{w_s \in \mathbb{N}\}$ which are required to satisfy the following conditions:

- **Polyhedral complex condition** At each pair of rational convex polyhedral domains, the intersection $\underline{V}_s \cap \underline{V}_t$ is either empty, or a boundary facet of both $\underline{V}_s$ and $\underline{V}_t$.
- **Balancing condition** At facets $\underline{W} \subset \underline{V}_s$, let $\underline{V}_1, \dots, \underline{V}_k$ be the rational polyhedral domains containing $\underline{W}$. Consider lattices $T_{\mathbb{Z}}\underline{W}$, each of which is a sublattice of $T_{\mathbb{Z}}\underline{V}_i$ for each $i \in \{1, \dots, k\}$. Select for each i a vector $\vec{v}_i \in T_{\mathbb{Z}}\underline{V}_i$ such that $T_{\mathbb{Z}}\underline{V}_i = T_{\mathbb{Z}}\underline{W} \oplus \langle \vec{v}_i \rangle$ as oriented lattices. We require that

$$\sum_i w_i \vec{v}_i \equiv 0 \in T_{\mathbb{Z}}Q / T_{\mathbb{Z}}\underline{W}.$$

Example 2.2.2 Consider the polyhedral domains in $Q = \mathbb{R}^2$

$$\underline{V}_1 = \{(-t, 0) \mid t \in \mathbb{R}_{\geq 0}\}, \quad \underline{V}_2 = \{(0, -t) \mid t \in \mathbb{R}_{\geq 0}\}, \quad \underline{V}_3 = \{(t, t) \mid t \in \mathbb{R}_{\geq 0}\}.$$

As the directions $\langle -1, 0\rangle$, $\langle 0, -1\rangle$, and $\langle 1, 1\rangle$ sum to zero this is balanced and gives us a tropical curve. The collection of these three polyhedral domains is called the *standard tropical pair of pants*. The curve $V_{\text{pants}} \subset \mathbb{R}^2$ is drawn in Figure 2, left.

We say that a tropical curve $V \subset \mathbb{R}^n$ is *smooth* if every 0-dimensional stratum is locally modeled after the pair of pants.

Notation 2.2.3 Given $V \subset Q$ a tropical subvariety, we will use $V^{(0)}$ to denote the union of the interiors of the $\underline{V}_s$, and $V^{(1)}$ to denote the union of the interiors of the boundaries of the $\underline{V}_s$; more generally we will use $V^{(i)}$ to denote the codimension-i linearity strata of V. For any $\underline{W} \subset V^{(i)}$, let $\operatorname{star}(\underline{W})$ be the set of all strata which contain $\underline{W}$. If V is a tropical curve, we will usually call the strata vertices and edges, and use v and w for vertices and e and f for edges.

2.3 B-tropicalization

B-tropicalization is the process of taking a subvariety of $X_B^{\mathbb{C}}$ and obtaining a tropical subvariety of Q. The first approach one considers is the image of $Y^{\mathbb{C}} \subset X_B^{\mathbb{C}}$ under the B-torus fibration

$$\pi_B : X_B^{\mathbb{C}} \to Q.$$

Under good conditions, $\pi_B(Y^{\mathbb{C}}) \subset Q$ approximates a tropical subvariety of Y; see for instance [40]. The image $\pi_B(Y^{\mathbb{C}})$ is called the *amoeba* of $Y^{\mathbb{C}}$, which computationally can be checked to approach the tropical curve (see Figure 2, right).

To obtain a theory where the tropicalization of a subvariety is a tropical subvariety, we look to non-Archimedean geometry. Let Λ be the Novikov field. Given M a rank-n lattice, denote by X_B the torus $\operatorname{Spec} \Lambda[M]$. The points of X_B can be identified with n-tuples of invertible elements of Λ, so we will frequently write

$$X_B = \{(z_1, \dots, z_n) \mid z_i \in \Lambda^*\}.$$

We build a tropicalization map by taking the valuation coordinatewise:

$$\operatorname{TropB} : X_B \to M \otimes \mathbb{R} = Q, \quad (z_1, \dots, z_n) \mapsto (\operatorname{val}(z_1), \dots, \operatorname{val}(z_n)).$$

Given a $Y \subset X_B$ a closed analytic subset, we call the image $\operatorname{TropA}(Y) \subset Q$ its tropicalization.

Example 2.3.1 Consider $M = \mathbb{R}^2$, and the closed analytic subset $Y \subset (\Lambda^*)^2$ given by

$$Y = \{(z_1, z_2) \mid 1 + z_1 + z_2 = 0\}.$$

We compute the valuation of such a point $(z_1, z_2) \in Y$. Since

$$\operatorname{val}(1 + z_1 + z_2) \geq \min(\operatorname{val}(1), \operatorname{val}(z_1), \operatorname{val}(z_2)),$$

with equality holding whenever the valuations differ, we obtain that for all $(z_1, z_2) \in Y$ at least one of the following equalities hold:

$$\operatorname{val}(z_1) = \operatorname{val}(z_2), \quad \operatorname{val}(z_1) = \operatorname{val}(1), \quad \operatorname{val}(z_2) = \operatorname{val}(1).$$

This means that the image of $\operatorname{TropB}(Y)$ agrees with $V_{\text{pants}} \subset \mathbb{R}^2$ from Example 2.2.2. It follows that V is B-realizable.

This phenomenon holds much more broadly:

Theorem 2.3.2 [11; 27] *Let $Y \subset X_B$ be an irreducible k-dimensional analytic subset. Then* $\operatorname{TropB}(Y)$ *is a k-dimensional polyhedral complex.*

It is expected that when Y is an irreducible k-dimensional analytic subset, $\operatorname{TropB}(Y)$ is a k-dimensional tropical subvariety. To our knowledge, this result has not appeared in the literature. A discussion on the current status of tropicalization for analytic subsets is included in [53, Section 5.3].

2.4 Floer cohomology and A-tropicalization

The definition of the A-tropicalization of a Lagrangian submanifold requires a little more exposition because we wish to do some computations of the A-tropicalization. Our goal is to replace the Lagrangian torus fibration map $\pi_A\colon X_A \to Q$ with a correspondence of subsets

$$\mathrm{TropA}\colon \{\text{Lagrangian branes}\} \to \{\text{subsets of } Q\}$$

which only depends on the Hamiltonian isotopy class of the Lagrangian brane.

2.4.1 Lagrangian intersection Floer cohomology Our main computational tool will be Lagrangian intersection Floer cohomology. We first equip a symplectic manifold (X, ω) with an ω-compatible choice of almost complex structure J.

Definition 2.4.1 [19] Suppose we have a pair of transversely intersecting Lagrangian submanifolds $L_0, L_1 \subset X$ and choice of almost complex structure J such that

(i) X, L_1, and L_2 are compact,

(ii) the symplectic area of all disks with boundary on L_i vanish $\omega(\pi_2(X, L_i)) = 0$,

(iii) the Lagrangians L_i are equipped with spin structures,

(iv) the Lagrangians L_i are graded (in the sense of [52]),

(v) the moduli spaces of J-holomorphic strips in (2) are regular.

Then the Lagrangian intersection Floer cohomology is a chain complex where:

- The generators are the points of intersection between L_0 and L_1, so that as a vector space

$$\mathrm{CF}^\bullet(L_0, L_1) := \bigoplus_{x \in L_0 \cap L_1} \Lambda_x,$$

where Λ is the *Novikov field*. The grading $\deg(x)$ of an intersection point $x \in L_0 \cap L_1$ is determined by the Maslov index.

- The differential on this complex is defined by a count of holomorphic strips with boundary on $L_0 \cup L_1$ and ends limiting to the intersection points. Namely, let $x_\pm \in L_0 \cap L_1$ be two intersection points, and $\beta \in H^2(X, L_0 \cup L_1)$. Let $\mathcal{M}_\beta(L_0, L_1, x_+, x_-)$ denote the moduli space

$$\left\{ u\colon \mathbb{R}_s \times [0,1]_t \to X_A \;\middle|\; \begin{matrix} u(s,0) \in L_0,\ u(s,1) \in L_1,\ \lim\limits_{s \to \pm\infty} u(s,t) = x_\pm, \\ \bar{\partial}_J u = 0,\ [u] = \beta \in H_2(X_A, L_0 \cup L_1) \end{matrix} \right\} \Big/ (s \mapsto s + c) \tag{2}$$

of holomorphic strips with ends limiting to $x^\pm$ in the relative homology class β, up to reparametrization of the strip along the s-coordinate. Using the grading data on L_0 and L_1, one can compute that

$$\dim(\mathcal{M}_\beta(L_0, L_1, x_+, x_-)) = \deg(x_-) - \deg(x_+) - 1.$$

The spin structures on L_0 and L_1 provide orientations for the spaces $\mathcal{M}_\beta(L_0, L_1, x_+, x_-)$; in particular if $\deg(x_+) + 1 = \deg(x_-)$, then $\dim(\mathcal{M}_\beta(L_0, L_1, x_+, x_-)) = 0$ and we can count the points in this moduli space with signs. The structure coefficients of the differential $d\colon \mathrm{CF}^\bullet(L_0, L_1) \to \mathrm{CF}^\bullet(L_0, L_1)$ are obtained by counting the elements in $\mathcal{M}_\beta(L_0, L_1, x_+, x_-)$,

$$\langle d(x_+), x_-\rangle = \sum_{\beta \in H^2(X, L_0 \cup L_1)} T^{\omega(\beta)} \# \mathcal{M}_\beta(L_0, L_1, x_+, x_-),$$

where # is the signed count of points with orientation and $T^{\omega(\beta)}$ records the symplectic area of the strip u whose homology class is β.

The proof that this is a chain complex proceeds in a similar method to Morse theory. Because of (i), one can use Gromov compactness to prove that the 1-dimensional moduli spaces of strips have compactifications whose boundaries are given by products of the 0-dimensional moduli spaces of strips

$$\partial \mathcal{M}_\beta(L_0, L_1, x_+, x_-) = \bigsqcup_{x_0 \in L_0 \cap L_1} \mathcal{M}_\beta(L_0, L_1, x_+, x_0) \times \mathcal{M}_\beta(L_0, L_1, x_0, x_-).$$

To ensure that the only broken configurations which show up in the compactification are given by strips breaking (as opposed to disk bubbling), we use (ii), which states that there are no holomorphic disks with boundary on either L_0 or L_1. The compactification is compatible with the orientations given to the moduli spaces of holomorphic strips. Since the signed count of boundary components of a 1-dimensional manifold is zero, $\langle d^2(x_+), x_-\rangle = 0$. Unless otherwise stated, all Lagrangians we consider will be $\mathbb{Z}$-graded and spin. A major feature of Lagrangian intersection Floer cohomology is its invariance under Hamiltonian isotopy.

Theorem 2.4.2 [19] *Let L_0 and L_1 be Lagrangian submanifolds of (X, ω) satisfying (i)–(v). Let $\phi\colon X \to X$ be a Hamiltonian isotopy. Suppose that L_0 and L_1 intersect transversely and we've picked ϕ in such a way that $\phi(L_0)$ and L_1 intersect transversely. Then* $\mathrm{HF}^\bullet(L_0, L_1) = \mathrm{HF}^\bullet(\phi(L_0), L_1)$.

For this reason, whenever L_0 and L_1 do not intersect transversely, we can compute their Floer cohomology by taking a Hamiltonian perturbation which makes their intersection transverse; the resulting cohomology groups are independent of the choice of perturbation taken. One can similarly show that it does not depend on the choice of an almost complex structure.

The conditions (i) and (ii) can be weakened. For example, (i) — which is required to prove that the moduli spaces of strips admit compactifications — can be replaced with the weaker condition of monomial admissibility (Definition 3.1.1) or W-admissibility (Definition A.0.1). Later we will look at weakening (ii) to *unobstructedness* (Section 4). We now drop (i) and compute the Lagrangian intersection Floer cohomology between two Lagrangians in a cotangent bundle. The computation we give is a direct generalization of [54, Example 3.1].

Example 2.4.3 (running example) Let $F_0 = T^n$ be the n-dimensional torus. Let $T^{n-k} \subset F_0$ be the subtorus spanning the first $n-k$ coordinates on T^n. Then T^*F_0 is an example of an exact symplectic manifold. The zero section F_0 and the conormal bundle N^*T^{n-k} are examples of exact Lagrangian submanifolds. Lagrangian intersection Floer cohomology requires that our Lagrangians intersect transversely, so we will apply a Hamiltonian perturbation to one of the Lagrangians to achieve transverse intersections. Pick $\lambda_0 \in \mathbb{R}_{>0}$. Consider the Hamiltonian function

$$H = \sum_{i=1}^{n-k} \lambda_0 \cos(\theta_i) \tag{3}$$

on T^*F_0. Let $\phi\colon T^*F_0 \to T^*F_0$ be the time-1 Hamiltonian flow of H. The resulting intersections of $\phi(N^*T^{n-k})$ with F_0 are the points

$$\phi(N^*T^{n-k}) \cap F_0 = \{(a_1\pi, \dots, a_{n-k}\pi, 0, \dots, 0) \mid a_i \in \{0,1\}\},$$

and the index of each intersection point x is given by $\deg(x) = \sum_{i=1}^{n-k} a_i$. We will call the corresponding generators of Floer cohomology x_I, where $a_i = 1$ whenever $i \in I \subset \{1, \dots, n-k\}$. Write $I \lessdot J$ if $I = J \cup \{x_i\}$ for some i. As a vector space $\mathrm{CF}^\bullet(\phi(N^*T^{n-k}), F_0)$ matches $\mathrm{CM}^\bullet(T^{n-k})$ for the Morse function H.

The differential on $\mathrm{CF}^\bullet(\phi(N^*T^{n-k}), F_q)$ is related to the Morse differential. Let $|I|+1 = |J|$, meaning that $\deg(x_I)$ and $\deg(x_J)$ differ by one. If I and J differ at more than two elements, then $\mathcal{M}_\beta(\phi(N^*T^{n-k}), F_0, x_I, x_J)$ has nonzero dimension. If $I \lessdot J$ differ at a single element j, then there are exactly two holomorphic strips traveling between x_I and x_J,

$$\mathcal{M}(\phi(N^*T^{n-k}), F_0, x_I, x_J) = \{u^+_{I\lessdot J}, u^-_{I\lessdot J}\},$$

which as points receive opposite orientations. By our choice of perturbation, the symplectic areas of the strips $u^-_{I\lessdot J}$ and $u^-_{I\lessdot J}$ agree (and are exactly λ_0). Therefore

$$\langle d(x_I), x_J\rangle = \begin{cases} T^{\omega(u^+_{I\lessdot J})} - T^{\omega(u^-_{I\lessdot J})} = 0 & \text{if } I \lessdot J, \\ 0 & \text{otherwise,} \end{cases}$$

and we conclude that

$$\mathrm{HF}^\bullet(\phi(N^*T^{n-k}), F_q) = \Lambda\langle x_I\rangle = \bigwedge_{i\in\{1,\dots,n-k\}} \Lambda\langle x_i\rangle.$$

The example relates to the discussion of tropicalization as T^*F_0 can be identified with $X_A = (\mathbb{C}^*)^n = T^*Q/T^*_{\mathbb{Z}}Q$. If T^{n-k} is a linear subtorus of F_q, it corresponds to an $(n-k)$-dimensional subspace of $\tilde{T}^{n-k} \subset T^*_0Q$; let $\underline{V} \subset T_0Q$ correspond to the set of vectors which are annihilated by $\tilde{T}^{n-k}$. By abuse of notation, we use $\underline{V}$ to denote the integral affine subspace of Q with prescribed tangent space at 0. Under this identification, $N^*T^{n-k} \subset T^*F_0$ is $L_{\underline{V}} \subset X_A$. Using that Lagrangian intersection Floer cohomology is invariant under symplectomorphisms, and noting that if $q \notin \underline{V}$ then $F_q \cap L_{\underline{V}} = \varnothing$, we have computed

$$\mathrm{HF}^\bullet(L_{\underline{V}}, F_q) = \begin{cases} \bigwedge_{i\in\{1,\dots,n-k\}} \Lambda\langle x_i\rangle & \text{if } q \in \underline{V}, \\ 0 & \text{if } q \notin \underline{V}. \end{cases}$$

2.4.2 Local systems Recall that the points of X_B are in bijection with pairs (F_q, ∇) of Lagrangian torus fibers equipped with local systems. We now discuss how to incorporate this data into Lagrangian intersection Floer cohomology. The *unitary Novikov elements*

$$U_\Lambda := \left\{ a_0 + \sum_{i=1}^{\infty} a_i T^{\lambda_i} \;\middle|\; \lim_{i\to\infty} \lambda_i = \infty, \lambda_i > 0, a_0 \in \mathbb{C}^*, a_i \in \mathbb{C} \right\}$$

are those elements whose nonzero lowest-order term is a constant. We now consider (L_i, ∇_i), which are Lagrangian submanifolds with the additional choice of a trivial Λ-line bundle E_i and a U_Λ local system ∇_i. Given L_0 and L_1 which intersect transversely satisfying (i)–(v) we define $\mathrm{CF}^\bullet((L_0, \nabla_0), (L_1, \nabla_1))$ to be the chain complex where:

- The underlying vector space is $\bigoplus_{x\in L_0\cap L_1} \hom((E_0)_x, (E_1)_x)$.
- The differential is given by taking a ∇_i-weighted count of the holomorphic strips with boundary in $L_0 \cup L_1$. More precisely, let $\partial_i u$ be the boundary of u contained in L_i, and let $P_\gamma^{\nabla_i} : (E_i)_{\gamma(0)} \to (E_i)_{\gamma(1)}$ be the parallel transport induced by the local system along a path $\gamma : [0, 1] \to L_i$.

As in the definition of Lagrangian intersection Floer cohomology without local systems, let $x_+, x_- \in L_0 \cap L_1$ be intersection points with $\deg(x_+) + 1 = \deg(x_-)$. Given $\phi_x \in \hom((E_0)_{x_+}, (E_1)_{x_+})$ and a holomorphic strip $u \in \mathcal{M}_\beta(L_0, L_1, x_+, x_-)$ we obtain a map between the fibers above x_0,

$$P^{\nabla_1}_{(\partial^1 u)} \circ \phi_{x_+} \circ P^{\nabla_0}_{(\partial^0 u)^{-1}} \in \hom((E_0)_{x_-}, (E_1)_{x_-}).$$

The differential on $\mathrm{CF}^\bullet(L_0, L_1)$ is defined by taking the contributions $u \cdot \phi_{x_+}$ over all holomorphic strips between x_+ and x_-, weighted by the symplectic area,

$$d_{\nabla_0,\nabla_1}(\phi_{x_+}) := \sum_{x_- \mid \deg(x_-)=\deg(x_+)+1} \; \sum_{u\in M_\beta(L_0,L_1,x_+,x_-)} \pm T^{\omega(\beta)} P^{\nabla_1}_{(\partial^1 u)} \circ \phi_{x_+} \circ P^{\nabla_0}_{(\partial^0 u)^{-1}},$$

where the sign is determined by the orientation of the moduli space.

When ∇_i are the trivial local systems, this recovers $\mathrm{CF}^\bullet(L_0, L_1)$.

Example 2.4.4 (running example, continued) We now return to Example 2.4.3. Fix coordinates on F_q, and let $\{c_1, \dots, c_n\}$ be generators of $H^1(F_q, \mathbb{Z})$ associated to the coordinate directions. A local system on F_q is determined completely by its monodromy on the c_i. Given a Λ-unitary local system ∇_1 on F_q, we write $z_i = P^{\nabla_1}_{c_i}$. Let ∇_0 be the trivial local system on $L_{\underline{V}}$. We now compute the differential on $\mathrm{CF}^\bullet((L_{\underline{V}}, \nabla_0), (F_q, \nabla_1))$. Given $\phi_{x_I} \in \hom((E_0)_{x_I}, (E_1)_{x_I})$, and $I \lessdot J$ an index which differs at one spot j, we have

$$d_{\nabla_0,\nabla_1}(\phi_I) = (T^{\omega(u^+_{I\lessdot J})} P^{\nabla_1}_{(\partial^1 u^+_{I\lessdot J})} \circ \phi_I \circ P^{\nabla_0}_{(\partial^0 u^+_{I\lessdot J})^{-1}}) - (T^{\omega(u^-_{I\lessdot J})} P^{\nabla_1}_{(\partial^1 u^-_{I\lessdot J})} \circ \phi_I \circ P^{\nabla_0}_{(\partial^0 u^-_{I\lessdot J})^{-1}}).$$

Recall that all of the holomorphic strips between intersection points differing in index by 1 have the same area $\lambda_0 = \omega(u^+_{I\lessdot J}) = \omega(u^-_{I\lessdot J})$. Using that ∇_0 is the trivial local system, we get

$$T^{\lambda_0} P^{\nabla_1}_{(\partial^1 u^+_{I\lessdot J})} (\mathrm{id} - P^{\nabla_1}_{c_j}) \circ \phi_I \circ P^{\mathrm{id}}_{\partial^0 u^+_{I\lessdot J}}.$$

This vanishes if and only if $P_{c_j}^{\nabla_1} = z_j = 1$ for all $1 \leq j \leq n-k$. We conclude that

$$\mathrm{HF}^\bullet((L_{\underline{V}}, \mathrm{id}), (F_q, \nabla_1)) = \begin{cases} H^\bullet(T^{n-k}) z_j = 1 & \text{for all } 1 \leq j \leq n-k, \\ 0 & \text{otherwise.} \end{cases}$$

Notation 2.4.5 Given two Lagrangians L_0 and L_1 which intersect transversely, we will pick at each intersection point $x \in L_0 \cap L_1$ an isomorphism in $\hom((E_0)_x, (E_1)_x)$; by abuse of notation, we will denote this isomorphism also by $x \in \hom((E_0)_x, (E_1)_x)$. We can in this way write

$$\mathrm{CF}^\bullet(L_0, L_1) = \Lambda\langle x\rangle,$$

and the differential on this complex will be given by the structure coefficients

$$\langle d(x), y\rangle = \sum_{u \in \mathcal{M}_\beta(L_0, L_1, x, y)} T^{\omega(\beta)} P_{\partial u}^{\nabla_1, \nabla_2},$$

where $P_{\partial u}^{\nabla_1,\nabla_2} \in U_\Lambda$ is a unitary element determined by $P_{\partial u}^{\nabla_1,\nabla_2} \cdot y = P_{(\partial^1 u)}^{\nabla_1} \circ x \circ P_{(\partial^0 u)^{-1}}^{\nabla_0}$. This allows us to use the simpler (and more commonly employed) notation from Definition 2.4.1.

2.4.3 ***A*-tropicalization** When considering a complex space $X_B^{\mathbb{C}}$ on the B side, we used the projection $\pi_B\colon X_B^{\mathbb{C}} \to Q$ to obtain from each subvariety of $X_B^{\mathbb{C}}$ an amoeba which approximated the tropical subvariety. Just as with the B-tropicalization, given a Lagrangian submanifold $L \subset Q$ we could consider the Lagrangian torus fibration image of a Lagrangian submanifold $\pi_A(L) \subset Q$. However, since even Hamiltonian isotopic Lagrangian submanifolds can have different projections to the base of the Lagrangian torus fibration, this does not provide a very good definition of A-tropicalization. Instead, we use Lagrangian intersection Floer theory to define the A-tropicalization.

Definition 2.4.6 (preliminary) Let $L \subset X_A$ be a Lagrangian submanifold satisfying the conditions of Definition 2.4.1. We define the *A-tropicalization* or Floer-theoretic support of L to be the set

$$\mathrm{TropA}(L) := \{q \in Q \mid \text{there exists a Lagrangian brane } (F_q, \nabla) \text{ with } \mathrm{HF}^\bullet(L, (F_q, \nabla)) \neq 0\}.$$

The A-tropicalization is a decategorification of a much more powerful invariant captured by family Floer theory due to [5; 20]. From this viewpoint, the chain complexes $\mathrm{CF}^\bullet(L, (F_q, \nabla))$ should be considered as the stalks of a sheaf which are appropriately bundled together into a sheaf on X_B. This viewpoint on tropicalization is also employed in [53]. The A-tropicalization is a refinement of projection to the base of the Lagrangian torus fibration in the following sense:

Proposition 2.4.7 *Let $L \subset X_A$ be a Lagrangian brane. Then* $\mathrm{TropA}(L) \subset \pi_A(L)$.

Proof Suppose that $q \notin \pi_A(L)$. Then $F_q = \pi_A^{-1}(q)$ is disjoint from L. As the Floer intersection complex is generated on the intersection points, $\mathrm{CF}^\bullet(L, (F_q, \nabla)) = 0$. ◻

While $\operatorname{TropA}(L) \subset \pi_A(L)$ always holds, it will rarely be the case that $\pi_A(L) \subset \operatorname{TropA}(L)$. By invariance of $\operatorname{TropA}(L)$ under Hamiltonian isotopies, we obtain that

$$\operatorname{TropA}(L) \subset \bigcap_{\phi \in \operatorname{Ham}(X_A)} (\pi_A(\phi(L))),$$

where $\operatorname{Ham}(X_A)$ is the set of Hamiltonian isotopies of X_A. However, there is no reason to expect even this to be equality. Section 5 proves that when L is a Lagrangian constructed from the data of a tropical subvariety of Q, the above inclusion becomes equality. We see a toy version of this statement below.

Example 2.4.8 (running example, continued) We now are able to compute the A-tropicalization of a Lagrangian submanifold. Let $\underline{V} \subset Q$ be an integral affine k-subspace, so that $L_{\underline{V}} \subset X_A$ is a $T^{n-k} \times \mathbb{R}^k$ Lagrangian submanifold. We now compute the A-tropicalization of q. Since $\pi_A(L_{\underline{V}}) = \underline{V}$, by Proposition 2.4.7 $\operatorname{TropA}(L_{\underline{V}}) \subset \underline{V}$. By Example 2.4.4, whenever $q \in \underline{V}$ there exists a local system such that $\operatorname{HF}^\bullet(L_{\underline{V}}, (F_q, \nabla_1)) \neq 0$. Therefore $\operatorname{TropA}(L_{\underline{V}}) = \underline{V}$.

In this example, we see there are three steps of the A-realizability problem.

(i) First, we constructed a *geometric* lift $L_{\underline{V}}$ of $\underline{V}$.

(ii) The second step is to show that we have well-defined Floer cohomology groups. In the example above, this follows from $\pi_2(X_A, L_{\underline{V}}) = 0$, but more generally amounts to showing that the Lagrangian $L_{\underline{V}}$ is *unobstructed*.

(iii) Finally, the computation of support from Example 2.4.4 proves that this is a *faithful* lift of $\underline{V}$.

In the example of the lift of $\underline{V}$, we can do slightly more than compute the tropicalization of $L_{\underline{V}}$. We compute the A-*support*, which is the set of pairs (F_q, ∇) which have nontrivial pairing with $L_{\underline{V}}$:

$$\operatorname{SuppA}(L_{\underline{V}}) = \{(F_q, \nabla) \mid q \in \underline{V},\ P_c^\nabla = 0 \text{ for } c \cdot \underline{V} = 0\}. \tag{4}$$

Here we identify $H_1(F_q, \mathbb{Z})$ with $T^*_{\mathbb{Z}}(Q)$. At each point $q \in \underline{V}$, there is a $(U_\Lambda)^k$ choice of local systems satisfying the above criteria. The support can be identified with the set $\operatorname{SuppA}(L_{\underline{V}}) = \underline{V} \times (U_\Lambda)^k = (\Lambda^*)^k \subset X_B$.

2.5 A-tropicalization for the pair of pants

In this subsection, we carry out the entire A-realizability process with the tropical curve V_{pants} from Example 2.2.2. This computation first appeared in unpublished work from [30, Section 4.3], and stems from a discussion with Diego Matessi. We use this example computation to outline the remainder of the paper.

Geometric realizability: Section 3 We first discuss the process of building a Lagrangian submanifold which geometrically is a lift of V in the sense that $\pi_A(L_{V_{\text{pants}}})$ approximates V_{pants}. In dimension 2, one can obtain Lagrangian submanifolds in $(\mathbb{C}^*)^2$ by hyper-Kähler rotation of complex curves.[2] We therefore can build a Lagrangian lift of V_{pants} by starting with the holomorphic lift $\{(z_1, z_2) \mid 1 + z_1 + z_2 = 0\} \subset (\mathbb{C}^*)^2$

[2] We only use this construction for the ease by which it builds a Lagrangian pair of pants in dimension 2; we emphasize at this juncture that hyper-Kähler rotation is *not* mirror symmetry.

and applying hyper-Kähler rotation. For every $\epsilon > 0$, we can find a Lagrangian submanifold $L^{\varepsilon}_{V_{\text{pants}}} \subset X_A$ Hamiltonian isotopic to our hyper-Kähler rotation with the following properties:

- When restricted to the complement of a neighborhood of $0 \in Q$, we have
$$L^{\varepsilon}_{V_{\text{pants}}} \setminus \pi_A^{-1}(B_\varepsilon(0)) = L_{\underline{V}_1} \cup L_{\underline{V}_2} \cup L_{\underline{V}_3} \setminus \pi_A^{-1}(B_\varepsilon(0)).$$
This is one of the properties which characterizes a Lagrangian lift of a tropical curve.
- Furthermore, we can construct this Lagrangian so that it is symmetric under the permutation of coordinates (z_1, z_2) on X_A.

Unobstructedness: Section 4 The next step to the A-realization process is to show that the Lagrangian submanifold one builds can be analyzed with Floer theory. In this example, $L_{V_{\text{pants}}}$ is exact and so $\omega(\pi_2(X_A, L_{V_{\text{pants}}}))$ vanishes. It follows that $\mathrm{HF}^\bullet(L_{V_{\text{pants}}}, F_q)$ will be well defined.

Faithfulness: Section 5 We now compute $\mathrm{TropA}(L_{V_{\text{pants}}})$. Consider the Lagrangian pair of pants $L_{V_{\text{pants}}}$, the Lagrangian fiber F_q, and the holomorphic cylinder $z_1 = z_2$ as drawn in Figure 3, left. We take Hamiltonian perturbations so that the Lagrangian submanifolds intersect transversely. Nearby the point q, the Lagrangian $L_{V_{\text{pants}}}$ agrees with $L_{\underline{V}_3}$; therefore $F_q \cap L_{V_{\text{pants}}} = F_q \cap L_{\underline{V}_3}$. Following the notation from Example 2.4.4, we call the degree-0 intersection point $x_\varnothing$, and the degree-1 intersection point x_1. In addition to the agreement of intersection points, there are two "small strips" contributing to the differential on $\mathrm{CF}^\bullet(L_{V_{\text{pants}}}, F_q)$ which match the strips in the differential of $\mathrm{CF}^\bullet(L_{\underline{V}_3}, F_q)$. We call these holomorphic strips $u^+_{x_\varnothing < x_1}$ and $u^-_{x_\varnothing < x_1}$.

From the symmetry of our setup, the Lagrangian $L_{V_{\text{pants}}}$ intersects the complex plane $z_1 = z_2$ cleanly along a curve. Furthermore, the holomorphic cylinder $z_1 = z_2$ intersects F_q along a circle; therefore the portion of $z_1 = z_2$ bounded by $L_{V_{\text{pants}}}$ and F_q gives an example of a holomorphic strip with boundary on $L_{V_{\text{pants}}}$ and F_q. The ends of this holomorphic strip limit toward $x_\varnothing$ and x_1. The valuation projection of this strip

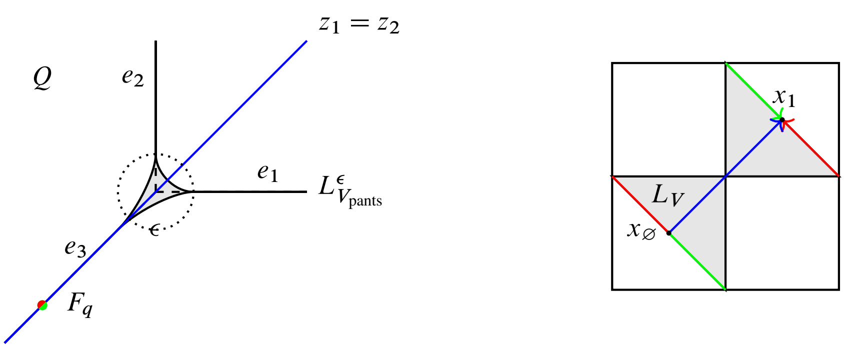


Figure 3: Left: the intersection of the blue holomorphic cylinder and the tropical Lagrangian pair of pants is clean, and gives a holomorphic strip with boundary on $L_{V_{\text{pants}}}$ and F_q. Right: the argument projection of $L_{V_{\text{pants}}}$ to F_q. The intersection points are labeled. The three holomorphic strips are denoted by the arrows, with u^{qv} drawn in blue.

is a line segment connecting the point q with the vertex of the tropical pair of pants. For this reason, we will call this holomorphic strip u^{qv}. The area of this strip is the length of the line segment corresponding to $\pi_A(u^{qv})$. The three holomorphic strips are more readily seen by considering the argument projection of $L_{V_{\text{pants}}}$ to F_q as in Figure 3, right.

We will think of u^{qv} as being a "big strip" as we can choose λ_0 small enough that $\lambda_0 = \omega(u^+_{x_\varnothing < x_1}) = \omega(u^-_{x_\varnothing < x_1}) \ll \omega(u^{qv})$. If no local systems are used, the differential on $\mathrm{CF}^\bullet(L_{V_{\text{pants}}}, F_q)$ is

$$d(x_\varnothing) = (T^{\omega(u^+_{x_\varnothing < x_1})} - T^{\omega(u^-_{x_\varnothing < x_1})} + T^{\omega(u_{qv})}) \cdot x.$$

This does not vanish, so $\mathrm{HF}^\bullet(L_{V_{\text{pants}}}, F_q) = 0$.

However, to compute the A-support we must compute Lagrangian intersection Floer cohomology where we equip F_q with a local system. We characterize the local system ∇ on F_q in terms of its holonomy along the $\arg(z_1)$ and $\arg(z_2)$ loops of F_q, giving us quantities $(\exp(b_1), \exp(b_2)) \in (U_\Lambda)^2$. We'll denote this nonunitary local system by ∇_{b_1,b_2}. Given a point $q = (-a, -a) \in \underline{V}_3$, we compute the quantities

$$\begin{aligned} &\omega(u^+_{x_\varnothing < x_1}) = \lambda_0, && \omega(u^-_{x_\varnothing < x_1}) = \lambda_0, && \omega(u^{qv}) = -a + \lambda_0, \\ &P^{\nabla_{b_1,b_2}}_{\partial u^+_{x_\varnothing < x_1}} = \exp\big(\tfrac{1}{2}(b_1 - b_2)\big), && P^{\nabla_{b_1,b_2}}_{\partial u^-_{x_\varnothing < x_1}} = \exp\big(\tfrac{1}{2}(b_2 - b_1)\big), && P^{\nabla}_{\partial u^{qv}} = \exp(\tfrac{1}{2}(b_1 + b_2)). \end{aligned}$$

The weights given by the local system are determined by the paths drawn in Figure 3, right, from which we obtain the differential on the $\mathrm{CF}^\bullet(L_{V_{\text{pants}}}, (F_q, \nabla_{b_1,b_2}))$:

$$\begin{aligned} \langle d_{\nabla_{b_1,b_2}}(x_\varnothing), x_1\rangle &= \overbrace{(P^{\nabla_{b_1,b_2}}_{\partial u^+_{x_\varnothing < x_1}} \cdot T^{\omega(u^+_{x_\varnothing < x_1})} - P^{\nabla_{b_1,b_2}}_{\partial u^-_{x_\varnothing < x_1}} \cdot T^{\omega(\partial u^-_{x_\varnothing < x_1})})}^{\text{small strips near } q} + \overbrace{P^{\nabla}_{\partial u^{qv}} \cdot T^{\omega(u^{qv})}}^{\text{large strips}} \\ &= T^\lambda_0\big(\exp\big(\tfrac{1}{2}(b_1 - b_2)\big) - \exp\big(\tfrac{1}{2}(b_2 - b_1)\big) + \exp\big(\tfrac{1}{2}(b_1 + b_2)\big) \cdot T^{-a}\big) \\ &= T^{-a+\lambda_0} \exp\big(-\tfrac{1}{2}(-b_1 - b_2)\big)(T^a \exp(b_1) - T^a \exp(b_2) + 1). \end{aligned}$$

This always has a U_Λ-worth of solutions obtained by setting $b_1 = \log(T^{-a}(T^a \exp(b_2) - 1))$. Therefore $(-a, -a) \in \mathrm{TropA}(L_{V_{\text{pants}}})$. From this we conclude that $\mathrm{TropA}(L_{V_{\text{pants}}}) = V_{\text{pants}}$.

This is one of the rare situations where we can compute the Floer-theoretic support explicitly: under the substitution $z_1 = T^{a_1}\exp(b_1), z_2 = T^{a_2}\exp(b_2)$, the Lagrangian tori $(F_{a_1,a_2}, \nabla_{b_1,b_2})$ belong to the support of $L_{V_{\text{pants}}}$ if and only if $z_1 - z_2 + 1 = 0$. This should be compared with the computation of the B-realization of V_{pants} from Example 2.2.2.

***B*-realizability: Section 6** The matching of the supports of the A- and B-realizations of V_{pants} can be captured in the language of homological mirror symmetry. This requires a description of the *Fukaya category* of a symplectic manifold. We define the *Fukaya precategory* of a compact symplectic manifold (X, ω):

- Objects are given by mutually transverse Lagrangian submanifolds $L \subset X$ which are graded, spin, and tautologically unobstructed (Section 4).

- For $L_0 \neq L_1$, the morphisms $\hom(L_0, L_1)$ are given by Lagrangian intersection Floer cochains $\mathrm{CF}^\bullet(L_0, L_1)$.
- k-compositions of morphisms
$$m^k\colon \bigotimes_{i=0}^{k-1} \hom^{g_i}(L_i, L_{i+1}) = \hom^{2-k+\sum g_i}(L_0, L_k)$$
are given by counts of holomorphic polygons with boundary on the L_k.

This is an A_∞ *precategory*, meaning that for every collection of objects $L_0, \dots, L_k$, the filtered A_∞ relations hold:
$$\sum_{j_1+j+j_2=k} (-1)^{\clubsuit} m^{j_1+1+j_2}(\mathrm{id}^{\otimes j_1} \otimes m^j \otimes \mathrm{id}^{\otimes j_2}) = 0.$$
Here $\clubsuit = j_1 + \sum_1^{j_1} g_i$, and $k \geq 1$.

The precategory can be appropriately completed to give a triangulated A_∞ category, the Fukaya category $\mathrm{Fuk}(X_A)$. Some of the hypotheses of the construction can be dropped or modified: for example, if X_A is a cotangent bundle (and not compact) there is a version of the Fukaya category (the wrapped Fukaya category, $\mathcal{W}(X_A)$) which can be defined with appropriate Lagrangian submanifolds. $X_A = (\mathbb{C}^*)^n = T^*F_0$ is one of these cases.

The homological mirror symmetry conjecture predicts that on mirror spaces the Fukaya category and derived category of coherent sheaves are derived equivalent.

Theorem *Let $X_A = (\mathbb{C}^*)^n$ and $X_B^{\mathbb{C}} = (\mathbb{C}^*)^n$. There is an equivalence of derived categories*
$$\mathcal{F}\colon \mathcal{W}(X_A) \to D^b_{dg}\operatorname{Coh}(X_B^{\mathbb{C}})$$
between the wrapped Fukaya category of exact admissible Lagrangian submanifolds of X_A and the bounded derived category of coherent sheaves on $X_B^{\mathbb{C}}$.

The proof of the theorem first shows that the zero section $L(0)$ of $\pi_A\colon X_A \to Q$ is a Lagrangian submanifold that generates $\mathrm{Fuk}(X_A)$. Then $\mathrm{HF}^\bullet(L(0), L(0))$ is shown to be the algebra $\mathbb{C}[(\mathbb{Z})^n] = \hom(\mathcal{O}_{(\mathbb{C}^*)^n}, \mathcal{O}_{(\mathbb{C}^*)^n})$. Since this generates $D^b_{dg}\operatorname{Coh}(X_B^{\mathbb{C}})$, these two categories are equivalent. However, this proof is nonconstructive: given an arbitrary exact Lagrangian submanifold $L \subset X_A$, there is no immediate way of determining the corresponding mirror sheaf in $D^b_{dg}\operatorname{Coh}(X_B^{\mathbb{C}})$. There are a few objects which we can match up under this functor. Let (F_q, ∇) be an exact fiber of the Lagrangian torus fibration. Then $\mathcal{F}(F_q, \nabla) \simeq \mathcal{O}_z$ for some $z \in X_B^{\mathbb{C}}$. From here, we obtain the following toy result, whose extension to the general V is the objective of the remainder of this paper.

Theorem 2.5.1 *$V_{\text{pants}} \subset \mathbb{R}^2$ is B-realizable.*

Proof From Section 2.5, we proved that V_{pants} is A-realizable by a Lagrangian $L_{V_{\text{pants}}}$. The support of the mirror sheaf $\mathcal{F}(L_{V_{\text{pants}}})$ is a B-realization of V_{pants}. □

Remark 2.5.2 There are other approaches to homological mirror symmetry which would yield the same theorem. A stronger result than what is given here would be to show that V_{pants} is realizable, and its realization compactifies to a subvariety (a line) in the projective plane. To prove this result, one would first show that L_V belongs to an appropriate "partially wrapped Fukaya category", and apply homological mirror symmetry theorems for toric varieties for the appropriate partially wrapped Fukaya category; see [37; 23] or [2; 28; 29]. Then one would need a mirror symmetry statement for the exact Lagrangian torus fiber equipped with nonunitary local systems, and replicate the argument of Section 2.5.

3 Geometric realization

The flexibility of Lagrangian submanifolds both complicates and simplifies the construction of a Lagrangian lift of a tropical subvariety. The additional flexibility means that we have a lot of wiggle room to construct a potential lift; however, identifying a Lagrangian submanifold as "the" lift of a tropical subvariety becomes impossible. For example, given any candidate lift L_V of a tropical subvariety V, one could apply a Hamiltonian isotopy to V to obtain a new Lagrangian submanifold. More generally, each potential Lagrangian lift L_V of V is supposed to represent the data of a sheaf on X_B whose support has tropicalization V; there are many such sheaves!

Despite all of this flexibility, we already have a good idea of what the Lagrangian lift L_V of V should look like from (1). Recall that $V^{(0)}$ is the union of the interiors of the top-dimensional polyhedral domains $\underline{V}$ defining V. At each component we can take the conormal torus construction to obtain a Lagrangian chain:

$$L_{V^{(0)}} := \bigcup_{\underline{V}\subset V^{(0)}} L_{\underline{V}}.$$

Intuitively, a geometric Lagrangian lift of V should approximate the chain $L_{V^{(0)}}$.

Remark 3.0.1 Fix an orientation on F_q, a fiber of the SYZ fibration. Then $L_{\underline{V}}$ inherits an orientation (which in local coordinates comes from $dq_1\wedge\cdots\wedge dq_k\wedge dp_{k+1}\wedge\cdots\wedge dp_n$). We will assume that we have fixed an orientation on F_q in advance so that $L_{\underline{V}}$ is equipped with a standard orientation.

We propose the following definition for a geometric Lagrangian lift of a tropical subvariety (which is similar to that proposed in [42, Definition 2.1]):

Definition 3.0.2 A family of oriented Lagrangian submanifolds L^ε_V for $\varepsilon > 0$ is a *geometric* Lagrangian lift of a weight-1 polyhedral complex $V \subset Q$ if the following conditions hold:

(i) The Lagrangians L^ε_V are all Hamiltonian isotopic,

(ii) Let $V^{(i)}$ be the collection of codimension-i strata of V. We require that

$$L^\varepsilon_V \setminus \pi_A^{-1}(B_\varepsilon(V^{(1)})) = L_{\underline{V}^{(0)}} \setminus \pi_A^{-1}(B_\varepsilon(V^{(1)})) \tag{5}$$

away from the codimension-1 strata, as oriented submanifolds.

(iii) The Lagrangians L^ε_V are embedded, graded, spin, and admissible (in the sense of Definition 3.1.1).

Remark 3.0.3 Definition 3.0.2 has two simplifying requirements; one is included due to current technical limitations in the definition of Floer cohomology, and the second is for convenience.

The requirement that L^{ϵ}_{V} is embedded is a technically needed assumption; we believe that this condition can be dropped without modifying our main results. Our reason for restricting ourselves to the embedded setting is that the Charest–Woodward pearly model as written does not include a description of Floer cohomology for immersed Lagrangian submanifolds.

While Definition 3.0.2 looks only at weight-1 polyhedral complexes, one can extend the story to weighted polyhedral complexes by asking that at each top-dimensional stratum $\underline{V} \subset \underline{V}^{(0)}$ with weight m, the realization $L_{\underline{V}}$ is m-disjoint copies of $N^*\underline{V}/N^*_{\mathbb{Z}}\underline{V}$. All results in this paper can be extended to the weighted setting.

The constructions from [30; 38; 39; 42] all satisfy Definition 3.0.2(i)–(ii). To prove that the previous definitions give examples of geometric Lagrangian lifts, we need to additionally show that they are admissible, graded, and spin. We prove these properties for certain examples of Lagrangian lifts in Sections 3.1–3.3.

While Definition 3.0.2 only asks that we take the lift of a weight-1 polyhedral complex, the only polyhedral complexes which admit such lifts are tropical ones.

Proposition 3.0.4 *Let V be a weight-1 rational polyhedral complex, and suppose that it has a Lagrangian lift L^{ε}_{V} satisfying Definition 3.0.2(i)–(ii). Then V is a tropical subvariety.*

Proof Select an interior point $r \in \underline{W} \subset V^{(1)}$ of the codimension-1 stratum of V. Pick $U_r \subset T_rQ$ a rational subspace such that $U_r \oplus T_rW = T_rQ$. Let $R \subset Q$ be a small polyhedral domain passing through r with tangent space U_r. Then $V|_R$ is a weight-1 rational polyhedral curve. By taking R small enough, $V|_R$ has a single vertex and edges pointing in directions $v_1, \dots, v_k$ corresponding to facets $F_1, \dots, F_k$ containing W. We need to prove that $\sum_{i=1}^{k} v_i = 0$; see Figure 4.

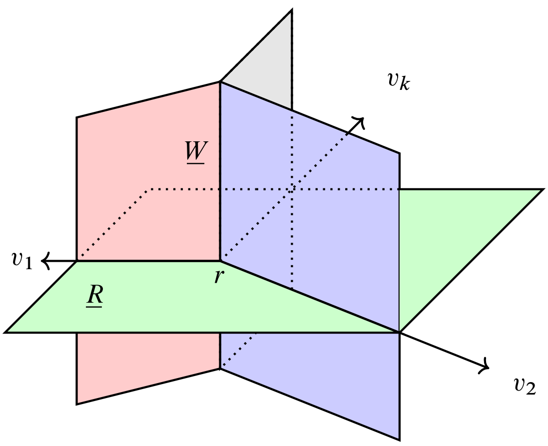


Figure 4: The Polyhedral complexes discussed in Proposition 3.0.4.

Consider the symplectic manifold $Y_A := T^*R/T^*_{\mathbb{Z}}R \subset T^*Q/T^*_{\mathbb{Z}}Q$, with the Lagrangian torus fibration $\pi_{Y_A}: Y_A \to R$. Let $i: R \to Q$ be the inclusion. Select ε small enough that $B^\varepsilon(W) \cap R$ is an interior set of R. Given a Lagrangian submanifold $L \subset T^*Q/T^*_{\mathbb{Z}}Q$, we can take a Hamiltonian perturbation of L so that

$$L_{i^*} \circ L := \{(r, i^*(p)) \mid (r, p) \in L,\ r \in R\}$$

is a Lagrangian submanifold of Y_A. See [29, Section 5.2] for a more general discussion of this construction from the perspective of Lagrangian correspondences. By definition $\pi_{Y_A}(L_{i^*} \circ L) = \pi_{Y_A}(L) \cap R$, so $L^\varepsilon_{V|_R} := L_{i^*} \circ L^\varepsilon_V$ is a geometric realization of $V|_R \subset R$. We therefore have reduced to the setting which is the lift of a tropical curve with a single vertex.

Given a tropical curve $V|_R \subset R$ with a single vertex v, the Lagrangian $L^\varepsilon_{V|_R}$ is a manifold with boundary. Consider the projection $\arg_R: Y_A \to F_r = T^*_r R/T^*_{\mathbb{Z},r}R$. Considering $\arg_R(L^\varepsilon_{V|_R})$ as a $\dim(F_r)-1$ chain, we obtain the relation in homology

$$0 = [\arg_R(\partial(L^\varepsilon_{V|_R}))] \in H_{\dim_{F_r} -1}(F_r).$$

There is an identification (as vector spaces) that sends an integral basis $e_1, \dots, e_n$ to the class of the perpendicular subtorus

$$T_r R \to H_{\dim_{F_r} -1}(F_r), \quad e_i \mapsto [\{\eta \in T^*_r R \mid \eta(e_i) = 0\}].$$

Since the boundary of $L^\varepsilon_{V|_R}$ lies in the region where (5) holds and we have an agreement of oriented submanifolds, we can therefore compute

$$[\arg_R(\partial(L^\varepsilon_{V|_R}))] = \sum_{i=1}^{k} [\{\eta \in T^*_r R \mid \eta(e_i) = 0\}],$$

proving that $\sum e_i = 0$. □

Notation 3.0.5 From here on, we will drop the ε in L^ε_V and simply write L_V for a Lagrangian which belongs to such a family.

3.1 Geometric Lagrangian lifts: admissibility

When Lagrangian submanifolds are noncompact, we need to place taming conditions on them so that they are Floer-theoretically well-behaved.

Definition 3.1.1 [28] Let $W_\Sigma: X_A \to \mathbb{C}$ be a Laurent polynomial whose monomials are indexed by A, the set of rays of a fan Σ. A *monomial division* Δ_Σ for $W_\Sigma = \sum_{\alpha \in A} c_\alpha z^\alpha$ is an assignment of a closed set $U_\alpha \subset Q$ to each monomial $\alpha \in A$ so that the following conditions hold:

- The sets U_α cover the complement of a compact subset of $Q = \mathbb{R}^n$.

- There exist constants $k_\alpha \in \mathbb{R}_{>0}$ such that for all z with $\mathrm{val}(z) \in U_\alpha$ the expression
$$\max_{\alpha\in A}(|c_\alpha z^\alpha|^{k_\alpha})$$
is always achieved by $|c_\alpha z^\alpha|^{k_\alpha}$.
- U_α is a subset of the open star of the ray α in the fan Σ.

A Lagrangian $L \subset X_A$ is Δ_Σ*-monomially admissible* if over $\pi_A^{-1}(U_\alpha)$ the argument of $c_\alpha z^\alpha$ restricted to L is zero outside of a compact set.

We will always assume that $\arg(c_\alpha) = 0$. An advantage of using the monomial admissibility condition for Lagrangian submanifolds is that it is a relatively simple check to see if a Lagrangian submanifold satisfies the condition.

Theorem [31, Theorem 3.1.7] *Suppose that V is the tropicalization of a hypersurface whose Newton polytope has dual fan Σ. Then the construction of L_V from [30] is Δ_Σ-monomially admissible.*

Let $V \subset Q$ be a tropical curve. We say that V is adapted to Σ if each semi-infinite edge of V points in the direction of a ray of Σ.

Claim 3.1.2 *Suppose that $V \subset \mathbb{R}^n$ is a weight-1 tropical curve adapted to Σ. Any Lagrangian lift L_V is Δ_Σ-monomially admissible.*

Proof Let $V_\infty^{(0)} = \{e_i\}_{i=1}^k$ denote the semi-infinite edges of V. We note that there exists a compact set $K \subset Q$ such that $L_C \setminus \pi_A^{-1}(K) = \bigsqcup_{e\in V_\infty^{(0)}} L_{\underline{e}} \setminus \pi_A^{-1}(K)$. Furthermore, K can be chosen so that $e \setminus K \subset U_\alpha$ if and only if e points in the direction $\alpha \in \Sigma$. Over this region, we observe that $\arg(z^\alpha)|_{N^*e/N^*_{\mathbb{Z}}e} = 0$. □

Remark 3.1.3 If some of the semi-infinite edges of V are weighted, we must replace the last condition in monomially admissible with "there exists a discrete set of values $\{\theta_i\}$ such that the argument of $c_\alpha z^\alpha|_{(L\cap C)\alpha}$ is a subset of $\{\theta_i\}$". The Floer-theoretic arguments in [28] can be applied to this setting as well (simply by letting θ_i be k-roots of unity, and replacing α with $k\alpha$).

3.2 Geometric Lagrangian lifts: homologically minimal and graded

The additional amount of flexibility that symplectic geometry affords us means that there are many geometric Lagrangian lifts of a single tropical subvariety. Some of these lifts differ for unimportant reasons: for instance, we could have included some extra topology in our Lagrangian by attaching a Lagrangian with vanishing Floer cohomology to a previously constructed lift. The following condition is imposed to weed out some of these worst offenders:

Definition 3.2.1 Let $j : L_{V^{(0)}} \setminus \pi_A^{-1}(B_\epsilon(V^{(1)})) \hookrightarrow L_V$ be the inclusion that is induced from the inclusion of the codimension-0 strata of V into V. We say that a lifting is *homologically minimal* if there exists a section $i : V \to L_V \subset X_A$ such that $H_1(L_V)$ is generated by the images of
$$(i)_*: H_1(V) \to H_1(L_V), \quad (j)_*: H_1(L_{V^{(0)}} \setminus \pi_A^{-1} B_\epsilon(V^{(1)})) \to H_1(L_V).$$

Let $i_{L_V}: L_V \to X_A$ be the inclusion of our Lagrangian submanifold. We say that L_V is an untwisted realization of V if the composition

$$V \xrightarrow{(i_{L_V} \circ i)} X_A \xrightarrow{\arg} F_q$$

is nullhomologous (for any choice of $q \in Q$).

Remark 3.2.2 For a fixed tropical subvariety V, there can be several geometric Lagrangian lifts of V which are meaningfully different. We expand on how these different choices of lifts correspond to tropical line bundles of V in Section 6.5.

The homologically minimal condition places some constraints on our Lagrangian submanifolds.

Lemma 3.2.3 *If L_V is homologically minimal and untwisted, then L_V is graded.*

Proof We recall the definition of graded from [52, Example 2.9]. Since $c_1(X_A) = 0$, we can take a section $\bigwedge_{i=1}^n (dq_i + i\,d\theta_i)^{\otimes 2}$ of $\Lambda^n(TX_A, J)^{\otimes 2}$. This determines a map

$$\det{}^2 \circ s_L : L \to S^1, \quad x \mapsto \Big(\bigwedge (dq_i + i\,d\theta_i)(T_x L)\Big)^2$$

A Lagrangian is $\mathbb{Z}$ graded if this map can be lifted to $\mathbb{R}$.

Consider a homologically minimal Lagrangian and untwisted Lagrangian L_V. There exist generators $\{[\alpha_k], [\beta_l]\}$ for $H_1(L_V)$ such that α_k is in the image of i and β_l is in the image of j. Since the compositions

$$\det{}^2 \circ s_{L_V} \circ i : V \to S^1, \quad \det{}^2 \circ s_{L_V} \circ j : (L_{V^{(0)}}) \to S^1$$

are constantly 0, it follows that there is no obstruction to lifting $\det^2 \circ s_{L_V} : L_V \to S^1$ to $\mathbb{R}$. □

Proposition 3.2.4 *Suppose that $V \subset \mathbb{R}^n$ is either a smooth tropical curve or a smooth tropical hypersurface. Then the construction of L_V given by [30; 39; 42] produces a homologically minimal Lagrangian lift L_V. The lifts are therefore graded.*

Proof In the cases of tropical curves, this follows from computing the homology of L_V from a cover given by $L_{\text{star}(v)}$. For hypersurfaces, this is proven in [31, Proposition 3.18]. □

Unless otherwise specified, the lift of a smooth tropical curve or hypersurface will always be the one given by [30; 39; 42].

3.3 Geometric Lagrangian lifts: spin

We start with a lemma on the topology of lifts of smooth genus-0 tropical curves.

Lemma 3.3.1 *Let $V \subset \mathbb{R}^n$ be a smooth genus-0 tropical curve.*

(i) *For any semi-infinite edge $f \in V_\infty^{(0)}$, the restriction map $\operatorname{res}_f^V : H^1(L_V) \to H^1(L_f)$ is a surjection.*

(ii) *For any semi-infinite edge f, the restriction map $\operatorname{res}^V_{V_\infty \setminus f} : H^2(L_V) \to \bigoplus_{g \neq f} H^2(L_g)$ is an injection.*

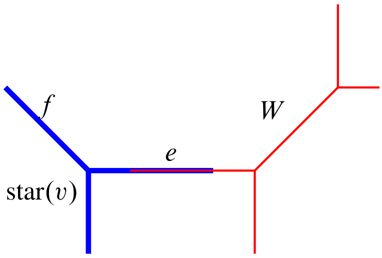


Figure 5: Covering our tropical curve V with two charts: W and a pair of pants $\operatorname{star}(v)$ centered at v.

Proof We prove (i) and (ii) by induction on the number of vertices in V.

Base case Suppose that V has one vertex. Then V is planar, and there exists a splitting of $(\mathbb{C}^*)^n = (\mathbb{C}^*)^2 \times (\mathbb{C}^*)^{n-2}$ such that $L_V = L_{\text{pants}} \times T^{n-2}$, where $L_{\text{pants}} \subset (\mathbb{C}^*)^2$ is the standard pair of pants. The boundary of the pair of pants is $S^1_{e_1} \cup S^1_{e_2} \cup S^1_{e_3}$, where e_1, e_2, and e_3 label the three edges of the pair of pants. A direct computation shows that

$$H^0(L_{\text{pants}}) \to H^0(S^1_{e_1}), \quad H^1(L_{\text{pants}}) \to H^1(S^1_{e_1}),$$

surjects, and that

$$H^0(L_{\text{pants}}) \to H^0(S^1_{e_1} \cup S^1_{e_2}), \quad H^1(L_{\text{pants}}) \to H^1(S^1_{e_1} \cup S^1_{e_2}),$$

injects. An application of the Künneth formula gives (i) and (ii) for L_V.

Inductive step Let $f \in V^{(0)}_\infty$ be any semi-infinite edge and let v be the vertex of V belonging to that edge. Let W be the tropical curve given by vertices not equal to v, so that $L_{\operatorname{star}(v)}$ and L_W cover L_V with intersection $L_{\operatorname{star}(v)} \cap L_W = L_e = T^{n-1} \times e$, as in Figure 5. This can be done because V is a tree.

(i) We use $L_{\operatorname{star}(v)}$ and L_W to compute the first cohomology of L_V using Mayer–Vietoris, and show that the red arrow in the diagram below is a surjection:

$$\begin{array}{ccccc}
H^1(L_V) & \xrightarrow{\operatorname{res}^V_{\operatorname{star}(v)} \oplus \operatorname{res}^V_W} & H^1(L_{\operatorname{star}(v)}) \oplus H^1(L_W) & \xrightarrow{\operatorname{res}^{\operatorname{star}(v)}_e - \operatorname{res}^W_e} & H^1(L_e) \\
\downarrow & & \downarrow{\scriptstyle \operatorname{res}^{\operatorname{star}(v)}_f \oplus 0} & & \\
H^1(L_f) & = & H^1(L_f) \oplus 0 & &
\end{array}$$

From the base case, given $\alpha \in H^1(L_f)$, there exists $\alpha' \in H^1(L_{\operatorname{star}(v)})$ with $\operatorname{res}^{\operatorname{star}(v)}_f(\alpha') = \alpha$. From the induction hypothesis, there exists $\beta' \in H^1(L_W)$ with $\operatorname{res}^W_e(\beta') = \operatorname{res}^{\operatorname{star}(v)}_e(\alpha')$. Therefore $(\alpha', \beta') \in \ker(\operatorname{res}^{\operatorname{star}(v)}_e - \operatorname{res}^W_e)$, and by exactness of the rows is in the image of $\operatorname{res}^V_{\operatorname{star}(v)} \oplus \operatorname{res}^V_W$. Let α'' be in the preimage of (α', β'). By commutativity of the below diagram, we conclude $\operatorname{res}^V_f(\alpha'') = \alpha$.

(ii) We compute $H^2(L_V)$ using Mayer–Vietoris, and show that the blue arrow of the following diagram is injective:

$$\begin{array}{ccccccc}
H^1(L_{\mathrm{star}(v)})\oplus H^1(L_W) & \twoheadrightarrow & H^1(L_e) & \xrightarrow{0} & H^2(L_V) & \xhookrightarrow{\mathrm{res}^V_{\mathrm{star}(v)}\oplus\,\mathrm{res}^V_W} & H^2(L_{\mathrm{star}(v)})\oplus H^2(L_W) \\
 & & & & \Big\downarrow {\scriptstyle \bigoplus_{g\neq f}\mathrm{res}^V_g} & & \Big\downarrow {\scriptstyle C\oplus D} \\
 & & & & \bigoplus_{\substack{g\in V^{(0)}_\infty \\ g\neq f}} H^2(L_g) & \hookrightarrow & \bigoplus_{\substack{g\in \mathrm{star}(v)^{(0)}_\infty \\ g\neq e,f}} H^2(L_g)\oplus\bigoplus_{\substack{g\in W^{(0)}_\infty \\ g\neq e}} H^2(L_g)
\end{array}$$

By (i), the leftmost arrow is surjective. By the exactness of the sequence, $\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W$ is injective on the second cohomology groups. Let $C=\bigoplus_{g\in\mathrm{star}(v)^{(0)}_\infty,g\neq e,f}\mathrm{res}^{\mathrm{star}(v)}_g$ and $D=\bigoplus_{g\in W^{(0)}_\infty,g\neq e}\mathrm{res}^W_g$. Now consider a class $\alpha\in H^2(L_V)$. Suppose that $\bigoplus_{g\neq f}\mathrm{res}^V_g(\alpha)=0$. We will show that $\alpha=0$. By commutativity of the diagram, $(C\oplus D)\circ(\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W)=0$. Because D is injective and $(\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W)$ is injective, $C\circ\mathrm{res}^V_{\mathrm{star}(v)}(\alpha)=0$ and $\mathrm{res}^V_W(\alpha)=0$. We now break into two cases:

Case I $\mathbf{res}^{V}_{\mathbf{star}(v)}(\alpha)=0$ This implies $(\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W)(\alpha)=0$, which by injectivity of $\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W$ tells us that $\alpha=0$.

Case II $\mathbf{res}^{V}_{\mathbf{star}(v)}(\alpha)\neq 0$ Observe that $(C\oplus\mathrm{res}^{\mathrm{star}(v)}_e)\circ\mathrm{res}^V_{\mathrm{star}(v)}$ is injective from the base case, so $\mathrm{res}^{\mathrm{star}(v)}_e\circ\mathrm{res}^V_{\mathrm{star}(v)}(\alpha)\neq 0$. Since $\mathrm{res}^V_W(\alpha)=0$, we obtain that

$$(\mathrm{res}^{\mathrm{star}(v)}_e\oplus\mathrm{res}^W_e)\circ(\mathrm{res}^V_{\mathrm{star}(v)}\oplus\mathrm{res}^V_W)(\alpha)\neq 0.$$

This violates the exactness of the top row, so Case II cannot occur. □

Proposition 3.3.2 *In the setting where $V\subset\mathbb{R}^n$ has genus* 0, *the constructions of* [30; 39; 42] *give homologically minimal untwisted geometric Lagrangian lifts L_V of V.*

Proof We prove that this Lagrangian submanifold is homologically minimal because the homology of the pair of pants is generated by the homology of the legs. If $n=2$, then L_V is a surface, and therefore spin.

To prove that the $n\geq 3$ cases are spin, we induct on the number of vertices in V. For the 1-vertex case, $L_{\mathrm{star}(v)}\simeq L_{\mathrm{pants}}\times T^{n-2}$. The manifolds $L_{\mathrm{pants},v}\times T^{n-2}$ have trivializations given by embedding $L_{\mathrm{pants},v}$ into $\mathbb{R}^2$, and is therefore spin.

As in the proof of Lemma 3.3.1, write $V=L_W\cup L_{\mathrm{star}(v)}$, where e is the common edge $L_W\cap L_{\mathrm{star}(v)}$. By the induction hypothesis we have a spin structure on L_W. By pullback, this gives a spin structure over L_e. Since $H^1(L_{\mathrm{star}(v)},\mathbb{Z}/2\mathbb{Z})\to H^1(L_e)$ surjects, there is no obstruction to picking a spin structure on $L_{\mathrm{star}(v)}$ agreeing with the prescribed spin structure on L_e. □

This method of proof can be extended to a slightly larger set of examples. We say that a smooth tropical curve V has planar genus if there exist cycles $c_1,\dots,c_k\subset V$ such that $\{[c_1],\dots,[c_k]\}$ generate $H_1(V)$, and there exist 2-dimensional planes $\underline{V}_k\subset\mathbb{R}^n$ such that $c_i\subset\underline{V}_k$.

Corollary 3.3.3 *If $V \subset \mathbb{R}^n$ is a smooth tropical curve V with planar genus, then L_V is spin.*

The other setting where tropical Lagrangian lifts have been studied is the setting of hypersurfaces.

Lemma 3.3.4 *If $V \subset \mathbb{R}^n$ is a smooth tropical hypersurface, the construction of [31; 39] of L_V is spin.*

Proof We break into several cases.

- If $n = 2$, then L_V is a punctured surface (and therefore spin).
- If $n = 3$, then L_V is an orientable 3-manifold (and therefore spin).
- If $n \geq 4$, then by [31] the Lagrangian L_V is the connected sum of two copies of $\mathbb{R}^n$ at several contractible regions U_α indexed by Δ, the Newton polytope of the defining tropical polynomial for V. Assume that $\dim(\Delta) = n \geq 4$ (as otherwise, we may reduce to one of the previous cases). Following [31, Proposition 3.18], we take two charts $L_r, L_s \simeq \mathbb{R}^n \setminus \bigcup_{\alpha\in\Delta} U_\alpha$ such that $L_V = L_r \cup L_s$. The L_r and L_s are homotopic to V. Then $L_r \cap L_s \simeq \bigcup_{\alpha\in\Delta} \partial U_\alpha$, where each ∂U_α is homotopic to either S^{n-1} or D^{n-1}. By Mayer–Vietoris, we compute

$$\bigoplus_{\alpha\in\Delta} H^1(\partial U_\alpha) \to H^2(L_V, \mathbb{Z}/2\mathbb{Z}) \to H^2(L_r, \mathbb{Z}/2\mathbb{Z}) \oplus H^2(L_r\mathbb{Z}/2\mathbb{Z}).$$

The left and right terms are zero when $n \geq 4$, so $H^2(L_V, \mathbb{Z}/2\mathbb{Z}) = 0$ and our Lagrangian is spin. □

4 Unobstructed Lagrangian lifts of tropical subvarieties

Since the geometric Lagrangian lifts L_V we construct will not be exact, to obtain a Lagrangian Floer *cohomology* theory we need to show that these Lagrangian submanifolds have Λ-filtered A_∞ algebra which can be unobstructed.

4.1 Pearly model for Floer cohomology

We will adopt the model employed in [12] to define $\mathrm{CF}^\bullet(L)$.

Theorem [12] *Let $L \subset X$ be a compact relative spin and graded Lagrangian submanifold inside a rational compact symplectic manifold X. Pick $h\colon L \to \mathbb{R}$ a Morse function, and $D \subset X \setminus L$ a stabilizing divisor. There exists a choice of perturbation datum $\mathcal{P}$ which defines a filtered A_∞ algebra $\mathrm{CF}^\bullet(L, h, \mathcal{P}, D)$ where:*

- *Chains are given by the Morse cochains of L, so that $\mathrm{CF}^\bullet(L, h, \mathcal{P}, D) = \Lambda\langle \mathrm{Crit}(h)\rangle$.*
- *Product structures come from counting configurations of treed disks. More precisely, given a collection of critical points $\underline{x} = (x_1, \dots, x_k)$, we define the structure coefficients*

$$\langle m^k(x_1 \otimes \cdots \otimes x_k), x_0\rangle = \sum_{\beta\in H_2(X,L)} (-1)^{\heartsuit}(\sigma(u)!)^{-1} T^{\omega(\beta)} \cdot \#\mathcal{M}_{\mathcal{P}}(X, L, D, \underline{x}, \beta)$$

which determine the A_∞ product structure. Here $\#\mathcal{M}_{\mathcal{P}}(X, L, D, \underline{x}, \beta)$ is the count of points in the moduli space of $\mathcal{P}$-perturbed pseudoholomorphic treed disks, $\sigma(u)$ denotes the number of stabilizing points on each of these treed disks, and $\heartsuit = \sum_{i=1}^{k} i\,|x_i|$.

The Λ-filtered A_∞ homotopy class does not depend on the choices of perturbation, divisor, and Morse function taken in the construction.

When the choices of $h, \mathcal{P}$, and D are unimportant, we will write $\mathrm{CF}^\bullet(L)$ instead of $\mathrm{CF}^\bullet(L, h, \mathcal{P}, D)$. The most visible difference between the tautologically unobstructed setting and this more general definition is that there now exists a curvature term $m^0 \colon \Lambda \to \mathrm{CF}^\bullet(L)$, which obstructs the squaring of the differential to zero. We say that L is *unobstructed* if $\mathrm{CF}^\bullet(L)$ has a bounding cochain $b \in \mathrm{CF}^\bullet(L)$; see Section B.1. When L is unobstructed, the deformed A_∞ structure on $\mathrm{CF}^\bullet(L, b)$ is a chain complex.

In this section we discuss whether a geometric Lagrangian lift L_V of a tropical subvariety is an unobstructed Lagrangian submanifold. We give an example computation in the pearly disk model to fix notation.

Example 4.1.1 (running example, continued) Returning to Example 2.4.4, we first examine $\mathrm{CF}^\bullet(F_q, \nabla_1)$. Since F_q bounds no topological disks, it does not bound any holomorphic disks. Therefore the Floer complex is the Morse-tree algebra of F_q. Give F_q the Morse function

$$f = \sum_{i=1}^{n} \cos(\theta_i). \tag{6}$$

We label the generators of

$$\mathrm{CF}^\bullet(F_q, \nabla_1) = \Lambda\langle y_I^1 \rangle,$$

where $I \subset \{1, \dots, n\}$. The differential is given by $m^1(y_I^1) = 0$, and for a particular set of perturbations the product structure is

$$m^2(y_I^1 \otimes y_J^1) = \begin{cases} \pm y_{I \cup J}^1 & \text{if } I \cap J = \varnothing, \\ 0 & \text{otherwise,} \end{cases}$$

where the sign is determined by the number of transpositions required to reorder $I \cup J$.

Remark 4.1.2 To our knowledge, it is unknown if there exists a perturbation scheme for Morse flow trees such that all higher products $m^k \colon \mathrm{CM}^\bullet(S^1)^{\otimes k} \to \mathrm{CM}^\bullet(S^1)[2-k]$ vanish.

To work in the setting where X_A is noncompact, we need to place restrictions on the noncompact behavior of the Lagrangian L to ensure that the moduli spaces of pseudoholomorphic treed-disks considered by [12] remain compact. A natural condition to impose is that $L \subset X_A$ is admissible (Definition A.0.1) with respect to a potential function $W \colon X_A \to \mathbb{C}$, so that the projection $W(L)$ fibers over the real axis $\mathbb{R}_{>0}$ outside of a compact set. Let $Y_A = W^{-1}(t)$ for $t \in \mathbb{R}_{\gg 0}$. Choices of different sufficiently large t yield fibers which are symplectomorphic. The restriction of L to a large fiber will be called $M := L|_{Y_A}$; this is a Lagrangian submanifold of Y_A. By Theorem A.0.2 there exists a treed-disk model for Lagrangian

Floer cohomology $\mathrm{CF}^\bullet(L)$ for W-admissible Lagrangians L. Furthermore, there exist compatible choices of perturbation data such the standard projection

$$\mathrm{CF}^\bullet(L) \to \mathrm{CF}^\bullet(M)$$

is a Λ-filtered map of A_∞ algebras.

A useful lemma of [28] states that when we have a monomially admissible Lagrangian L, there exists a potential function W such that L is W-admissible. From the data of a fan and $t \in \mathbb{R}$, [3] constructs *tropicalized potential*, which is a symplectic fibration $W_\Sigma^{t,1} : (\mathbb{C}^*)^n \to \mathbb{C}$ outside of a compact set.

Lemma 4.1.3 [28, Section 4.4; 29, Remark 2.10] *Suppose that L is a Lagrangian submanifold that is monomially admissible with respect to a monomial division adapted to Σ (in the sense of Definition 3.1.1). Then L can be made admissible for the tropicalized potential.*

4.2 Geometric Lagrangians versus Lagrangian branes

Definition 4.2.1 We say that an unobstructed Lagrangian submanifold (L_V, b) is a *Lagrangian brane lift* of V if L_V is a geometric Lagrangian lift of V.

Before developing constructions of bounding cochains for geometric Lagrangian lifts, we give some examples of geometric Lagrangian lifts which are known to be unobstructed (or tautologically unobstructed) Lagrangian submanifolds.

Example 4.2.2 (Lagrangian pair of pants) In [39], it was shown that the tropical pair of pants centered at the origin is an exact Lagrangian submanifold; a similar proof was given in [33], which showed that all tropical Lagrangian submanifolds constructed from the data of a dimer are exact.

Claim 4.2.3 *Let $V \subset \mathbb{R}^n$ be a tropical variety such that $0 \in \underline{V}_i$ for all facets $\underline{V}_i \subset V$. Let L_V^ε be a homologically minimal lift of V. Then L_V^ε is exact.*

Proof Let $\eta = p\,dq$ be the primitive for ω on $X_A = T^*T^n$. We need to show that η is exact on L_V^ε; equivalently we show that $\eta(\gamma) = 0$ for all $[\gamma] \in H_1(L_V^\varepsilon)$. Observe that L_V^ε retracts onto $L_V^\varepsilon \cap \pi_A^{-1}(B_\epsilon(0))$. Therefore, for every loop $\gamma \in H_1(L_V^\varepsilon)$, there exists γ' which is homotopic to γ and lives within in $\pi_A^{-1}(B_\epsilon(0))$; by letting $\epsilon \to 0$ we obtain $[\gamma'] = [\gamma]$ and $\gamma' \subset F_0$. As F_0 is exact, $\eta(\gamma') = 0$. □

Since these Lagrangians are exact, they are tautologically unobstructed and we can conclude that L_V^ε is a tropical Lagrangian brane.

In some cases, one obtains tautological unobstructedness (or unobstructedness) of the Lagrangian submanifold for free.

Example 4.2.4 We can obtain tautological unobstructedness for curves $V \subset \mathbb{R}^2$. Since L_V is a graded Lagrangian submanifold, the only holomorphic curves which might cause us difficulty are Maslov index 0 curves. However, the expected dimension of Maslov index 0 disks with boundary on a 2-dimensional Lagrangian is negative, therefore for a generic choice of almost complex structure these disks disappear and L_V^ε is tautologically unobstructed.

It is possible for nonregular Maslov index 0 disks to appear with boundary on L_V^ε, even in simple examples (see Example 4.2.11). More generally, [33] shows that there exists a "wall-crossing" phenomenon which occurs for isotopies between tropical Lagrangian submanifolds, and that the count of these Maslov index 0 holomorphic disks play a crucial role in understanding coordinates on the moduli space of tropical Lagrangian submanifolds.

Example 4.2.5 We now examine a setting outside of the mirrors to toric varieties. Let Q be any tropical abelian surface; then $X_A := T^*Q/T_{\mathbb{Z}}Q$ is a symplectic 4-torus. Given any tropical curve $V \subset Q$, there is a Lagrangian surface $L_V^\varepsilon \subset X_A$. By the same reasoning as above, L_V is tautologically unobstructed for generic choice of almost complex structure.

Example 4.2.6 We also know unobstructedness for geometric Lagrangian lifts is [38]. In that setting, the base of the Lagrangian torus fibration has nontrivial discriminant locus, and the tropical Lagrangians constructed are lifts of compact genus-0 tropical curves in the base. Mak and Ruddat show that the associated tropical Lagrangians are homology spheres and therefore are always unobstructed by a choice of bounding cochain [22].

In general, other techniques are required to prove that a geometric Lagrangian lift of a tropical subvariety is unobstructed.

Example 4.2.7 Given any smooth tropical hypersurface $V \subset \mathbb{R}^n$, [30] shows that the tropical Lagrangian lift can be equipped with a bounding cochain so that (L_V, b) is an unobstructed Lagrangian submanifold of $(\mathbb{C}^*)^n$. The proof uses that L_V can be constructed as a mapping cone of two Lagrangian sections in the Fukaya category; as these sections bound no holomorphic strips or disks, one expects that their Lagrangian connected sum can be equipped with a bounding cochain. In practice, the process of constructing the bounding cochain is delicate.

We furthermore expect that similar methods should show that given $V = V_1 \cap \cdots \cap V_k$ a transverse intersection of tropical hypersurfaces V_i, there exists L_V an unobstructed Lagrangian lift of V. The Lagrangian L_V is constructed by using the fiberwise sum of the lifts [29; 57], so that

$$L_V = L_{V_1} +_Q \cdots +_Q L_{V_k}.$$

While the resulting Lagrangian submanifold L_V may be immersed, over the top-dimensional stratum of V the Lagrangian submanifold L_V satisfies Definition 3.0.2. This provides the geometric realization. To obtain unobstructedness, we can also write L_V as the geometric composition of unobstructed Lagrangian correspondences (each giving the fiberwise sum with L_{V_i}). It is expected (from [21; 60]) that the geometric composition of unobstructed Lagrangian correspondences is unobstructed in this setting, from which it follows that L_V is unobstructed by the pushforward bounding cochain.

Example 4.2.8 Given a smooth tropical hypersurface V of a tropical abelian variety $Q = \mathbb{R}^n / M_{\mathbb{Z}}$, [30, Example 5.2.0.7] constructs an unobstructed Lagrangian lift (L_V, b) inside the symplectic torus $T^*Q / T^*_{\mathbb{Z}} Q$. The proof of unobstructedness is easier than the hypersurface setting (as one does not need to worry about issues of compactness).

The next two examples were suggested by Dhruv Ranganathan.

Example 4.2.9 Let $L_1 \subset X_1$ and $L_2 \subset X_2$ be tautologically unobstructed Lagrangian submanifolds. Then $L_1 \times L_2 \subset X_1 \times X_2$ is again a tautologically unobstructed Lagrangian submanifold. Furthermore, if the methods in [7] can be adapted to the Charest–Woodward model of Floer cohomology that we use, then the product of unobstructed Lagrangians is unobstructed. It follows that when $V_i \subset Q_i$ have Lagrangian brane lifts, then so does $V_1 \times V_2 \subset Q_1 \times Q_2$.

Example 4.2.10 Suppose for $t \in [0,1]$ we have geometric Lagrangian lifts L_{V_t} of a family of tropical subvarieties V_t. Furthermore, suppose that for $t \in [0,1)$ the lift is a Lagrangian brane lift. Then L_{V_1} is a Lagrangian brane lift of V_1. The proof uses Fukaya's trick to choose perturbation data so that for t close to 1, the L_{V_t} all have the same moduli spaces of pseudoholomorphic disks. Then there exists a subsequence of bounding cochains for the L_t which converge to a bounding cochain on L_1.

There are few general criteria for determining if a Lagrangian submanifold is unobstructed. To highlight some of the subtlety of the problems, we exhibit a tropical Lagrangian which bounds a nonregular Maslov index 0 disk.

Example 4.2.11 Consider the projection to the base of the Lagrangian torus fibration of the tropical Lagrangian submanifold L_{V_4} drawn in Figure 6. Let ℓ be the dashed red line. Take the standard metric on $\mathbb{R}^2$ so we may identify $T\mathbb{R}^2$ with $T^*\mathbb{R}^2$. Provided that one takes a symmetric construction of the Lagrangian pairs of pants (for example, using the construction of [39]), the holomorphic cylinder $T\ell / T_{\mathbb{Z}}\ell$ tropicalizing to the line ℓ intersects the Lagrangian L_{V_4} cleanly along an S^1. This yields an isolated holomorphic disk with boundary on L_{V_4}. This is *not* a regular holomorphic disk.

Further examples of nonregular Maslov index 0 disks are given in [32]. In Section 6.3, we give examples of geometric Lagrangians lifts which are unobstructed, but not tautologically unobstructed for any choice of admissible almost complex structure on $(\mathbb{C}^*)^n$. In Section 6.4, we show that there exists V such that L_V is obstructed.

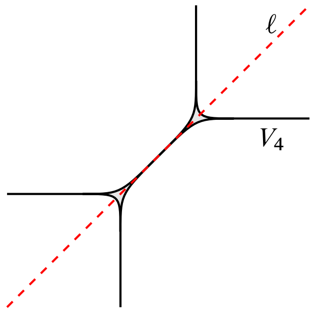


Figure 6: The projection to the base of the Lagrangian torus fibration of a tropical Lagrangian. The holomorphic cylinder above the red dashed line cleanly intersects the tropical Lagrangian.

4.3 Unobstructedness at the boundary

We now give a method for constructing a bounding cochain for a Lagrangian which is W-admissible. By Lemma 4.1.3, for every L a Δ_Σ-monomially admissible Lagrangian there exists a tropicalized potential $W_\Sigma^{t,1}: (\mathbb{C}^*)^n \to \mathbb{C}$ such that L is $W_\Sigma^{t,1}$-admissible. We state our results for W-admissible Lagrangians as the methods may be of interest beyond the monomially admissible setting.

Theorem 4.3.1 *Let $W: X \to \mathbb{C}$ be a potential function, and suppose that L is a W-admissible Lagrangian submanifold whose restriction to a large fiber is $M = L \cap (W^{-1}(t))$, where $t \in \mathbb{R}_{\gg 0}$. Suppose that there exists $M_0 \subset M$ a union of connected components of M with the property that*

(i) *the Lagrangian M_0 bounds no holomorphic disks, and*

(ii) *the map $H^1(M_0) \to H^2(L, M_0)$ is surjective.*

Then L is unobstructed.

The idea of the proof is to construct the bounding cochain for L by "lifting the curvature term of L to the boundary M_0". The condition that $H^1(M_0) \to H^2(L, M_0)$ shows that the curvature term (which takes values in the subcomplex $H^2(L, M)$) is the coboundary of something coming from the boundary M_0 of L. The algebraic content of this statement is Lemma B.2.8.

Proof We show that the A_∞ algebras $A = \mathrm{CF}^\bullet(L, M_0)$, $B = \mathrm{CF}^\bullet(L)$, and $C = \mathrm{CF}^\bullet(M_0)$ satisfy Lemma B.2.8(i)–(iii). From Theorem A.0.2, the sequence $A \to B \to C$ is exact, and A is an A_∞ ideal. Since M_0 bounds no holomorphic disks, C is tautologically unobstructed and A is a strong ideal, giving us Lemma B.2.8(i). Because M_0 bounds no holomorphic disks, $\mathrm{CF}^\bullet(M_0) = \mathrm{CM}^\bullet(M_0)$, which is quasi-isomorphic to $\Omega^\bullet(M)$. Thus we have Lemma B.2.8(ii). Finally, the hypothesis that $H^1(M_0) \to H^2(L, M_0)$ surjects is exactly Lemma B.2.8(iii). □

We give an example that relates to the discussion in [18, Section 5.2]:

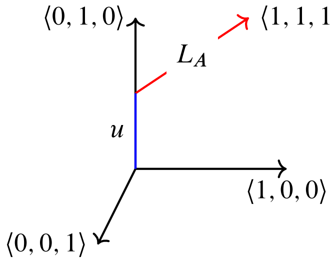


Figure 7: The projection of the AV Lagrangian L_A to the base $(\mathbb{R}_{\geq 0})^3$ of the Lagrangian torus fibration of the Aganagic–Vafa Lagrangian $L_A \subset \mathbb{C}^3$. We also draw the projection of the single simple holomorphic disk which contributes to a bounding cochain for L_A.

Example 4.3.2 (Aganagic–Vafa brane) Let $A \in \mathbb{R}_{>0}$ be some constant. The Aganagic–Vafa (AV) brane is a Lagrangian submanifold $L_A \subset \mathbb{C}^3$ parametrized by

$$D^2 \times S^1 \to \mathbb{C}^3, \quad (r_1, \theta_1, \theta_2) \mapsto ((\sqrt{A^2 + r^2}) e^{-i(\theta_1+\theta_2)}, re^{i\theta_1}, re^{i\theta_1}).$$

The Lagrangian L_A is admissible for the potential function $W(z_1, z_2, z_3) = z_1 z_2 z_3$. The restriction to the fiber $M_A \subset W^{-1}(s) = (\mathbb{C}^*) \times (\mathbb{C}^*)$ is a product-type torus, so it bounds no holomorphic disks, and we may apply Theorem 4.3.1 to conclude that this Lagrangian is unobstructed by a bounding cochain.

The bounding cochain corrects this Lagrangian submanifold so that it agrees with predictions from mirror symmetry. By application of the open mapping theorem, the only holomorphic disks with boundary on L_A for the standard complex structure must lie in the fiber $W^{-1}(0)$; in fact, the only simple holomorphic disk with boundary on L_A is parametrized by

$$u\colon (D^2, \partial D^2) \to (\mathbb{C}^2, L_A), \quad z \mapsto (Az, 0, 0).$$

A computation shows that the partial Maslov indices of this disk are $(2, -1, -1)$ and therefore this is a regular Maslov index 0 disk by [46]. This shows that the bounding cochain constructed by Theorem 4.3.1 is nontrivial.

Under an additional assumption [34, Assumption 5.2.3] one can compute the m^0-term, which counts the multiple covers of the disk u with an appropriate weight. The bounding cochain is $\sum_{k=1}^{\infty} (1/k) T^{k\omega(u)} x$, where $x \in \mathrm{CM}^\bullet(M_A)$ is a meridional class of the torus.

We remark that the Lagrangian L_A is an example of a tropical Lagrangian submanifold considered in [42], and the projection of L_A under the moment map $\mathbb{C}^3 \to Q = \mathbb{R}^3_{\geq 0}$ is the ray $(|A|^2, 0, 0) + t\langle 1, 1, 1\rangle$.

Corollary 4.3.3 *Let $V \subset Q$ be a genus-0 smooth tropical curve. Let L_V be a homologically minimal geometric Lagrangian lift of V. Then L_V is unobstructed, so there exists (L_V, b) a Lagrangian brane lift of V.*

Proof We show that the Lagrangian L_V satisfies the criteria of Theorem 4.3.1. Let $V_\infty^{(0)} \subset V$ be the set of semi-infinite edges of V. The boundary of this tropical Lagrangian realization $M \subset Y_A$ is contained within the lift of the semi-infinite edges $\bigsqcup_{e\in V_\infty^{(0)}} L_e = T_e^{n-1} \times e$. Therefore M is the disjoint union of tori indexed by the semi-infinite edges of V, $\bigcup_{e\in V_\infty^{(0)}} T_e^{n-1}$. At each edge, we see that $\pi_2(X_A, L_e) = 0$. It follows that $M \subset Y_A$ bounds no holomorphic disks, so we satisfy Theorem 4.3.1(i). Select $f \in V_\infty^{(0)}$ any edge, and let $M_0 = \bigcup_{g\in V_\infty^{(0)}, g\neq f} T^{n-1}$.

It remains to prove Theorem 4.3.1(ii), that the image of $H^1(M_0)$ generates $H^2(L_V, M_0)$. From Lemma 3.3.1, for any semi-infinite edge f of V, $\mathrm{res}^V_{V_\infty\setminus f}\colon H^2(L_V) \to \bigoplus_{g\in V_\infty^{(0)}, g\neq f} H^2(L_g)$ is an injection. From the long exact sequence for relative cohomology,

$$\bigoplus_{\substack{g\in V_\infty^{(0)}\\ g\neq f}} H^1(L_g) \to H^2(L_V, M_0) \xrightarrow{0} H^2(L_V) \hookrightarrow \bigoplus_{\substack{g\in V_\infty^{(0)}\\ g\neq f}} H^2(L_g),$$

the leftmost arrow surjects. □

5 Faithfulness: unobstructed lifts as A-realizations

Given a Lagrangian torus fibration $X_A \to Q$, the *A-tropicalization* of a tautologically unobstructed Lagrangian submanifold L is the set of points $q \in Q$ such that $\mathrm{HF}^\bullet(L, (F_q, \nabla)) \neq 0$ for some choice of local system ∇ on F_q; see (4). We now describe the A-tropicalization when L is unobstructed by a bounding cochain. Because we again work in the scenario where the space X is noncompact, we must apply a taming condition at infinity to study Floer cohomology. By the same arguments as for Theorem A.0.2, whenever L_0 is admissible and L_1 is compact for a potential $W\colon X_A \to \mathbb{C}$, there exists a well-defined $\mathrm{CF}^\bullet(L_0, \nabla_0) - \mathrm{CF}^\bullet(L_1, \nabla_1)$ bimodule $\mathrm{CF}^\bullet((L_0, \nabla_0), (L_1, \nabla_1))$ given by [12]. As in the setting of Definition 2.4.1, $\mathrm{CF}^\bullet((L_0, \nabla_0), (L_1, \nabla_1))$ is generated on the transverse intersections between L_0 and L_1. The A_∞ bimodule structure comes from counting pseudoholomorphic treed strips. We require L_1 to be compact to avoid issues of determining how to apply wrapping Hamiltonians in the definition.

Example 5.0.1 (running example, continued) We return to Example 4.1.1. Now we bring in the second Lagrangian $(L_{\underline{V}}, \nabla_0)$, which we give the trivial local system. The bimodule $\mathrm{CF}^\bullet((L_{\underline{V}}, \nabla_0), (F_q, \nabla_1))$ has the same generators as Example 2.4.4; following Notation 2.4.5, we call these generators x_J^{01}, where $J \subset \{0, \dots, n-k\}$. Since neither F_q nor $L_{\underline{V}}$ bounds a holomorphic disk, the differential agrees with Example 2.4.4,

$$m^1(x_I^{01}) = \sum_{I \lessdot J} \pm T^{\lambda_0}(\mathrm{id} - P_{c_j}^{\nabla_1}) x_J.$$

where λ_0 is the area of the small holomorphic strips.

We now describe the module product structure. This is given by counts of configurations of a Morse flow-line on F_q which are incident to a strip with boundary on $F_q \cup L_{\underline{V}}$. Recall that the Hamiltonian pushoff for $L_{\underline{V}}$ is given by (3) while the Morse function for F_q is given by (6). As before, we use

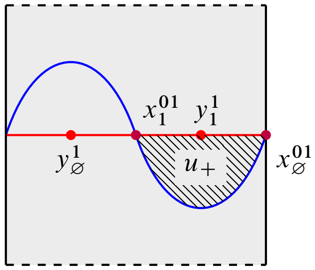


Figure 8: The treed strip contributing to the bimodule product $m^2(x_\varnothing^{01}\otimes y_1^1)=x_1^{01}$.

$\{y_I^1\}_{I\subset\{1,\dots,n\}}$ to label the critical points of $f: F_q\to\mathbb{R}$. The moduli space of strips from $\underline{x}_I^{01}$ and $\underline{x}_J$ is nonempty when $I<J$; the boundary of the strips sweep out the subtorus spanned by the indices of $J\setminus I$. The downward flow space of y_K is the subtorus spanned by indices $\{1,\dots,n\}\setminus K$. These two subtori intersect transversely only when $(J\setminus I)\sqcup(\{1,\dots,n\}\setminus K)=\{1,\dots,n\}$, which can be rephrased as

$$J=K\cup I,\quad K\cap I=\varnothing.$$

See Figure 8 for a treed strip that contributes to the product. From this, it follows that the module product structure is given by

$$m^2(x_I^{01}\otimes y_J^1)=\begin{cases}P_{\partial u^+}^{\nabla_1}T^{|J|\lambda_0}x_{I\cup J}^{01} & \text{if } I\cap J=\varnothing \text{ and } I\cup J\subset\{0,\dots,n-k\},\\ 0 & \text{otherwise.}\end{cases}$$

Here $|J|\lambda_0$ is the area of a holomorphic strip from x_i^{01} to $x_{I\cup J}^{01}$, and $P_{\partial u}^{\nabla_1}$ is the holonomy of the local system along the F_q boundary of the strip. We remark that when $J=\varnothing$, the same formula holds (simply that u^+ is regarded as the constant strip at x_I^{01}). The map

$$m^2(x_\varnothing^{01},-)\colon \mathrm{HF}^1((F_q,\nabla_0))\to T^{\lambda_0}\,\mathrm{HF}^1((L_{\underline{V}},\nabla_0),(F_q,\nabla_1))$$

surjects whenever the local system ∇_1 has holonomy of the form $\mathrm{id}+T^{\lambda_1}A$ along all the F_q boundary of all strips.

5.1 Definition of support

When Lagrangians (L_0,∇_0) and (L_1,∇_1) are unobstructed by bounding cochains b_0 and b_1, we can deform the Lagrangian intersection Floer cohomology $\mathrm{CF}^\bullet((L_0,\nabla_0),(L_1,\nabla_1))$ by these bounding cochains to obtain $\mathrm{CF}^\bullet((L_0,\nabla_0,b_0),(L_1,\nabla_1,b_1))$, a $\mathrm{CF}^\bullet(L_0,\nabla_0,b_0)$–$\mathrm{CF}^\bullet(L_1,\nabla_1,b_1)$ bimodule. Since $\mathrm{CF}^\bullet(L_j,\nabla_0,b_i)$ has no curvature, the differential

$$m^1\colon \mathrm{CF}^\bullet((L_0,\nabla_0,b_0),(L_1,\nabla_1,b_1))\to\mathrm{CF}^\bullet((L_0,\nabla_0,b_0),(L_1,\nabla_1,b_1))$$

squares to zero, giving us cohomology groups which we can study.

Definition 5.1.1 Let $(L,\nabla,b)\subset X_A$ be an admissible Lagrangian brane. The *A-tropicalization* of (L,∇,b) is the set

$$\mathrm{TropA}(L,\nabla,b):=\{q\mid\exists(F_q,\nabla')\text{ such that }\mathrm{HF}^\bullet((L,\nabla,b),(F_q,\nabla'))\neq 0\}.$$

Remark 5.1.2 Suppose that there is a bounding cochain b' for F_q such that

$$\mathrm{HF}^\bullet((L,\nabla,b),(F_q,\nabla',b')\neq 0.$$

As F_q is tautologically unobstructed, a general principle of Lagrangian Floer cohomology (the divisor axiom) states that there exists a local system called ∇'' such that

$$\mathrm{HF}^\bullet((L,\nabla,b),(F_q,\nabla''))=\mathrm{HF}^\bullet((L,\nabla,b),(F_q,\nabla',b')). \tag{7}$$

The local system ∇'' is usually denoted by $\exp(b')$. To our knowledge, the divisor axiom has not been proven for the [12] model of Lagrangian intersection Floer cohomology. In [9] a proof of the divisor axiom was given for the de Rham version of open Gromov–Witten invariants. The central idea of the proof is that the coefficients in the exponential function make an appearance through the application of the "forgetting boundary points" relation between moduli spaces of holomorphic disks. The coefficients $1/k!$ in the expansion of the exponential function show up via the number of ways one can forget boundary marked points. Under the assumptions that Auroux uses, the forgetful axiom for pseudoholomorphic disks holds. In the Charest–Woodward model for $\mathrm{CF}^\bullet(L_0,L_1)$ we do not expect that perturbations for Morse theory admit a "forgetting marked point" axiom. In our setting (where $\omega(\pi_2(X_A,F_q))=0$) the arguments used in Lemma 5.2.2 show that for all (F_q,∇',b') there exists (F_q,∇'') such that the identity on (F_q,∇',b') factors through (F_q,∇'') and vice-versa. Provided that a Charest–Woodward model of the Fukaya *category* with homotopy unit exists, this would prove (7) (although not give the closed-form expression for ∇' as the exponential of the bounding cochain, as in the de Rham version). From the divisor axiom, it follows that the A-tropicalization can be rewritten as

$$\mathrm{TropA}(L,\nabla,b)=\{q\mid \exists(F_q,\nabla',b')\text{ with }\mathrm{HF}^\bullet((L,\nabla,b),(F_q,\nabla',b'))\neq 0\}. \tag{8}$$

There remain some subtle differences between bounding cochains and local systems in general. It is clear that we can only expect to replace bounding cochains with local systems in the setting where L is tautologically unobstructed. Furthermore, we do not expect that when L is tautologically unobstructed that we can replace (L,∇) with (L,b). This is because $\mathrm{val}(b)>0$, so it can only be expected to represent local systems whose holonomy is of the form $\mathrm{id}+T^\lambda A$ where $A\in U_\Lambda$ and $\lambda>0$. In the specialization to Lagrangian tori in a Lagrangian torus fibration, we believe that the requirement that $\mathrm{val}(b)>0$ may be loosened to $\mathrm{val}(b)\geq 0$ by application of the reverse isoperimetric inequality (in the same way that the reverse isoperimetric inequality is used to prove that the family Floer sheaf has structure coefficients defined over an affinoid algebra).

5.2 A-tropicalization of tropical Lagrangian lifts

In general, it is difficult to compute $\mathrm{TropA}(L,b)$, as it requires having a very good understanding of the differential on $\mathrm{CF}^\bullet((L,b),(F_q,b'))$. In Section 2.5 we performed this computation for the pair of pants $V\subset\mathbb{R}^2$. Computation of the A-tropicalization is more tractable when the Lagrangian L_V is a geometric lift of a tropical subvariety because we have good control of leading-order contributions to the differential.

The main tool that we use to compute the A-tropicalization is the following lemma:

Lemma 5.2.1 *Let L_V be a Lagrangian lift of a tropical curve, and let $U \subset Q$ be an open set such that $L|_{\pi_A^{-1}(U)} = L_{\underline{V}}|_{\pi_A^{-1}(U)}$. For $q \in U$, let $R(q)$ be the distance from q to $Q \setminus U$. There exists a function $A_{L_V,U} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that every holomorphic strip u with boundary on $L \cup F_q$ with $q \in U$ is either*

- *"small" and has image contained within $\pi_A^{-1}(U)$, and therefore describes a holomorphic strip with boundary on $L_{\underline{V}} \cup F_q$, or*
- *"large" and has symplectic energy greater than $A_{L_V,U}(R(q))$.*

Furthermore, there exists a constant $C_{L_V,U}$ such that

$$\lim_{R \to \infty} \frac{A_{L_V,U}(R)}{R} = 2C_{L_V,U}.$$

Additionally, we may replace F_q with a small Hamiltonian pushoff of F_q while preserving the bound.

Proof The lemma is an application of the reverse isoperimetric inequality from [25]. We use the proof for holomorphic strips which is employed by [13] following [6; 16]. Recall that the reverse isoperimetric inequality states that given a Lagrangian L and choice of almost complex structure J there exists a constant $A_{L,J}$ such that we can lower-bound the energy of pseudoholomorphic disks u with boundary on L by

$$A_{L,J}\,\ell(\partial u) \leq \int_u \omega, \tag{9}$$

where ℓ is the length as computed by the metric determined by J and ω. The reverse isoperimetric inequality for pseudoholomorphic strips requires the intersections of our Lagrangians L_V and F_q to be "locally standard" [13, Definition II.1]. Any F_q and $L_{\underline{V}}$ satisfy this criterion, so whenever $q \in U$, the Lagrangian submanifolds L_V and F_q have locally standard intersection. The reverse isoperimetric inequality from [13] can be stated as

$$sA_{L_V,F_q}\,\ell(\partial u \cap B(C)^c) \leq \omega(u \cap \widetilde{U}_s), \tag{10}$$

where C is chosen so that the radius-s normal neighborhoods $N_s(L_V)$ and $N_s(F_q)$ are defined for all $s < C$, and

$$\widetilde{U}_s = N_s(L_V) \cup N_s(F_q), \quad B(C) = N_C(L_V) \cap N_C(F_q).$$

We obtain a weaker but more applicable bound by making the replacement $B(C) := \pi_A^{-1}(B_C(q))$, for which F_q is a subset. With this substitution, the left-hand side of (10) only depends on the length of the boundary of u in L_V. As the excluded neighborhood $B(C)$ is monotonic in C, if we choose $C(R) < \min\left(\frac{1}{4}R, C_{L_V,U}\right)$ where $C_{L_V,U}$ is the injectivity radius of L_V we can impose the additional condition that $\pi_A(B(C(R))) \subset U$. We now bound the constant A_{L_V,F_q}, called K in [13, Corollary II.11]. It is the product of the constants

- C_1 and C_2^{-1} from [13, Proposition II.8], which provide constants of domination between the pseudometric given by a particular plurisubharmonic function h and the standard metric, and
- A, which provides a bound for $|\mathrm{grad}\, h|$ over $\widetilde{U}_s$.

A special feature of tropical Lagrangian submanifolds is that over the region $\pi_A^{-1}(U)\setminus B(C)$, the function h agrees with the distance to $L_{\underline{V}}$. As L_V is totally geodesic over $\pi_A^{-1}(U)$, we obtain that $dd^c h(-,\sqrt{-1}-)$ agrees with the metric induced by the standard metric over this chart. Therefore the constants C_1, C_2^{-1}, and A are all 1 over this region. By restricting the integral on the penultimate line of [13, (9)] to the region $\pi_A^{-1}(U)$, we may replace A_{L_V,F_q} with 1 to obtain the bound $C(R)\ell(\partial u\cap B(C(R))^c\cap\pi_A^{-1}(U))\le\omega(u)$.

We now show that every strip is either "small" or "large":

- Suppose that $\partial u\subset\pi_A^{-1}(U)$. Then u describes a strip with boundary on $L_{\underline{V}}\cup F_0$; we know that all such strips are contained within $\pi_A^{-1}(U)$ and have an upper bound for their energy.
- Otherwise $\partial u\not\subset\pi_A^{-1}(U)$. Let ℓ_Q be distance as measured on Q. Observe that for any path γ with one endpoint in F_q and another endpoint in $X\setminus\pi_A^{-1}(U)$ we have the bound

$$R-C(R)\le\ell_Q(\pi_A(\gamma\cap B(C)^c))\le\ell(\gamma\cap B(C)^c).$$

Since the boundary of u must have at least two such paths,

$$A_{L_V,U}(R):=2C(R)(R-C(R))<\omega(u).$$

As $R\to\infty$, we have that $C(R)\to C_{L_V,U}$, from which we obtain the asymptotic behavior of $A_{L_V,U}$. □

The constant $C_{L_V,U}$ giving the injectivity radius of $L_{\underline{V}}$ can be computed from the tropical data of $\underline{V}$. In the 2-dimensional setting, we obtain the following nice relation. At a top-dimensional stratum (edge) e with integral primitive direction $\vec{v}$, the constant $C_{L_V,U}$ in Lemma 5.2.1 is $1/(2|\vec{v}|)$. The bound for the holomorphic energy of the strips becomes

$$2C(R-C_{L_V,U})=\frac{R-C_{L_V,U}}{|\vec{v}|}.$$

We observe that $R/|\vec{v}|$ is the *affine radius* of the neighborhood around the point q. This can be observed in Section 2.5 and Example 6.3.2, where the affine lengths of edges in tropical curves govern the areas of holomorphic disks and strips which appear in those computations.

Lemma 5.2.2 *Let $U\subset Q$ be a neighborhood of q. Suppose that $(L,\nabla,b_0)\subset X_A$ is a Lagrangian brane whose restriction to $\pi_A^{-1}(U)$ is*

$$L|_{\pi_A^{-1}(U)}=L_{\underline{V},m}|_{\pi_A^{-1}(U)},$$

where $\underline{V}\subset U$ is a k-dimensional linear subspace, and m the multiplicity. Then there exists a choice of bounding cochain and local system on F_q such that

$$\mathrm{HF}^0((L,\nabla_0,b_0),(F_q,\nabla,b))=\Lambda.$$

Proof To reduce notation in the proof, we will take the same simplifying assumptions as in Lemma 5.2.1. Additionally, we assume that the local system ∇_0 and bounding cochain b_0 on L are trivial.

We see that $L|_{\pi_A^{-1}(U)} \cap F_q$ cleanly intersect along a $T^{n-k} \subset F_0$; morally we now apply the spectral sequence of [48; 51] to compute the Floer cohomology of $\mathrm{CF}^\bullet(L, F_0)$ as a deformation of $C^\bullet(T^{n-k})$. Because $L|_{\pi_A^{-1}(U)} = L_{\underline{V}}|_{\pi_A^{-1}(U)}$, we can apply Lemma 5.2.1. Following Example 5.0.1, apply a Hamiltonian isotopy to L so that L and F_q intersect transversely. Take the perturbation small enough that the area of the holomorphic strips λ_0 is less than the bound $\lambda_1 := A_{L_V,U}(R)$ provided by Lemma 5.2.1. By Lemma 5.2.1, the map $m^2\colon \mathrm{CF}^\bullet(L_V, F_q) \otimes \mathrm{CF}^\bullet(F_q) \to \mathrm{CF}^\bullet(L_V, F_q)$ agrees with Example 5.0.1 at valuation less than λ_1:

$$m^2(x_I^{01} \otimes x_J^1) \equiv \begin{cases} T^{\lambda_0} x_{I\cup J}^{01} & \text{if } I \cap J = \varnothing \text{ and } I \cup J \subset \{0, \dots, n-k\}, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{mod } T^{\lambda_1}. \tag{11}$$

Let $\mathrm{CF}^\bullet(F_q, \Lambda_{\geq 0})$ and $\mathrm{CF}^\bullet(L_V, F_q, \Lambda_{\geq 0})$ be the filtered A_∞ algebra and bimodule where we use $\Lambda_{\geq 0}$ rather than Λ-coefficients. It follows that the map on chains

$$m^2\colon (x_\varnothing^{01}) \otimes \mathrm{CF}^1(F_q, \Lambda_{\geq 0}) \to \mathrm{CF}^1(L_V, F_q, \Lambda_{\geq \lambda_0})/\mathrm{CF}^1(L_V, F_q, \Lambda_{\geq \lambda_1})$$

surjects. Hence $\mathrm{CF}^\bullet(L_V, F_q, \Lambda_{\geq 0})$ as a right $\mathrm{CF}^\bullet(F_q, \Lambda_{\geq 0})$ module satisfies the criterion of Lemma B.3.1 and there exists $b \in \mathrm{CF}^\bullet(F_q)$ such that $\mathrm{HF}^0(L_V, (F_q, b)) \neq 0$.

To extend to the setting where L has a local system ∇_0, we simply require that F_q be equipped with a local system ∇_1 which agrees with ∇_0 on the torus spanned by the classes $\{c_1, \dots, c_{n-k}\}$. □

Remark 5.2.3 The constant λ_0 can be taken to zero provided that one works with a model of $\mathrm{CF}^\bullet(L_V, F_q)$ which allows for clean intersections between L_V and F_q; the proof of Lemma B.3.1 becomes slightly simpler in that setting. The pearly model developed by [12] allows for such configurations of Lagrangian submanifolds.

Corollary 5.2.4 *Let (L, ∇_0, b_0) and (F_q, ∇, b) be as above. Then*

$$\mathrm{HF}^\bullet((L, \nabla_0, b_0), (F_q, \nabla, b)) = \bigwedge_{i \in \{1, \dots, n-k\}} \Lambda\langle x_i \rangle.$$

Proof Again for expositional purposes, we assume that ∇_0 and ∇ are trivial local systems, assume that the multiplicity of the local model $\underline{V}$ is 1, and suppress the bounding cochain on L. On chains, the action of $\mathrm{CF}^\bullet(F_q, b)$ on $\mathrm{CF}^\bullet(L, (F_q, b))$ is a deformation of the action of $\mathrm{CF}^\bullet(F_q, b)$ on $\mathrm{CF}^\bullet(L, (F_q, b))$. By using an argument on filtration similar to the one above, the map

$$m^2(x_\varnothing, -)\colon \mathrm{CF}^\bullet((F_q, b)) \to \mathrm{CF}^\bullet(L, (F_q, b))$$

is a surjection. As every class in $\mathrm{CF}^\bullet((F_q, b))$ is closed and we've proven that $x_\varnothing$ is closed, every element in $\mathrm{CF}^\bullet(L, (F_q, b))$ is closed. This proves that $m^1_{\mathrm{CF}^\bullet(L,(F_q,b'))} = 0$, and that $\mathrm{HF}^\bullet(L, (F_q, b)) = \mathrm{CF}^\bullet(L, (F_q, b)) = \bigwedge_{i\in\{1,\dots,n-k\}} \Lambda\langle x_i \rangle$. □

Corollary 5.2.5 *Let (L_V, b) be an unobstructed geometric Lagrangian lift of V. Then $V^{(0)} \setminus V^{(1)} \subset \mathrm{TropA}(L_V, b)$.*

If we assume (8), this immediately follows.

Proof The proof of Lemma 5.2.2 can be modified to replace the bounding systems everywhere with local systems. The needed observation is that the map from the space of local systems

$$H^1(F_q, U_\Lambda) \to \mathrm{CF}^1(L_V, F_q, \Lambda_{\geq \lambda_0})/\mathrm{CF}^1(L_V, F_q, \Lambda_{\geq \lambda_1}), \quad \nabla \mapsto m^1_{(F_q,\nabla)}(x_\varnothing),$$

is surjective. The same argument as in Lemma B.3.1 can be used to construct a local system term by term so that $m^1_{(F_q,\nabla)}(x_\varnothing) = 0$. See also [53, Proposition 5.13], which proves a similar statement for tropical curves using the implicit function theorem [1, Section 10.8]. □

6 B-realizability and unobstructedness

6.1 HMS for $(\mathbb{C}^*)^n$

6.1.1 Construction of the mirror space Given $\pi_A\colon X_A \to Q$ a Lagrangian torus fibration, there is a rigid analytic space X_B with a tropicalization map $\mathrm{TropB}\colon X_B \to Q$. As a set, X_B is the set of Lagrangian torus fibers equipped with a U_Λ local system,

$$X_B := \{(F_q, \nabla)\}$$

which comes with a map $\pi_B\colon X_B \to Q$ given by $(F_q, \nabla) \mapsto q$. When $Q = \mathbb{R}^n$, the points of X_B are in bijection with $(\Lambda^*)^n$. We now describe, following [4; 17], how this can be realized as the set of points of a rigid analytic space. We also recommend the discussion in [53, Section 5.1].

The *Tate algebra* in n-variables over Λ is the set of formal power series

$$T_n := \left\{ \sum_{A \in \mathbb{Z}^n} f_A z^A \;\middle|\; f_A \in \Lambda,\ \mathrm{val}(f_A) \to \infty \text{ as } |A| \to \infty \right\},$$

which is equipped with the *sup-norm*

$$\left\| \sum_{A \in \mathbb{Z}^n} f_A z^A \right\| := \max_A |f_A| \geq 0.$$

We note that the maximal ideals of T_n are $\{(f_1, \dots, f_n) \mid \mathrm{val}(f_i) \leq 1\}$.

To build our spaces we will glue together *affinoid algebras*, which are quotients of the Tate algebra. The affinoid algebras we will look at are the polytope algebras. Given a bounded rational polytope $P \subset \mathbb{R}^n$, define

$$\mathcal{O}_P := \left\{ \sum_{A \in \mathbb{Z}^n} f_A z^A \;\middle|\; \mathrm{val}(f_A) + Ap \to \infty \text{ as } \|A\| \to \infty \text{ for all } p \in P \right\}.$$

This is the affinoid algebra. The elements of this affinoid algebra have the property that they converge when evaluated on $z \in (\Lambda^*)^n$ with $\mathrm{val}(z) \in P$. Furthermore, the points of $\mathcal{O}_P$ are seen to be in bijection with the points of $\pi_B^{-1}(P)$. When Q is compact X_B can be covered by finitely many sets $\pi_B^{-1}(P)$, giving X_B the structure of a rigid analytic space.

6.1.2 From Lagrangians to coherent sheaves Due to the limitations on currently existing constructions for Fukaya categories, we do not have homological mirror symmetry for a category of nonexact Lagrangian submanifolds in $(\mathbb{C}^*)^n$. However, different aspects of this homological mirror symmetry statement exist in the literature with strengthened hypotheses.

- The family Floer functor associates to a compact Lagrangian torus fibration $\pi_A\colon X_A \to Q$ a rigid analytic space $X_B \to Q$ whose points are in bijection with Lagrangian tori $F_q \subset X_A$ equipped with a U_Λ local system. Furthermore, [5, Theorem 2.10] constructs a faithful A_∞ functor $\mathcal{F}\colon \mathrm{Fuk}^{\mathrm{taut}}(X_A) \to \mathrm{Perf}(X_B)$. Here $\mathrm{Fuk}^{\mathrm{taut}}(X_A)$ is the Fukaya category of tautologically unobstructed Lagrangian submanifolds.

- In the exact setting, we have a complete proof of homological mirror symmetry for $(\mathbb{C}^*)^n$. The proof comes from recasting a section $L(0)$ of the fibration $\pi_A\colon (\mathbb{C}^*)^n \to Q$ as a cotangent fiber in T^*T^n, which is known to generate the exact Fukaya category. A computation shows that the A_∞ algebra $\mathrm{CF}^\bullet(L(0), L(0))$ is homotopy equivalent to $\hom(\mathcal{O}_{\mathbb{C}^n}, \mathcal{O}_{\mathbb{C}^n})$. In fact, we have a little bit more: it is known that the partially wrapped Fukaya category is mirror to the derived category of coherent sheaves on a toric variety [2; 37].

For this paper, we will only compute $\mathrm{CF}^\bullet((L_V, b), (F_q, \nabla))$, which means that we need substantially less than an HMS functor of [5].

Theorem 6.1.1 [4] *Consider the Lagrangian torus fibration $\pi_A\colon X_A \to Q$, with Q* ***compact****. From this data we can construct a rigid analytic mirror space X_B whose points z are in bijection with pairs (F_q, ∇). For any* ***tautologically unobstructed*** *Lagrangian brane $L \subset X_A$, there exists a coherent sheaf $\mathcal{F}(L)$ on X_B such that*

$$\hom(\mathcal{F}(L), \mathcal{O}_z) = \mathrm{HF}^0(L, (F_q, \nabla)).$$

Assumption 6.1.2 Theorem 6.1.1 still holds under the following weakened assumptions:

(∗) The base is allowed to be $Q = \mathbb{R}^n$, and we additionally require that the Lagrangian L be monomially admissible.

(∗∗) The Lagrangian L is allowed to be unobstructed by bounding cochains, in which case there exists a coherent sheaf $\mathcal{F}(L, b)$ on X_B such that

$$\hom(\mathcal{F}(L, b), \mathcal{O}_z) = \mathrm{HF}^0((L, b), (F_q, \nabla)).$$

We now discuss the difficulties, expectations, and progress of proving the assumption. The primary difficulties arise from noncompactness and unobstructedness.

Noncompactness presents three immediate issues. The first is Gromov compactness. We expect that after one places appropriate taming conditions on our Lagrangian submanifolds (as in Appendix A) the moduli spaces needed to construct the family Floer functor can be given appropriate compactifications.

The second more difficult issue regards the role that wrapping plays in computing the Floer cohomology between two noncompact Lagrangians. In the exact setting, the morphism space between two Lagrangians is computed as the limit of $\mathrm{CF}^\bullet(\phi^i(L_0), L^i)$, where ϕ^i is a wrapping Hamiltonian, and the limit is taken over continuation maps. In the nonexact setting, these continuation maps have a nonzero valuation, and only have inverses defined over the Novikov field (with possibly negative valuation). To our knowledge, this version of the Fukaya category has not been constructed. However, since for our application we only need to compute Floer cohomology against Lagrangian torus fibers (which are compact), we can ignore the issues of the wrapping Hamiltonian.

Finally, there is the issue of coherence of $\mathcal{F}(L,b)$. Here we use the monomial admissibility condition. We recall the proof of coherence when Q is compact. The sheaf $\mathcal{F}(L,b)$ is constructed by defining it over affinoid domains on the mirror, which correspond to convex domains $U \subset Q$. The convex domain U is "small enough" if there exists a Hamiltonian isotopy of L such that it intersects all Lagrangian torus fibers F_q with $q \in U$ transversely. Over each small enough U, the sheaf is computed by $\mathrm{CF}^\bullet((L,b),(F_q,\nabla)) \otimes \mathcal{O}_U$, where $\mathcal{O}_U$ is the affinoid ring of the affinoid domain $X_{U,B}$ associated to the convex domain U. Since $\mathrm{CF}^\bullet((L,b),(F_q,\nabla))$ is finitely generated, and (in the compact setting) we can cover Q with finitely many such U, we obtain that the mirror sheaf is coherent. If we drop the condition of Q being compact, and impose the condition that L is monomially admissible, we can still cover Q with a finite set of convex (possibly noncompact) small enough domains $U \subset Q$ by using invariance of the Lagrangian submanifold under symplectic flow in the direction of the monomial ray over each monomial region.

We now remark upon the difficulty of unobstructedness. Remark 1.1 of [5] states that the "tautologically unobstructed" hypothesis for construction of the family Floer functor is technical in nature, and it is expected that the family Floer functor should carry through using unobstructed Lagrangian submanifolds. As we do not require functoriality, such an adaptation of family Floer cohomology to the Charest–Woodward model would not require studying moduli spaces beyond those already studied in [12]. We believe the main items left to prove for this construction are the following:

- It must be shown that "Fukaya's trick" for pulling back a perturbation datum between Lagrangian fibers over sufficiently small convex domains can be worked out in the more technically challenging setting of domain-dependent perturbations. This does not appear to present a problem when working with the setup of [12].
- It must be shown that "homotopies of continuation maps" exist in the version of Lagrangian intersection Floer cohomology one is working with. In [12], continuation maps are constructed using holomorphic quilts. There is also an additional challenge of showing that one can construct homotopies of continuation maps corresponding to changes in the choice of stabilizing divisor.

Finally, we note that the work in progress of Abouzaid, Groman, and Varolgunes generalizing [24; 59] to the Fukaya category will prove homological mirror symmetry for unobstructed Lagrangian submanifolds of $(\mathbb{C}^*)^n$, giving us Assumption 6.1.2.

Remark 6.1.3 A different approach that would bypass family Floer theory would be to expand homological mirror symmetry for toric varieties [2; 37] in the nonexact setting. This would involve developing [23] to the nonexact setting. While there is no clear obstruction to expanding the Liouville sector framework to include obstructed Lagrangian submanifolds that are geometrically bounded, there are at least two technical and challenging issues that would need to be overcome. Many of the arguments used in [23] would have to be carefully redone by replacing geometric bounds obtained by energy and exactness with other methods for bounding holomorphic disks. In fact, these techniques are already employed in a limited capacity in [23] for the proof of the Künneth formula (as products of cylindrical Lagrangians are usually not cylindrical). The second issue is understanding how to incorporate curvature into the homological algebra constructions employed by [23]. One possible workaround would be to first construct the partially wrapped precategory of Lagrangian branes that are equipped with bounding cochains (which is an uncurved filtered A_∞ precategory) and localize at continuation maps to construct the partially wrapped category. This already requires some care, as it is not immediately clear how the filtration would play a role in this localization (the continuation maps would have positive energy, so there may be convergence issues). The second, more ambitious approach would be to attempt to construct a "curved partially wrapped Fukaya category", by starting with a partially wrapped Fukaya precategory whose objects are (potentially obstructed) Lagrangian submanifolds. This second approach would require one to understand what a filtered A_∞ precategory is and also to construct localizations of these categories.

6.2 Unobstructed Lagrangian lift implies B-realizability

By employing [6] (with the possible extensions stated in Assumption 6.1.2) we can associate to each Lagrangian brane (L_V, b) a closed analytic subset of X_B:

$$Y(L_V, b) := \mathrm{Supp}(H^0(\mathcal{F}(L_V, b))).$$

Corollary 6.2.1 *Consider the Lagrangian torus fibration $\pi_A\colon (\mathbb{C}^*)^n = X_A \to Q$ and a tropical subvariety $V \subset Q$. Suppose that (L_V, b) is a Lagrangian brane lift of V. Then*

- *(L_V, b) is an A-realization of V in the sense that* $\mathrm{TropA}(L_V, b) = V$,
- *V is B-realizable.*

Proof By Assumption 6.1.2, $\mathrm{TropA}(L_V, b) = \mathrm{TropB}(Y(L_V, b))$. In Corollary 5.2.5, we proved that $V^{(0)} \subset \mathrm{TropA}(L_V, b) \subset V$. Since $Y(L_V, b)$ is a closed analytic subset, $\mathrm{TropB}(Y(L_V, b))$ is the union of closed rational polyhedra in $N_\mathbb{R}$ [27, Proposition 5.2]. As a result, TropB is closed and contains $\overline{V^{(0)}} = V$. It follows that $Y(L_V, b)$ is a closed analytic subset of X_B which realizes V. □

Corollary 6.2.2 *Assuming Assumption 6.1.2(∗)–(∗∗), let V be a smooth hypersurface or a smooth genus-0 tropical curve in $\mathbb{R}^n$. Then V is B-realizable.*

Corollary 6.2.3 *Assuming Assumption 6.1.2(∗∗), let V be a smooth tropical hypersurface of a tropical abelian variety $Q = \mathbb{R}^n / M_\mathbb{Z}$. Then V is B-realizable.*

Corollary 6.2.4 ***Without*** *assuming any portion of Assumption 6.1.2, let V be a 3-valent tropical curve in a tropical abelian surface Q. Then V is B-realizable.*

Proof The condition of 3-valency comes from using

- [35] to build an affine dimer model associated to each 3-valent vertex, and
- [33] to build tropical Lagrangian lifts from a dimer model.

We now address why Assumption 6.1.2 may be dropped. Since Q is a tropical abelian surface (and is therefore compact), the symplectic manifold X_A is compact. Since the Lagrangian lift L_V is graded of dimension 2, it is tautologically unobstructed for a generic choice of almost complex structure (as Maslov index 0 disks appear in expected dimension -1). $\square$

6.3 Nonplanar tropical curves do not have tautologically unobstructed lifts

Even in the setting where V is a genus-0 tropical curve, it is rare for the Lagrangian lift L_V to be a tautologically unobstructed Lagrangian submanifold.

Before constructing an example, we observe that the valuations of the "big-strips" in Lemma 5.2.2 are dictated by the radius of the neighborhood U_q that we can construct around the point q which is disjoint from $V^{(1)}$. In particular, this can be applied to [53, Proposition 5.10] to show that tautologically unobstructed Lagrangian lifts of tropical curves have supports that extend to an appropriate toric compactification of the mirror algebraic torus.

Proposition 6.3.1 *Let Σ be a fan. Suppose that V is a tropical curve with semi-infinite edges in the directions of the rays of Σ. Suppose the fan of Σ has the additional property that $\langle \alpha, \beta\rangle \leq 0$ for all 1-dimensional cones $\alpha \neq \beta$ and $\langle -,-\rangle$ is the standard inner product. Then $Y((L_V, b), 0)$ compactifies to a rigid analytic space inside $X_B(\Sigma)$, the rigid analytic toric variety with fan Σ.*

Proof We first describe the rigid analytic structure on $X_B(\Sigma)$ given by [49]. From [47], the space $X_B(\Sigma)$ comes with a fibration $\mathrm{TropB}\colon X_B(\Sigma) \to Q(\Sigma)$, which is a partial compactification of Q; see [49, Definition 3.6]. Rabinoff then covers $X_B(\Sigma)$ with charts given by the max-spec of affinoid algebras.

Let $P_\sigma \subset Q$ denote a convex set which can be written as $P' + \sigma$ for $\sigma \in \Sigma$ and some convex compact polytope $P' \subset Q$. Associated to P_σ is a subset $\overline{P}_\sigma \subset Q(\Sigma)$, and an affinoid algebra

$$\mathcal{O}_{P_\sigma} := \left\{ \sum_{A \in (\sigma^\vee \cap \mathbb{Z}^n)} f_A z^A \;\middle|\; \mathrm{val}(f_a) + ap \to \infty \text{ as } \|A\| \to \infty \text{ for all } p \in P_\sigma \right\}.$$

We can cover $X_B(\Sigma)$ with charts given by the max-spec of $\mathcal{O}_{P_\sigma}$ (which covers $\mathrm{TropB}^{-1}(\overline{P}_\sigma)$).

We now unpack what it means for a Lagrangian submanifold (L, b) constructed via family Floer theory to give a coherent sheaf $\mathcal{F}((L, b))$ on the rigid analytic space $X_B(\Sigma)$. In the family Floer construction, for a sufficiently small convex polytope P in the base of Q, one takes a Hamiltonian perturbation L_P of L so that $L_P|_{\pi_A^{-1}(P)}$ is a disjoint set of flat sections of $\pi_A^{-1}(P) \to P$, and that the bounding cochain is similarly parallel to the flat section. As a result, we may identify the chains $\mathrm{CF}^\bullet((L_P, b), (F_q, \nabla))$ for all $q \in P$. Additionally, for $q \in P$, one can appropriately choose almost complex structures (using Fukaya's trick) so that the moduli spaces of strips contributing to the differential on $\mathrm{CF}^\bullet((L, b), (F_q, \nabla))$ does not depend on q. Because the bounding cochain on L is parallel to the flat section, the contribution of b to the differential on $\mathrm{CF}^\bullet((L, b), (F_q, \nabla))$ does not depend on $q \in P$. As a consequence, the dependence of the structure coefficients $\langle m^1_{(L,b),(F_q,\nabla)}(x), y\rangle$ on (F_q, ∇) factors through the flux homomorphism. Pick a basepoint x_0 on F_q and for each $x \in L \cap F_q$ a path γ_x from x_0 to x. By identifying (F_q, ∇) with a point $z \in \mathrm{TropB}^{-1}(P)$, we obtain that

$$\langle m^1_z(x), y\rangle = \sum_{a \in H_1(F_q)} c_a z^a,$$

where c_a is the area and local-system weighted count of pseudoholomorphic strips u such that

$$[\gamma_x \partial_{F_q} u \gamma_y^{-1}] = a.$$

The Lagrangian L defines a complex of sheaves $\mathcal{F}(L)$ over $(X_B|_P)$ if these structure coefficients belong to $\mathcal{O}_P$. The restriction maps compose up to homotopy of the chain complex. This is proven using the reverse isoperimetric inequality to bound the area of holomorphic strips u (which govern convergence) below by the winding of the F_q boundary component of U (which governs the exponent appearing in z^a). To obtain a coherent sheaf of complexes on X_B, one must be able to cover Q with finitely many sufficiently small sets P. When Q is compact, this is always possible. In the setting we study, we must take some of the sets P to be of the form P_σ in order to construct a finite cover.

We now perform this construction for L_V our tautologically unobstructed Lagrangian lift of a tropical curve V. Let e be a semi-infinite edge of V pointing in the α direction, where $\alpha \in \Sigma$ is a 1-dimensional cone. Then there exists a P_α such that $V|_{P_\alpha}$ is a 1-dimensional ray. Since $\langle \alpha, \beta\rangle < 0$ for all 1-dimensional rays $\beta \neq \alpha$, the projection $\chi^\alpha : X_A \to \mathbb{C}$ given by the α-monomial has the property that the $\chi^\alpha|_{L_V} : L_V \to \mathbb{C}$ fibers over a real ray outside of a compact set. This is the main input needed in [53, Proposition 5.10]

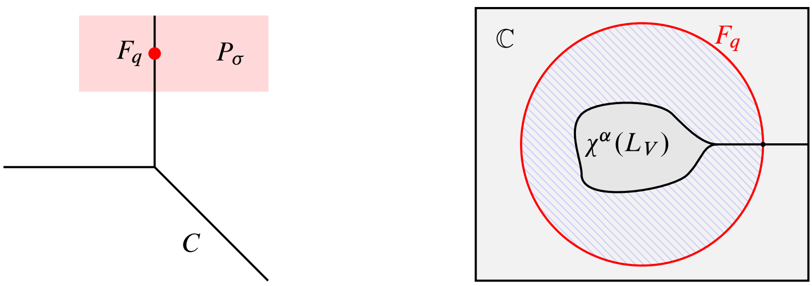


Figure 9: Monomial admissibility forces strips with $\langle \partial u, \alpha\rangle > 0$ to have large symplectic area. Left: condition on a tropical curve. Right: projection on $\chi^\alpha : X_A \to \mathbb{C}$.

to show that the differential on $\mathrm{CF}^\bullet(L_V, (F_q, \nabla))$ is of the form $\sum_{a\in H_1(F_q),\langle a,\alpha\rangle\geq 0} c_a z^a$, and that $\mathrm{val}(c_a)+ap\to\infty$ as $\|A\|\to\infty$ for all $p\in P_\alpha$. It follows that $\langle m_z^1(x), y\rangle \in \mathcal{O}_{P_\alpha}$.

We can choose a finite cover of Q by sets of the form P_σ so that $\pi_A(L_V)\subset P_\sigma$ if and only if $|\sigma|\leq 1$. It follows that $\mathcal{F}(L_V)$ defines a sheaf on $X_B(\Sigma)$. □

In the setting above (where L_V is tautologically unobstructed and equipped with the trivial local system), the above computation not only shows that $\mathcal{F}(L_V)$ extends to $X_B(\Sigma)$ but also shows that we can compute the points in the compactifying locus. For a semi-infinite edge e, let $P_\alpha = P+\langle\alpha\rangle$ be a convex polytope whose only intersection with V is along the edge e. Without loss of generality, we will assume that the edge e is of the form $(t,0,\dots,0)\subset Q=\mathbb{R}^n$, with t tending to ∞. We can write the max-spec of P_α as

$$\{(z_1,\dots,z_n)\in\Lambda\times(\Lambda^*)^{n-1}\mid \mathrm{val}(z_1,\dots,z_n)\in\overline{P}_\alpha\subset(\mathbb{R}\cup\infty)\times\mathbb{R}^{n-1}\}.$$

We prove that $(0,1,\dots,1)\in\mathrm{Supp}(\mathcal{F}(L_V))$. The $\langle a,\alpha\rangle=0$ terms of $\langle m_z^1(x_\varnothing), x_I\rangle$ agree with holomorphic strips for the differential on $\mathrm{CF}^\bullet(L_{\underline{e}}, (F_q,\nabla))$, so we can write

$$\langle m_z^1(x_\varnothing), x_I\rangle = (1-z^{\langle I,a\rangle}) + \sum_{a\in H_1(F_q),\langle a,\alpha\rangle>0} c_a z^a.$$

When we have a sequence of points $\{z^k\}_{k\in\mathbb{N}}$ with the property that $m^1_{z^k}(x_\varnothing)=0$ (ie $z^k\in\mathrm{Supp}(\mathcal{F}(L_V))$) and $\lim_{k\to\infty}\mathrm{val}(z_1^k)=\infty$ (so that the limit belongs to the compactifying toric divisor), the above equation states that $\lim_{k\to\infty}\mathrm{val}(z_i^k)=1$ for all $i\neq 1$. We conclude that the closure of $\mathrm{Supp}(\mathcal{F}(L_V))$ inside of $X_B(\Sigma)$ contains the point $(0,1,\dots,1)$.

We now construct an example of a Lagrangian brane lift of a tropical curve that is unobstructed, but not tautologically unobstructed.

Example 6.3.2 Consider the tropical line $V_c\in\mathbb{R}^3$ drawn in Figure 10. The tropical line V_c has two pants centered at the points $(0,0,0)$ and $(-c,-c,0)$, whose legs at $(0,0,0)$ point in the directions

$$e_1=\langle 1,0,0\rangle,\quad e_2=\langle 0,1,0\rangle,\quad e_c=\langle -1,-1,0\rangle,$$

and whose legs at $(-c,-c,0)$ point in the directions

$$e_3=\langle 0,0,1\rangle,\quad e_4=\langle -1,-1,-1\rangle,\quad -e_c=\langle 1,1,0\rangle.$$

We prove that L_{V_c} bounds a holomorphic disk for all but at most 1 value of c.

Assume for contradiction that for all values of c the Lagrangian submanifold L_{V_c} is tautologically unobstructed, and requires no bounding cochain. Then the Lagrangians L_{V_c} satisfy the conditions of Proposition 6.3.1, so each $Y_{V_c}:=\mathrm{Supp}(\mathcal{F}(L_{V_c}))$ compactifies to give a curve inside of $\mathbb{P}^3$. Since this curve intersects each of the toric divisors at a single point, we conclude that every Y_{V_c} is a line in $\mathbb{P}^3$. Furthermore, every one of these lines contains the points $(1:0:0:0)$ and $(0:1:0:0)$ in $\mathbb{P}^3$. Since a line in $\mathbb{P}^3$ is determined by two points, this implies that $Y_{V_c}=Y_{V_{c'}}$. However, as $V_c\neq V_{c'}$, they cannot be realized by the same subvariety, a contradiction.

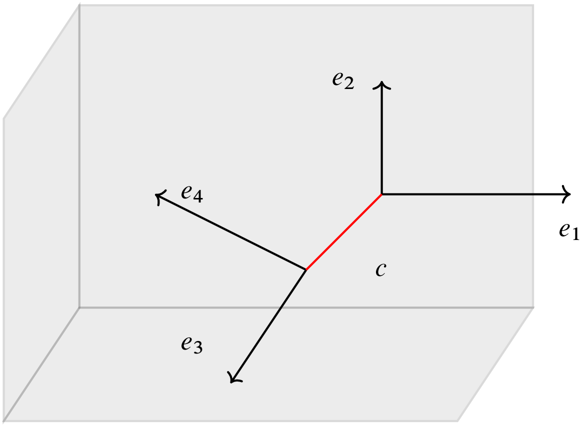


Figure 10: A tropical line V_c. The Lagrangian lift L_{V_c} necessarily bounds a holomorphic disk; we conjecture that the projection to the base of the Lagrangian torus fibration of this holomorphic disk lives over the red edge and has area controlled by the affine length c.

This doesn't contradict the realizability of V_c. Indeed, by Corollary 4.3.3, the bounding cochain on L_{V_c} need only be supported on three of the four legs of L_{V_c}. However, the above argument shows that one cannot construct a bounding cochain for L_{V_c} which restricts to zero on the two semi-infinite edges which share a vertex (which implies that the bounding cochain cannot be zero).

Using mirror symmetry, we can "back solve" for the valuation of the holomorphic disk, which necessitates the use of a bounding cochain on L_{V_c}. We may assume that the bounding cochain has trivial restriction to the e_1 edge. It follows that the tropical line Y_{V_c} may intersect toric divisors at the points $(0, 1, 1)$, $(1 + \exp(b_1), 0, 1 + \exp(b_3))$ and $(z^{-c} + \exp(c_1), z^{-c} + \exp(c_2), 0)$. Since these have to satisfy the equation of a line, there exists t such that

$$(1-t)(0,1,1) + t(\exp(b_1), 0, \exp(b_2)) = (z^{-c}(\exp(c_1)), z^{-c}(\exp(c_2)), 0).$$

From examining the third term, $t = (1-\exp(b_2))^{-1}$, we already see that $b_2 \neq 0$. From examining the third term,

$$(1-\exp(b_2))^{-1}\exp(b_1) = z^c(\exp(c_1)),$$

from which we see that $\operatorname{val}(b_2) = c$. From this, we conclude that there exists a pseudoholomorphic disk of energy c on L_{V_c}.

6.4 Speculation on Speyer's well-spacedness criterion

Corollary 6.2.1 proves the forward direction of Conjecture 1.1.1. To investigate the reverse direction, we look at an example of a nonrealizable tropical curve. In [40] it was observed that every cubic curve in $\mathbb{CP}^3$ is planar (Figure 11, left). Consequently, the example drawn in Figure 11, right — a tropical cubic which is not contained within any tropical plane — cannot arise as the tropicalization of any curve in $\mathbb{CP}^3$.

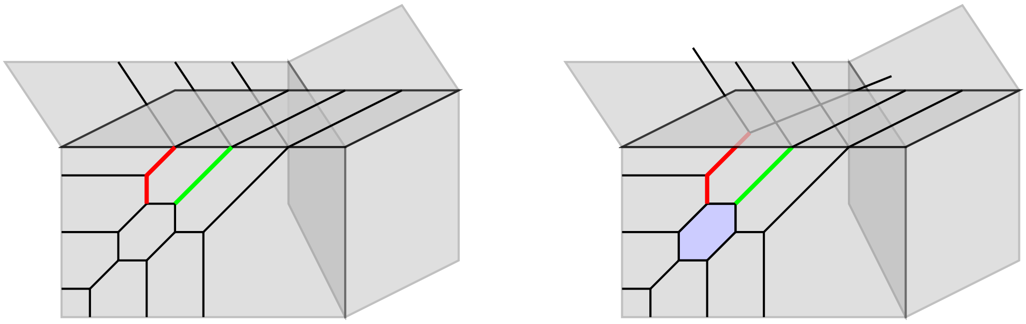


Figure 11: Left: A well-spaced tropical curve. The affine lengths of the red and green match. Right: A nonrealizable tropical curve. The affine length of the green segment is uniquely minimal among all paths from the cycle to the nonlinearity locus of the curve.

Corollary 6.4.1 *Let V be the tropical curve from* [40, Example 5.12]. *Then the standard lift of L_V is an obstructed Lagrangian.*

A general criterion for understanding this phenomenon was stated in [55]:

Theorem (Speyer's well-spacedness) *Let V be a genus-*1 *tropical curve whose cycle is contained within a linear subspace H. Let $d_1, \dots, d_k$ be the affine lengths of paths along the edges of V to the boundary of $V \cap H$. If the minimal distance occurs at least twice*, *the curve V is realizable.*

We now speculate on how Speyer's well-spacedness criterion can be understood in terms of holomorphic disks with boundary on L_V. For L_V to be unobstructed, it is necessary for the lowest-energy terms in m^0 to be nullhomologous. In particular, the set

$$\Big\{u \;\Big|\; \omega(u) \leq \min_{0 \neq [\partial u'] \in H^2(L,M)} \omega(u')\Big\}$$

of minimal-area nonnullhomologous disks must contain at least two elements. This matches the "two minimal distance" criterion of Speyer's well-spacedness theorem.

In [32], we saw that tropical cycles on $W \subset \mathbb{R}^2$ are related to nonregular Maslov index 0 disks with boundaries on the Lagrangian lifts L_W; it was speculated that these Maslov index 0 disks could appear regularly if they were glued onto a regular holomorphic disk or strip. In Example 6.3.2 we saw that the Lagrangian brane lift of a small neighborhood of the green segment in Figure 11, right, must have a regular disk with energy given by the affine length of the edge.

In the example given by Figure 11, right, we conjecture that there are regular holomorphic disks with boundaries on L_V whose projections under the moment map are:

- the union of the blue hexagon (a nonregular disk) and green path (a regular disk), call this speculative disk u_1, and
- the union of the blue hexagon (a nonregular disk) and red path (a regular disk), call this speculative disk u_2.

Using that the area of homology classes of disks with boundary on L_V correspond to affine length, the disks u_1 and u_2 have matching symplectic area if the affine lengths of the green and blue paths match. In this case, the homology class of $[\partial u_1]-[\partial u_2]$ doesn't wrap around the portion of the homology of L_V which arises from V, and by a similar argument to that used in Corollary 4.3.3 we see that $[\partial u_1]-[\partial u_2]\subset H_1(L_{V_\infty^{(0)}})$. We could then apply the methods used in the proof of Corollary 4.3.3 to conclude that L_V is unobstructed.

In the event that $\omega(u_2)$ is uniquely minimal, the boundary of $\partial(u_2)$ is a nontrivial homology class in $H_1(L_V)$, suggesting that the contribution to $m^0\in \mathrm{CF}^\bullet(L_V)$ is a nonremovable obstruction.

6.5 Deformations, superabundance, and not-wide

6.5.1 Geometric deformations of L and $(V,\mathcal{L})$ Given $V\subset\mathbb{R}^n$ a tropical subvariety, a Lagrangian L_V should correspond to a lift of V equipped with a line bundle. In this section, we examine how the deformations of L_V up to Hamiltonian isotopy match deformations of a tropical curve equipped with a line bundle $(V,\mathcal{L})$.

Given a fixed tropical line bundle $\mathcal{L}\to V$ we can identify deformations of $\mathcal{L}$ with $H^1(V,\mathbb{R})$; this is because deformations of invertible locally integral affine functions from U to $\mathbb{R}$ correspond to constant differences. Similarly, the deformations of $V\subset\mathbb{R}^n$ as a smooth tropical subvariety can be computed sheaf-theoretically. We choose a cover conducive to this computation. For each $v\in V$, let $\mathrm{star}(v)$ be the union of the edges that contain v. We allow v to be a leaf (at the end of a semi-infinite edge). Then the $\mathrm{star}(v)$ form a cover of V, with $\mathrm{star}(v)\cap\mathrm{star}(w)=\underline{V}_{vw}$ whenever vw is an edge. There are two types of vertices v that we must consider:

- If v is an internal vertex, then the deformations of $\mathrm{star}(v)$ are identified with the integral affine space $Nv=T_v\mathbb{R}^n=\mathbb{R}^n$.
- If v_∞ is a boundary vertex incident to edge e, then the deformations of $\mathrm{star}(v_\infty)$ are identified with the integral affine space $\mathbb{R}^{n-1}$ perpendicular to the semi-infinite edge attached to v_∞.

Over each edge e, the deformations of the tropical curve are given by the normal bundle to e. In summary, let Def_V be the sheaf of deformations of the tropical embedding of V, and let $\mathrm{Def}_{\mathcal{L}}$ be the deformations of a fixed line bundle $\mathcal{L}$ over V. We have:

$$\mathrm{Def}_V(\mathrm{star}(v))=\mathbb{R}^n,\quad \mathrm{Def}_V(\mathrm{star}(v_\infty))=e^\perp_{v_\infty},\quad \mathrm{Def}_V(\mathrm{star}(e))=e^\perp.$$

For compact Lagrangian L, the infinitesimal deformations of L up to Hamiltonian isotopy are described by classes in $H^1(L,\mathbb{R})$. Since L_V is noncompact, we only consider the *admissible* deformations of noncompact L_V which preserve the condition in Definition 3.1.1. Let $\Omega^1_{\mathrm{admis}}(L_V,\mathbb{R})$ be the 1-forms on L_V with the property that

- for each monomial region U_α, the 1-form $\eta|_{L_V\cap U_\alpha}$ is invariant under the flow in the α-direction,
- $\eta(\alpha)=0$.

We let $\Omega^0_{\text{admis}}(L_V, \mathbb{R})$ be those functions which, outside of a compact set, are invariant under the flow in the α direction of the corresponding monomial region from Definition 3.1.1.

We can similarly decompose L_V into sets $L_{\text{star}(v)}$, which we will take to be

- the standard Lagrangian pair of pants when v in an interior vertex such that $L_{\text{star}(v)} \cap L_{\text{star}(w)} = L_{\underline{V}_{vw}}$, in which case $\Omega^i_{\text{admis}}(L_V, \mathbb{R}) = \Omega^i(L_V, \mathbb{R})$,
- a noncompact cylinder extending to the boundary whenever w is a vertex at a noncompact edge.

We then compute $H^1(\Omega^\bullet_{\text{admis}}(L_V))$. The cohomology is the same as the first cohomology of the total complex; the first page in the spectral sequence is

$$\begin{array}{ccc}
\bigoplus_{v\in V} H^0(\Omega^\bullet(L_v)) & \bigoplus_{v\in V} H^1(\Omega^\bullet(L_v)) & \cdots \\
\downarrow & \downarrow & \\
\bigoplus_{e\in V} H^0(\Omega^\bullet(L_e)) & \bigoplus_{e\in V} H^1(\Omega^\bullet(L_e)) & \cdots \\
\downarrow & \downarrow & \\
0 & 0 &
\end{array}$$

We now start to identify these with deformations of tropical curves,

$$\begin{aligned}
\mathrm{Def}_V(\mathrm{star}(v)) &= H^1(\Omega^\bullet(L_v)), & \mathrm{Def}_V(\mathrm{star}(e)) &= H^1(\Omega^\bullet(L_e)), \\
\mathbb{R} &= H^0(\Omega^\bullet(L_v)), & \mathbb{R} &= H^0(\Omega^\bullet(L_e)),
\end{aligned}$$

turning the first page of the spectral sequence into

$$\begin{array}{ccc}
\bigoplus_{v\in V} \mathbb{R} & \bigoplus_{v\in V} \mathrm{Def}_V(e) & \cdots \\
\downarrow & \downarrow & \\
\bigoplus_{e\in V} \mathbb{R} & \bigoplus_{e\in V} \mathrm{Def}_V & \cdots \\
\downarrow & \downarrow & \\
0 & 0 &
\end{array}$$

The spectral sequence for $H^1(\Omega^\bullet(L))$ converges at the second page for this covering, so

$$H^1(\Omega^\bullet(L_V)) = H^0(V, \mathrm{Def}_V)) \oplus H^1(V, \mathbb{R}) = H^0(V, \mathrm{Def}_V) \oplus H^0(V, \mathrm{Def}_{\mathcal{L}}).$$

In general, understanding the moduli space of Lagrangian submanifolds isotopic to L_V modulo Hamiltonian isotopy is a difficult question. In the setting of Lagrangian torus fibrations, there is a smaller class of isotopies that we can hope to understand. We say that a Lagrangian isotopy $i_t \colon L_V \to X_A$ is a fiberwise isotopy if $\pi_A(i_t(q))$ is constant for all $q \in L_V$.

Conjecture 6.5.1 *Let L_V be a homologically minimal Lagrangian lift of a tropical curve V. Then the subspace of $H^1(L_V, \mathbb{R})$ arising from the flux classes of fiberwise Lagrangian isotopies is identified with $H^0(V, \mathrm{Def}_{\mathcal{L}})$. Additionally,*

$$\{\text{fiberwise isotopies}\}/\{\text{fiberwise Hamiltonian isotopies}\} \simeq H^1(V, \mathrm{Aff}^*_V)_0,$$

where $\mathrm{Aff}^*_V(U)$ *is the sheaf of invertible locally integral affine functions from* U *to* $\mathbb{R}$*, and* $H^1(V, \mathrm{Aff}^*_V)_0$ *is the connected component of the group which contains the identity.*

Previous work of Mikhalkin [43] identifies the tropical Picard group of V with $H^1(V, \mathrm{Aff}^*_V)$. Mirror symmetry, therefore, identifies fiberwise isotopies of a tropical Lagrangian L_V with modifications of the line bundle on the mirror curve Y_V.

6.5.2 Not-wide and superabundance A tropical curve is called *superabundant* if the space of deformations Def_V has a higher dimension than the expected dimension of deformations of the B-realization. Superabundance is a computable criterion that indicates that a curve may not be realizable. For example, the tropical curve examined in Section 6.4 is a superabundant curve. It is known in certain cases [14] that nonsuperabundant implies realizable.

In symplectic geometry there are two ways to make sense of deformations of Lagrangian submanifolds. The first is the deformation of geometric Lagrangian submanifolds up to Hamiltonian isotopy. The infinitesimal deformations of Lagrangian submanifolds modulo Hamiltonian isotopy are given by $H^1(L, \mathbb{R})$. The second is the component of the moduli space of objects at L [58]. The tangent space to this moduli space is $\mathrm{HF}^1(L)$. We note that as $\mathrm{CF}^\bullet(L)$ is a deformation of $C^\bullet(L)$, we have that $\dim \mathrm{HF}^1(L) \leq \mathrm{HF}^1(L)$. If $\dim \mathrm{HF}(L) = \dim H(L)$, then the Lagrangian L is called *wide*.

As the previous section identifies infinitesimal deformations of the pair $(V, \mathcal{L})$ with $H^1(L_V)$, we are led to conjecture:

Conjecture 6.5.2 *Let* V *be a smooth tropical curve, and let* L_V *be its Lagrangian lift. Then* V *is superabundant if and only if* L_V *is not wide.*

Appendix A The pearly model in symplectic fibrations

Given a compact, spin, and graded Lagrangian L inside of a rational compact symplectic manifold X, [12] constructs a filtered A_∞ algebra $\mathrm{CF}^\bullet(L, h, \mathcal{P}, D)$. In [12] it is assumed that the space X is compact. In this appendix, we outline how to extend [12] to the setting where X is noncompact and equipped with a potential function $W: X \to \mathbb{C}$, and L is a Lagrangian submanifold which is admissible with respect to W.

Definition A.0.1 Let X be a symplectic manifold, and let $W: X \to \mathbb{C}$ be a function. We say that W is a potential if there exists a compact subset $U \subset \mathbb{C}$ such that

- $W^{-1}(U)$ is compact, and
- the restriction $W: X \setminus W^{-1}(U) \to \mathbb{C} \setminus U$ is a symplectic fibration with compact fibers.

We say that a Lagrangian L is *W-admissible* if there exists $R \in \mathbb{R}$ such that $W(L) \cap \{z \mid |z| > R\} \subset \mathbb{R}_{>R}$.

Given a W-admissible L, we say that a Morse function $h\colon L \to \mathbb{R}$ is admissible if there exists $R' > R$ such that

$$W(\mathrm{Crit}(h)) \cap \{z \mid |z| > R\} \subset \{R'\},$$

and $\operatorname{grad} h$ points outwards from R' under the projection W.

Let $Y = W^{-1}(R')$. Given a W-admissible Lagrangian submanifold L, the restriction to the fiber $M := L \cap Y$ is a Lagrangian submanifold of Y. Because h points outwards along the collar $M \times \mathbb{R}_{>R'} \subset L$, the Morse complex $\mathrm{CM}^\bullet(L, h)$ is well defined. The compatibility of the Morse function with the potential function means that $h^+ := h|_M$ is a Morse function for M and that we have a map of A_∞ algebras

$$\underline{\pi}\colon \mathrm{CM}^\bullet(L, h) \to \mathrm{CM}^\bullet(M, h^+).$$

This should be interpreted as the pullback map of the inclusion of the boundary.

We show that [12] extends to the setting of W-admissible Lagrangian submanifolds:

Theorem A.0.2 *Let $W\colon X \to \mathbb{C}$ be a potential function. Let L be a W-admissible Lagrangian submanifold whose restriction to a large fiber is $M \subset Y = W^{-1}(t)$. Let $h\colon L \to \mathbb{R}$ and $h^+ := h|_M\colon M \to \mathbb{R}$ be admissible Morse functions. There exist*

- *stabilizing symplectic divisors $D_X \subset X$, $D_Y \subset Y$, and*
- *regular choices for perturbation systems $\mathcal{P}_L$ and $\mathcal{P}_M$ for L and M,*

such that the construction of [12] can be applied to give a well-defined A_∞ algebra $\mathrm{CF}^\bullet(L, h, \mathcal{P}_L, D_X)$. Furthermore, the choices of perturbations and divisors can be taken so that the projection on chains

$$\pi\colon \mathrm{CF}^\bullet(L, h, \mathcal{P}_L, D_X) \to \mathrm{CF}^\bullet(M, h^+, \mathcal{P}_M, D_Y)$$

is a Λ-filtered A_∞ algebra homomorphism.

The theorem consists of two statements: the construction of a pearly model of stabilized treed disks in the setting of Lagrangians which are admissible for a potential function, and the compatibility between the pearly model of total space of the fibration and the pearly model of the fiber. These are analogous to [34, Corollary C.4.2 and Theorem C.5.1] which handle the setting where $X = Y \times \mathbb{C}$ and $W\colon X \to \mathbb{C}$ is projection to the second factor. In this appendix, we prove that $\mathrm{CF}^\bullet(L, h, \mathcal{P}_L, D_X)$ is well defined; the existence of the projection $\pi\colon \mathrm{CF}^\bullet(L, h, \mathcal{P}_L, D_X) \to \mathrm{CF}^\bullet(M, h^+, \mathcal{P}_M, D_Y)$ is the same as the proof of [34, Theorem C.5.1].

To construct $\mathrm{CF}^\bullet(L, h, \mathcal{P}_L, D_X)$ one needs to

(1) construct a stabilizing divisor for X which is suitably compatible with the potential $W\colon X \to \mathbb{C}$,

(2) show that we can pick perturbations for the almost complex structure so that the map $W\colon X \to \mathbb{C}$ is holomorphic outside of a compact set, and

(3) prove that for such choices of perturbations the moduli spaces have appropriate Gromov compactifications.

Item (1): constructing a stabilizing divisor

Pick R sufficiently large so that outside of $U = B_R(0) \subset \mathbb{C}$ the Lagrangian submanifold L fibers over the positive real ray, and the map $X \setminus W^{-1}(U) \to \mathbb{C} \setminus U$ is a symplectic fibration. For $\theta \in [0, 2\pi]$ and $r \geq R$ we take a path

$$\gamma_{\theta,r}(t) = \begin{cases} Re^{i\theta(2t)} & t \in \left[0, \frac{1}{2}\right), \\ (R + (2t-1)(r-R))e^{i\theta} & t \in \left[\frac{1}{2}, 1\right), \end{cases}$$

which travels first in the angular, then radial direction from R to $re^{i\theta}$. For every path $\gamma(t) : I \to \mathbb{C}_{|z|>R}$ we have a symplectic parallel transport map $P_\gamma : Y_{\gamma(0)} \to Y_{\gamma(1)}$. Consider the monodromy $P_{\gamma_{2\pi,R}} : Y_R \to Y_R$ given by parallel transport around the loop $Re^{i\theta}$ in the positive direction. Pick a path of ω_Y-tamed almost complex structures $J_{Y_R,\theta} : [0, 2\pi] \to J_\tau(Y_R, \omega_Y)$ such that $P^*_{\gamma_{2\pi,R}} J_{2\pi} = J_0$. This gives us an endomorphism of the subbundle of the tangent spaces to the fibers

$$J_{re^{i\theta}} : TY_{re^{i\theta}} \to TY_{re^{i\theta}}, \quad J_{re^{i\theta}} = P^*_{\gamma_{\theta,r}} J_{Y_R,\theta}.$$

Since over every point with $|z| > R$ we have a splitting $T_{(y,z)}X = T_yY \oplus T_z\mathbb{C}$, we can give $T(X \setminus W^{-1}(U))$ the tame almost complex structure locally defined by $J_{re^{i\theta}} \oplus j_{\mathbb{C}}$.

Definition A.0.3 We say that an ω-tame almost complex structure on X is W-admissible if, when restricted to $W^{-1}(U)$, it can be written as $J_{re^{i\theta}} \oplus j_{\mathbb{C}}$ for some path of almost complex structures $J_{Y_R,\theta} \in \mathcal{J}_\tau(Y_R, \omega)$. We denote the space of such almost complex structures $\mathcal{J}_{\tau,W,R}(X, \omega_X)$.

The goal will to be to construct a stabilizing divisor $D_X \subset X$ in such a way that D_X is transverse to all $Y_{re^{i\theta}}$ with $r \geq R$, and subsequently show that there exists an open dense set of almost complex structures belonging to $\mathcal{J}_{\tau,W,R}(X, \omega_X)$ which are E-stabilized by D_X [12, Definition 4.24]. In the setting of Lagrangian cobordisms, the comparable statements are proven in [34, Appendix C.3 and Lemma C.1.3].

We first construct the divisor D_X. Take $E_X \to X$ a vector bundle whose first Chern class is $\frac{1}{2}\pi[\omega_X]$, so that the pullback $E_Y \to Y$ is a vector bundle whose first Chern class is $\frac{1}{2}\pi[\omega_Y]$. Pick a family of Hermitian structures on $E_{Y,\theta} \to Y_{Re^{i\theta}}$ depending on θ so that the curvature is $-i\omega_Y$ and so that $P^*_{\gamma_{2\pi,R}} E_{Y,2\pi} = E_{Y,0}$ as Hermitian line bundles. Let $i_{\theta_0} : Y_{Re^{i\theta_0}} \to X$ be the inclusion of the fiber over $Re^{i\theta_0}$. Take a Hermitian structure on $E_X \to X$ with curvature $-i\omega_X$ and the property that $i^*\theta_0 P^*_{\gamma_{\theta_0,r}|_{[1/2,1]}} E_X = E_{Y,\theta_0}$ as Hermitian line bundles.

We will construct the stabilizing divisor D_X as the zero locus of an asymptotically holomorphic section $s_{k,X} : X \to E_X^k$. First, using [10] we can pick asymptotically holomorphic sections $s_{k,Y} : Y \to E_Y^k$ with the property that $s_{k,Y}^{-1}(0)$ is disjoint from M. We obtain a second asymptotically holomorphic section by pullback $P^*_{\gamma_{2\pi,R}} s_{k,Y}$. By [8] we can find a family $s_{k,Y,\theta}$ of such sections satisfying $s_{k,Y,0} = s_{k,Y}$ and $s_{k,Y,2\pi} = P^*_{\gamma_{2\pi,R}} s_{k,Y}$.

Using this family of sections, we create an asymptotically holomorphic section $s_{k,X,\text{out}} : X \to E_X^k$ which is given by

$$s_{k,X,\text{out}} := \rho_{k,R+1}(|z|) P^*_{\gamma_{\theta,r}} s_{k,Y,\theta},$$

where $\rho_{k,R+1}(|z|)\colon \mathbb{C} \to \mathbb{R}$ is a function which is concentrated (in the sense of [10, Definition 2]) at the circle of radius $R+1$. The zero set $s^{-1}_{k,X,\mathrm{out}}(0)$ enjoys the properties that

- $s^{-1}_{k,X,\mathrm{out}}(0)$ is disjoint from L,
- for k sufficiently large, $s^{-1}_{k,X,\mathrm{out}}(0)$ is a symplectic divisor in $W^{-1}(\{z \mid |z| > R\})$, and
- $s^{-1}_{k,X,\mathrm{out}}(0)$ intersects $Y_{re^{i\theta}}$ transversely for all $r > R$.

This constructs the sections taking the place of $s_{k,X\times\mathbb{C},\mathrm{out}}$ in [34, Appendix C.3.2]. The remainder of the construction of D_X involves subsequently perturbing this section over the region $W^{-1}(U)$, which exactly follows [34, Appendix C.3.2].

Item (2): finding perturbations

The construction of an open dense set of E-stabilized almost complex structures proceeds in the same fashion as [34, Section C.3.3], which is itself based on the argument of [12, Section 4.5]. The main tool needed for the argument to run is to show that the space of almost complex structures regularizing holomorphic disks of energy up to E is dense in $\mathcal{J}_{\tau,W,R}(X,\omega_X)$. By application of the open mapping principle to W, every pseudoholomorphic disk in consideration must either

- pass through $W^{-1}(U)$, where they can be made regular through perturbations confined to the region $W^{-1}(U)$ by application of [15, Lemma 5.6], or
- be confined to a fiber $W^{-1}(t)$ with $t \in U$, in which case they can be made regular through perturbations constrained in the fiberwise direction, and since the fiber is compact, the set of such perturbations is open and dense.

Item (3): compactness of moduli spaces

The proof that the moduli spaces of pseudoholomorphic treed disks considered are compact uses that we may apply open-mapping principle type arguments for perturbations chosen from $\mathcal{J}_{\tau,W,R}(X,\omega_X)$, and that the Morse flow line components of treed disks point outwards at the boundary [34, Proposition C.4.1].

Remark A.0.4 In the examples we consider (potentials coming from tropicalized superpotentials associated to a monomial admissibility data), the fibers of the potential will in general not be compact. However, the monomially admissible condition ensures that the restriction of monomially admissible $L \subset X$ to $M \subset Y$ will be compact. As a result, all pseudoholomorphic disks contributing to treed disks will have boundary contained within a compact subset of X; we conclude that the moduli space of treed disks has compactification given by broken treed disks.

Appendix B Auxiliary results for filtered A_∞ algebras and modules

In this section, we give some background for filtered A_∞ algebras and bimodules, as well as provide some methods for constructing bounding cochains using the filtration on the A_∞ algebra.

B.1 A short review of bounding cochains

The *Novikov ring* with $\mathbb{C}$-coefficients is the ring of formal power series

$$\Lambda_{\geq 0} := \left\{ \sum_{i=0}^{\infty} a_i T^{\lambda_i} \;\middle|\; a_i \in \mathbb{C}, \lambda \in \mathbb{R}_{\geq 0}, \lim_{i\to\infty} \lambda_i = \infty \right\}.$$

The field of fractions is the Novikov field Λ. A *filtered* A_∞ algebra (A, m_A^k) is a free graded $\Lambda_{\geq 0}$-module $A^\bullet$ equipped with $\Lambda_{\geq 0}$-linear products

$$m^k : (A^\bullet)^{\otimes k} \to (A^{\bullet+2-k})$$

for each $k \geq 0$. These are required to satisfy the axioms of [22, Definition 3.2.20]. Among these axioms are

- the quadratic filtered A_∞ relationship
$$0 = \sum_{k_1+k'+k_2=k} (-1)^{\clubsuit(\underline{x},k_1)} (m^{k_1+1+k_2}) \circ (\mathrm{id}^{\otimes k_1} \otimes m^{k'} \otimes id^{\otimes k_2})(x_1, \dots, x_k),$$
where the sign is determined by $\clubsuit(\underline{x}, k_1) := k_1 + \sum_{j=1}^{k_1} \deg(x_j)$,
- each A^i has a filtration $F^\lambda A^i$ respecting that of $\Lambda_{\geq 0}$, and a basis belonging to $F^0(A^i) \setminus \bigcup_{\lambda>0} F^\lambda A^i$.

Given a filtered A_∞ algebra, we can also consider the Λ-linear products on $A \otimes_{\Lambda_{\geq 0}} \Lambda$. We call $A \otimes_{\Lambda_{\geq 0}} \Lambda$ a Λ-filtered A_∞ algebra.

Let (A, m_A^k) and (B, m_B^k) be A_∞ algebras. A *filtered A_∞ homomorphism* from A to B is a sequence of filtered graded maps

$$f^k : A^{\otimes k} \to B$$

satisfying the *quadratic A_∞ homomorphism relations*

$$\sum_{k_1+k'+k_2=k} (-1)^{\clubsuit(\underline{x},k_1)} f^{k_1+1+k_2} \circ (\mathrm{id}^{\otimes k_1} \otimes m_A^{k'} \otimes \mathrm{id}^{k_2}) = \sum_{i_1+\cdots+i_j=k} m_B^j \circ (f^{i_1} \otimes \cdots \otimes f^{i_j}).$$

There similarly exists a notion of a homotopy between filtered A_∞ homomorphisms.

The main difficulty with filtered A_∞ algebras is that they do not have cohomology groups, as

$$(m_A^1)^2 = m_A^2(m_A^0 \otimes \mathrm{id}) \pm m_A^2(\mathrm{id} \otimes m_A^0).$$

When $m_A^0 = 0$, the right-hand side of the relation is zero and we say that A is a *tautologically unobstructed* A_∞ algebra.

It is desirable to work with tautologically unobstructed A_∞ algebras as they can be studied with the standard tools employed for cochain complexes. Therefore, one might restrict one's study to tautologically unobstructed filtered A_∞ algebras. Problematically, tautologically unobstructed filtered A_∞ algebras are not closed under the relation of filtered A_∞ homotopy equivalence. This can be remedied by considering filtered A_∞ algebras equipped with bounding cochains.

Let A be a filtered A_∞ algebra. A deforming cochain is an element $d \in A^1$ with $\operatorname{val}(d) > 0$. The d-deformation of A is the filtered A_∞ algebra (A, d) whose

- underlying chain groups agree with A, and
- composition maps are given by the d-deformed A_∞ products,

$$m^k_{(A,d)} = \sum_{l=0}^{\infty} \sum_{j_0+\cdots+j_k=l} m_A^{k+l}(d^{\otimes j_0} \otimes \mathrm{id} \otimes d^{\otimes j_1} \otimes \cdots \otimes \mathrm{id} \otimes d^{\otimes j_k}). \tag{12}$$

Definition B.1.1 When $m^0_{(A,b)} = 0$, we say that b is a *bounding cochain*, and we say that the algebra A is *unobstructed.*

Given $f : A \to B$ a filtered A_∞ homomorphism and $b \in A$ a bounding cochain, there is a pushforward bounding cochain $f_*(b) \in B$ such that $(B, f_*(b))$ is unobstructed. When $f^k = 0$ for $k \neq 1$, then $f_*(b) = f(b)$. The existence of a pushforward bounding cochain shows that unobstructedness is a property of filtered A_∞ algebras which is preserved under the equivalence relation of filtered A_∞ homotopy equivalence.

In applications, we use Λ-filtered A_∞ algebras as opposed to filtered A_∞ algebras.[3] However, the homological algebra of filtered A_∞ algebras is notationally easier to describe (as there exist elements living in a minimal filtration level). A computation allows us to understand deformations and bounding cochains for the former (defined using (12)) in terms of the latter.

Claim B.1.2 *Suppose that A is a filtered A_∞ algebra and b a bounding cochain for A. Then $b \otimes 1 \in A \otimes_{\Lambda_{\geq 0}} \Lambda$ is a bounding cochain for the Λ-filtered A_∞ algebra $A \otimes_{\Lambda_{\geq 0}} \Lambda$.*

B.2 Extending an unobstructed ideal

Following ideas from [22], we will provide a method for constructing bounding cochains by inducting on the valuation. In order to do this, we need a slight refinement of a filtered A_∞ algebra which states that the valuation of the structure coefficients is ordered by a monoid. A gapped A_∞ algebra is a filtered A_∞ algebra for which there exists a finitely generated monoid G and a monoid homomorphism $\omega : G \to \mathbb{R}_{\geq 0}$ such that $\omega(\beta) = 0$ implies that $\beta = 0$, and we have the decomposition

$$m^k = \sum_{\beta \in G} T^{\omega(\beta)} m^{k,\beta},$$

where $m^{k,\beta}$ are graded with respect to the filtration. We say that it satisfies the gapped A_∞ relations if for all $\beta \in G$,

$$\sum_{\beta_1+\beta_2=\beta} \sum_{j_1+j+j_2=k} (-1)^{\clubsuit} m^{j_1+1+j_2,\beta_1}(\mathrm{id}^{\otimes j_1} \otimes m^{j,\beta_2} \otimes \mathrm{id}^{\otimes j_2}) = 0.$$

[3] This is because the continuation maps in Lagrangian intersection Floer cohomology are usually only weakly filtered.

Given $b = \sum_{\beta\in G\setminus\{0\}} b_\beta$, we can deform the product structure by

$$m^{k,\beta}_{(B,b)} = \sum_{\substack{\beta_0+\cdots+\beta_k=\beta \\ \beta_i=\sum_{j=1}^{l_i}\beta_{i,j}}} m_B^{k+l}(\beta_{0,0}\otimes\cdots\otimes\beta_{0,l_0}\otimes \mathrm{id}\otimes\cdots\otimes\mathrm{id}\otimes\beta_{k,0}\otimes\cdots\otimes\beta_{k,l_k}),$$

so that $m^k_{(B,b)} := \sum_{\beta\in G} T^{\omega(\beta)} m_B^{k,\beta}$ gives a G-gapped A_∞ algebra satisfying the gapped A_∞ relations. There similarly exist G-gapped filtered A_∞ homomorphisms, which also contain the data of a morphism of monoids $\phi\colon G_A \to G_B$.

We will also need some basic statements about ideals in filtered A_∞ algebras:

Definition B.2.1 A subspace $A\subset B$ is a weak A_∞ ideal if for all $k = k_1+1+k_2 > 0$, the map

$$m^k\colon B^{\otimes k_1}\otimes A\otimes B^{\otimes k_2}\to B$$

has image contained in A.

Notably, we *do not* require that the curvature term m_A^0 be an element of A. As a result, it is not necessarily the case that A is itself a filtered A_∞ algebra. We say that A is a *strong A_∞ ideal* if additionally $m_B^0\in A$.

Claim B.2.2 *Let $A\subset B$ be an A_∞ ideal. The quotient $C = A/B$ inherits a filtered A_∞ structure. A is a strong A_∞ ideal if and only if C is tautologically unobstructed.*

Proof The filtered A_∞ structure is the natural one:

$$m_C^k([x_1]\otimes\cdots\otimes[x_k]) := [m_B^k(x_1\otimes\cdots\otimes x_k)].$$

Because the m_B^k are multilinear, we see that if $[x_i]=[x_i']$, then $m_C^k([x_1]\otimes\cdots\otimes[x_i]\otimes\cdots\otimes[x_k]) = m_C^k([x_1]\otimes\cdots\otimes[x_i']\otimes\cdots\otimes[x_k])$. A is a strong A_∞ ideal if and only if $m_C^0 = [m_B^0] = [0]$. □

Example B.2.3 Given a formal filtered A_∞ morphism $f\colon B\to C$ (so that $f^k = 0$ for all $k\neq 1$) the kernel of f is a weak A_∞ ideal.

Example B.2.4 Given a filtered A_∞ algebra A, the set $A_{>0}$ of positively filtered elements is an example of a strong A_∞ ideal. The quotient $\underline{A} := A/A_{>0}$ is an example of a tautologically unobstructed A_∞ algebra. A relevant example comes from Lagrangian Floer cohomology, where $\underline{\mathrm{CF}^\bullet(L)} = \mathrm{CM}^\bullet(L)$.

Claim B.2.5 *Suppose that $A\subset B$ is an A_∞ ideal, and $d\in B$ is a deforming cochain. Then A is an A_∞ ideal of (B,d). If $A\subset B$ is a strong A_∞ ideal, and $m_d^0\in A$, then A is a strong A_∞ ideal of B. In particular, if $d\in A$ then A is a strong A_∞ ideal of (B,d).*

Proof Suppose that $a\in A$ is some element. Then

$$m^k_{(B,d)}(x_1\otimes\cdots\otimes a\otimes\cdots\otimes x_k) = \sum_{l=0}^{\infty}\sum_{j_0+\cdots+j_k=l} m_A^{k+l}(d^{\otimes j_1}\otimes\mathrm{id}\otimes d^{\otimes j_1}\otimes\cdots\otimes a\otimes\cdots\otimes\mathrm{id}\otimes d^{\otimes j_k})\in A,$$

proving that A is an A_∞ ideal of (B,d). □

The vector space $H^1(A)$ is a lowest-order approximation to the space of bounding cochains. When $\overline{C}$ is an anticommutative differential graded algebra, elements of $H^1(\overline{C})$ are bounding cochains.

Claim B.2.6 *Suppose that C is tautologically unobstructed. Suppose that $f: C \to \overline{C}$ is an A_∞ map with gapped A_∞ homotopy inverse $g: \overline{C} \to C$. Assume that $\overline{C}$ is an anticommutative differential graded algebra. Then for every class $[c] \in H^1(C)$ with $\operatorname{val}(c) > 0$, there exists a bounding cochain $c' \in C$ and $\lambda > \operatorname{val}(c')$ with $[c'] = [c] \in H^1(C/T^\lambda C)$.*

Proof Since C and $\overline{C}$ are gapped, we can select $\lambda > \operatorname{val}(c)$ such that $\omega(\beta) < \lambda$ implies $\omega(\beta) \leq \operatorname{val}(c)$. We observe that $f(c) \in \overline{C}$ is closed, and therefore provides a bounding cochain for $\overline{C}$, as

$$m^0_{(\overline{C}, f(c))} = m^1_{\overline{C}}(f(c)) + m^2_{\overline{C}}(f(c), f(c)) = 0.$$

We then take c' to be the pushforward bounding cochain

$$g_*(f(c)) = \sum_{k=1}^{\infty} g^k((f(c))^{\otimes k}).$$

Since $c' = (g \circ f)(c) \bmod T^\lambda$, we obtain that $[c] = [c'] \in H^1(C/T^\lambda C)$. □

Claim B.2.7 [34, Claim A.4.8] *Suppose that $A' = (A, a)$. Given a deforming cochain $a' \in A'$, the chain $a'' = a + a' \in A$ is a deforming cochain such that $(A', a') = (A, a'')$.*

We now come to the main lemma of this appendix. Suppose that we have an exact sequence (on the chain level) $A \to B \to C$. If A is a strong A_∞ ideal containing the curvature of B, then we prove that there is no obstruction to finding a bounding cochain for B. The argument is in the style of [22, Theorem 3.6.18].

Lemma B.2.8 *Consider a G-gapped A_∞ algebra B satisfying the gapped A_∞ relations. Suppose that*:

(i) *A is a strong A_∞ ideal of B and $C = B/A$, giving us an exact sequence $A \xrightarrow{i} B \xrightarrow{\pi} C$ of gapped A_∞ algebras,*

(ii) *there exists $\overline{C}$ which is A_∞ homotopic to C and is an anticommutative DGA,*

(iii) *the connecting map $\delta: H^1(\underline{C}) \to H^2(\underline{A})$ surjects.*

Then for every $\lambda > 0$ there exists a deforming cochain $b = \sum_{\beta \in G \setminus \{0\}} b_\beta$ for B such that for all β with $\omega(\beta) \leq \lambda$, we have $m^{0,\beta}_{(B,b)} = 0$.

Proof Because A,, B, and C are gapped A_∞ algebras, there exists $\{\lambda_i\}_{i=1}^n$ an ordering of the image $\omega(G) \in [0, \lambda]$.

We prove the statement by induction on λ_i. Suppose that $A' = (A, a_{i-1})$, $B' = (B, b_{i-1})$, and $C' = (C, c_{i-1})$ are G-gapped A_∞ algebras satisfying (i)–(iii) and additionally

(iv) the curvature has large valuation, $\operatorname{val}(m^0_{B'}) > \lambda_{i-1}$.

The inductive step will construct deforming cochains a', b', and c' such that the algebras (A', a'), (B', b'), and (C', c') satisfy (i)–(iv), where λ_{i-1} is replaced with λ_i. By Claim B.2.7, we can then construct the A_∞ algebras (A, a_i), (B, b_i), and (C, c_i).

Write $m^0_{B'} = \sum_{j=i}^{\infty} \sum_{\omega(\beta)=\lambda_j} \underline{b}_{j,\beta} T^{\lambda_j}$, where the $\underline{b}_{j,\beta}$ are elements of $\underline{B}' = \underline{B}$ of degree 2. Because A is a strong A_∞ ideal, we can find $\underline{a}_{i,\beta} \in \underline{A}$ with $i(\underline{a}_{i,\beta}) = \underline{b}_{i,\beta}$.

We examine the lowest-order terms of the A_∞ relation $m^1_{A'} \circ m^0_{A'} = 0$, and obtain

$$\underline{m}^1_{A'}(\underline{a}_{i,\beta}) = 0.$$

Since $[\underline{a}_{i,\beta}] \in H^2(\underline{A})$, by (iii) $[\underline{b}_{i,\beta}] = 0$. Therefore there exists $\hat{\underline{b}}_{i,\beta}$ such that $\underline{m}^1_{B'}(\hat{\underline{b}}_{i,\beta}) = \underline{b}_{i,\beta}$. The class $\underline{c}_{i,\beta} := \pi(\hat{\underline{b}}_{i,\beta})$ is closed. Using Claim B.2.6, we can find $\underline{c}'_{j,\beta}$ with $j \geq i$ such that $c' = \sum_{j=i}^{\infty} \sum_{\beta | \omega(\beta)=\lambda_i} \underline{c}'_{j,\beta} T^{\lambda_j}$ is a bounding cochain for C' with the property that $[c'_{i,\beta}] = [c_{i,\beta}] \in H^1(C'/T^{\lambda_{i+1}} C')$.

Because $\pi : B \to C$ surjects, we can find for all $j \geq i$ cochains $\underline{b}'_{\beta,j} \in \underline{B}$ with $\pi(\underline{b}'_{j,\beta}) = \underline{c}''_{j,\beta}$. Let

$$b' = -\sum_{j=i}^{\infty} \sum_{\beta | \omega(\beta)=\lambda_i} \underline{b}'_{i,\beta} T^{\lambda_i}.$$

This constructed b' satisfies the property

$$m^1_{B'}(b') \equiv -m^0_{B'} \mod T^{\lambda_{i+1}}.$$

Since π is a filtered A_∞ homomorphism without higher terms, the pushforward $\pi_*(b')$ equals c' and

$$\pi \circ m^0_{(B,b')} = m^0_{(C,c')} = 0.$$

Therefore $m^0_{(B',b')}$ is contained in A', and we write a' for the corresponding element in A'. Claim B.2.5 states that (A', a') is a strong A_∞ ideal of (B', b'), whose quotient is (C', c').

This gives us the G-gapped A_∞ algebras (A', b'), (B', b'), and (C', c'), which we've shown satisfy (i). We now show that these algebras satisfy (ii)–(iv). For (ii), observe that deformations by Maurer–Cartan classes preserve having an anticommutative model. Since the deformation occurs at valuation greater than 0, the map $H^1(\underline{C}) \to H^2(\underline{A})$ continues to surject, proving (iii).

To check (iv),

$$\operatorname{val}(m^0_{(B',b')}) = \operatorname{val}\left(\sum_{k=0}^{\infty} m^k_{B'}((b')^{\otimes k})\right) \geq \min\left(\operatorname{val}(m^0_{B'} + m^1_{B'}(b')), \sum_{k=2}^{\infty} m^k_{B'}((b')^{\otimes k})\right).$$

Given that $m^0_{B'} \equiv m^1_{B'}(b') \mod T^{\lambda_i}$,

$$\operatorname{val}(m^0_{(B',b')}) \geq \lambda_{i+1}. \qquad \square$$

Corollary B.2.9 *Let A, B, and C be A_∞ algebras as in Lemma B.2.8. Then there exists a bounding cochain for B.*

Proof The deforming cochains constructed in the above proof satisfy the condition that

$$b_i \equiv b_{i+1} \mod T^{\lambda_i}.$$

It follows that if we use the inductive procedure to build a sequence of deforming cochains $\{b_i\}_{i=0}^{\infty}$ such that $\mathrm{val}(m^0_{(B,b_i)}) > \lambda_i$, the limit $\lim_{i\to\infty} b_i$ is a bounding cochain. □

B.3 A_∞ bimodules and bounding cochains

Let A and B be A_∞ algebras. An (A,B)- *bimodule* is a filtered graded $\Lambda_{\geq 0}$-module M, along with a set of maps for all $k_1,k_2 \geq 0$,

$$m^{k_1|1|k_2}_{A|M|B} \colon A^{\otimes k_1} \otimes M \otimes B^{\otimes k_2} \to M,$$

satisfying filtered quadratic A_∞ module relations for each triple $(k_1|1|k_2)$:

$$\begin{aligned}0 = \sum_{\substack{j_1+j+j_2=k_1+1+k_2\\ j_1+j\leq k_1}} & m^{k_1-j+1|1|k_2}_{A|M|B} \circ (\mathrm{id}_A^{\otimes j_1} \otimes m^j_A \otimes \mathrm{id}^{\otimes k_1-j_1-j} \otimes \mathrm{id}_M \otimes \mathrm{id}_B^{k_2})\\ &+ \sum_{\substack{j_1+j+j_2=k_1+1+k_2\\ j_1\leq k_1\leq j_1+j-1}} m^{j_1|1|j_2}_{A|M|B} \circ (\mathrm{id}_A^{\otimes j_1} \otimes m^{k_1-j_1|1|k_2-j_2}_{A|M|B} \otimes \mathrm{id}_B^{\otimes j_2})\\ &\quad + \sum_{\substack{j_1+j+j_2=k_1+1+k_2\\ k_1+1<j_1}} m^{k_1|1|k_2-j+1}_{A|M|B} \circ (\mathrm{id}_A^{\otimes k_1} \otimes \mathrm{id}_M \otimes \mathrm{id}_B^{k_2-j_2-j} \otimes m^j_B \otimes \mathrm{id}_B^{\otimes j_2}).\end{aligned}$$

There is a G-gapped version of an A_∞ bimodule, where we have the data of a map of monoids $\omega \colon G_M \to \mathbb{R}_{\geq 0}$ and our A_∞ bimodule products can be decomposed as $m^{k_1|1|k_2,\beta}_{A|M|B}$; we also have morphisms $\phi_{A/B} \colon G_{A/B} \to G_M$ which intertwine with ω.

If M is a filtered (A,B) bimodule, and $a \in A$ and $b \in B$ are deforming cochains, then the filtered A_∞ bimodule products on M can be deformed to give it the structure of an $((A,a),(B,b))$ bimodule. As in the setting of Λ-filtered A_∞ algebras, we can define Λ-filtered A_∞ bimodules.

Lemma B.3.1 *Let M be a G-gapped (A,B) bimodule. Suppose that A and B are tautologically unobstructed and that A has an anticommutative DGA model $\bar{A}$ as in Claim B.2.6. Suppose that there exist $\lambda_0 < \lambda_1 \in \mathbb{R}$ such that*

(i) *the maps $m^{k|1|0}_{A|M|B} \colon A^{\otimes k} \otimes M \to M$ all have image contained within $T^{\lambda_0}M$, and*

(ii) *there exists $[\underline{e}] \in H^1(\underline{M})$ an element such that the map*

$$H^1(A) \to H^1(T^{\lambda_0}M/T^{\lambda_1}M), \quad [a] \mapsto [m^{1|1|0}_{A|M|B}(a \otimes \underline{e})]$$

is surjective.

Then there exists a choice of bounding cochain $a \in A^1$ and element $e \in M^0$ such that $m^{0|1|0}_{(A,a)|M|B}(e) = 0$.

Proof We again use the gapped structure and induct on valuations. For simplicity of exposition, we will assume that the monoid G is $\mathbb{N}$, so that $\omega(G) = \{n\lambda \mid n \in \mathbb{N}\} \subset \mathbb{R}$. We will construct a sequence of bounding cochains a_i and elements $e_i \in M^0$ with the property that

(iii) a_i are bounding cochains,

(iv) $m^{0|1|0}_{(A,a_i)|M|B}(e_i) \in T^{\lambda_0+i\lambda_1} M$, and

(v) for $i > 1$, $a_i - a_{i-1} \in T^{\lambda_0+i\lambda_1} A$ and $e_i - e_{i-1} \in T^{\lambda_0+i\lambda_1} M$.

Base case Let $a_0 = 0$ and $e_0 = e$. Items (i) and (ii) are given by the hypothesis, (iii) is trivial, (iv) follows from the gapped structure, and (v) has no content.

Inductive step Suppose we have constructed a_i and e_i satisfying the induction hypothesis. By (iv), we can write $m^{0|1|0}_{(A,a_i)|M|B}(e_i) \equiv c_i \mod T^{\lambda_0+(i+1)\lambda_1}$, where $c_i \in T^{\lambda_0+i\lambda_1} M$. At order $T^{\lambda_0+(i+1)\lambda_1}$,

$$\underline{m}^{0|1|0}_{A|M|B}(c_i) \equiv m^{0|1|0}_{(A,a_i)|M|B}(c_i) \equiv m^{0|1|0}_{(A,a_i)|M|B} \circ m^{0|1|0}_{(A,a_i)|M|B}(e_i) = 0 \mod T^{\lambda_0+(i+1)\lambda_1}.$$

We therefore obtain a class $[c_i] \in T^{\lambda_0+i\lambda_1} H^1(\underline{M})$. Using (ii), we have a homology class $\underline{a} \in T^{\lambda_0+i\lambda_1} A$ with

$$[m^{1|1|0}_{A|M|B}(\underline{a} \otimes \underline{e})] \equiv [c_i] \mod T^{\lambda_0+(i+1)\lambda_1}.$$

By Claim B.2.6, there exists a bounding cochain $a' \in T^{i\lambda_1} A^1$ for the product structures $m^k_{(A,a_i)}$ satisfying

$$a' \equiv \underline{a} \mod T^{\lambda_0+(i+1)\lambda_1}, \qquad [m^{1|1|0}_{A|M|B}(a' \otimes \underline{e})] = [c_i] \quad \text{in } H^1(T^{\lambda_0+i\lambda_1} M / T^{\lambda_0+(i+1)\lambda_1} M).$$

Write $m^{1|1|0}_{A|M|B}(a' \otimes \underline{e}) = c_i + m^{0|1|0}_{A|M|B} e'$, where $e' \in T^{\lambda_0+i\lambda_1}$. Then let $e_{i+1} = e_i + e'$ and let $a_{i+1} = a_i - a'$. By construction, we satisfy (v). By Claim B.2.7, a_{i+1} is a bounding cochain for A, and we therefore obtain (iii). Conditions (i) and (ii) are unchanged by deformations. It remains to prove (iv):

$$\begin{aligned} &m^{1|1|0}_{(A,a_{i+1})|M|B}(e_{i+1}) \\ &\quad \equiv m^{0|1|0}_{(A,a_i)|M|B}(e_i) - m^{1|1|0}_{A|M|B}(a', \underline{e}) + m^{0|1|0}_{A|M|B}(e') - m^{1|1|0}_{A|M|B}(a', e') \mod T^{\lambda_0+(i+1)\lambda_1} \\ &\quad \equiv 0 \mod T^{\lambda_0+(i+1)\lambda_1}. \end{aligned}$$

To complete the proof of the lemma, we can take the bounding cochain a and element e to be

$$a = \lim_{i\to\infty} a_i, \qquad e = \lim_{i\to\infty} e_i. \qquad \square$$

References

[1] **S S Abhyankar**, *Local analytic geometry*, Pure Appl. Math. 14, Academic, New York (1964) MR Zbl

[2] **M Abouzaid**, *Homogeneous coordinate rings and mirror symmetry for toric varieties*, Geom. Topol. 10 (2006) 1097–1156 MR Zbl

[3] **M Abouzaid**, *Morse homology, tropical geometry, and homological mirror symmetry for toric varieties*, Selecta Math. 15 (2009) 189–270 MR Zbl

[4] **M Abouzaid**, *Family Floer cohomology and mirror symmetry*, from "Proceedings of the International Congress of Mathematicians, II" (S Y Jang, Y R Kim, D-W Lee, I Ye, editors), Kyung Moon Sa, Seoul (2014) 813–836 MR Zbl

[5] **M Abouzaid**, *The family Floer functor is faithful*, J. Eur. Math. Soc. 19 (2017) 2139–2217 MR Zbl

[6] **M Abouzaid**, *Homological mirror symmetry without correction*, J. Amer. Math. Soc. 34 (2021) 1059–1173 MR Zbl

[7] **L Amorim**, *The Künneth theorem for the Fukaya algebra of a product of Lagrangians*, Int. J. Math. 28 (2017) art. id. 1750026 MR Zbl

[8] **D Auroux**, *Asymptotically holomorphic families of symplectic submanifolds*, Geom. Funct. Anal. 7 (1997) 971–995 MR Zbl

[9] **D Auroux**, *Special Lagrangian fibrations, wall-crossing, and mirror symmetry*, from "Surveys in differential geometry, XIII: Geometry, analysis, and algebraic geometry" (H-D Cao, S-T Yau, editors), International, Somerville, MA (2009) 1–47 MR Zbl

[10] **D Auroux**, **D Gayet**, **J-P Mohsen**, *Symplectic hypersurfaces in the complement of an isotropic submanifold*, Math. Ann. 321 (2001) 739–754 MR Zbl

[11] **R Bieri**, **J R J Groves**, *The geometry of the set of characters induced by valuations*, J. Reine Angew. Math. 347 (1984) 168–195 MR Zbl

[12] **F Charest**, **C T Woodward**, *Floer cohomology and flips*, Mem. Amer. Math. Soc. 1372, Amer. Math. Soc., Providence, RI (2022) MR Zbl

[13] **J-P Chassé**, **J Hicks**, **Y J N Nho**, *Reverse isoperimetric inequalities for Lagrangian intersection Floer theory*, preprint (2023) arXiv 2306.04761 To appear in Bull. Lond. Math. Soc.

[14] **M-W Cheung**, **L Fantini**, **J Park**, **M Ulirsch**, *Faithful realizability of tropical curves*, Int. Math. Res. Not. 2016 (2016) 4706–4727 MR Zbl

[15] **K Cieliebak**, **K Mohnke**, *Symplectic hypersurfaces and transversality in Gromov–Witten theory*, J. Symplectic Geom. 5 (2007) 281–356 MR Zbl

[16] **J Duval**, *On a result by Y Groman and J P Solomon*, Math. Ann. 364 (2016) 1361–1363 MR Zbl

[17] **M Einsiedler**, **M Kapranov**, **D Lind**, *Non-Archimedean amoebas and tropical varieties*, J. Reine Angew. Math. 601 (2006) 139–157 MR Zbl

[18] **T Ekholm**, *Notes on topological strings and knot contact homology*, from "Proceedings of the Gökova Geometry/Topology Conference" (S Akbulut, D Auroux, T Önder, editors), Gökova Geometry/Topology Conf. (2014) 1–32 MR Zbl

[19] **A Floer**, *Morse theory for Lagrangian intersections*, J. Differential Geom. 28 (1988) 513–547 MR Zbl

[20] **K Fukaya**, *Floer homology for families: a progress report*, from "Integrable systems, topology, and physics" (M Guest, R Miyaoka, Y Ohnita, editors), Contemp. Math. 309, Amer. Math. Soc., Providence, RI (2002) 33–68 MR Zbl

[21] **K Fukaya**, *Unobstructed immersed Lagrangian correspondence and filtered A_∞ functor*, preprint (2017) arXiv 1706.02131

[22] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, I*, AMS/IP Stud. Adv. Math. 46, Amer. Math. Soc., Providence, RI (2009) MR Zbl

[23] **S Ganatra**, **J Pardon**, **V Shende**, *Sectorial descent for wrapped Fukaya categories*, J. Amer. Math. Soc. 37 (2024) 499–635 MR Zbl

[24] **Y Groman**, *The wrapped Fukaya category for semi-toric SYZ fibrations*, preprint (2018) arXiv 1805.03635 To appear in J. Eur. Math. Soc.

[25] **Y Groman**, **J P Solomon**, *A reverse isoperimetric inequality for J-holomorphic curves*, Geom. Funct. Anal. 24 (2014) 1448–1515 MR Zbl

[26] **M Gross**, *Topological mirror symmetry*, Invent. Math. 144 (2001) 75–137 MR Zbl

[27] **W Gubler**, *Tropical varieties for non-Archimedean analytic spaces*, Invent. Math. 169 (2007) 321–376 MR Zbl

[28] **A Hanlon**, *Monodromy of monomially admissible Fukaya–Seidel categories mirror to toric varieties*, Adv. Math. 350 (2019) 662–746 MR Zbl

[29] **A Hanlon**, **J Hicks**, *Aspects of functoriality in homological mirror symmetry for toric varieties*, Adv. Math. 401 (2022) art. id. 108317 MR Zbl

[30] **J S Hicks**, *Tropical Lagrangians and homological mirror symmetry*, PhD thesis, University of California, Berkeley (2019) `https://www.proquest.com/docview/2296778726`

[31] **J Hicks**, *Tropical Lagrangian hypersurfaces are unobstructed*, J. Topol. 13 (2020) 1409–1454 MR Zbl

[32] **J Hicks**, *Observations on disks with tropical Lagrangian boundary*, from "2019–20 MATRIX annals" (D R Wood, J de Gier, C E Praeger, T Tao, editors), MATRIX Book Ser. 4, Springer (2021) 603–607 MR Zbl

[33] **J Hicks**, *Tropical Lagrangians in toric del-Pezzo surfaces*, Selecta Math. 27 (2021) art. id. 3 MR Zbl

[34] **J Hicks**, *Wall-crossing from Lagrangian cobordisms*, Algebr. Geom. Topol. 24 (2024) 3069–3138 MR Zbl

[35] **D Holmes**, *Affine dimers from characteristic polygons*, PUMP J. Undergrad. Res. 5 (2022) 24–51 MR Zbl

[36] **M Kontsevich**, *Homological algebra of mirror symmetry*, from "Proceedings of the International Congress of Mathematicians, I" (S D Chatterji, editor), Birkhäuser (1995) 120–139 MR Zbl

[37] **T Kuwagaki**, *The nonequivariant coherent-constructible correspondence for toric stacks*, Duke Math. J. 169 (2020) 2125–2197 MR Zbl

[38] **C Y Mak**, **H Ruddat**, *Tropically constructed Lagrangians in mirror quintic threefolds*, Forum Math. Sigma 8 (2020) art. id. e58 MR Zbl

[39] **D Matessi**, *Lagrangian pairs of pants*, Int. Math. Res. Not. 2021 (2021) 11306–11356 MR Zbl

[40] **G Mikhalkin**, *Amoebas of algebraic varieties and tropical geometry*, from "Different faces of geometry" (S Donaldson, Y Eliashberg, M Gromov, editors), Int. Math. Ser. (N. Y.) 3, Kluwer, New York (2004) 257–300 MR Zbl

[41] **G Mikhalkin**, *Enumerative tropical algebraic geometry in $\mathbb{R}^2$*, J. Amer. Math. Soc. 18 (2005) 313–377 MR Zbl

[42] **G Mikhalkin**, *Examples of tropical-to-Lagrangian correspondence*, Eur. J. Math. 5 (2019) 1033–1066 MR Zbl

[43] **G Mikhalkin**, **J Rau**, *Tropical geometry*, preprint (2018) `https://math.uniandes.edu.co/~j.rau/downloads/main.pdf`

[44] **T Nishinou**, *Realization of tropical curves in abelian surfaces*, preprint (2020) arXiv 2007.16148

[45] **T Nishinou**, **B Siebert**, *Toric degenerations of toric varieties and tropical curves*, Duke Math. J. 135 (2006) 1–51 MR Zbl

[46] **Y-G Oh**, *Riemann–Hilbert problem and application to the perturbation theory of analytic discs*, Kyungpook Math. J. 35 (1995) 39–75 MR Zbl

[47] **S Payne**, *Fibers of tropicalization*, Math. Z. 262 (2009) 301–311 MR Zbl

[48] **M Poźniak**, *Floer homology*, *Novikov rings and clean intersections*, PhD thesis, University of Warwick (1994)

[49] **J Rabinoff**, *Tropical analytic geometry, Newton polygons, and tropical intersections*, Adv. Math. 229 (2012) 3192–3255 MR Zbl

[50] **D Ranganathan**, *Skeletons of stable maps, I: Rational curves in toric varieties*, J. Lond. Math. Soc. 95 (2017) 804–832 MR Zbl

[51] **F Schmäschke**, *Floer homology of Lagrangians in clean intersection*, preprint (2016) arXiv 1606.05327

[52] **P Seidel**, *Graded Lagrangian submanifolds*, Bull. Soc. Math. France 128 (2000) 103–149 MR Zbl

[53] **N Sheridan**, **I Smith**, *Lagrangian cobordism and tropical curves*, J. Reine Angew. Math. 774 (2021) 219–265 MR Zbl

[54] **I Smith**, *A symplectic prolegomenon*, Bull. Amer. Math. Soc. 52 (2015) 415–464 MR Zbl

[55] **D E Speyer**, *Parameterizing tropical curves, I: Curves of genus zero and one*, Algebra Number Theory 8 (2014) 963–998 MR Zbl

[56] **A Strominger**, **S-T Yau**, **E Zaslow**, *Mirror symmetry is T-duality*, Nuclear Phys. B 479 (1996) 243–259 MR

[57] **A Subotic**, *A monoidal structure for the Fukaya category*, PhD thesis, Harvard University (2010) `https://www.proquest.com/docview/614155983`

[58] **B Toën**, **M Vaquié**, *Moduli of objects in dg-categories*, Ann. Sci. École Norm. Sup. 40 (2007) 387–444 MR Zbl

[59] **U Varolgunes**, *Mayer–Vietoris property for relative symplectic cohomology*, Geom. Topol. 25 (2021) 547–642 MR Zbl

[60] **K Wehrheim**, **C T Woodward**, *Functoriality for Lagrangian correspondences in Floer theory*, Quantum Topol. 1 (2010) 129–170 MR Zbl

School of Mathematics, University of Edinburgh
Edinburgh, United Kingdom

`jeff.hicks@ed.ac.uk`

`https://www.maths.ed.ac.uk/~jhicks2`

Proposed: Leonid Polterovich
Seconded: David Fisher, Mark Gross

Received: 15 May 2022
Revised: 10 December 2023

Geometry & Topology **29**:4 (2025) 1975–2046
DOI: 10.2140/gt.2025.29.1975
Published: 27 June 2025

Relations in singular instanton homology

PETER B KRONHEIMER
TOMASZ S MROWKA

We calculate the singular instanton homology with local coefficients for the simplest n-strand braids in $S^1 \times S^2$ for all odd n, describing these homology groups and their module structures in terms of the coordinate rings of explicit algebraic curves. The calculation is expected to be equivalent to computing the quantum cohomology ring of a certain Fano variety, namely a moduli space of stable parabolic bundles on a sphere with n marked points.

57R58; 14H60

1 Introduction

1.1 Background

A pair (Y, K), consisting of a closed, oriented 3-manifold and an embedded link, gives rise to a 3-dimensional orbifold $Z = Z(Y, K)$ whose underlying topology is that of Y and whose singular locus consists of the locus K where the orbifold structure has local stabilizers of order 2. The pair (Y, K), or the orbifold Z, is *admissible* if $[K]$ has odd pairing with some integer homology class. To an admissible orbifold Z, there is associated its *singular instanton homology* (Kronheimer and Mrowka [20]), constructed from the Morse theory of the Chern–Simons functional on the space of SO(3) orbifold connections modulo a determinant-1 gauge group. With rational coefficients, we denote the singular instanton homology by $I(Z; \mathbb{Q})$.

A deformation of this instanton homology is described in [21]. It can be viewed as an instanton homology group with values in a local coefficient system on the space of connections modulo gauge, and it appears in this paper as $I(Z;\Gamma)$, where Γ denotes a local system of free rank-1 modules over the ring of Laurent polynomials

$$\mathcal{R} = \mathbb{Q}[\tau^{\pm 1}].$$

The variable τ should be seen as a deformation parameter, with the specialization $\tau = 1$ recovering the original case of $\mathbb{Q}$ coefficients. (See Section 2.2.)

A choice of a 2-dimensional homology class in Z gives rise to an operator α, on both $I(Z;\mathbb{Q})$ and $I(Z;\Gamma)$. For each choice of basepoint $p \in K$, there is also an operator δ_p, depending on the connected component of K on which p lies and a choice of local orientation at p. These operators commute, and make $I(Z;\mathbb{Q})$ and $I(Z;\Gamma)$ into modules over the rings $\mathbb{Q}[\alpha,\delta_1,\dots,\delta_n]$ and $\mathcal{R}[\alpha,\delta_1,\dots,\delta_n]$ respectively, where n is the number of connected components of K.

In [32], Street completely described the instanton homology $I(Z;\mathbb{Q})$ and its module structure in the case that Z is the product

$$Z_n = S^1 \times S^2_n.$$

Here S^2_n denotes the 2-sphere with n orbifold points. An extension of Street's result to the case of $S^1 \times \Sigma_{g,n}$ was obtained by Xie and Zhang [36], and an earlier model for both of these calculations is the work of Muñoz [28; 27] on the case of $S^1 \times \Sigma_g$ (where the orbifold locus is empty).

The purpose of this paper is to extend Street's calculation to the case of instanton homology with local coefficients Γ. Alongside Z_n, a closely related calculation is for the instanton homology of an orbifold we call $Z_{n,1}$. If the n orbifold points in S^2_n are arranged symmetrically around a circle, then a rotation h through $2\pi/n$ is an automorphism of S^2_n which permutes the orbifold points, and we write $Z_{n,1}$ for its mapping torus:

$$Z_{n,1} = M_h, \quad h\colon S^2_n \to S^2_n.$$

Since the orbifold locus in $Z_{n,1}$ is connected, there is only one operator $\delta = \delta_p$ in this case, and $I(Z_{n,1};\Gamma)$ is a module for an algebra $\mathcal{R}[\alpha,\delta]$, where $\mathcal{R}$ is again a ring of Laurent polynomials. We can summarize the main theme of this paper as the solution to the following.

Problem $(\star)$ *Describe $I(Z_n;\Gamma)$ and $I(Z_{n,1};\Gamma)$ explicitly as modules for the algebras $\mathcal{R}[\alpha,\delta_1,\dots,\delta_n]$ and $\mathcal{R}[\alpha,\delta]$ respectively.*

The motivation for studying this question came from a desire to calculate a variant of the singular instanton homology of torus knots, $I^\natural(T_{n,q};\Gamma)$, as studied in our paper [24], and the related knot concordance invariants of these. In [24], the base ring always had characteristic 2, as necessitated by the construction there. An alternative formulation allows characteristic 0, and the results of this paper are a main step. We return to this discussion briefly in Section 7.

1.2 Statement of the result

We shall give a complete answer to ($\star$), and to give a flavor of the result here, we describe $I(Z_{n,1};\Gamma)$. First, there is an involution on the configuration space of connections on both of these orbifolds, defined by multiplying the holonomy on the S^1 factor in $S^1 \times S^2$ by $-1 \in \mathrm{SU}(2)$. This gives rise to an operator ϵ on instanton homology, and there is therefore a decomposition

$$I(Z_{n,1};\Gamma) = I(Z_{n,1};\Gamma)^+ \oplus I(Z_{n,1};\Gamma)^-$$

into the eigenspaces of ϵ. As modules, these two are related by changing the variable $\tau \in \mathcal{R}$ to $-\tau$. Each of the two summands is a cyclic module for $\mathcal{R}[\alpha,\delta]$ and they are therefore characterized by their ideals of relations, $J^{\pm}_{n,1}$ in the algebra:

$$I(Z_{n,1};\Gamma)^+ \cong \mathcal{R}[\alpha,\delta]/J^+_{n,1}, \quad I(Z_{n,1};\Gamma)^- \cong \mathcal{R}[\alpha,\delta]/J^-_{n,1}.$$

Over the field $\mathbb{C}$, we can regard $J^+_{n,1}$ and $J^-_{n,1}$ as the defining ideals of possibly nonreduced curves

$$D^+_n, D^-_n \subset \mathbb{C}^* \times \mathbb{C} \times \mathbb{C}$$

with coordinates (τ,α,δ). Our final description of these curves is as determinantal varieties: they are the loci of points where particular $m \times (m+1)$ matrices S^+ and S^- with entries in $\mathcal{R}[\alpha,\delta]$ fail to have full rank. Here $m = \frac{1}{2}(n-1)$. Equivalently, $J^{\pm}_{n,1}$ is the ideal generated by the $m \times m$ minors of $S^{\pm}$. Explicitly when $n = 11$ and $m = 5$, the matrix $S^{\pm}$ is given by $S_0 \pm S_1$, where S_0 is the matrix

$$\begin{pmatrix} -\alpha-\delta/2 & \alpha-19\delta/2 & 0 & 0 & 0 & 0 \\ 0 & -\alpha-5\delta/2 & \alpha-15\delta/2 & 0 & 0 & 0 \\ 0 & 0 & -\alpha-9\delta/2 & \alpha-11\delta/2 & 0 & 0 \\ 0 & 0 & 0 & -\alpha-13\delta/2 & \alpha-7\delta/2 & 0 \\ 0 & 0 & 0 & 0 & -\alpha-17\delta/2 & \alpha-3\delta/2 \end{pmatrix}$$

and S_1 is the matrix

$$\begin{pmatrix} \tau^7 & 0 & 0 & 0 & 0 \\ 0 & \tau^3 & 0 & 0 & 0 \\ 0 & 0 & 1/\tau & 0 & 0 \\ 0 & 0 & 0 & 1/\tau^5 & 0 \\ 0 & 0 & 0 & 0 & 1/\tau^9 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 & 0 & -9 & 5\tau^4+4 \\ 0 & 0 & 0 & -7 & 3\tau^4+2 & 2\tau^4 \\ 0 & 0 & -5 & \tau^4 & 4\tau^4 & 0 \\ 0 & -3 & -\tau^4-2 & 6\tau^4 & 0 & 0 \\ -1 & -3\tau^4-4 & 8\tau^4 & 0 & 0 & 0 \end{pmatrix}.$$

Although the matrices may look elaborate at first glance, they follow a fairly simple pattern that is readily described for general n. (See Section 6.3.) Note in particular that S_0 is a 2-band matrix with entries that are linear forms in (α,δ), while the entries of S_1 depend only on τ. On setting $\tau = 1$ in S_0 above, one recovers generators for the ideal that is identified by Street in [32]. For a general fixed value of τ, the corresponding locus is a subscheme of the (α,δ)-plane of length $m(m+1)$. A picture of the real locus of $D^{\pm}_n$ for $n = 7$ is given in Figure 1, together with the set of points on $D^{\pm}_n$ where $\tau = 0.6$.

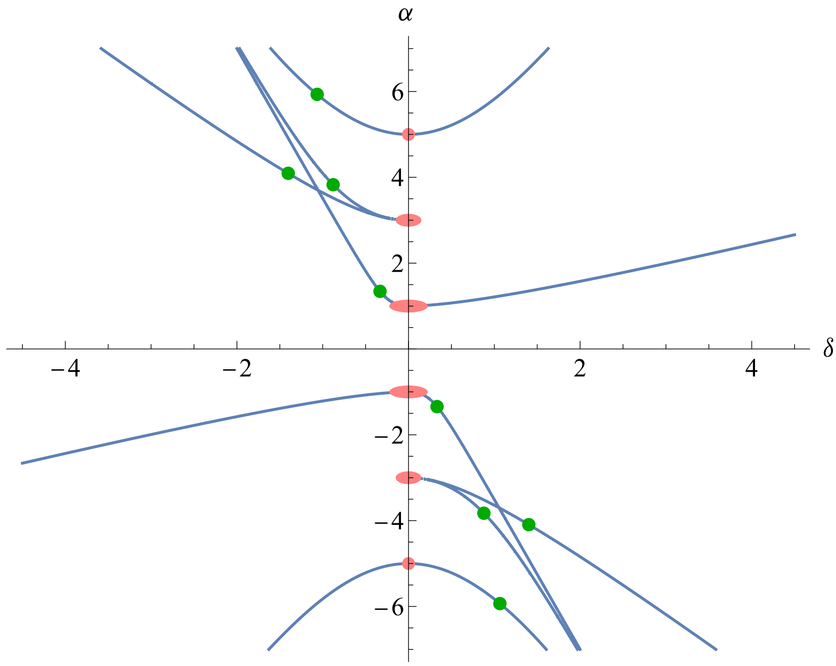


Figure 1: The blue curve is the projection of the real locus of $D_n^{\pm}$ to the (δ, α) plane for $n = 7$. The green points are the points where $\tau = 0.6$, showing the simultaneous eigenvalues of the operators δ and α for this value of τ. There are 12 of these, only 8 of which are real. The pink points indicate the subscheme of total length 12 defined by the minors of $S_0 \pm S_1$ when $\tau = 1$, which is the case described by Street [32]. Although the real curve looks rather smooth at $\alpha = \pm 1$, it has a uni-branch triple point there: in local analytic coordinates, the equation of the curve has the form $y^3 = x^7$.

Remark This description of $D_n^{\pm}$ as a determinantal variety means that the corresponding ideal $J_{n,1}^{\pm}$ is generated by $m+1$ elements, for this is the number of $m \times m$ minors. We shall see in fact that each of these ideals can be generated by just two of the minors.

As in Muñoz [28; 27], Street [32] and Xie and Zhang [36], the starting point for the calculation is an explicit generating set for the ideal of relations in the ordinary cohomology of a representation variety: in our case, as in [36], these are the "Mumford relations" in the cohomology of the representation variety associated to S_n^2. (See Earl and Kirwan [8] for example.) We obtain simple explicit formulae for these relations as products of linear forms in the variables α and δ_i. The matrix S_0 above arises as a matrix of syzygies for the Mumford relations. To compute the deforming term S_1, it is only necessary to understand the contributions of moduli spaces of instantons on $\mathbb{R} \times Z_n$ of smallest nonzero action (action $\frac{1}{4}$ in the normalization where the standard instanton on $\mathbb{R}^4$ has action 1). The contributions of these moduli spaces can be understood quite explicitly by a wall-crossing argument. A closely related phenomenon is present in [27].

1.3 Outline

In Section 2 we recall the definition of singular instanton homology with local coefficients and the construction of the operators that act on it in general. (Note that from Section 2 onwards, we simply write $I(Z)$ for the homology group referred to as $I(Z;\Gamma)$ above, without explicit mention of the local coefficients.) In Section 3, we introduce Z_n and $Z_{n,\pm 1}$ and study the ordinary cohomology of the relevant representation varieties and instanton homologies, enough to show that these can be described as cyclic modules for the algebra of operators which act on them. This material is quite standard.

In Section 4, we describe the Mumford relations in the ordinary cohomology of the representation variety of Z_n. We derive a very explicit formula for generators of the ideal of relations in these cohomology groups. The relations in the ordinary cohomology ring of the representation variety of Z_n admit a deformation which yields relations in the instanton homology $I(Z_n)$. The existence of this deformation is established in Section 5 together with a calculation of the subleading term using a wall-crossing calculation rather as in [27].

Knowledge of the subleading term turns out to be sufficient to obtain a complete answer, and the description of $I(Z_{n,1})$ (or equivalently $I(Z_{n,-1})$) that is outlined earlier in this introduction is derived in Section 6. Some further remarks are contained in Section 7 at the end of the paper.

Acknowledgements The work of Kronheimer was supported by the National Science Foundation through NSF grants DMS-1707924 and DMS-2005310. The work of Mrowka was supported by NSF grants DMS-1808794 and DMS-2105512. Both authors were supported by a Simons Foundation Award #994330 (Simons Collaboration on *New structures in low-dimensional topology*). This paper was completed while Mrowka was in residence at the Simons Laufer Mathematical Sciences Institute as a Clay Senior Scholar and supported by NSF grant DMS-1928930. The authors are grateful to the referee for improvements and corrections.

2 A version of singular instanton homology

In this section we review the construction of instanton homology with local coefficients, for admissible bifolds. General references include [20] and [23].

2.1 Bifolds and their Floer homology

For economy of notation, we will typically write simply Z for a pair consisting of a connected, oriented 3-manifold Y and an embedded (unoriented) link $K = K(Z) \subset Y$. Following [20] and [19], we will regard Z as determining an orbifold (a *bifold* in the terminology of [22]) whose underlying topological space is Y and whose singular set is $K(Z)$. The local stabilizer of the orbifold geometry at points of $K(Z)$ is of order 2. When talking of (for example) Riemannian metrics on Z, we will always mean orbifold Riemannian metrics. A bifold Z is *admissible* if there is an element of $H^1(Y;\mathbb{Z})$ which has nonzero mod-2 pairing with the class $[K(Z)] \in H_1(Y;\mathbb{Z}/2)$.

Associated to a 3-dimensional bifold Z, we have a space of bifold connections $\mathcal{B}(Z)$. In this paper, $\mathcal{B}(Z)$ will always consist of the bifold SO(3) connections with $w_2 = 0$ modulo the determinant-1 gauge group. In the language of [23, Section 2], this is the space of marked bifold connections in which the marking region is the complement of the singular set $K(Z)$ and the bundle has $w_2 = 0$ on the marking region.

Remark The space $\mathcal{B}(Z)$ can be identified with the space of gauge equivalence classes of SU(2) connections on the complement of the singular set $K(Z)$ such that the associated SO(3) bundle extends to an orbifold SO(3) bundle on Z with nontrivial monodromy (of order 2) at the singular points. When interpreted as SU(2) connections in this way, the limiting holonomy of the SU(2) connections on small loops linking the singular locus has order 4. This is the viewpoint adopted, for example, in [17; 18].

Definition 2.1 We write $\operatorname{Rep}(Z) \subset \mathcal{B}(Z)$ for the space of flat bifold connections modulo the determinant-1 gauge group. If Z is admissible, then $\operatorname{Rep}(Z)$ consists only of irreducible connections.

2.2 A local coefficient system

For each component $K^i \subset K(Z)$, after choosing a framing, we obtain a map to S^1,

$$h_i \colon \mathcal{B}(Z) \to S^1,$$

as in [20] and [23, Section 2.2]. Specifically, following [20], given $[A] \in \mathcal{B}(Z)$, we may restrict the connection $[A]$ to the boundary of the framed ϵ-tubular neighborhood of K^i and obtain, in the limit as $\epsilon \to 0$, a flat SO(3) connection on the torus, whose structure group reduces to SO(2). The holonomy of the SO(2) connection along the longitude defines $h_i([A])$.

An orientation of K^i is not needed here, because the orientation of the SO(2) bundle also depends on an orientation of K^i. (That is, the orientation of K^i is used twice in this construction.) The framing is also inessential, as a change of framing will change h_i by a half-period.

Taking the product over the set of all components of K, we define a single map $h \colon \mathcal{B}(Z) \to S^1$ by

$$h = \times_i h_i.$$

Over the circle S^1, there is a standard local system with fiber the ring of finite Laurent series

$$\mathcal{R} = \mathbb{Q}[\tau^{\pm 1}] \tag{1}$$

such that the monodromy of the local system around the positive generator of S^1 is multiplication by τ. Then by pulling back this local system by the map h, we obtain a local system Γ on $\mathcal{B}(Z)$. We summarize this construction with a definition.

Definition 2.2 Unless otherwise stated, the notation $\mathcal{R}$ will denote the ring $\mathbb{Q}[\tau^{\pm 1}]$, and Γ will denote the corresponding local system of free rank-1 $\mathcal{R}$-modules over $\mathcal{B}(Z)$, for any 3-dimensional bifold Z.

If Z is admissible, then by the standard construction (see [20; 21]), we obtain an instanton homology group for admissible bifolds:

Definition 2.3 Let Z be an admissible bifold of dimension 3. After choosing a Riemannian metric and perturbation to achieve a Morse–Smale condition for the gradient flow of the Chern–Simons functional on $\mathcal{B}(Z)$, we obtain an instanton Floer complex $CI(Z;\Gamma)$ of free $\mathcal{R}$-modules whose homology $I(Z;\Gamma)$ is the instanton homology of Z. We will generally write $I(Z)$ and omit Γ from the notation, unless the context demands otherwise. This is a $\mathbb{Z}/4$ graded module.

2.3 Functoriality and operators

We consider 4-dimensional bifolds W as cobordisms between 3-dimensional bifolds. In the context of this paper, the singular locus $\Sigma = \Sigma(W)$ of the orbifold W will always be an embedded surface (not necessarily orientable). In particular, we do not consider foams — singular surfaces — as in [22]. The Floer homology groups $I(Z)$ are functorial in the sense that a bifold cobordism W from Z^0 to Z^1 gives rise to a map

$$I(W)\colon I(Z^0) \to I(Z^1)$$

compatible with compositions.

The map $I(W)$ is obtained from suitable weighted counts of solutions to the perturbed anti-self-duality equations on the bifold W, after attaching cylindrical ends. This construction initially gives rise only to a projective functor, in that the overall sign of $I(W)$ is ambiguous. When $\Sigma(W)$ is oriented, the sign ambiguity can be resolved by choosing a homology orientation for W in the sense of [20]. In the case that $\Sigma(W)$ is not necessarily orientable, an appropriate substitute is the notion of an $\imath$-orientation introduced in [19]. (The sign ambiguity in the nonorientable case will not particularly concern us in this paper.)

Recall that in the present context $I(Z)$ denotes the instanton homology with coefficients in the local system Γ. That being so, the solutions A to the perturbed anti-self-duality equations on W are counted not just with signs ± 1, but with weights that are units in the ring $\mathcal{R}$. More precisely, if ρ_0 and ρ_1 are critical points of the perturbed Chern–Simons functional in $\mathcal{B}(Z^0)$ and $\mathcal{B}(Z^1)$, and if $[A]$ is a solution of the perturbed equations on W with cylindrical ends, asymptotic to ρ_0 and ρ_1, then $[A]$ contributes to the matrix entry of the map $I(W)$ at the chain level with a contribution $\pm\Gamma(A)$, where $\Gamma(A)\colon \Gamma(\rho_0) \to \Gamma(\rho_1)$ is given by

$$\Gamma(A) = \tau^{\nu(A) + \frac{1}{2}(\Sigma\cdot\Sigma)}. \tag{2}$$

Here ν is obtained from a curvature integral on the 2-dimensional singular set $\Sigma = \Sigma(W)$, and the self-intersection number $\Sigma\cdot\Sigma$ is computed relative to chosen framings of the singular sets $K(Z^0)$ and $K(Z^1)$. The expression on the right-hand side of (2) is not an element of $\mathcal{R}$ itself, because the exponent is not generally an integer. It is, however, a homomorphism between the rank-1 $\mathcal{R}$-modules $\Gamma(\rho_0) \to \Gamma(\rho_1)$ in a natural way. For details of this construction see, for example, [20, Section 3.9] and [23]. As explained there, the choice of framings is essentially immaterial. Consistent with our notation $I(Z)$ in which the local coefficient system Γ is implied, we will continue to write simply $I(W)$ for the $\mathcal{R}$-module homomorphism between these instanton homology groups.

As well as the map $I(W)$ above, we have the generalizations obtained by cutting down the moduli spaces on W by cohomology classes in the configuration space of bifold connections $\mathcal{B}(W)$. Here $\mathcal{B}(W)$ is a space of SO(3) bifold connections modulo the determinant-1 gauge group, and in the language of [23], this is the space of marked bifold connections in which the marking region is the complement of the singular set $\Sigma(W)$ and the bundle has $w_2 = 0$ on the marking region.

To describe the relevant cohomology classes more specifically, and to fix conventions, there is a universal orbifold SO(3) bundle,

$$\mathbb{E} \to \mathcal{B}^*(W) \times W,$$

which has an orbifold Pontryagin class,

$$p_1^{\mathrm{orb}}(\mathbb{E}) \in H^4(\mathcal{B}^*(W) \times W; \mathbb{Q}).$$

We adopt the convention that our preferred 4-dimensional characteristic class is $-\frac{1}{4} p_1^{\mathrm{orb}}(\mathbb{E})$, which coincides with $c_2^{\mathrm{orb}}(\tilde{\mathbb{E}})$ in the case that there is a lift to an SU(2) bundle $\tilde{\mathbb{E}}$. Given a class γ in $H^2(W; \mathbb{Q})$ or $H^0(W; \mathbb{Q})$, we obtain classes

$$-\tfrac{1}{4} p_1^{\mathrm{orb}}(\mathbb{E})/[\gamma] \tag{3}$$

in $H^2(\mathcal{B}^*(W); \mathbb{Q})$ or $H^4(\mathcal{B}^*(W); \mathbb{Q})$ respectively.

In addition to the classes (3), if p is a point of the orbifold locus $\Sigma(W)$, then the restriction of $\mathbb{E}$ to $\mathcal{B}^*(W) \times \{p\}$ has a decomposition

$$\mathbb{E}_p = \mathbb{R} \oplus \mathbb{V}_p,$$

where $\mathbb{V}_p$ is a 2-plane bundle. An orientation of $\mathbb{V}_p$ depends on a choice of normal orientation to the orbifold locus at p. Having chosen such an orientation, a class $\delta_p \in H^2(\mathcal{B}^*(W); \mathbb{Q})$ is then defined as

$$\delta_p = \tfrac{1}{2} e(\mathbb{V}_p). \tag{4}$$

We can regard δ here as depending on a choice of an element in $H_0(\Sigma(W); O)$, where O is the orientation bundle of $\Sigma(W)$ with rational coefficients.

Combining the classes (3) for $\gamma \in H^i(W; \mathbb{Q})$ and the classes δ_p, we obtain homomorphisms of $\mathcal{R}$-modules

$$I(W, a)\colon I(Z^0) \to I(Z^1) \tag{5}$$

depending linearly on

$$a \in \mathrm{Sym}_*\big(H_2(W; \mathbb{Q}) \oplus H_0(W; \mathbb{Q}) \oplus H_0(\Sigma(W); O)\big). \tag{6}$$

Since $I(Z^0)$ and $I(Z^1)$ are $\mathcal{R}$-modules, we may extend linearly over $\mathcal{R}$ to allow also

$$a \in \mathrm{Sym}_*\big(H_2(W; \mathbb{Q}) \oplus H_0(W; \mathbb{Q}) \oplus H_0(\Sigma(W); O)\big) \otimes \mathcal{R}. \tag{7}$$

The construction of the operators $I(W,a)$ is suitably functorial. In particular, this means for us that, in the case that W is a cylinder $[0,1]\times Z$, we have

$$I(W,a_1a_2)=I(W,a_1)I(W,a_2).$$

We will always be dealing with the case that W is connected, so there is only one class $[w]$ in $H_0(W;\mathbb{Q})$. From [16; 20], we note the following relation among the homomorphisms $I(W,a)$.

Proposition 2.4 *Let p a point in $\Sigma(W)$ with a chosen orientation of $T_p\Sigma(W)$, representing a class in $H_0(\Sigma(W);O)$ in the algebra (6). Let w be a point in W, representing a class in $H_0(W;\mathbb{Q})$. Then we have a relation*

$$I\big(W,(p^2+w-\tau^2-\tau^{-2})b\big)=0$$

for any b in the algebra (6).

Corollary 2.5 *The map $I(W,p^2b)$ is independent of the choice of oriented point $p\in\Sigma(W)$.*

Remark The relation in Proposition 2.4 reflects (in part) a relation in the cohomology ring of $\mathcal{B}^*(W)$, where we have a 2-dimensional class δ_p and a 4-dimensional class $-\frac{1}{4}p_1^{\mathrm{orb}}(\mathbb{E})/[w]$. From their construction as characteristic classes, these satisfy

(8)

The extra terms $\tau^2+\tau^{-2}$ in the proposition arise from instanton bubbling contributions [16].

Proposition 2.4 also tells that the generator corresponding to $[w]\in H_0(W;\mathbb{Q})$ is redundant. We obtain the most general homomorphism $I(W,a)$ if we only take a in the smaller algebra

$$\mathrm{Sym}_*\big(H_2(W;\mathbb{Q})\oplus H_0(\Sigma(W);O)\big). \tag{9}$$

There is an additional construction we can make if we are given a distinguished class $e\in H_2(W;\mathbb{Z})$. We consider the space $\mathcal{B}(W)^e$ of marked bifold SO(3) connections on W where the marking region is again the complement of $\Sigma(W)$ and where the marking data has

$$w_2=\mathrm{PD}(e)|_{W\setminus\Sigma(W)} \bmod 2.$$

After attaching cylindrical ends, the instantons in $\mathcal{B}(W)^e$ provide us with maps

$$I(W,a)^e\colon I(Z^0)\to I(Z^1). \tag{10}$$

When the singular set $\Sigma(W)$ is oriented, the integer lift e in homology, together with the homology-orientation of W, is used to orient the moduli spaces and determines the overall sign of the map $I(W,a)^e$. If $e-e'=2v$, so that e and e' define the same mod 2 class, then (as in [6]) we have

$$I(W,a)^{e'}=(-1)^{v\cdot v}I(W,a)^e. \tag{11}$$

Remark As discussed for example in [19], one can more generally consider the case that e is a relative class so that $\partial e\in H_1(\Sigma(W))$, but the more restrictive version here is required because we wish to use the local coefficient system Γ, which is otherwise not defined. See also [23, Section 2.2].

3 Torus braids in $S^1 \times S^2$

3.1 The torus braids

The following examples play an important role for us.

Definition 3.1 Let $\pi = \{p_1, \dots, p_n\}$ be n points arranged symmetrically around the equator of S^2. We write Z_n for the bifold whose underlying 3-manifold Y is the product $S^1 \times S^2$ and whose singular locus K is the n-component link

$$K_n = S^1 \times \pi \subset S^1 \times S^2.$$

Definition 3.2 For any $q \in \mathbb{Z}$, we define a bifold $Z_{n,q}$ as follows. The 3-manifold Y is again $S^1 \times S^2$. If $\varphi \in \mathbb{R}/(2\pi\mathbb{Z})$ denotes an angular coordinate on the equator of S^2, and θ a coordinate on the S^1 factor, then $K = K_{n,q}$ will be the link determined by $n\varphi = q\theta \pmod{2\pi}$.

The bifold $Z_{n,q}$ is admissible when n is odd. The link $K_{n,q} \subset S^1 \times S^2$ is connected (a knot) when n and q are coprime. When $q = 0$, the orbifold $Z_{n,0}$ coincides with Z_n above.

When needed, we orient the singular set $K_n \subset Z_n$ as the boundary of n disks in the product 4-manifold $D^2 \times S^2$, and we orient $K_{n,q}$ similarly using the fact that they have the same infinite cyclic cover.

It is evident from the definitions that the orbifold $Z_{n,q}$ is isomorphic to $Z_{n,-q}$ by an orientation-reversing map. With a little more thought, one can see that there is also an orientation-*preserving* isomorphism:

Lemma 3.3 *The link $K_{n,q}$ is isotopic in $S^1 \times S^2$ to the link $K_{n,-q}$. As a consequence, there is an orientation-preserving isomorphism of bifolds from $Z_{n,q}$ to $Z_{n,-q}$, preserving the orientation of the singular set.*

Proof Let L be an oriented axis in $\mathbb{R}^3$ passing through two points of the equatorial circle in the above description of $K_{n,q}$. Let ρ_t be the rotation of S^2 about this axis through angle $2\pi t$, and let $1 \times \rho_t$ be the resulting map $S^1 \times S^2 \to S^1 \times S^2$. Then the link

$$K_t = (1 \times \rho_t)(K_{n,-q}) \subset S^1 \times S^2$$

coincides with $K_{n,-q}$ when $t = 0$ and with $K_{n,q}$ when $t = \frac{1}{2}$. □

We aim to give a description of $I(Z_n)$ (the instanton homology with local coefficients) as an $\mathcal{R}$-module, together with a description of the operators

$$I([0,1] \times Z_n, a) \colon I(Z_n) \to I(Z_n) \quad \text{and} \quad I([0,1] \times Z_n, a)^e \colon I(Z_n) \to I(Z_n)$$

arising from classes a by the general construction (5) and (10), where e is the 2-dimensional class in $H_2(Z_n; \mathbb{Q})$.

3.2 The representation variety of S^2_n

Let us assume henceforth that n is odd, so that the orbifold Z_n described above is admissible. We may describe Z_n as a product $S^1 \times S^2_n$, where S^2_n is a 2-dimensional bifold of genus 0, and we begin with some observations about the representation variety $\mathrm{Rep}(S^2_n)$, drawn from [3; 35; 32]. Note that we can identify $\mathrm{Rep}(S^2_n)$ with the space of flat SU(2) connections on the complement of the n singular points such that the monodromy at each puncture has order 4. (See the remark in Section 2.1.)

First, as n is odd, the variety $\mathrm{Rep}(S^2_n)$ consists entirely of irreducible connections. It is a smooth, compact, connected manifold of dimension $2n-6$ for $n \geq 3$, and is empty for $n=1$. We have no need for a detailed description of their topology, but we record the fact that $\mathrm{Rep}(S^2_3)$ is a single point and $\mathrm{Rep}(S^2_5)$ is diffeomorphic to the blow up of $\mathbb{CP}^2$ at 5 points. It will be convenient to make use of the following result, which the authors believe has the status of folklore. The statement and proof are very minor adaptations of the main result of [13]. See also [33].

Lemma 3.4 *For any odd n, the manifold $\mathrm{Rep}(S^2_n)$ admits a Morse function with critical points only in even index.*

Proof Following [13], we present a proof by induction on n. So assume the result is true for a particular n, and consider $\mathrm{Rep}(S^2_{n+2})$. Let $C \subset \mathrm{SU}(2)$ be the subset of elements of order 4, ie the unit sphere of imaginary quaternions. Let $\widetilde{R} \subset C^{n+2}$ be the locus

$$\{(i_1,\dots,i_{n+2}) \in C^{n+2} \mid i_1 i_2 \cdots i_{n+2} = 1\},$$

so that the representation variety $\mathrm{Rep}(S^2_{n+1})$ is the quotient of $\widetilde{R}$ by conjugation. For $\boldsymbol{i} \in \widetilde{R}$, there is a unique $\theta \in [0,\pi]$ such that

$$i_{n+1} i_{n+2} \sim \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

and we have a smooth function

$$h = \cos(\theta) = \tfrac{1}{2}\,\mathrm{tr}(i_{n+1} i_{n+2}),$$

which descends to a smooth function

$$h\colon \mathrm{Rep}(S^2_{n+2}) \to [-1,1].$$

We consider separately the loci $h^{-1}(1)$, $h^{-1}(-1)$ and $h^{-1}((-1,1))$.

If $\boldsymbol{i} \in h^{-1}(1)$, then $i_{n+1} i_{n+2} = 1$, and it follows that $i_1 i_2 \cdots i_n = 1$. So these remaining n points define a point in $\mathrm{Rep}(S^2_n)$. The remaining choice of i_{n+1} exhibits $h^{-1}(1)$ as a 2-sphere bundle over $\mathrm{Rep}(S^2_n)$. As in [13], we may use the induction hypothesis to show that a perturbation of h has critical points only of even index near $h=1$. The situation at $h^{-1}(-1)$ is essentially the same: multiplying i_1 and i_{n+2} by -1 interchanges these two loci.

On the locus $h^{-1}((-1,1))$, the function h itself is Morse and its critical points can be described as follows. Let i, j, k in C be the standard unit quaternions with $ijk=-1$. Given any element of $h^{-1}((-1,1))$ we can use the action of conjugation to uniquely put in standard form with $i_{n+1}=i$ and i_{n+2} lying in the interior of the semicircle γ which joins i to $-i$ and passes through j. In this standard form, there is a circle action on $h^{-1}((-1,1))$ which fixes i_{n+1} and i_{n+2} and rotates the points $i_1,\dots,i_n$ about the axis through k. The function $\theta=\cos^{-1}(h)$ is smooth on this locus and is the moment map of the circle action. The critical points of h are therefore precisely the fixed points of this circle action. These fixed points are the points which in standard form have $i_{n+1}=i$, $i_{n+2}=j$ and $i_m=\pm k$ for all other m. The constraint $i_1 i_2\cdots i_{n+1}=1$ means that $i_m=-k$ for an even number of indices m in the range $1,\dots,n$. As a general property of moment maps, because these fixed points are isolated, they are Morse critical points for h, of even index. □

Remark The proof of the lemma above gives a little bit more, for we can easily identify the indices of the critical points, and hence establish the recursive formula for the Poincaré polynomial of $\mathrm{Rep}(S^2_n)$ which is given in [32]. The loci $h^{-1}(1)$ and $h^{-1}(-1)$, which are the 2-sphere bundles over $\mathrm{Rep}(S^2_n)$ inside $\mathrm{Rep}(S^2_{n+2})$, are the minima and maxima of h and together make a contribution

$$(1+t^2)^2 P_n(t)$$

to the Poincaré polynomial P_{n+2} for $\mathrm{Rep}(S^2_{n+2})$. Using the symmetries of $\mathrm{Rep}(S^2_{n+2})$ obtained by multiplying an even number of the i_l by -1, it is easy to see that the remaining critical points in $h^{-1}((-1,1))$ all have the same index and that this index is the middle dimension $(n-1)$. There are 2^{n-1} of these critical points, so we recover the recursive formula from [32],

$$P_{n+2}(t)=(1+t^2)^2 P_n(t)+(2t)^{n-1}. \tag{12}$$

Atiyah and Bott [1] described standard generators for the cohomology ring of representation varieties of surfaces in the nonorbifold case (a smooth surface of genus g), and there is an extension of those techniques for the orbifold case, developed in [2]. For the specific case of S^2_n, the results are given in [32]. In this description, the generators of the cohomology ring $H^*(\mathrm{Rep}(S^2_n;\mathbb{Q}))$ are classes

$$\alpha\in H^2(\mathrm{Rep}(S^2_n);\mathbb{Q}),\qquad \beta\in H^4(\mathrm{Rep}(S^2_n);\mathbb{Q}),\qquad \delta_p\in H^2(\mathrm{Rep}(S^2_n);\mathbb{Q})\quad\text{for } p\in\pi, \tag{13}$$

which are the restrictions to $\mathrm{Rep}(S^2_n)$ of classes defined on the space of irreducible bifold connections, $\mathcal{B}^*(S^2_n)$, arising from the slant product construction (3). More specifically, the classes α and β arise from the fundamental 2-dimensional class $[S^2_n]\in H_2(S^2_n)$ and the point class $[w]\in H_0(S^2_n)$ respectively, while δ_p is defined as in (4):

$$\alpha=-\tfrac14 p_1^{\mathrm{orb}}(\mathbb{E})/[S^2_n],\quad \beta=-\tfrac14 p_1^{\mathrm{orb}}(\mathbb{E})/[w],\quad \delta_p=\tfrac12 e(\mathbb{V}_p). \tag{14}$$

We will sometimes write

$$\delta_1,\dots,\delta_n$$

for the classes δ_{p_i} as p_i runs through π.

The classes α and β can also be seen as arising from the Künneth decomposition in $H^4(\mathcal{B}^*(S^2_n)\times S^2_n;\mathbb{Q})$,

$$-\tfrac{1}{4}p_1^{\mathrm{orb}}(\mathbb{E}) = \beta\times 1+\alpha\times v,$$

where v is the generator of $H^2(S^2_n;\mathbb{Q})$. The generator β is redundant, because of the relation

$$\delta_p^2 = -\beta \quad \text{for all } \ p\in\pi,$$

which is a restatement of (8) in the current situation.

In the rational cohomology ring of $\mathcal{B}^*(S^2_n)$, there are no further relations: the cohomology ring is the algebra

$$H^*(\mathcal{B}^*(S^2_n);\mathbb{Q}) = \mathbb{Q}[\alpha,\delta_1,\dots,\delta_n]/\langle\delta_k^2-\delta_l^2\rangle_{k,l}. \tag{15}$$

We have a surjective homomorphism

$$\varphi\colon H^*(\mathcal{B}^*(S^2_n);\mathbb{Q})\to H^*(\mathrm{Rep}(S^2_n);\mathbb{Q}). \tag{16}$$

Definition 3.5 We write A_n for the algebra

$$A_n = H^*(\mathcal{B}^*(S^2_n);\mathbb{Q}) = \mathbb{Q}[\alpha,\delta_1,\dots,\delta_n]/\langle\delta_k^2-\delta_l^2\rangle_{k,l},$$

and we write

$$j_n\subset A_n$$

for the kernel of the surjective homomorphism φ.

Generators for the ideal j_n are described in detail in [32], which leads to a complete description of the cohomology ring,

$$H^*(\mathrm{Rep}(S^2_n);\mathbb{Q}) = A_n/j_n. \tag{17}$$

See also Proposition 4.8.

3.3 The representation variety of Z_n

The flat bifold connections on Z_n are of two sorts, which we call the "plus" and "minus" components, which can be distinguished by examining the holonomy of the flat connection along the S^1 factor in $Z_n = S^1\times S^2_n$. The representations in the plus component are pulled back from S^2_n. The representations in the minus component are obtained from these by multiplication by a flat real line bundle with holonomy -1 on the S^1 factor. Thus we have

$$\mathrm{Rep}(Z_n) = \mathrm{Rep}(Z_n)_+\cup\mathrm{Rep}(Z_n)_- = \mathrm{Rep}(S^2_n)\cup\mathrm{Rep}(S^2_n). \tag{18}$$

Because of this, the description (17) of the cohomology ring of $\mathrm{Rep}(S^2_n)$ leads immediately to a description of the cohomology of $\mathrm{Rep}(Z_n)$. We are also eventually interested in the cohomology of the representation variety with constant coefficients $\mathcal{R}$ rather than $\mathbb{Q}$ (because of our interest in instanton homology with local coefficients Γ). With this in mind, let

$$\epsilon\colon H^*(\mathrm{Rep}(Z_n);\mathcal{R})\to H^*(\mathrm{Rep}(Z_n);\mathcal{R})$$

be the map obtained from interchanging the two copies, so that $\epsilon^2 = 1$. We write $\mathcal{A}_n$ for the algebra

(19) $$\mathcal{A}_n = \mathcal{R}[\alpha, \delta_1, \dots, \delta_n, \epsilon] / \langle \epsilon^2 - 1, \delta_k^2 - \delta_l^2 \rangle_{k,l}.$$

That is, we extend the coefficient ring of the algebra (17) from $\mathbb{Q}$ to $\mathcal{R}$, and we adjoin the element ϵ with square 1. This provides us with the following description. In the statement below, we write

$$1_+ \in H^0(\mathrm{Rep}(Z_n))$$

for the element Poincaré dual to the fundamental class of the component $\mathrm{Rep}(Z_n)_+$.

Proposition 3.6 *The cohomology of the representation variety* $\mathrm{Rep}(Z_n)$ *with coefficients in* $\mathcal{R}$ *is a cyclic module for the algebra* $\mathcal{A}_n$ *with generator the element* $1 \in H^0(\mathrm{Rep}(Z_n); \mathcal{R})$. *We have*

(20) $$H^*(\mathrm{Rep}(Z_n); \mathcal{R}) \cong \mathcal{A}_n / J_n, \quad \text{where } J_n = (j_n + \epsilon j_n) \otimes \mathcal{R},$$

and j_n *is the ideal in* (17). *Using Poincaré duality, the homology* $H_*(\mathrm{Rep}(Z_n); \mathcal{R})$ *can equivalently be described as a cyclic* $\mathcal{A}_n$*-module with generator the class* $[\mathrm{Rep}(Z_n)_+]$, *with the classes* α *and* δ_k *acting by cap product.*

We regard $\mathcal{A}_n$ as a graded algebra with the generators α and δ_k in grading 1 (not 2) and ϵ in grading 0. From the grading, $\mathcal{A}_n$ obtains an increasing filtration, which for future reference we record as

(21) $$\mathcal{A}_n^{(0)} \subset \mathcal{A}_n^{(1)} \subset \mathcal{A}_n^{(2)} \subset \cdots \subset \mathcal{A}_n,$$

where $\mathcal{A}_n^{(s)}$ is the $\mathcal{R}$-submodule generated by elements in grading less than or equal to s.

From the explicit description of the generators of j_n given in [32] (for rational coefficients), we can read off that there are no relations between the generators up to the middle dimension of $\mathrm{Rep}(Z_n)$:

Proposition 3.7 *For* $s \le (n-3)/2$, *we have* $J_n \cap \mathcal{A}_n^{(s)} = \{0\}$. □

3.4 The instanton homology of Z_n

The instanton homology $I(Z_n; \mathbb{Q})$ with rational coefficients was described, together with its ring structure, by Street [32] drawing on work of Boden [3] and Weitsman [35]. We summarize part of these results here, adapted to the case of $I(Z_n)$ (by which we continue to mean the instanton homology with local coefficients).

The representation variety $\mathrm{Rep}(Z_n)$ is a Morse–Bott critical locus for the Chern–Simons functional. By Lemma 3.4, there is a Morse function on $\mathrm{Rep}(Z_n)$ with critical points only in even index. The proof of that lemma allows one to construct such a Morse function as a linear combination of traces of holonomies around loops in Z_n. We may use such a Morse function as a holonomy perturbation for the Chern–Simons functional, so that the critical points of the perturbed Chern–Simons functional correspond to the critical points of the Morse function on $\mathrm{Rep}(Z_n)$. After making such a perturbation, the set of critical points forms a natural basis both for the ordinary homology of $\mathrm{Rep}(Z_n)$ as a $\mathbb{Q}$-vector space, and for the instanton homology $I(Z_n)$ as an $\mathcal{R}$-module. We therefore obtain an isomorphism

$$I(Z_n) = H_*(\mathrm{Rep}(Z_n)) \otimes \mathcal{R}.$$

In the $\mathbb{Z}/4$ grading of the instanton homology, the minus component $\mathrm{Rep}(S^2_n)_-$ is shifted by 2 relative to the plus component. This is established in [32] for rational coefficients, but the argument extends to any coefficients, including our local coefficient system Γ. We record this in the following proposition.

Proposition 3.8 *As $\mathcal{R}$-modules with $\mathbb{Z}/4$ grading, we have an isomorphism,*

$$\Lambda\colon I_*(Z_n) = H_*(\mathrm{Rep}(S^2_n);\mathcal{R}) \oplus H_*(\mathrm{Rep}(S^2_n);\mathcal{R})[2]$$

for all odd $n \geq 1$. In particular, the instanton homology is a free $\mathcal{R}$-module and is nonzero only in even degrees mod 4.

The isomorphism Λ in the above proposition depends on the choice of perturbation (at least a priori), because the isomorphism goes by identifying both sides with the free $\mathcal{R}$-module generated by the critical points. The following two propositions add some additional structure. In the statement of the first proposition below, we write $\mathbb{1}_+ \in I(Z_n)$ for the relative invariant of the 4-dimensional orbifold $D^2 \times S^2_n$ with boundary Z^n:

$$\mathbb{1}_+ = I(D^2 \times S^2_n).$$

Proposition 3.9 *The instanton homology $I(Z_n)$ is a cyclic module for the filtered algebra $\mathcal{A}_n$ in (19), with cyclic generator the element $\mathbb{1}_+$.*

This proposition (whose proof is given below) prompts the following definition.

Definition 3.10 We write $\mathcal{J}_n \subset \mathcal{A}_n$ for the annihilator of the cyclic module $I(Z_n)$, so that

$$I(Z_n) \cong \mathcal{A}_n/\mathcal{J}_n.$$

From this description, the instanton homology $I(Z_n)$ inherits an increasing filtration from the filtration of $\mathcal{A}_n$:

$$I(Z_n)^{(m)} = (\mathcal{A}_n^{(m)} + \mathcal{J}_n)/\mathcal{J}_n.$$

Proposition 3.11 *The isomorphism Λ of Proposition 3.8 respects the filtrations, and the isomorphism on the associated graded is an isomorphism of $\mathcal{A}_n$-modules, independent of the choice of perturbations.*

We begin the proof of the two propositions above by describing the $\mathcal{A}_n$-module structure of $I(Z_n)$. Recall that the $\mathcal{A}_n$-module structure of $H_*(\mathrm{Rep}(Z_n);\mathcal{R})$ arises from operators α, $\delta_1, \dots, \delta_n$ (acting by cap product) and ϵ. The instanton homology $I(Z_n)$ carries parallel operators which we now make explicit.

First, the classes α, β and δ_p in $H^*(\mathcal{B}^*(Z_n);\mathbb{Q})$ correspond to operators on the Floer homology $I(Z_n)$ by the general construction (5). We write these operators as

$$(22)\qquad \tilde{\alpha}\colon I_*(Z_n)\to I_{*-2}(Z_n),\quad \tilde{\beta}\colon I_*(Z_n)\to I_{*-4}(Z_n)=I_*(Z_n),\quad \tilde{\delta}_p\colon I_*(Z_n)\to I_{*-2}(Z_n),$$

where the subscripts denote the mod 4 grading. In the notation of (5), these are the operators

$$\begin{aligned}\tilde{\alpha}&=I([0,1]\times Z_n,[S^2_n]),\\ \tilde{\beta}&=I([0,1]\times Z_n,[w])\quad\text{for } [w]\in H_0([0,1]\times Z_n),\\ \tilde{\delta}_p&=I([0,1]\times Z_n,[p])\quad\text{for } [p]\in H_0([0,1]\times K_n).\end{aligned}$$

Remark According to the results of [16], the operator $2\tilde{\delta}_p$ can be realized as the map corresponding to a cobordism W_1 from Z to Z, derived from the product cobordism $I\times Z$ by summing a standard torus to $I\times K$ at the point $\left(\frac{1}{2},p\right)$. The local orientation of K is used to fix a homology orientation of the torus.

The counterpart of the operator ϵ is a special case of the construction of $I(W,a)^e$. Specifically, following Street [32], it is the map (10) in the special case that W is the cylindrical cobordism, the element a is 1, and e is the class $[\{\text{point}\}\times S^2_n]$:

$$\tilde{\epsilon}=I([0,1]\times S^2_n)^e.$$

In order for the operators $\tilde{\alpha}$, $\tilde{\delta}_p$ and $\tilde{\epsilon}$ to make the instanton homology $I(Z_n)$ into a module over the algebra $\mathcal{A}_n$, we need to see that they satisfy the relations that are baked into the definition of $\mathcal{A}_n$. We turn to this next. The relation in Proposition 2.4 specializes to the following:

Lemma 3.12 *With $\mathcal{R}=\mathbb{Q}[\tau^{\pm1}]$ as usual, the actions of the operators $\tilde{\delta}_p$ and $\tilde{\beta}$ on the $\mathcal{R}$-module $I(Z_n)$ are related by*

$$\tilde{\delta}_p^2=-\tilde{\beta}+\tau^2+\tau^{-2}.$$

In particular, $\tilde{\delta}_p^2$ is independent of the chosen point p on the singular set of Z_n. □

The element ϵ in $\mathcal{A}_n$ has square 1 by definition, so we need the following lemma also.

Lemma 3.13 *The operator $\tilde{\epsilon}\colon I(Z_n)\to I(Z_n)$ has square 1, and under the isomorphism of Proposition 3.8 it corresponds to the interchange of the two summands.*

Proof This is proved in [32] for rational coefficients, except that an ambiguity in the orientation of the moduli spaces left the sign of $\tilde{\epsilon}^2$ unresolved there. (See also the proof of Proposition 3.14 below.) In our present context we have

$$\tilde{\epsilon}^2=I([0,1]\times S^2_n)^e\circ I([0,1]\times S^2_n)^e=I([0,1]\times S^2_n)^{2e}=(-1)^{e\cdot e}I([0,1]\times S^2_n)=1,$$

where the second equality is by functoriality and the third equality is from (11). □

The relations in Lemmas 3.12 and 3.13 are the same relations satisfied by the elements ϵ and δ_k in the algebra $\mathcal{A}_n$, so we can indeed use these operators to define an $\mathcal{A}_n$-module structure on $I(Z_n)$ by

$$(23)\qquad \alpha\mapsto\tilde{\alpha},\qquad \delta_i\mapsto\tilde{\delta}_i\quad\text{for } i=1,\dots,n,\qquad \epsilon\mapsto\tilde{\epsilon}.$$

Having described the module structure of $I(Z_n)$, the fact that it is a cyclic module generated by $\mathbb{1}_+$ (Proposition 3.9) and the assertions of Proposition 3.11 are both consequences of the fact that, under the isomorphism of Proposition 3.8, the operators $\widetilde{\alpha}$, $\widetilde{\delta}_p$ and $\widetilde{\epsilon}$ agree with the operators α, δ_p and ϵ on $H_*(\mathrm{Rep}(S^2_n))$ in their leading terms. This is the assertion of the proposition below, which is the final proposition of this subsection.

Proposition 3.14 *Let Λ be the isomorphism of Proposition 3.8. Then for any $\xi \in I(Z_n)^{(m)}$ and $u \in \mathcal{A}_n^{(k)}$, we have*

$$\Lambda(u\xi) = u\Lambda(\xi) \mod I(Z_n)^{(m+k-1)}, \tag{24}$$

and $\Lambda(\mathbb{1}_+) = 1_+$.

Proof It is enough to verify (24) in the case that u is one of the generators, α, δ_p or ϵ. The essential point is that $u\xi$ is defined using instantons on the cylinder $\mathbb{R} \times Z_n$ and that the leading term is defined by (perturbations of) the flat connections, while the nonleading terms are defined by instantons with positive action.

In more detail, let us write $\mathrm{Rep}(Z_n) = R_+ \cup R_-$, as an abbreviation for the components $\mathrm{Rep}(Z_n)_\pm$. Before any perturbations are made, we have seen that the two components $R_+ \cup R_-$ are copies of the representation variety $\mathrm{Rep}(S^2_n)$ of the orbifold sphere (equation (18)). For each $\kappa > 0$, let us write

$$M_\kappa(R_\pm, R_\pm)$$

for the moduli space of (unperturbed) instanton trajectories from one component of $\mathrm{Rep}(Z_n)$ to another, with action κ.

Lemma 3.15 (i) *The moduli spaces $M_\kappa(R_+, R_+)$ and $M_\kappa(R_-, R_-)$ are nonempty only for $\kappa \in \frac{1}{2}\mathbb{Z}$.*

(ii) *The moduli spaces $M_\kappa(R_+, R_-)$ and $M_\kappa(R_-, R_+)$ are nonempty only for $\kappa \in \frac{1}{2}\mathbb{Z} + \frac{1}{4}$.*

(iii) *The formal dimension of the moduli space, in every case, is $8\kappa + (2n-6)$.*

Proof The moduli spaces $M_\kappa(R_+, R_+)$ and $M_\kappa(R_-, R_-)$ are nonempty when $\kappa = 0$, consisting then of constant trajectories on the cylinder and forming a regular moduli space of dimension $2n-6$ (the dimension of the representation variety). For other values of κ, these moduli spaces are related to each other by gluing in instantons and monopoles, which will change κ by multiples of $\frac{1}{2}$ while always changing the formal dimension by 8κ; see [17; 20].

The formal dimension and action κ for the moduli spaces $M_\kappa(R_+, R_-)$ and $M_\kappa(R_-, R_+)$ are the same as for moduli spaces on the closed bifold $S^1 \times Z_n = T^2 \times S^2_n$ for a bundle with marking data where $w_2(E)$ is dual to the class $T^2 \times \{\text{point}\}$. The action in this case is equal to $\frac{1}{4}n$ modulo $\frac{1}{2}$, or in other words belongs to $\frac{1}{4} + \frac{1}{2}\mathbb{Z}$ since n is odd. (In the language of [17], the monopole number on each of the n components of the singular set is a half-integer.) The formula for the formal dimension in terms of the action κ is unchanged. □

After perturbation of the Chern–Simons functional, the manifolds R_+ and R_- each become a finite set of nondegenerate critical points, $\mathfrak{C}_+$ and $\mathfrak{C}_-$. The action of the perturbed instantons will be close to integer multiples of $\frac{1}{4}$ if the perturbation is small, so for critical points c and c' and $\kappa \in \frac{1}{4}\mathbb{Z}$ we continue to write $M_\kappa(c,c')$ for the perturbed moduli spaces. We have the dimension formula

$$\dim M_\kappa(c,c') = 8\kappa + \operatorname{index}(c) - \operatorname{index}(c'),$$

where index denotes the ordinary Morse index for the Morse function on $R_\pm$. Furthermore, the moduli space is nonempty only if $\kappa \in \frac{1}{2}\mathbb{Z}$ in the case that c, c' both belong to $\mathfrak{C}_+$ or to $\mathfrak{C}_-$, and only if $\kappa \in \frac{1}{4} + \frac{1}{2}\mathbb{Z}$ otherwise.

Consider now the operator $\tilde{\alpha}$ for example. (The case of $\tilde{\delta}_p$ is no different.) When $\kappa = 0$, the moduli space $M_0(c,c')$ between critical points $c, c' \in \mathfrak{C}_+$ or $c, c' \in \mathfrak{C}_-$ coincides with a perturbation of the space of ordinary Morse trajectories between the critical points in $R_\pm$. The construction of $\tilde{\alpha}$ means that we can write it as a sum

$$\tilde{\alpha} = \sum_{\kappa \in \frac{1}{4}\mathbb{Z},\, \kappa \geq 0} \tilde{\alpha}_{(\kappa)} \tag{25}$$

according to the contributions of the different moduli spaces M_κ. The matrix entry of $\tilde{\alpha}_{(0)}$ is the evaluation of the cohomology class α on the Morse trajectory space $M_0(c,c')$ between critical points on R_+ or R_- with $\operatorname{index}(c) - \operatorname{index}(c') = 2$. This is the cap product by the class α, under the isomorphism between Morse homology and singular homology. Thus we have

$$\Lambda(\tilde{\alpha}_{(0)}\xi) = \alpha\Lambda(\xi),$$

where ξ is the class corresponding to the critical point c. The dimension formula shows that the remaining terms $\Lambda(\tilde{\alpha}_{(\kappa)}\xi)$ for positive κ correspond to 2-dimensional moduli spaces $M_\kappa(c,c'')$ where the index difference $\operatorname{index}(c) - \operatorname{index}(c'')$ is 4 or more.

In the case of $\tilde{\epsilon}$, the equality (24) holds exactly. This is the content of Lemma 3.13. In the present context it can be understood by the same argument as applies to $\tilde{\alpha}$ and $\tilde{\delta}_p$, but with the additional observation that the moduli spaces of positive action contribute zero because of the action of translation on these moduli spaces. □

If we keep track of the difference between R_+ and R_- which is highlighted in part (ii) of Lemma 3.15, then we can extract a slightly more detailed statement from the proof of the proposition above. Recall that $J_n \subset \mathcal{A}_n$ is the annihilator of $H_*(\operatorname{Rep}(Z_n))$. (See Proposition 3.6.) In the following corollary, we also write

$$\mathcal{A}_n^+ \subset \mathcal{A}_n$$

for the subalgebra generated over $\mathcal{R}$ by α and $\delta_1, \dots, \delta_n$, so that

$$\mathcal{A}_n = \mathcal{A}_n^+ + \epsilon\mathcal{A}_n^+.$$

Corollary 3.16 *For any element $w \in J_n \cap \mathcal{A}_n^{(m)}$, there exists $\omega \in \mathcal{J}_n \cap \mathcal{A}_n^{(m)}$ with*

$$\omega - w \in \mathcal{A}_n^{(m-1)}.$$

More particularly, if w is a homogeneous element of degree m in the graded algebra $\mathcal{A}_n$, then ω can be taken to have the form

$$\omega = w(0) + w(2) + w(4) + \cdots + \epsilon\bigl(w(1) + w(3) + \cdots\bigr),$$

where $w(0) = w$ and $w(i) \in \mathcal{A}_n^{(m-i)} \cap \mathcal{A}_n^{+}$ is homogeneous of degree $m-i$ for all i. Furthermore, if $m \le \frac{1}{2}(n-1)$, then ω is uniquely determined by w.

Proof This follows from the proposition above and Proposition 3.7. □

3.5 The instanton homology of $Z_{n,-1}$

We now examine the bifold $Z_{n,-1}$; see Definition 3.2. The singular locus $K(Z_{n,-1})$ in this case is a knot in $S^1 \times S^2$, with winding number n. We still require n to be odd, so that this is an admissible bifold.

Proposition 3.17 *The representation variety of $Z_{n,-1}$ is nondegenerate and consists of $\frac{1}{4}(n^2-1)$ points.*

Proof The orbifold $Z_{n,-1}$ is a fiber bundle over the circle, with fiber the orbifold sphere S^2_n. The restriction map to the fiber,

$$\mathrm{Rep}(Z_{n,-1}) \to \mathrm{Rep}(S^2_n),$$

has image the set of representations in $\mathrm{Rep}(S^2_n)$ which are invariant under the action h_* of the monodromy of the circle bundle, $h\colon S^2_n \to S^2_n$. The latter is the map which rotates the sphere through $2\pi/n$. The restriction map is two-to-one, just as it is for Z_n, and for the same reason.

The fixed points of h_* are representations of the orbifold fundamental group of the quotient $\Sigma = S^2_n/\langle h\rangle$. This orbifold surface has one orbifold point of order 2 and two orbifold points of order n. For a spherical orbifold with three singular points, the representation variety consists of isolated points, and this is essentially the situation considered in [10] (for example). The enumeration of representations, as in [10], becomes an enumeration of lattice points in a region. (The same conclusion can also be reached by identifying the representations with stable parabolic bundles on a curve of genus 0 with appropriate parabolic structure at the orbifold points. See Section 4.1.) In this particular case, the number of representations of the orbifold fundamental group of $S^2_n/\langle h\rangle$ is $\frac{1}{8}(n^2-1)$, and $\mathrm{Rep}(Z_{n,-1})$ therefore consists of $\frac{1}{4}(n^2-1)$ points. The nondegeneracy of the former leads to the nondegeneracy of the latter. □

We can view $K(Z_{n,-1})$ as the closure of a braid in $S^1 \times D^2 \subset S^1 \times S^2$ whose braid diagram has $n-1$ negative crossings. There is therefore a cobordism W of bifolds, from $Z_{n,-1}$ to Z_n, obtained by smoothing each of the crossings. We can write W as a composite of $n-1$ cobordisms, $W_1, \dots, W_{n-1}$,

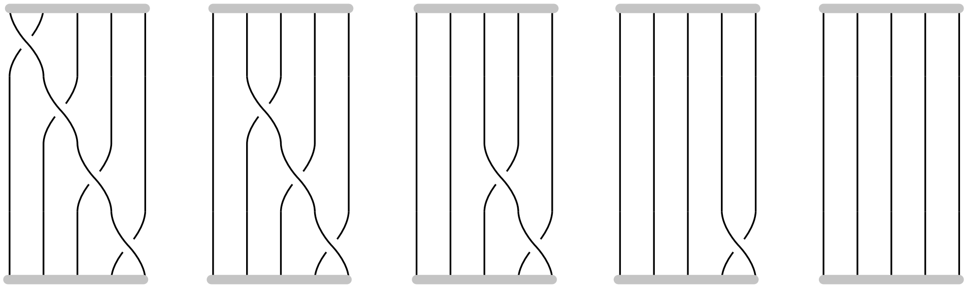


Figure 2: The composite cobordism from $Z_{n,-1}$ to Z_n, illustrated for $n = 5$.

in the order illustrated in Figure 2. The intermediate bifolds each correspond to braids with k "straight" strands and $n-k$ braided strands: a side-by-side juxtaposition of Z_k and $Z_{n-k,-1}$, which we temporarily denote by $Z_k * Z_{n-k,-1}$ (with the understanding that Z_0 is $S^1 \times S^2$ with an empty link). So we have

$$I(W_k)\colon I(Z_{k-1} * Z_{n-k+1,-1}) \to I(Z_k * Z_{n-k,-1}) \quad \text{for } k = 1, \dots, n-1.$$

(Note that, when $k = n-1$, we have $Z_k * Z_{n-k,-1} \cong Z_n$.)

Proposition 3.18 *For each odd n and each $k \le n-1$, the induced map $I(W_k)$ is an inclusion of one free $\mathcal{R}$-module in another, as a direct summand.*

Proof The cobordism W_k is one map in a skein exact triangle [23; 19], in which the third instanton homology group is $I(X_{n,k})$, where $X_{n,k}$ is a braid as shown in Figure 3. Thus,

(26) $$\dots \xrightarrow{c} I(Z_{k-1} * Z_{n-k+1,-1}) \to I(Z_k * Z_{n-k,-1}) \to I(X_{n,k}) \xrightarrow{c} \cdots$$

is a long exact sequence.

After an isotopy, we have, for $k \le n-2$,

(27) $$X_{n,k} = Z_{k-1} * Z_{n-2-k+1,-1}.$$

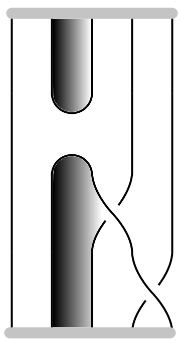


Figure 3: The third braid $X_{n,k}$ in the exact triangle, illustrated in the case $n = 5$ and $k = 2$. The shaded region (which is connected in a projection of $S^1 \times S^2$) can be eliminated by a Reidemeister I move.

The case $k = n-1$ is slightly different: in this case $X_{n,n-1}$ is the connected sum of Z_{n-2} and the bifold obtained from an unknot in S^3. (See Figure 3 again.) From another application of the skein triangle, we have an exact sequence

$$\cdots \to I(Z_{n-2}) \to I(X_{n,n-1}) \to I(Z_{n-2}) \to \cdots .$$

The connecting homomorphism $I(Z_{n-2}) \to I(Z_{n-2})$ has odd degree in the mod 2 grading, while $I(Z_{n-2})$ is supported in even gradings only. So the connecting homomorphism is zero and so

$$\operatorname{rank}_{\mathcal{R}} I(X_{n,n-1}) = 2 \operatorname{rank}_{\mathcal{R}} I(Z_{n-2}).$$

For brevity, let us write

$$f(n,k) = \operatorname{rank}_{\mathcal{R}} I(Z_k * Z_{n-k,-1}) \quad \text{and} \quad x(n,k) = \operatorname{rank}_{\mathcal{R}} I(X_{n-k}).$$

From the long exact sequence (26), we obtain

$$f(n,k) \le f(n,k-1) + x(n,k)$$

with equality if and only if the connecting homomorphism c has rank zero. We have also seen that

$$x(n,k) = \begin{cases} f(n-2,k-1) & \text{if } k \le n-2, \\ 2f(n-2,n-2) & \text{if } k = n-1. \end{cases}$$

From these we inductively obtain

$$f(n,n) \le \sum_{p=0}^{(n-1)/2} \binom{n}{p} f(n-2p,0). \tag{28}$$

The quantities $f(n-2p,0)$ are the ranks of the instanton homologies of $Z_{n-2p,-1}$, which are bounded above by the number of generators, which in turn can be read off from Proposition 3.17:

$$f(n-2p,0) \le \tfrac{1}{4}((n-2p)^2-1). \tag{29}$$

Combining this with the previous inequality we have

$$f(n,n) \le \sum_{p=0}^{(n-1)/2} \binom{n}{p} \cdot \tfrac{1}{4}((n-2p)^2-1). \tag{30}$$

On the other hand, we know what $f(n,n)$ is: it is twice $P(1)$, where P is the Poincaré polynomial of the representation variety of S^2_n, given by the recursive formula (12). From that recursion, we can verify the closed formula

$$f(n,n) = 2^{n-3}(n-1). \tag{31}$$

But the sum on the right-hand side of (30) is also $2^{n-3}(n-1)$, as is easily verified by comparing it to the second derivative of $(t+t^{-1})^n$ at $t=1$. It follows that the inequalities above are all equalities. In particular, from the equality in (30), we learn that $I(Z_{n,-1})$ is a free module of rank $(n^2-1)/4$. It follows that the connecting homomorphisms c in the exact sequences (26) all have rank zero. An inductive argument now shows the modules in the exact sequences are all free $\mathcal{R}$-modules and the connecting homomorphisms are all zero. The proposition follows. □

The following corollary summarizes the conclusions of the previous propositions.

Corollary 3.19 *The instanton homology $I(Z_{n,-1})$ with local coefficients is a free $\mathcal{R}$-module of rank $(n^2-1)/4$, supported in even degrees mod 4. The cobordism $W: Z_{n,-1} \to Z_n$ induces a map $I(W)$ on instanton homology with local coefficients,*

$$I(W)\colon I(Z_{n,-1}) \to I(Z_n),$$

which is an inclusion of this free $\mathcal{R}$-module as a direct summand. □

The bifold obtained from $Z_{n,-k}$ by reversing the orientation is $Z_{n,k}$, and by dualizing the above corollary we obtain:

Corollary 3.20 *The instanton homology $I(Z_{n,1})$ with local coefficients is also a free $\mathcal{R}$-module of rank $(n^2-1)/4$. The cobordism $W^\dagger: Z_n \to Z_{n,1}$ induces a surjective map $I(W^\dagger)$ on these free modules.* □

On the other hand, we have Lemma 3.3 which identifies $Z_{n,-1}$ and $Z_{n,1}$ in an orientation-preserving manner by an isotopy. So we have another variant of the corollary:

Corollary 3.21 *There is a surjective homomorphism of free $\mathcal{R}$-modules from $I(Z_n)$ to $I(Z_{n,-1})$ obtained from a cobordism between the links $K(Z_n)$ and $K(Z_{n,-1})$ inside $[0,1]\times S^1\times S^2$.* □

Like Z_n, the bifold $Z_{n,-1}$ contains a copy S of the orbifold sphere S^2_n intersecting the singular locus in n points. By the general constructions of Section 2.3, this gives rise to operators $\tilde{\alpha}$, $\tilde{\delta}_1, \dots, \tilde{\delta}_n$ and $\tilde{\epsilon}$, acting on $I(Z_{n,-1})$ just as in the case of $I(Z_n)$, making $I(Z_{n,-1})$ also an $\mathcal{A}_n$-module. Note that the n points of intersection with S all lie on the same component of the singular locus $K(Z_{n,-1})$ (which is now a knot, not a link). The operators $\tilde{\delta}_p$ are therefore all equal on $I(Z_{n,-1})$, and we will sometimes write this operator as $\tilde{\delta}$.

Proposition 3.22 *With the instanton module structure in which $\alpha, \delta_i, \epsilon \in \mathcal{A}_n$ act by the operators $\tilde{\alpha}$, $\tilde{\delta}$ and $\tilde{\epsilon}$, the instanton homology $I(Z_{n,-1})$ is a cyclic module for the algebra $\mathcal{A}_n$ and can therefore be described as a quotient,*

$$I(Z_{n,-1}) \cong \mathcal{A}_n/\mathcal{J}_{n,-1}.$$

The ideal $\mathcal{J}_{n,-1}$ contains the ideal $\mathcal{J}_n$ as well as the elements $\delta_i - \delta_j$.

Proof We have seen that there is a cobordism from Z_n to $Z_{n,-1}$ inducing a surjection on instanton homology (Corollary 3.21). The proposition follows from this and the above remark that the actions of the $\tilde{\delta}_i$ are all equal. □

It is helpful here to introduce the smaller algebra

$$\bar{\mathcal{A}} = \mathcal{A}_n / \langle \delta_i - \delta_j \rangle_{i,j},$$

which we can write simply as

$$\bar{\mathcal{A}} = \mathcal{R}[\alpha, \delta, \epsilon]/\langle \epsilon^2 - 1 \rangle, \tag{32}$$

where δ denotes the image of the δ_i in the quotient ring. The algebra $\bar{\mathcal{A}}$ described this way is independent of n. The above proposition then can be recast as

$$I(Z_{n,-1}) \cong \bar{\mathcal{A}}/\bar{\mathcal{J}}_{n,-1}, \tag{33}$$

where $\bar{\mathcal{J}}_{n,-1}$ is the image of $\mathcal{J}_{n,-1}$ in $\bar{\mathcal{A}}$.

Our main goal in this paper is to identify $I(Z_n)$ and $I(Z_{n,-1})$ completely, by describing the ideals $\mathcal{J}_n \subset \mathcal{A}_n$ and $\bar{\mathcal{J}}_{n,-1} \subset \bar{\mathcal{A}}$. In particular, as described in the introduction, we will eventually provide a set of generators of $\bar{\mathcal{J}}_{n,-1}$ in closed form, as minors of an explicit matrix.

4 Relations in ordinary cohomology

4.1 Loci in families of parabolic bundles on S^2_n

Recall from Proposition 3.6 the description of the cohomology ring of the representation variety

$$\mathrm{Rep}(Z_n) = \mathrm{Rep}(S^2_n) \cup \mathrm{Rep}(S^2_n)$$

as a quotient $\mathcal{A}_n/J_n$, where J_n is an ideal. (The coefficient ring here, as in Proposition 3.6, is $\mathcal{R}$, though at this point our calculations will involve only $\mathbb{Q}$, so rational coefficients would suffice.) The Betti numbers of $\mathrm{Rep}(S^2_n)$ were calculated recursively by Boden [3], and a full presentation of the cohomology ring (in a more general case) is described in [8]. Generators for the ideal of relations in the specific case of $\mathrm{Rep}(S^2_n)$ are given by Street [32]. We shall describe a particular source of such relations, arising from a mechanism first pointed out by Mumford in the smooth case [1]. (In [8] it is shown that essentially the same mechanism gives rise to a complete set of relations in the orbifold case.)

As stated earlier, although we have taken SO(3) connections as our starting point, the representation variety $\mathrm{Rep}(S^2_n)$ can be identified with the space of flat SU(2) connections having monodromy of order 4 at each of the n punctures. In turn, this representation variety can be identified with a moduli space of stable parabolic bundles by the results of [25]. We adopt the following conventions to make this more specific in the rank-2 case, following [17; 18].

We consider a compact Riemann surface S equipped with a set of distinguished points $\pi = \{p_1, \dots, p_n\}$, and a parameter $\alpha \in \left(0, \tfrac{1}{2}\right)$. Given a fixed holomorphic line bundle $\Theta \to S$ (usually trivial in our case), we study rank-2 holomorphic bundles $\mathcal{E} \to S$ with $\Lambda^2\mathcal{E} = \Theta$, together with a filtration of the rank-2 fiber at each $p \in \pi$ determined by a choice of a one-dimensional subspace (a line) $\mathcal{L}_p \subset \mathcal{E}_p$. The data $(\mathcal{E}, \mathcal{L}_{p_1}, \dots, \mathcal{L}_{p_n}, \alpha)$ is a bundle with parabolic structure. Given a line subbundle $\mathcal{F} \subset \mathcal{E}$, the *parabolic degree* of $\mathcal{F}$ is defined by

$$\text{par-deg}\,\mathcal{F} = c_1(\mathcal{F})[S] + \sum_{\pi} \pm\alpha, \tag{34}$$

where we take $+\alpha$ in the sum when $\mathcal{F}$ contains $\mathcal{L}_p$ at p and $-\alpha$ when it does not. The parabolic bundle is *semistable* if

$$\text{par-deg}\,\mathcal{F} \le \tfrac{1}{2}\deg\Theta$$

for every line subbundle $\mathcal{F}$, and is *stable* if strict inequality holds. At present we will take Θ to be trivial and *we are only concerned with the special case* $\alpha = \frac{1}{4}$. In this case, when n is odd, all semistable bundles are strictly stable, and the moduli space of stable parabolic bundles is a projective variety of complex dimension $3g-3+n$. In the case of genus 0, we write $\mathcal{M}(S^2_n)$ for this projective variety: the moduli space of stable parabolic bundles, with parabolic structure at the n marked points and $\alpha = \frac{1}{4}$.

With this notation understood, the theorem of [25] identifies the representation variety $\text{Rep}(S^2_n)$ for odd n with the moduli space of stable parabolic bundles:

$$\text{Rep}(S^2_n) \cong \mathcal{M}(S^2_n).$$

Suppose now that we have a family of parabolic bundles on S^2_n parametrized by a space T. This means that we have a rank-2 bundle,

$$\boldsymbol{\mathcal{E}} \to T \times S^2,$$

with $\Lambda^2\boldsymbol{\mathcal{E}} \cong \Phi \boxtimes \Theta$ (with Θ still trivial on S^2 at the moment, but Φ a nontrivial line bundle on the base T), together with line subbundles

$$\boldsymbol{\mathcal{L}}_p \subset \boldsymbol{\mathcal{E}}|_{T\times p} \quad \text{for} \quad p \in \pi.$$

The bundle $\boldsymbol{\mathcal{E}}$ is equipped with a holomorphic structure on each $\{t\} \times S^2$, giving rise to parabolic bundles $\mathcal{E}_t$.

In such a family over T, we can consider the locus of those $t \in T$ where the parabolic bundle $\mathcal{E}_t$ is unstable (for $\alpha = \frac{1}{4}$). From the definition at (34), being unstable means the following.

(i) We have a holomorphic line bundle $\mathcal{F} \to S^2$, of degree f say, necessarily the bundle $\mathcal{O}(f)$.

(ii) We have a subset $\eta \subset \pi$, whose cardinality we denote by h.

(iii) There is a nonzero holomorphic map $\iota\colon \mathcal{F} \to \mathcal{E}_t$ such that $\iota(\mathcal{F}|_p) \subset \mathcal{L}_t|_p$ for all $p \in \eta$.

(iv) We have $f + \frac{1}{4}(2h-n) > 0$.

Altering this slightly, given any $\lambda \in \mathbb{R}$, we make the following definition.

Definition 4.1 Let $\eta \subset \pi = \{p_1, \dots, p_n\}$ be any subset, and write $h = |\eta|$ for its cardinality. Let λ be an odd multiple of $\frac{1}{4}$ satisfying the additional constraint that

$$h = \tfrac{1}{2}(n - 4\lambda) \pmod 2. \tag{35}$$

This being so, there is $f \in \mathbb{Z}$ such that

$$f + \tfrac{1}{4}(2h - n) = -\lambda. \tag{36}$$

Let $\mathcal{F} \to S^2$ be the line bundle $\mathcal{O}(f)$. Given a family of parabolic bundles on S^2_n parametrized by T as above, we define

$$T^\eta_\lambda \subset T \tag{37}$$

to be the locus of points $t \in T$ such that there is a nonzero holomorphic map $\iota\colon \mathcal{F} \to \mathcal{E}_t$ with $\iota(\mathcal{F}|_p) \subset \mathcal{L}_t|_p$ for all $p \in \eta$.

This definition is set up so that the unstable locus is the union

$$\bigcup_{\eta;\, \lambda \le -\frac{1}{4}} T^\eta_\lambda.$$

The definition of the locus T^η_λ is readily rephrased as the statement that a certain Fredholm operator P_t (defined below at (41), and determined by the parabolic bundle $\mathcal{E}_t$ and the choice of λ and η) has nonzero kernel. If we suppose that the resulting map

$$P : T \to \mathrm{Fred}$$

is transverse to the stratification of the space of Fredholm operators by the dimension of the kernel, then the locus $T^\eta_\lambda \subset T$ will itself be a stratified space whose Poincaré dual is a cohomology class that one can calculate using the index theorem for families. With slight abuse of notation, we write (37) as

$$T^\eta_\lambda = T \cap U^\eta_\lambda,$$

where U^η_λ denotes the locus where the Fredholm operator has kernel. It will also be useful to group together the different subsets η according to their size $h = |\eta|$, so that we write (with a slight further abuse of notation),

$$U^h_\lambda = \bigcup_{|\eta| = h} U^\eta_\lambda \quad \text{and} \quad T^h_\lambda = T \cap U^h_\lambda.$$

Again, this locus is nonempty only if h satisfies the parity condition (35).

We now compute the Chern classes of the index of the family of operators P in order to derive a formula for the class dual to the stratum T^η_λ. Note that if P is a family of complex Fredholm operators of index $-k+1$, then (assuming transversality) the locus where P_t has kernel is dual to

$$c_k(-\mathrm{index}(P)) \in H^{2k}(T). \tag{38}$$

(This is the first case of Porteous's formula in the case of Fredholm maps [29; 15].)

It is evident from the definition that the locus T^{η}_{λ} is unchanged if the family of bundles $\mathcal{E}$ is modified by tensoring with a line bundle pulled back from the base T. Recall that we have written $\Lambda^2\mathcal{E} = \Phi \boxtimes \Theta$, where $\Phi \to T$ is a line bundle and Θ is taken to be trivial. If Φ has a square root, we may tensor by $\Phi^{-1/2}$ to make $c_1(\mathcal{E}) = 0$. Although a square root will not exist in general, the calculation below is not invalidated by assuming that $c_1(\mathcal{E}) = 0$, and we will make this simplification from here on. This means in particular that $c_2(\mathcal{E}) = -p_1(\operatorname{ad}\mathcal{E})/4$. Let us then write

$$c_2(\mathcal{E}) = \beta \times 1 + \hat{\alpha} \times v \in H^4(T \times S^2),$$

where v is the unit volume form on S^2. From the binomial theorem, we have

$$c_2(\mathcal{E})^r = \beta^r \times 1 + r\hat{\alpha}\beta^{r-1} \times v. \tag{39}$$

The class $\hat{\alpha}$ here does not quite correspond to the class α in (14), because the latter was defined using the orbifold Pontryagin class. The relation between the two is:

$$\hat{\alpha} = \alpha - \frac{1}{2}\sum_{p\in\pi} \delta_p. \tag{40}$$

For each $p \in \pi$ we also have the line subbundle $\mathcal{L}_p$ and the quotient line bundle $\mathcal{Q}_p = (\mathcal{E}|_{T\times p})/\mathcal{L}_p$, and from these we obtain the cohomology class

$$\delta_p = \tfrac{1}{2}(c_1(\mathcal{Q}_p) - c_1(\mathcal{L}_p)).$$

The definition is set up so that $2\delta_p$ coincides with the Euler class of the oriented rank-2 subbundle of $\operatorname{ad}(\mathcal{E}|_{T\times p})$ determined by $\mathcal{L}_p$.

Fix a holomorphic line bundle $\mathcal{F} \cong \mathcal{O}(f)$ on S^2. We are seeking a nonzero holomorphic map $\iota\colon \mathcal{F} \to \mathcal{E}_t$ such that the composite with the quotient map,

$$\mathcal{F} \to \mathcal{E}_t \xrightarrow{\pi_p} \mathcal{Q}_{(t,p)},$$

vanishes for all $p \in \eta$. That is, $\iota \in \Omega^{0,0}(\mathcal{F}^* \otimes \mathcal{E}_t)$ lies in the kernel of the map

$$P_t\colon \Omega^{0,0}(\mathcal{F}^* \otimes \mathcal{E}_t) \to \Omega^{0,1}(\mathcal{F}^* \otimes \mathcal{E}_t) \oplus \Bigl(\bigoplus_{p\in\eta} \mathcal{Q}_{(t,p)}\Bigr) \tag{41}$$

given by $\iota \mapsto \bigl(\bar\partial\iota, \sum_{p\in\eta} \pi_p \circ \iota(p)\bigr)$. So, for the family of Fredholm operators P that we are interested in,

$$\operatorname{index}(P) = \operatorname{index}(\bar\partial_{\mathcal{F}^*\otimes\mathcal{E}}) - \sum_{p\in\eta}[\mathcal{Q}_p],$$

where the first part is the ordinary family $\bar\partial$ operators. From the index theorem for families, we have

$$\operatorname{ch}(\operatorname{index}(P)) = \bigl(\bigl(\operatorname{Todd}(S^2) \cup \operatorname{ch}(\mathcal{F}^* \otimes \mathcal{E})\bigr)/[S^2]\bigr) - \sum_{p\in\eta} \operatorname{ch}(\mathcal{Q}_p). \tag{42}$$

To compute the Chern characters that appear on the right-hand side of this formula, we introduce formal Chern roots $\pm\rho \in H^2(T \times S^2; \mathbb{Q})$ so that $c_2(\mathcal{E}) = -\rho^2$. Then we can write

$$\operatorname{ch}(\mathcal{E}) = e^{-\rho} + e^{\rho} = 2\cosh(\sqrt{-c_2(\mathcal{E})}),$$

and a short calculation using (39) yields

$$\operatorname{ch}(\mathcal{E}) = 2\cosh(\sqrt{-\beta}) - v\frac{\sinh(\sqrt{-\beta})}{\sqrt{-\beta}}\widehat{\alpha}.$$

We also have

$$\operatorname{ch}(\mathcal{F}^*) = 1 - fv \quad \text{and} \quad \operatorname{ch}(\mathcal{Q}_p) = e^{\delta_p}.$$

Finally, on the right-hand side of (42) we have $\operatorname{Todd}(S^2) = 1 + v$. Assembling these and calculating the slant product by $[S^2]$, we find

$$\operatorname{ch}(\operatorname{index}(P)) = (2 - 2f - h)\cosh(\sqrt{-\beta}) - \frac{\sinh(\sqrt{-\beta})}{\sqrt{-\beta}}\Big(\widehat{\alpha} + \sum_{p\in\eta}\delta_p\Big),$$

where h is the number of elements of η. If we use the equality of (36), and if we substitute α for $\widehat{\alpha}$ using the relation (40), we obtain:

$$\operatorname{ch}(-\operatorname{index}(P)) = \big(\tfrac{1}{2}n - 2\lambda - 2\big)\cosh(\sqrt{-\beta}) + \frac{\sinh(\sqrt{-\beta})}{\sqrt{-\beta}}\Big(\alpha + \frac{1}{2}\sum_{p\in\eta}\delta_p - \frac{1}{2}\sum_{p\notin\eta}\delta_p\Big). \tag{43}$$

If we recall that $\delta_p^2 = -\beta$ for all p, then we can equivalently write this formula as

$$\operatorname{ch}(-\operatorname{index}(P)) = \big(\tfrac{1}{2}n - 2\lambda - 2\big)\cosh(\delta_1) + \frac{\sinh(\delta_1)}{\delta_1}\Big(\alpha + \frac{1}{2}\sum_{p\in\eta}\delta_p - \frac{1}{2}\sum_{p\notin\eta}\delta_p\Big), \tag{44}$$

or in abbreviated form as

$$\operatorname{ch}(-\operatorname{index}(P)) = i_\lambda\cosh(\delta_1) + \frac{\sinh(\delta_1)}{\delta_1}B_\eta, \tag{45}$$

where i_λ and B_η are the indicated subexpressions of (44). Note that i_λ is minus the numerical index of P.

The above formula defines a graded infinite sum of elements of the algebra

$$A_n = \mathbb{Q}[\alpha, \delta_1, \dots, \delta_n]/\langle\delta_i^2 - \delta_j^2\rangle_{i,j} = H^*(\mathcal{B}^*(S^2_n); \mathbb{Q})$$

(see Definition 3.5), thus an element of the formal completion

$$\widehat{H}^*(\mathcal{B}^*(S^2_n); \mathbb{Q}) \supset H^*(\mathcal{B}^*(S^2_n); \mathbb{Q}).$$

By the usual formulae expressing elementary symmetric polynomials in terms of power sums, there is a map

$$\mathfrak{c}_k : \widehat{H}^*(\mathcal{B}^*(S^2_n); \mathbb{Q}) \to H^{2k}(\mathcal{B}^*(S^2_n); \mathbb{Q})$$

such that $\mathfrak{c}_k(\operatorname{ch}(V)) = c_k(V)$ for any V, and so we have explicit formulae for

$$c_k(-\operatorname{index}(P)) \in H^*(\mathcal{B}^*(S^2_n); \mathbb{Q}),$$

given as $\mathfrak{c}_k(r)$, where r is the right-hand side of (44). The case we are interested in from (38) is the Chern class c_k, where $-k+1$ is the numerical index of P. From the constant term in the formula for the Chern character above, we read

$$k = \tfrac{1}{2}n - 2\lambda - 1. \tag{46}$$

So we make the following definition.

Definition 4.2 Given λ an odd multiple of $\frac{1}{4}$ and given a subset $\eta \subset \pi = \{p_1, \dots, p_n\}$ of size h, where h satisfies the parity condition (35), let k be the integer given by (46), and denote by

$$w^k_{n,\eta} \in H^*(\mathcal{B}^*(S^2_n); \mathbb{Q}) \subset \mathcal{A}_n$$

the element $\mathfrak{c}_k(r)$, where r is the right-hand side of (44).

To illustrate the general shape of the answers here, we take $n = 5$. When $\lambda = -\frac{1}{4}$, the value of k is 2. The parity condition allows the size of η to be 1, 3 or 5, and we have

$$w^2_{5,\eta} = \tfrac{1}{2}\bigl(\bigl(\alpha + \tfrac{1}{2}(\pm\delta_1 \pm \delta_2 \pm \delta_3 \pm \delta_4 \pm \delta_5)\bigr)^2 - \delta_1^2\bigr),$$

where the sign is $+$ when $p_i \in \eta$ and $-$ otherwise. When $\lambda = \frac{1}{4}$, the value of k is 1, and the parity condition allows the size of η to be 0, 2 or 4. We have

$$w^1_{5,\eta} = \alpha + \tfrac{1}{2}(\pm\delta_1 \pm \delta_2 \pm \delta_3 \pm \delta_4 \pm \delta_5).$$

Our definition means that, in $H^*(T;\mathbb{Q})$, we have $c_k(-\text{index } P) = \varphi(w^k_{n,\eta})$, where $\varphi\colon A_n \to H^*(T;\mathbb{Q})$ is the natural map (given, with slight abuse of notation, by $\alpha \mapsto \alpha$ and $\delta_p \mapsto \delta_p$).

Corollary 4.3 *Let $(\mathcal{E}, \mathcal{L}) \to T \times S^2$ be a family of parabolic bundles on S^2_n parametrized by T. Let λ and η be given, satisfying the conditions in Definition 4.2, and let $T^\eta_\lambda \subset T$ be the locus defined by (37). Assume that the corresponding family of Fredholm operators P is transverse to the stratification by the dimension of the kernel. Then the cohomology class dual to this stratum is given by*

$$\mathrm{PD}[T^\eta_\lambda] = \varphi(w^k_{n,\eta}),$$

where φ is the natural linear map $A_n \to H^*(T;\mathbb{Q})$, and k is given in terms of n and λ by (46).

Remarks In Definition 4.1, the loci T^η_λ are characterized by the existence of a holomorphic map $\iota\colon \mathcal{F} \to \mathcal{E}$ satisfying additional constraints at the distinguished points $\eta \subset \pi$. In the language of parabolic bundles, we can regard $\mathcal{F}$ as a line bundle with parabolic structure described by a subsheaf $\mathcal{F}_1 \subset \mathcal{F}$ such that in a neighborhood $\mathcal{U}_p$ of each $p \in \pi$ we have

$$\begin{aligned} \mathcal{F}_1|_{\mathcal{U}_p} &= \mathcal{F}|_{\mathcal{U}_p} && \text{if } p \in \eta, \\ \mathcal{F}_1|_{\mathcal{U}_p} &= (\mathcal{F} \otimes \mathcal{O}[-p])|_{\mathcal{U}_p} && \text{if } p \notin \eta. \end{aligned}$$

In these terms, what T^η_λ describes is the existence of a map $\mathcal{F} \to \mathcal{E}$ of parabolic bundles: ie a map which respects the filtrations. When regarded as a line bundle with parabolic structure in this way, we shall call $\eta \subset \pi$ the set of "hits" for $\mathcal{F}$.

4.2 The Mumford relations

As a consequence of Corollary 4.3, we have the following statement, which is the essential mechanism in Mumford's relations. (See the discussion in [1] for the earlier history of such relations.)

Proposition 4.4 *Let $(\mathcal{E}, \mathcal{L})$ be a family of parabolic bundles on S^2_n parametrized by a space T as in the previous subsection. Suppose that for every $t \in T$ the parabolic bundle $(\mathcal{E}_t, \mathcal{L}_t)$ on S^2_n is stable (with $\alpha = \frac{1}{4}$ as always). Then for any λ and η satisfying the conditions in Definition 4.2, with $\lambda < 0$, we have*

$$\varphi(w^k_{n,\eta}) = 0 \quad \text{in } H^{2k}(T; \mathbb{Q}),$$

where $k = \frac{1}{2}n - 2\lambda - 1$ and $\varphi\colon H^*(\mathcal{B}^*(S^2_n); \mathbb{Q}) \to H^*(T; \mathbb{Q})$ is the natural map determined by the characteristic classes of $\mathcal{E}$ and $\mathcal{L}$.

Proof When $\lambda < 0$, the stratum T^η_λ consists of unstable parabolic bundles, so the hypothesis of the proposition means that such strata are empty. The transversality condition is then vacuously satisfied and the result follows from Corollary 4.3. □

Proposition 4.5 *Let $\lambda = -\frac{1}{4}$ and let $\eta \subset \pi = \{p_1, \dots, p_n\}$ be a subset whose size h satisfies*

$$h = \tfrac{1}{2}(n+1) \bmod 2 \quad \text{and} \quad 0 \le h \le n. \tag{47}$$

(The first condition is the parity condition (35) for $\lambda = -\frac{1}{4}$.) *As in Definition 3.5, let j_n be the kernel of the restriction map to cohomology of the representation variety, $H^*(\mathrm{Rep}(S^2_n); \mathbb{Q})$. Then we have*

$$w^m_{n,\eta} \in j_n$$

for $m = \frac{1}{2}(n-1)$. That is, $w^m_{n,\eta}$ is a relation in the cohomology ring of $\mathrm{Rep}(S^2_n)$.

Proof This follows from the previous proposition by specializing to the case $\lambda = -\frac{1}{4}$, because $\mathrm{Rep}(S^2_n) \cong \mathcal{M}(S^2_n)$ parametrizes a family of stable parabolic bundles. □

Definition 4.6 Let $j_n \subset A_n$ be again the ideal of relations in the cohomology of $\mathrm{Rep}(S^2_n)$. With $m = \frac{1}{2}(n-1)$ and $\eta \subset \pi$ a subset whose size h satisfies the parity condition (47), we refer to the relation $w^m_{n,\eta} \in j_n$ as a *Mumford relation*. The collection of all these, as η varies, are the *Mumford relations* in the cohomology ring of $\mathrm{Rep}(S^2_n)$.

4.3 Explicit formulae

The elements $w^m_{n,\eta} \in A_n$ appearing as the Mumford relations, and more generally the cohomology classes $w^k_{n,\eta}$, have been described using a characterization that does not immediately yield explicit formulae. In particular, $w^k_{n,\eta}$ is defined in terms of a Chern *class* of an index element, while the explicit formula (44) provides the Chern *character* in closed form.

As a first step towards a closed formula for $w^k_{n,\eta}$, as in [36] for example, and following [37], a formula for the total Chern class can be derived as the formal series

$$\sum_{k=0}^{\infty} c_k(-\mathrm{index}(P)) = (1+\beta)^{i_\lambda/2} \left(\frac{1+\delta_1}{1-\delta_1}\right)^{B_\eta/(2\delta_1)}, \tag{48}$$

where i_λ and B_η are as in (44):

$$i_\lambda = \left(\tfrac{1}{2}n - 2\lambda - 2\right) \quad \text{and} \quad B_\eta = \alpha + \frac{1}{2}\sum_{p\in\eta}\delta_p - \frac{1}{2}\sum_{p\notin\eta}\delta_p. \tag{49}$$

(See [37] for the interpretation of the right-hand side of this formula.) We can therefore write

$$w^k_{n,\eta} = \frac{1}{k!}\left(\frac{d^k}{dt^k}\right)\bigg|_{t=0}\left((1+t^2\beta)^{i_\lambda/2}\left(\frac{1+t\delta_1}{1-t\delta_1}\right)^{B_\eta/(2\delta_1)}\right). \tag{50}$$

Note here that i_λ is minus the numerical index of P, and that in the definition of $w^k_{n,\eta}$ the integer k is $-\mathrm{index}(P)+1$, so we can write

$$w^k_{n,\eta} = \frac{1}{k!}\left(\frac{d^k}{dt^k}\right)\bigg|_{t=0}\left((1+t^2\beta)^{(k-1)/2}\left(\frac{1+t\delta_1}{1-t\delta_1}\right)^{B_\eta/(2\delta_1)}\right). \tag{51}$$

The following proposition gives a closed formula for this k^{th} term in the power series.

Proposition 4.7
$$k!\,w^k_{n,\eta} = \prod_{\substack{j=-k+1 \\ j=-k+1 \bmod 2}}^{k-1} (B_\eta + j\delta_1).$$

Proof Let us write

$$C_k = k!\,w^k_{n,\eta} = \left(\frac{d^k}{dt^k}\right)\bigg|_{t=0} G_{k-1}(t), \quad \text{where } G_{k-1}(t) = (1+t^2\beta)^{(k-1)/2}\left(\frac{1+t\delta}{1-t\delta}\right)^{B/(2\delta)},$$

and we have abbreviated

$$B = B_\eta \quad \text{and} \quad \delta = \delta_1$$

to streamline the notation.

Let $\widehat{C}_k$ denote the right-hand side in the proposition,

$$\widehat{C}_k = \prod_{\substack{j=-k+1 \\ j=-k+1 \bmod 2}}^{k-1} (B + j\delta),$$

so that what we aim to prove is that C_k and $\widehat{C}_k$ are equal. We shall prove in fact that

$$\frac{d^k}{dt^k} G_{k-1}(t) = \widehat{C}_k G_{-k-1}(t), \tag{52}$$

which yields the desired equality $C_k = \widehat{C}_k$ on substituting $t=0$, since $G_l(0)=1$ for all l.

We prove (52) by induction on k: specifically, assuming that (52) holds for k, we prove the result for $k+2$. The seed cases, $k=0,1$, are clear. Note first that $\widehat{C}_k$ satisfies a recurrence relation

$$\widehat{C}_{k+2} = (B^2 + (k+1)^2\beta)\widehat{C}_k = (B^2 - (k+1)^2\delta^2)\widehat{C}_k. \tag{53}$$

Next we examine the first two derivatives of $G_k(t)$: by induction on k and using the fact that $G_k(t) = (1+t^2\beta)G_{k-2}$, we obtain the following identity for the first derivative:

$$\frac{d}{dt}G_k(t) = (B - k\delta^2 t)G_{k-2}(t). \tag{54}$$

Applying this twice, we obtain an identity for the second derivative:

$$\frac{d^2}{dt^2}G_k(t) = \bigl(B^2 - k\delta^2 - 2(k-1)B\delta^2 t + k(k-1)\delta^4 t^2\bigr)G_{k-4}(t). \tag{55}$$

Using these identities for the first two derivatives, together with the induction hypothesis (52) and the recurrence relation (53), we compute

$$\begin{aligned}
&\frac{d^{k+2}}{dt^{k+2}}G_{k+1}(t)\\
&= \frac{d^{k+2}(1-\delta^2t^2)G_{k-1}(t)}{dt^{k+2}}\\
&= (1-\delta^2t^2)\frac{d^{k+2}G_{k-1}(t)}{dt^{k+2}} - 2(k+2)\delta^2 t\frac{d^{k+1}G_{k-1}(t)}{dt^{k+1}} - (k+2)(k+1)\delta^2\frac{d^kG_{k-1}(t)}{dt^k}\\
&= \biggl((1-\delta^2t^2)\frac{d^2}{dt^2} - 2(k+2)\delta^2 t\frac{d}{dt} - (k+2)(k+1)\delta^2\biggr)\frac{d^kG_{k-1}(t)}{dt^k}\\
&= \biggl((1-\delta^2t^2)\frac{d^2}{dt^2} - 2(k+2)\delta^2 t\frac{d}{dt} - (k+2)(k+1)\delta^2\biggr)\widehat{C}_k G_{-k-1}(t)\\
&= \bigl((B^2+\delta^2(k+1)+2(k+2)B\delta^2t+(k+1)(k+2)\delta^4t^2) - 2(k+2)\delta^2t(B+(k+1)\delta^2t)\\
&\qquad\qquad - (k+2)(k+1)\delta^2(1-\delta^2t^2)\bigr)\widehat{C}_k G_{-k-3}(t)\\
&= (B^2-(k+1)^2\delta^2)\widehat{C}_k G_{-k-3}(t) = \widehat{C}_{k+2}G_{-k-3}(t),
\end{aligned}$$

as required. □

4.4 The Mumford relations as generators of the ideal

In [32], a presentation of the cohomology ring of $\mathrm{Rep}(S_n^2)$ is given by exhibiting a complete set of generators for the ideal of relations $j_n \subset A_n$, all of which have degree $m = \frac{1}{2}(n-1)$. We now show that the elements $w_{n,\eta}^m$ also generate the ideal, by relating them to the generators in [32].

Remark The statement that the elements $w_{n,\eta}^s$, for $s \geq m$, generate the ideal is a counterpart of Kirwan's result [14] in the case of a (nonorbifold) surface of genus g. Kirwan's result was strengthened by Kiem [12], who showed that the relations in the middle dimension (ie the case $s = m$ in our context) are sufficient. The results of [14] were extended to the case of parabolic bundles on surfaces of genus $g \geq 2$ with one marked point by Earl and Kirwan [8].

Proposition 4.8 *Fix $n \geq 3$ odd, and write $n = 2m+1$. Then as η runs through all subsets of π whose size $h = |\eta|$ satisfies the parity condition (47), the elements $w^m_{n,\eta} \in A_n$ form a set of generators of the ideal j_n. That is, the elements $w^m_{n,\eta}$ form a complete set of relations for the cohomology of* $\mathrm{Rep}(S^2_n)$ *as a quotient of the algebra A_n.*

Proof From the results of [32], in the ideal j_n, there is an element r^m that has degree m and belongs to the subalgebra $\mathbb{Q}[\alpha,\beta] \subset A_n$, where $\beta = -\delta_p^2$. The element r^m is unique up to scale. According to [32, Corollary 1.6.2], the ideal j_n is generated by the elements

$$R^m_\zeta = r^{m-|\zeta|}(\alpha,\beta) \prod_{p\in\zeta} \delta_p,$$

where ζ runs through all subsets of π of size $0 \leq |\zeta| \leq m$. These elements all have degree m.

As we vary η, we obtain 2^{n-1} expressions $w^m_{n,\eta}$, all of which are elements of j_n of degree m. Because m is the lowest degree in which relations exist, each $w^m_{n,\eta}$ is a $\mathbb{Q}$-linear combination of the generators R^m_ζ. The number of generators R^m_ζ is also 2^{n-1}; so in order to see that the elements $w^m_{n,\eta}$ generate the ideal j_n, it will be enough if we show that they are linearly independent over $\mathbb{Q}$.

The fact that the elements $w^m_{n,\eta}$ are linearly independent can be deduced through a direct examination of the formulae which define it, as follows. Let us specialize the formulae by setting $\beta = 0$, in which case the expression (48) for the total Chern class of $-\mathrm{index}(P)$ simplifies to

$$(1+2\delta_1)^{B_\eta/(2\delta_1)} = \exp B_\eta$$

because $\delta_1^2 = 0$. The element $w^m_{n,\eta}$ therefore specializes to $B^m_\eta/m!$, or if we further specialize by setting $\alpha = 1$, to

$$\frac{1}{m!}\biggl(1 + \sum_p \eta_p \delta_p\biggr)^m,$$

where $\eta_p = 1$ for $p \in \eta$ and $\eta_p = -1$ otherwise. We can expand this as

$$\sum_{|\zeta|\leq m} C_{\eta,\zeta}\biggl(\prod_{p\in\zeta}\delta_p\biggr),$$

where the rational coefficient $C_{\eta,\zeta}$ is given by

$$C_{\eta,\zeta} = \frac{1}{(m-|\zeta|)!}\biggl(\prod_{p\in\zeta}\eta_p\biggr).$$

We wish to see that the matrix $C = (C_{\eta,\zeta})$, which is square of size 2^{n-1}, is nonsingular. To do this, we compute the dot product of the columns of C corresponding to different subsets ζ_1 and ζ_2. For fixed η we have

$$C_{\eta,\zeta_1} C_{\eta,\zeta_2} = \frac{1}{(m-|\zeta_1|)!\,(m-|\zeta_2|)!} \times \begin{cases} +1 & \text{if } |\eta \cap (\zeta_1 \ominus \zeta_2)^c| \text{ is even}, \\ -1 & \text{if } |\eta \cap (\zeta_1 \ominus \zeta_2)^c| \text{ is odd}, \end{cases}$$

where the superscript c denotes the complement and $\ominus$ means the symmetric difference. Since ζ_1 and ζ_2

are distinct subsets of size strictly less than $n/2$, their symmetric difference is a nonempty proper subset of π. The number of subsets η of a given parity for which the intersection is even and the number for which it is odd are therefore equal, and we see that

$$\sum_{\eta} C_{\eta,\zeta_1} C_{\eta,\zeta_2} = 0.$$

The columns are therefore orthogonal, showing that the square matrix C is nonsingular, as required. □

Remarks An alternative verification of the linear independence of the elements $w^m_{n,\eta}$, not depending on an examination of the formula, will be seen later, in the remarks at the end of Section 5.4.

5 Relations in instanton homology

5.1 Deforming the Mumford relation with instanton terms

The element $w^m_{n,\eta} \in j_n$ in Proposition 4.5 is a relation in the ordinary cohomology ring $H^*(\mathrm{Rep}(S^2_n);\mathbb{Q})$. Via its description in terms of the multiplicative generators α and δ_p, as an explicit element of the ring

$$\mathbb{Q}[\alpha, \delta_1, \dots, \delta_n]/\langle \delta_i^2 - \delta_j^2 \rangle_{i,j},$$

we may regard $w^m_{n,\eta}$ also as an element of the ideal $J_n \subset \mathcal{A}_n$ of Proposition 3.6, where it is a relation in the ordinary cohomology ring $H^*(\mathrm{Rep}(Z_n);\mathcal{R})$. As η varies over all subsets of π satisfying the parity condition, the elements $w^m_{n,\eta} \in J_n$ form a set of generators of the ideal, as follows immediately from the corresponding statement for $\mathrm{Rep}(S^2_n)$ (Proposition 4.8).

The following proposition promotes $w^m_{n,\eta}$ to a relation between the corresponding operators on the instanton homology $I(Z_n)$ by adding terms of lower degree. Recall that $\mathcal{J}_n \subset \mathcal{A}_n$ is the annihilator of $I(Z_n)$ as an $\mathcal{A}_n$-module.

Proposition 5.1 *Let n be odd and let $\eta \subset \pi$ be a subset whose size h satisfies the parity condition (47). Write $m = \frac{1}{2}(n-1)$ and let $w^m_{n,\eta} \in j_n \subset J_n$ be as in Proposition 4.5, regarded as a relation in the ordinary cohomology of the representation variety $\mathrm{Rep}(Z_n)$. Then there is a unique element $W^m_\eta \in \mathcal{J}_n \subset \mathcal{A}_n$ in filtration degree m whose leading term is $w^m_{n,\eta}$:*

$$W^m_\eta = w^m_{n,\eta} \pmod{\mathcal{A}_n^{(m-1)}}.$$

As η varies over all subsets satisfying the parity condition, these elements W^m_η form a set of generators for the ideal of relations $\mathcal{J}_n$.

Remark Our notation for W^m_η does not include n, since n is always related to m in this context by $n = 2m + 1$.

Proof of Proposition 5.1 Corollary 3.16 gives the existence of $W^m_\eta \in \mathcal{J}_n$ with leading term $w^m_{n,\eta}$. The uniqueness assertion is a consequence of Proposition 3.7. The fact that these are a complete set of generators for the ideal follows from the corresponding statement for the elements $w^m_{n,\eta} \in J_n$ together with the fact that $\mathcal{A}_n/J_n$ and $\mathcal{A}_n/\mathcal{J}_n$ are free modules of the same rank, because they are respectively the ordinary homology of $\operatorname{Rep}(Z_n)$ and the instanton homology of Z_n (Proposition 3.8). □

We aim to give an algorithm for computing W^m_η as a deformation of $w^m_{n,\eta}$, and our first main step will be to determine the subleading term. That is, Corollary 3.16 provides the existence of $w(1)$ with

$$W^m_\eta = w^m_{n,\eta} + \epsilon w(1) \pmod{\mathcal{A}_n^{(m-2)}},$$

and we wish to determine $w(1)$.

Proposition 5.2 *The subleading term of W^m_η is given by $\epsilon\tau^{n-2h} w^{m-1}_{n,\eta'}$, where η' is the complement $\pi \setminus \eta$ and $h = |\eta|$, so*

$$W^m_\eta = w^m_{n,\eta} + \epsilon\tau^{n-2h} w^{m-1}_{n,\eta'} \pmod{\mathcal{A}_n^{(m-2)}}.$$

The proof of this proposition will require some preparation. To understand how to characterize the subleading term $\epsilon w(1)$, we draw on the mechanism behind Proposition 3.14 and Corollary 3.16. Let $\mathbb{1}_+$ again be the standard cyclic generator of $I(Z_n)$ from Proposition 3.9, and let $\mathbb{1}_- = \tilde{\epsilon}\mathbb{1}_+$. Let Λ be the $\mathcal{R}$-module isomorphism in Proposition 3.14, and let $1_\pm = \Lambda(\mathbb{1}_\pm) \in H_*(\operatorname{Rep}(Z_n);\mathcal{R})$. Then $w(1)$ is a homogeneous polynomial of degree $m-1$ in α and the δ_p, with coefficients in $\mathcal{R}$, such that

$$\Lambda(w^m_{n,\eta}\mathbb{1}_-) - w^m_{n,\eta}1_- = w(1)1_+ \mod \bigoplus_{k\le 2(m-2)} H^k(\operatorname{Rep}(Z_n);\mathcal{R}).$$

(The right-hand side is the $(m-2)^{\text{th}}$ step of the increasing filtration of $H^*(\operatorname{Rep}(Z_n;\mathcal{R}).)$

Recall next we have an expansion of the operator $\tilde{\alpha}$ according to the action $\kappa \in \frac{1}{4}\mathbb{Z}$, as in (25) and Proposition 3.14. There is a similar expansion of each $\tilde{\delta}_p$. This gives a κ-expansion of any monomial in $\tilde{\alpha}$ and the $\tilde{\delta}_p$, and therefore of the multiplication operator of any $u \in \mathcal{A}_n$ acting on $I(Z_n)$. That is, we may write

$$u\xi = \sum_{\kappa\in\frac{1}{4}\mathbb{Z}} u *_\kappa \xi.$$

This description is set up so that if $u \in \mathcal{A}_n$ is in grading k and $\Lambda(\xi) \in H^{2l}(\operatorname{Rep}(Z_n);\mathcal{R})$, then

$$\Lambda(u *_\kappa \xi) \in H^{2(l+k)-8\kappa}(\operatorname{Rep}(Z_n);\mathcal{R}).$$

The description of $w(1)$ then becomes

$$w(1)1_+ = \Lambda(w^m_{n,\eta} *_{1/4} \mathbb{1}_-). \tag{56}$$

Computation of $w(1)$ in this form therefore depends directly on understanding the instantons on the cylinder $\mathbb{R}\times Z_n$ with action $\frac{1}{4}$. We address this calculation in the following subsection, where the proof of Proposition 5.2 will be completed.

5.2 Characterizing the subleading term

From the discussion above, we are interested in the moduli space $M_\kappa(\mathbb{R}\times Z_n)$ of anti-self-dual bifold SU(2) connections on the cylinder, particularly for $\kappa = \frac{1}{4}$. By attaching a copy of the bifold $D^2 \times S^2_n$ to each of the two ends, we form from the cylinder a compact bifold

$$X = S^2 \times S^2_n.$$

For clarity in distinguishing the two factors here, we will write

$$X = B \times C,$$

where B is S^2 and C is the bifold S^2_n. We write $M_\kappa(X)$ for the moduli space of anti-self-dual SU(2) connections on the bifold X, with action κ, and we write $M^e_\kappa(X)$ for the moduli space corresponding to $w_2 = [e]$, where $[e] = \{b\} \times C$. The moduli spaces depend, of course, on a choice of conformal structure on X. The moduli space $M_\kappa(X)$ is nonempty only if $\kappa \in \frac{1}{2}\mathbb{Z}$, while $M^e_\kappa(X)$ is nonempty only if $\kappa \in \frac{1}{2}\mathbb{Z} + \frac{1}{4}$. The moduli spaces have formal dimension

$$d(\kappa) = 8\kappa + 2n - 6.$$

For any element $u \in \mathcal{A}_n$ of degree $d(\kappa)/2$ in the variables α and δ_i, we can seek to evaluate a Donaldson polynomial invariant by evaluating the corresponding cohomology class on $M_\kappa(X)$ or $M^e_\kappa(X)$. Because we are working with local coefficients Γ, our Donaldson invariants should also involve $\mathcal{R}$-valued weights. By the formula (2), the local system Γ defines a locally constant function

$$\Gamma\colon M_\kappa(X) \to \mathcal{R}^\times, \tag{57}$$

and so the moduli spaces are a collection of oriented, weighted manifolds.

However, the bifold X has $b_2^+ = 1$, so the appearance of reducibles in one-parameter families means that the Donaldson invariant depends on a choice of chamber in the space of Riemannian metrics on X. We consider a product metric in which the area of B is very large compared to the area of C, and we call this the *B-chamber*. (This means that the self-dual 2-form for the Riemannian metric on X is nearly Poincaré dual to a multiple of PD$[C]$.) Similarly there is a distinguished chamber, the *C-chamber*, in which the area of C is very large compared to B. There is then a well-defined Donaldson invariant q^B_κ in the B-chamber,

$$u \mapsto q^B_\kappa(X; u), \quad \mathcal{A}_n \to \mathcal{R},$$

calculated using either the moduli space $M_\kappa(X)$ or the moduli space $M^e_\kappa(X)$, depending on whether 4κ is even or odd respectively. Our notation again makes no explicit mention of the local coefficient system, but the contributions of the various components of the moduli spaces are to be weighted by the locally constant function (57).

These Donaldson invariants of X are related to the action of u on $I(Z_n)$ by a gluing argument, because of the description of X as the union of the cylinder $[-1,1]\times Z_n$ and the two copies of $D^2 \times S^2_n$.

More specifically, let $\mathbb{1}_+ \in I(Z_n)$ be once more the cyclic generator obtained as the relative invariant of the manifold $D^2 \times S^2_n$, and let $\mathbb{1}^\dagger_+$ be the element of the instanton *co*homology group $I^*(Z_n)$ obtained by regarding $D^2 \times Z_n$ as a manifold with boundary $-Z_n$. Then for $\kappa \in \frac{1}{2}\mathbb{Z}$ and $u \in \mathcal{A}^+_n$, we can write

$$q^B_\kappa(X; u) = \langle u *_\kappa \mathbb{1}_+, \mathbb{1}^\dagger_+ \rangle,$$

where the pairing on the right is the $\mathcal{R}$-valued pairing between $I(Z_n)$ and $I^*(Z_n)$. For $\kappa \in \frac{1}{4} + \frac{1}{2}\mathbb{Z}$, we have

$$q^B_\kappa(X; \epsilon u) = \langle u *_\kappa \mathbb{1}_+, \mathbb{1}^\dagger_- \rangle.$$

From this relationship and Poincaré duality, it follows that (56) is equivalent to

$$q^B_{1/4}(w(0)v) = q^B_0(\epsilon w(1)v) \tag{58}$$

for all $v \in \mathcal{A}_n$ of degree

$$\deg(v) = \tfrac{1}{2} d\left(\tfrac{1}{4}\right) - \deg(w(0)) = n - 2 - m = m - 1,$$

where $n = 2m + 1$ as usual.

The situation is somewhat simplified now because the moduli spaces $M_0(X)$ and $M^e_{1/4}(X)$ are *compact*. This is because noncompactness of the moduli space arises only from bubbling, and bubbles decrease κ by multiples of $\frac{1}{2}$. So for $\kappa \le \frac{1}{4}$, the Donaldson invariants are simply evaluations on $[M_\kappa(X)]$ or $[M^e_\kappa(X)]$ of ordinary cohomology classes in $H^*(\mathcal{B}^*(X); \mathcal{R})$, weighted by the function locally constant (57). We will write $[\Gamma \cdot M_\kappa(X)]$ and $[\Gamma \cdot M^e_\kappa(X)]$ for these weighted fundamental classes, as elements of the ordinary homology $H_*(\mathcal{B}^*(X); \mathcal{R})$.

Via the relationship between $\mathcal{A}_n$ and $H^*(\mathcal{B}^*(Z_n); \mathcal{R})$, we have an inclusion

$$\mathcal{A}_n \hookrightarrow H^*(\mathcal{B}^*(X); \mathcal{R}).$$

The relation (58) can therefore be stated in terms of ordinary pairings, between these cohomology classes and the fundamental classes of the moduli spaces:

$$\langle w(0)v, [\Gamma \cdot M^e_{1/4}(X)] \rangle = \langle w(1)v, [\Gamma \cdot M_0(X)] \rangle.$$

The assertion in Proposition 5.2 concerning the value of the subleading term $w(1)$ can therefore be restated as the following proposition.

Proposition 5.3 *Let $n = 2m + 1$ as usual let $v \in \mathcal{A}_n$ be any element of degree $m - 1$. Let $w^k_{n,\eta} \in \mathcal{A}_n$ be the explicit polynomials described in Definition 4.2. Then we have*

$$\langle w^m_{n,\eta} v, [\Gamma \cdot M^e_{1/4}(X)] \rangle = \langle \tau^{n-2|\eta|} w^{m-1}_{n,\eta'} v, [\Gamma \cdot M_0(X)] \rangle,$$

where the (compact) moduli space $M^e_{1/4}(X)$ is computed using a metric on X in the B-chamber, and $M_0(X)$ is the moduli space of flat bifold connections, a copy of $\mathrm{Rep}(S^2_n)$.

The proof of Proposition 5.3 is given in Section 5.4, after a digression on the wall-crossing behavior of moduli spaces on X.

5.3 A wall-crossing argument

The structure of our argument up to this point is closely related to the work of Muñoz [27], in which a key step is the calculation of the contribution of the first nonflat moduli space (our $M^e_{1/4}(X)$ in the present context). In [27], the relevant moduli space was of the form $M^e_{1/2}(S^2 \times \Sigma_g)$ for a smooth surface Σ_g, and the key observation is that this moduli space is empty in one chamber (when the area of the S^2 factor is small, corresponding to the C-chamber in our notation) and undergoes a single wall-crossing where the metric passes to the B-chamber. (See [27, Proposition 2].) The description of the wall-crossing for $S^2 \times \Sigma_g$ leads to a description of the moduli space on the B side of the wall as a bundle over the Jacobian $J(\Sigma_g)$ with fiber a complex projective space.

Such a description has an exact parallel in our orbifold context, with the Jacobian $J(\Sigma_g)$ in Muñoz's situation replaced now by the finite set of bifold line bundles on S^2_n of a fixed bifold degree. That is, the wall-crossing contributes to $M^e_{1/4}(X)$ a finite number of copies of a complex projective space, where an explicit understanding of the cohomology classes allows a calculation of the Donaldson invariant. We now turn to the details of this calculation.

Lemma 5.4 *In the C-chamber, the Donaldson invariants $q^C_\kappa(u)$ are zero when κ is in $\frac{1}{4} + \frac{1}{2}\mathbb{Z}$.*

Proof The bifold X decomposes into two parts along a copy of $B \times S^1 \subset B \times C$, ie an $S^2 \times S^1$. The bundle has w_2 nonzero on this $S^2 \times S^1$ when κ is in $\frac{1}{2}\mathbb{Z} + \frac{1}{4}$, so there are no flat connections on $B \times S^1$. A stretching argument therefore shows that the anti-self-dual moduli space is empty when the metric on X contains a long neck $[-T, T] \times B \times S^1$. A metric with such a long neck lies in the C-chamber, so the invariant in this chamber is zero. □

Lemma 5.5 *For the moduli spaces $M^e_\kappa(X)$ with $\kappa \le \frac{1}{4}$, in a 1-parameter family of product metrics on $X = B \times C$ passing from the C-chamber to the B-chamber, exactly one wall is crossed.*

Proof The only nonempty moduli space $M^e_\kappa(X)$ with $\kappa \le \frac{1}{4}$ is the moduli space $M^e_{1/4}(X)$, and a wall is crossed when the Riemannian metric allows the existence of a reducible anti-self-dual connection in this moduli space. We are therefore looking for a reduction of the bifold adjoint SO(3) bundle as $\mathbb{R} \oplus K$, where K is a bifold 2-plane bundle. Let us write the bifold Euler class $\mathrm{eul}(K)$ as

$$\mathrm{PD}\,\mathrm{eul}(K) = x[B] + y[C].$$

Here y is an odd integer because $\mathrm{eul}(K)[B]$ is odd. On the curve C, the bundle K has n bifold points, and n is odd; so $2x$ is also an odd integer. For a given Riemannian metric, let us write the class of the self-dual 2-form as

$$\mathrm{PD}[\omega^+] = [B] + t[C],$$

suitably normalized. The condition that the curvature of K is anti-self-dual imposes the constraint that $\mathrm{eul}(K)$ and $[\omega^+]$ are orthogonal, which is to say

$$y = -tx.$$

The action κ is $-\mathsf{eul}(K)^2/4$ which is $-xy/2$. Using the orthogonality condition, we write this as $\kappa = tx^2/2$. With $\kappa = \frac{1}{4}$, our constraints therefore become

(i) tx and $2x$ are odd integers, and

(ii) $tx^2 = \frac{1}{2}$.

These constraints force $x = \pm\frac{1}{2}$ and $t = 2$. The orientation of K is indeterminate, and the sign of x can therefore be taken to be positive. A path of Riemannian metrics passing from the C chamber to the B chamber is a path in which t begins close to 0 and ends close to $+\infty$, and the wall is crossed at $t = 2$. □

The proof the lemma shows that the wall-crossing occurs when there is an orbifold 2-plane bundle K with

$$\mathrm{PD}\,\mathsf{eul}(K) = \tfrac{1}{2}[B] - [C].$$

The degree of K on $C = S^2_n$ is thus $\frac{1}{2}$. In terms of an SU(2) lift on the curve $\{b\} \times C$ then, we can write the bundle as

$$F \oplus F^{-1},$$

where F is a complex line bundle with limiting holonomy $\pm i$ on the linking circles at the n singular points. We orient K as F^{-2}. The Chern–Weil integral for the first Chern class of the singular connection on F is $-\frac{1}{4}$. As a parabolic bundle on S^2_n we can write the underlying rank-2 vector bundle as $\mathcal{E} = \mathcal{F} \oplus \mathcal{F}^{-1}$, and for each $p \in \pi$ the distinguished line $\mathcal{L}_p \subset \mathcal{E}_p$ is the summand $\mathcal{F}_p$ if the limiting holonomy is $-i$, and $\mathcal{F}_p^{-1}$ otherwise. Write $\xi \subset \pi$ for the set where the holonomy is $-i$. Then

$$c_1(\mathcal{F})[C] + \tfrac{1}{4}|\xi| - \tfrac{1}{4}(n - |\xi|) = -\tfrac{1}{4}.$$

This constraint imposes the parity condition $|\xi| = \frac{1}{2}(n-1) \bmod 2$, which allows 2^{n-1} possibilities for ξ. We summarize this with another lemma.

Lemma 5.6 *When the Riemannian metric on $X = B \times C$ lies on the wall between the two chambers, the moduli space $M^e_\kappa(X)$ consists of 2^{n-1} reducible anti-self-dual connections, corresponding to the subsets $\xi \subset \pi$ whose size $|\xi|$ has the same parity as $\frac{1}{2}(n-1)$.*

Let A_0 denote any one of the reducible connections described in the lemma. The formal dimension of the moduli space $M^e_{1/4}(X)$ is $2n-4$. If we write the orbifold adjoint bundle as $\mathbb{R} \oplus K$ now on the whole of X, then in the deformation theory of A_0 we have a contribution of 1 to the dimension of $H^0_{A_0}$ coming from the $\mathbb{R}$ summand because A_0 is reducible, and there is a similar contribution of 1 to the dimension of $H^2_{A_0}$ from the $\mathbb{R}$ summand because $b^+_2 = 1$. If we assume that the deformation theory is otherwise unobstructed (an assumption which we will see later is justified for product metrics on $B \times C$, without the need for perturbing the equations), then it follows that $H^1_{A_0}$ has dimension $2n-2$ and that this comes from the K summand of the adjoint bundle. With this in place, the standard model for wall-crossing describes the moduli space $M^e_{1/4}(X; g_t)$ for a Riemannian metric g_t whose conformal parameter t is $2+\epsilon$ for small ϵ as a copy of $\mathbb{CP}^{n-2}$ in a neighborhood of each reducible A_0. We therefore have the following proposition.

Proposition 5.7 *For a product metric on X which lies in the B-chamber and is close to the wall, the moduli space $M^e_{1/4}(X)$ consists of 2^{n-1} copies of $\mathbb{CP}^{n-2}$.*

As mentioned earlier, this is a close counterpart to the result [27, Proposition 2], where the corresponding description of the moduli space of smallest positive action is a bundle of projective spaces over the Jacobian of a smooth curve.

5.4 A proof of Proposition 5.3

From their definition, $w^m_{n,\eta}$ and $w^{m-1}_{n,\eta'}$ represent cohomology classes dual to loci $U^\eta_{-1/4}$ and $U^{\eta'}_{1/4}$ in the space of bifold connections $\mathcal{B}^*(S^2_n)$. If we select a fiber

$$\{b_0\} \times S^2_n \subset B \times S^2_n = X,$$

then we obtain by restriction corresponding loci in the spaces of bifold connections on X:

$$U^\eta_{-1/4}(b_0) \subset \mathcal{B}^*(X)^e, \quad U^{\eta'}_{1/4}(b_0) \subset \mathcal{B}^*(X).$$

In this way we can interpret the equality to be proved in Proposition 5.3 as

$$\langle v, [\Gamma \cdot M^e_{1/4}(X) \cap U^\eta_{-1/4}(b_0)]\rangle = \tau^{n-2|\eta|}\langle v, [\Gamma \cdot M_0(X) \cap U^{\eta'}_{1/4}(b_0)]\rangle, \tag{59}$$

provided that the loci are transverse to the filtration of the space of Fredholm operators by the dimension of the kernel. The moduli spaces on X should be obtained from metrics in the B-chamber as always.

We can obtain more information about $M^e_{1/4}(X)$ and the loci on both sides of (59) by interpreting the moduli space of bifold anti-self-dual connections as a moduli space of stable parabolic bundles on the pair (X, Σ) where Σ is the singular locus $B \times \pi \subset X$. To this end, we adopt the notation and results of [18] to identify $M^e_{1/4}(X)$ with the moduli space of parabolic bundles $(\mathcal{E}, \mathcal{L})$ with $\kappa = \frac{1}{4}$ satisfying the parabolic stability condition with parameter $\alpha = \frac{1}{4}$. Here we can write κ as $k + l/2$ following [17; 18], where in this case

$$k = \big(c_2(\mathcal{E}) - \tfrac{1}{4}c_1(\mathcal{E})^2\big)[X], \quad l = \big(\tfrac{1}{2}c_1(\mathcal{E}) - c_1(\mathcal{L})\big)[\Sigma]. \tag{60}$$

(The quantities k and l are the "instanton number" and "monopole number" in the notation of [17].) The rank-2 bundle $\mathcal{E}$ should have $c_1(\mathcal{E})[B]$ odd, so we take

$$\Lambda^2(\mathcal{E}) = \mathcal{O}(1, 0),$$

by which we mean the holomorphic line bundle with degree 1 on B. The moduli space $M_0(X)$ is similarly a moduli space of stable parabolic bundles on X, now with $\Lambda^2(\mathcal{E}) = \mathcal{O}$ and $\kappa = 0$. These bundles are the pullbacks of the stable parabolic bundles on the curve $C = S^2_n$.

The loci on either side of (59) have the following interpretations. Let $\mathcal{F} \to C$ be the parabolic line bundle whose set of hits is η and whose parabolic degree is par-deg $\mathcal{F} = \frac{1}{4}$. (See the remarks at the end of Section 4.1.) The dual parabolic bundle $\mathcal{F}^*$ has parabolic degree $-\frac{1}{4}$ and its set of hits is $\eta' = \pi \setminus \eta$. Given a stable parabolic bundle $\mathcal{E}$ on X, let $\mathcal{E}_b$ be the parabolic bundle obtained by restriction to $\{b\} \times C$.

Lemma 5.8 *Let $\mathcal{F}$ be the parabolic line bundle described above and $\mathcal{F}^*$ its dual. Then:*

(i) *The locus $M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0)$ is the locus of stable parabolic bundles $\mathcal{E} \in M^e_{1/4}(X)$ such that there exists a nonzero holomorphic map of parabolic bundles*

$$\mathcal{F} \to \mathcal{E}_{b_0}.$$

(ii) *The locus $M_0(X) \cap U^{\eta'}_{1/4}(b_0)$ is the locus of stable parabolic bundles $\mathcal{E} \in M_0(X)$ such that there exists a nonzero holomorphic map of parabolic bundles*

$$\mathcal{F}^* \to \mathcal{E}_{b_0}.$$

Proof These statements follow directly from the definitions. □

Going beyond the above lemma, we have the following constructions for the relevant bundles.

Lemma 5.9 (i) *The locus $M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0)$ consists of parabolic bundles $\mathcal{E} \to X = B \times C$ which are nonsplit extensions*

$$\mathcal{O}(1) \boxtimes \mathcal{F}^* \to \mathcal{E} \to \mathcal{F}$$

such that the extension class vanishes on $\{b_0\} \times C$.

(ii) *The locus $M_0(X) \cap U^{\eta'}_{1/4}(b_0)$ is the locus of parabolic bundles $\mathcal{E} \in M_0(X)$ which are nonsplit extensions*

$$\mathcal{F}^* \to \mathcal{E} \to \mathcal{F}.$$

In both cases, all bundles obtained as such extensions are stable in the B-chamber on X.

Proof In (ii), the bundles in $M_0(X)$ are pulled from the stable parabolic bundles on C, and the existence of a nonzero map of parabolic bundles $\iota\colon \mathcal{F}^* \to \mathcal{E}$ is the definition of the locus $U^{\eta'}_{1/4}$. The map ι must be an inclusion of a parabolic line subbundle, for otherwise this map would destabilize $\mathcal{E}$. So $\mathcal{E}$ is an extension of parabolic line bundles as described. The extension must be nonsplit, for otherwise $\mathcal{E}$ is destabilized by ι.

For (i), the first task is to verify that every stable parabolic bundle in $M^e_{1/4}(X)$ in the B-chamber is a nonsplit extension

$$\mathcal{O}(1) \boxtimes \mathcal{G}^* \to \mathcal{E} \to \mathcal{G}, \tag{61}$$

where $\operatorname{par-deg} \mathcal{G} = -\frac{1}{4}$ and the set of hits for $\mathcal{G}$ is a subset $\xi \subset \pi$ which is arbitrary, except for the parity constraint (35). There are 2^{n-1} choices for ξ, and once ξ is given, the nonsplit extensions are parametrized by a projective space, in this case of dimension $n-2$. In this way we find 2^{n-1} copies of $\mathbb{CP}^{n-2}$ in $M^e_{1/4}$, and it is straightforward to see that these are disjoint, because a given bundle $\mathcal{E}$ cannot be presented as an extension of this sort in two different ways. The verification that these 2^{n-1} copies of $\mathbb{CP}^{n-2}$ comprise the *entire* moduli space $M^e_{1/4}(X)$ in the B-chamber is the holomorphic analog of wall-crossing result described in Proposition 5.7, and it is proved in essentially the same way. This is also the content of [27, Proposition 2] in the slightly different context of that paper, which serves the same purpose there.

For an extension such as (61), the restriction to $\{b_0\} \times C$ is an extension of parabolic line bundles on C,

$$\mathcal{G}^* \to \mathcal{E}_{b_0} \to \mathcal{G},$$

and because $\text{par-deg}(\mathcal{F}) = \text{par-deg}(\mathcal{G}) > \text{par-deg}(\mathcal{G}^*)$, there can be a nonzero map $\mathcal{F} \to \mathcal{E}_{b_0}$ only if $\mathcal{F} = \mathcal{G}$ and the extension class is zero on $\{b_0\} \times C$. □

The extensions that arise in (ii) are parametrized by the projective space

$$\mathbb{P}\big(H^1(C; (\mathcal{F}^*)^{\otimes 2})\big), \tag{62}$$

where the cohomology group is interpreted as the cohomology of a sheaf on a bifold. The extensions that arise in (i) are parametrized by the subset of the projective space

$$\mathbb{P}\big(H^0(B; \mathcal{O}(1)) \otimes H^1(C; (\mathcal{F}^*)^{\otimes 2})\big)$$

corresponding to elements vanishing at b_0. If $Z_{b_0} \subset H^0(B; \mathcal{O}(1))$ is the one-dimensional space of sections vanishing at b_0, then this is the space

$$\mathbb{P}\big(Z_{b_0} \otimes H^1(C; (\mathcal{F}^*)^{\otimes 2})\big),$$

which is canonically identified with (62). Both spaces of extensions are copies of $\mathbb{CP}^{m-1}$.

We have now seen that there is a canonical identification of the two loci,

$$M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0) = M_0(X) \cap U^{\eta'}_{1/4}(b_0),$$

both of which are projective spaces. Furthermore, for any $b \neq b_0$ in B, the restrictions of the corresponding bundles in these loci to $\{b\} \times C$ agree. Indeed they are the same family of nonsplit extensions of $\mathcal{F}$ by $\mathcal{F}^*$ on C. The cohomology classes v arising from elements of the algebra $\mathcal{A}_n$ can be regarded as being pulled back via the restriction to $\{b\} \times C$, so it follows that the evaluation of such a class v is the same in the two cases.

Before accounting for the weights arising from the local system Γ, we therefore have an equality

$$\langle v, [M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0)] \rangle = \langle v, [M_0(X) \cap U^{\eta'}_{1/4}(b_0)] \rangle. \tag{63}$$

However, while $M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0)$ and $M_0(X) \cap U^{\eta'}_{1/4}(b_0)$ are both copies of $\mathbb{CP}^{m-1}$ and are canonically identified, the (constant) functions

$$\Gamma\colon M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0) \to \mathcal{R} \quad \text{and} \quad \Gamma\colon M_0(X) \cap U^{\eta'}_{1/4}(b_0) \to \mathcal{R}$$

are different. The next lemma provides these values.

Lemma 5.10 (i) *On $M_0(X) \cap U^{\eta'}_{1/4}(b_0)$, the value of Γ is* 1.

(ii) *On $M^e_{1/4}(X) \cap U^{\eta}_{-1/4}(b_0)$, the value of Γ is $\tau^{n-2|\eta|}$.*

Proof The singular set $\Sigma \subset X$ is a collection of spheres with trivial normal bundle, so there is no self-intersection term in the formula (2), and we simply have

$$\Gamma(A) = \tau^{\nu(A)},$$

where $\nu(A)$ is a 2-dimensional Chern–Weil integral on Σ. In the case of $M_0(X)$, the connections are flat and $\nu(A)=0$. So $\Gamma=1$ in this case, as stated in the first item of the lemma.

In the case of a closed manifold, the value $\nu(A)$ is $-2l$, where l is the "monopole number" of the bundle (60). The bundles that contribute to the moduli space $M_{1/4}(X)^e \cap U^{\eta}_{-1/4}(b_0)$ are described in Lemma 5.9. From there we read off that $c_1(\mathcal{E})[\Sigma_p]=1$ for each of the n components $\Sigma_p \subset \Sigma$, so that $c_1(\mathcal{E})[\Sigma]=n$. For $p\in\eta'$, the distinguished line subbundle $\mathcal{L}\subset\mathcal{E}|_{\Sigma_p}$ coincides with the image of the subbundle $\mathcal{O}(1)\boxtimes\mathcal{F}^*$ on Σ_p, which has degree 1. For $p\in\eta$, the distinguished line subbundle $\mathcal{L}$ on Σ_p maps isomorphically to the restriction of $\mathcal{F}$ in the extension in Lemma 5.9, so has degree 0. In all then,

$$c_1(\mathcal{L})[\Sigma]=|\eta'|.$$

The formula for the monopole number l in (60) therefore gives $(n/2)-|\eta'|$, which is $|\eta|-(n/2)$. Since $\nu(A)=-2l$, we have $\nu(A)=n-2|\eta|$, as the lemma claims. □

From the lemma, we see that

$$[\Gamma\cdot M^e_{1/4}(X)\cap U^{\eta}_{-1/4}(b_0)]=\tau^{n-2h}[M^e_{1/4}(X)\cap U^{\eta}_{-1/4}(b_0)],$$

while

$$[\Gamma\cdot M_0(X)\cap U^{\eta'}_{1/4}(b_0)]=[M_0(X)\cap U^{\eta'}_{1/4}(b_0)].$$

The equality to be proved in Proposition 5.3 now follows from the unweighted equality (63), and this completes the proof of the proposition.

Remark In the course of these arguments, we have seen first that $M^e_{1/4}(X)$ is a disjoint union of 2^{n-1} copies of $\mathbb{CP}^{n-2}$ and second that the class $w^m_{n,\eta}$ restricts to be nonzero on exactly one of them, being dual to a $\mathbb{CP}^{m-1}$ in exactly one of the copies of $\mathbb{CP}^{n-1}$. The components $\mathbb{CP}^{n-2}$ of $M^e_{1/4}(X)$ are in one-to-one correspondence with the subsets $\eta\subset\pi$ of the correct parity, so let us write them as $\mathbb{CP}^{n-2}_\eta$. If we choose a class v which has nonzero pairing (say 1) with each $\mathbb{CP}^{m-1}\subset\mathbb{CP}^{n-2}_\eta$, then we have

$$\langle w^m_{n,\eta}\cup v,[\mathbb{CP}^{n-2}_\xi]\rangle=\begin{cases}1 & \text{if } \eta=\xi,\\ 0 & \text{otherwise,}\end{cases}$$

from which it follows that the classes $w^m_{n,\eta}$ are linearly independent in $\mathcal{A}_n$. This provides an alternative verification of the result used in the proof of Proposition 4.8.

5.5 Changing the orientation of the singular set

Recall that in defining the bifold Z_n we gave a standard orientation to the n circles comprising the singular set K_n. Let Z^*_n denote the same bifold but with some of the circles of K_n equipped with the opposite orientation. Let f be the number of search circles. The construction of the operators δ_p depends on an orientation of the singular set at p, so in a straightforward way the corresponding operators δ^*_p on $I(Z^*_n)$ differ in sign from the operators δ_p if we define δ^*_p using the new orientations. But there is also a more subtle way in which the module structures differ.

Both $I(Z_n)$ and $I(Z_n^*)$ are $\mathcal{A}_n$-modules by our constructions. Let us continue to denote by α, δ_p and ϵ the operators on $I(Z_n)$, and let us denote by α^*, δ_p^* and ϵ^* the operators which define the $\mathcal{A}_n$-module structure of $I(Z_n^*)$.

Proposition 5.11 *There is an identification of the $\mathcal{R}$-modules $I(Z_n^*)$ and $I(Z_n)$ which is canonical up to overall sign-change. Under either one of this canonical pairs of identifications, the operators α^* etc on $I(Z_n^*)$ are related to the operators on $I(Z_n)$ by:*

(i) $\alpha^* = \alpha$;

(ii) $\delta_p^* = \pm\delta_p$, *according to whether or not the corresponding circles of K_n have the standard orientation in Z_n^*;*

(iii) $\epsilon^* = (-1)^f \epsilon$, *where f as above is the number of circles which have the nonstandard orientation.*

Proof First let us recall that the SU(2) instanton moduli spaces $M_k(X)$ for a closed Riemannian manifold X are orientable and are oriented by a choice of an element from a 2-element set $\Lambda(X)$, which can be identified with the set of homology orientations of X. In the case of a closed bifold with orientable singular set, if we regard the moduli space as the space of singular SU(2) connections $M_{k,l}(X,\Sigma)$ in the sense of [17; 18] and [20], then an element of the 2-element orientation set $\Lambda(X,\Sigma)$ can be specified by a choice of homology orientation of X together with an orientation of Σ. Changing the orientation of Σ changes the sign of the element of $\Lambda(X,\Sigma)$ by $(-1)^{\chi/2}$, where χ is the Euler number. (See [18].) To briefly explain why this is so, the conventions of [18] identify the difference between $\Lambda(X)$ and $\Lambda(X,\Sigma)$ as the set $\Lambda(\Sigma)$ of orientations of the real determinant line of the index of the $\bar\partial$ operator on Σ coupled to a line bundle of degree $2l$. The index of the $\bar\partial$ operator is $2l-\chi(\Sigma)/2$. Changing the orientation of Σ changes the index element to its complex conjugate and therefore changes the orientation of the real determinant line by $(-1)^{2l-\chi/2} = (-1)^{\chi/2}$. A similar formula holds if the orientation of Σ is changed only on certain components.

In the case that (X,Σ) is a product $S^1 \times (Y,K)$, there is a canonical homology orientation for X and the components of Σ are tori; so there is a canonical element of $\Lambda(X,\Sigma)$ in this case, independent of the orientation of the components of Σ.

Continuing with the closed case, we consider next the moduli space $M_{k,l}(X,\Sigma)^e$ corresponding to an SO(3) bundle whose w_2 has an integer lift e. As usual the gauge group is the determinant-1 gauge group. In the absence of Σ, the orientation set $\Lambda(X)^e$ is still canonically identified with the set of homology orientations of X, as in [6]. The difference between $\Lambda(X)^e$ and $\Lambda(X,\Sigma)^e$ is again identified with the real determinant line of the same $\bar\partial$ operator. The difference now however is that the monopole number l is in $\frac{1}{2}+\mathbb{Z}$ on any component of Σ having odd pairing with e. Changing the orientation of a component $\Sigma_1 \subset \Sigma$ therefore changes the orientation element in $\Lambda(X,\Sigma)^e$ by $(-1)^{g(\Sigma_1)-1}$ if e has even pairing with Σ_1 and by $(-1)^{g(\Sigma_1)}$ if the pairing is odd. In the special case that $(X,\Sigma) = S^1 \times (Y,K)$, an orientation of the moduli spaces $M(X,\Sigma)^e$ therefore depends on the orientation of the components of the singular set if and only if they have odd pairing with e. If the pairings are even, all orientations of moduli spaces are canonical and do not depend on the orientation of the singular set.

To apply these observations to the instanton homology, we recall the standard approach to orientations in Floer homology, described for example in [20, Section 3.6]. Let $\mathcal{B}^*$ denote the space of irreducible singular SU(2) connections on the bifold (Y, K). To each pair of points $a, b \in \mathcal{B}^*$ and each path ζ joining them, we may associate a 2-element set $\Lambda_\zeta(a, b)$ as the set of orientations of the determinant line of a Fredholm operator $P(a, b)$, in such a way that $\Lambda_\zeta(a, b)$ orients the moduli space of trajectories if a and b are nondegenerate critical points. If ζ_1 and ζ_2 are two different paths, then $\Lambda_{\zeta_1}(a, b)$ and $\Lambda_{\zeta_2}(a, b)$ are canonically identified, because their difference can be identified with $\Lambda(X, \Sigma)$, where $(X, \Sigma) = S^1 \times (Y, K)$. We can therefore define $\Lambda(a, b)$ without issue. No orientation of K is needed here.

Given a basepoint θ in $\mathcal{B}^*$, one may then define $\Lambda(a) = \Lambda(a, \theta)$ for all a. The chain complex for the singular instanton homology with local coefficients Γ is then

$$\bigoplus_a (\mathbb{Z}\Lambda(a)) \otimes \Gamma_a,$$

where the sum is over perturbed flat connections in a Morse perturbation of the Chern–Simons functional. Two different choices of basepoints θ and θ' will give rise to complexes which are identified up to an overall factor of -1: that ambiguity is a choice of element from $\Lambda(\theta, \theta')$. A canonical choice of basepoint is possible when K is oriented, as described in [20], making $I(Y, K)$ well-defined up to canonical isomorphism. The modules $I(Z_n)$ and $I(Z'_n)$ are identified only up to overall sign, because the basepoints are different.

The remaining interesting point is the final assertion of the proposition. To determine the sign of the matrix entries of the endomorphism ϵ between critical points a and b, one uses the canonical orientation of the product

$$\mathbb{Z}\Lambda(a, b)^e \otimes \mathbb{Z}\Lambda(a) \otimes \mathbb{Z}\Lambda(b).$$

Orienting this product is equivalent to orienting the moduli spaces $M(X, \Sigma)^e$ for the product $(X, \Sigma) = S^1 \times (Y, K)$. We have described above how these moduli spaces are canonically oriented once one has an orientation of the components of Σ (which are tori). Here the class e has pairing 1 with each of the components. So changing the orientation of any component changes the canonical orientation of $M(X, \Sigma)^e$ and changes the sign of all the matrix entries of ϵ. □

5.6 Passing to $Z_{n,-1}$

Recall that the algebra $\bar{\mathcal{A}}$ is defined as the quotient of $\mathcal{A}_n$ in which all the δ_i are equal (see equation (32)), and let $w^k_{n,\eta} \in \mathcal{A}_n$ be the elements from Definition 4.2. The image of $w^k_{n,\eta}$ in $\bar{\mathcal{A}}$ depends only on the cardinality of the subset $\eta \subset \pi$, not otherwise on its elements, and we write this element of $\bar{\mathcal{A}}$ as

$$\bar{w}^k_{n,h} = w^k_{n,\eta} + \langle \delta_i - \delta_j \rangle_{i,j} \in \bar{\mathcal{A}} \tag{64}$$

when $|\eta| = h$. Recall from (33) that we can write $I(Z_{n,-1})$ as $\mathcal{A}_n / \mathcal{J}_{n,-1}$ or as $\bar{\mathcal{A}}/\bar{\mathcal{J}}_{n,-1}$ and that $\mathcal{J}_{n,-1}$ contains $\mathcal{J}_n$ (Proposition 3.22). Propositions 5.1 and 5.2 therefore yield the following version for $Z_{n,-1}$.

Proposition 5.12 *Write $n = 2m+1$, let h be an integer satisfying the conditions (47), and let $\bar{w}^m_{n,h}$ be defined as above. Then there is an element $\overline{W}^m_h \in \bar{\mathcal{J}}_{n,-1}$ of the filtered algebra $\bar{\mathcal{A}}$ in filtration degree m whose leading term is $\bar{w}^m_{n,h}$. The subleading term of $\overline{W}^m_h$ is given by $\epsilon\tau^{n-2h}\bar{w}^{m-1}_{n,h'}$, where $h' = n-h$. Thus*

$$\overline{W}^m_h = \bar{w}^m_{n,h} + \epsilon\tau^{n-2h}\bar{w}^{m-1}_{n,h'} \pmod{\bar{\mathcal{A}}^{(m-2)}}.$$

The element $\overline{W}^m_h$ in $\bar{\mathcal{A}}$ is the image of $W^m_\eta \in \mathcal{J}_n$ under the quotient map $\mathcal{A}_n \to \bar{\mathcal{A}}$. □

We have not yet established that $\bar{\mathcal{J}}_{n,-1}$ is the image of $\mathcal{J}_n$, so we do not know yet that the elements $\overline{W}^m_h$ generate the ideal of relations $\bar{\mathcal{J}}_{n,-1}$ for $I(Z_{n,-1})$. We turn to this next.

Proposition 5.13 *When $n = 2m+1$, the elements $\overline{W}^m_h$ for h in the range $0 \le h \le n$ with $h = \frac{1}{2}(n+1)$ mod 2 are a set of generators for the ideal $\bar{\mathcal{J}}_{n,-1} \subset \bar{\mathcal{A}}$. In particular, $\bar{\mathcal{J}}_{n,-1}$ is the image of $\mathcal{J}_n$ in $\bar{\mathcal{A}}$.*

Proof The quotient $\bar{\mathcal{A}}/\bar{\mathcal{J}}_{n,-1}$ is $I(Z_{n,-1})$ which we know to be a free $\mathcal{R}$-module of rank $\frac{1}{4}(n^2-1)$ by Corollary 3.19. If $\mathcal{J}' \subset \bar{\mathcal{J}}_{n,-1}$ denotes the ideal generated by the elements $\overline{W}^m_h$, then the desired equality $\mathcal{J}' = \bar{\mathcal{J}}_{n,-1}$ will follow if we can prove that $\bar{\mathcal{A}}/\mathcal{J}'$ has the same rank. The leading m^{th}-degree terms of the elements $\overline{W}^m_h$ are the elements $\bar{w}^m_{n,h}$, so let us denote by $\bar{J}_n \subset \bar{\mathcal{A}}$ the ideal generated by these leading terms. (This is the image in $\bar{\mathcal{A}}$ of the ideal of relations $J_n \subset \mathcal{A}_n$ for the ordinary cohomology ring $H^*(\mathrm{Rep}(Z_n);\mathcal{R})$ in (20).) It will therefore suffice to show that $\mathcal{A}/\bar{J}_n$ has rank $\frac{1}{4}(n^2-1)$, and this is the content of the lemma below, which completes the proof. □

Lemma 5.14 *Write $n = 2m+1$ again and let $\bar{J}_n \subset \bar{\mathcal{A}}$ be as above, generated by the elements $\bar{w}^m_{n,h}$. Then $\bar{J}_n$ is the m^{th} power $\langle\alpha,\delta\rangle^m$ of the ideal $\langle\alpha,\delta\rangle$. In particular, the rank of $\bar{\mathcal{A}}/\bar{J}_n$ is $m(m+1)$, which is also equal to $\frac{1}{4}(n^2-1)$.*

Remark The quotient of a polynomial algebra in two variables by the m^{th} power of the maximal ideal at 0 has rank $\frac{1}{2}m(m+1)$. The extra factor of two in the lemma arises because of the extra generator ϵ in the algebra $\bar{\mathcal{A}}$.

Proof Recall that $w^m_{n,\eta}$ arises from the formal computation of $c_m(-\mathrm{index}(P))$, where P is a family of Fredholm operators, Definition 4.2. The formula (44) for the Chern character of $-\mathrm{index}(P)$ becomes the following, after passing to the formal completion of the quotient ring $\bar{\mathcal{A}}$ in which all the δ_i are equal:

$$(m-1)\cosh(\delta) + \frac{\sinh(\delta)}{\delta}\bigl(\alpha + \bigl(h - \tfrac{1}{2}n\bigr)\delta\bigr). \tag{65}$$

Passing from the Chern character to the m^{th} Chern class, we find that the image of $c_m(-\mathrm{index}(P))$ in $\bar{\mathcal{A}}$ has the form

$$V_m(B_h,\delta),$$

where $V_m(x,y)$ is a homogeneous polynomial of degree m in two variables and $B_h = \alpha + \bigl(-h + \frac{1}{2}n\bigr)\delta$. Furthermore the coefficient of x^m in V_m is $1/m!$.

Thus $\bar{J}_n$ is generated by the elements $V_m(B_h,\delta)$, for h in the range $0 \le h \le n$ with $h = \frac{1}{2}(n+1)$ mod 2. The assertion of the lemma is equivalent to the statement that the homogeneous polynomials $V_m\bigl(x + \bigl(h - \frac{1}{2}n\bigr)y, y\bigr)$ in $\mathbb{Q}[x,y]$ span the space of homogeneous degree-m polynomials. This in turn is true because $h - \frac{1}{2}n$ runs through $m+1$ distinct values in $\mathbb{Q}$ as h runs through its allowed range. (This is the same assertion as the statement that any $m+1$ distinct translates of a polynomial $f(x)$ of degree m are necessarily independent.) □

6 Calculation of the ideals

6.1 Hilbert schemes of points in the plane

We present here and in Section 6.2 below some material on Hilbert schemes of points in the plane, specialized to the particular situation for which we have application. General references are [26] for Section 6.1 and [9] for Section 6.2.

Let A be the algebra $k[x,y]$, with k a field, which we may take to be $\mathbb{C}$. Let $A_n \subset A$ be the subspace of homogeneous polynomials of degree n, and let $A^{(n)} = \bigoplus_{k \le n} A_k$. Let $\mathfrak{m} \subset A$ be the maximal ideal $\langle x, y\rangle$, and consider the m^{th} power $\mathfrak{m}^m$, which has generators

$$\mathfrak{m}^m = \langle x^m, x^{m-1}y, \dots, y^m\rangle. \tag{66}$$

The colength of $\mathfrak{m}^m$ (the dimension of the quotient $A/\mathfrak{m}^m$ as a k-vector space) is $N = \frac{1}{2}m(m+1)$, and a vector space complement is the linear subspace $A^{(m-1)}$:

$$A = \mathfrak{m}^m \oplus A^{(m-1)}.$$

We can consider $\mathfrak{m}^m$ as defining a point in the Hilbert scheme $\mathcal{H}^N$ which parametrizes all ideals of colength N in A. In the Hilbert scheme, there is an open neighborhood $U \ni \mathfrak{m}^m$ defined as

$$U = \{ I \in \mathcal{H}^N \mid A = I \oplus A^{(m-1)} \}. \tag{67}$$

For $I \in U$, there is the projection to the second factor, $A \to A^{(m-1)}$ with kernel I:

$$\varphi_I : A \to A^{(m-1)}.$$

It is an elementary matter to check that the restriction of φ_I to A_m completely determines I, and that I is in fact generated by

$$I = \langle a - \varphi_I(a) \mid a \in A_m \rangle.$$

We have in particular $a = \varphi_I(a)$ mod I for all $a \in A_m$.

The map $\varphi = \varphi_I$ is constrained by the condition that its kernel is an ideal rather than just a codimension-N linear subspace in A. To see how, consider elements $a, a' \in A_m$ with

$$xa = ya'.$$

We have $a = \varphi(a) \bmod I$, and therefore $xa = x\varphi(a)$, and applying φ again

$$xa = \varphi(x\varphi(a)) \pmod{I}.$$

Similarly with ya' so $\varphi(y\varphi(a')) = \varphi(x\varphi(a)) \bmod I$. However both sides of the last equality lie in the complementary subspace $A^{(m-1)}$, so in fact

$$\varphi(y\varphi(a')) = \varphi(x\varphi(a)). \tag{68}$$

Conversely, if we are given a linear map $\psi\colon A_m \to A^{(m-1)}$ satisfying the constraint (68), then there exists a unique (well-defined) extension to a linear map $\varphi\colon A \to A^{(m-1)}$ characterized by $\varphi(x^i y^j a) = \varphi(x^i y^j \varphi(a))$, and the kernel of φ is then an ideal I belonging to $U \subset \mathcal{H}^N$.

To expand on the constraint (68), write

$$\varphi|_{A_m} = \varphi_1 + \varphi_2 + \dots + \varphi_m,$$

where $\varphi_r : A_m \to A_{m-r}$, and use the fact that $\varphi|_{A_k} = 1$ for $k < m$ to obtain

$$\varphi(y\varphi_1(a')) + y\varphi_2(a') + \dots + y\varphi_m(a') = \varphi(x\varphi_1(a)) + x\varphi_2(a) + \dots + x\varphi_m(a).$$

Finally compare terms of like degree to see that

$$y\varphi_{r+1}(a') - x\varphi_{r+1}(a) = -\varphi_r(y\varphi_1(a')) + \varphi_r(x\varphi_1(a)) \tag{69}$$

for all $r \geq 1$ and all $a, a' \in A_m$ with $ya' = xa$. If we write $a' = xb$ and $a = yb$ for $b \in A_{m-1}$, the constraint becomes

$$y\varphi_{r+1}(xb) - x\varphi_{r+1}(yb) = -\varphi_r(y\varphi_1(xb)) + \varphi_r(x\varphi_1(yb)),$$

which we can express as

$$L_r(\varphi_{r+1}) = Q_r(\varphi_1, \varphi_r), \tag{70}$$

where $L_r : \mathrm{Hom}(A_m, A_{m-r-1}) \to \mathrm{Hom}(A_{m-1}, A_{m-r})$ is a linear map and Q_r is a bilinear expression. It is easy to verify that the operator L_r is injective (see below), so the constraints determine φ_{r+1} once φ_r and φ_1 are known.

We have shown:

Lemma 6.1 *Given a k-linear map $\varphi_1 : A_m \to A_{m-1}$, there exists at most one linear map $\varphi = \varphi_1 + \varphi_2 + \dots + \varphi_m$, with $\varphi_r : A_m \to A_{m-r}$, such that constraints (69) hold. The ideal I generated by the elements $\{a - \varphi(a) \mid a \in A_m\}$ then belongs to the open set $U \subset \mathcal{H}^N$. Every ideal in U arises in this way.* □

The lemma exhibits U as a closed subset of the vector space $\mathrm{Hom}_k(A_m, A_{m-1})$, which has dimension $m(m+1) = 2N$. This subset is also invariant under the action by scalars. It will follow that $U \cong \mathrm{Hom}_k(A_m, A_{m-1})$ if it can be shown that U has dimension $2N$. To do this, one can show that U contains an ideal I whose zero set consists of N distinct points in the plane k^2. Such an ideal can be realized as the "distraction" of $\mathfrak{m}^m$. This is the ideal I generated by the elements

$$u_h = \Big(\prod_{0\le j<h}(x-j)\Big)\Big(\prod_{0\le l<m-h}(y-l)\Big) \quad \text{for } h=0,\dots,m,$$

(allowing that one of the products may be empty). Its zero-set is the set of lattice points (j,l) in the first quadrant with $j+l<m$.

Proposition 6.2 *Given a k-linear map $\varphi_1 : A_m \to A_{m-1}$ there exists exactly one linear map $\varphi = \varphi_1 + \varphi_2 + \cdots + \varphi_m$, with $\varphi_r : A_m \to A_{m-r}$, such that the ideal I generated by the elements $\{a-\varphi(a) \mid a \in A_m\}$ has colength N. The matrix entries of φ_r for $r \ge 2$ can be expressed as polynomials in the matrix entries of φ_1.* □

The proposition tells us that, at each stage r in the equations (70), the right-hand side $Q_r(\varphi_1,\varphi_r)$ is in the image of the linear operator L_r. If we choose a right-inverse P_r for L_r, then we can express the iterative solution as

$$\varphi_{r+1} = P_r Q_r(\varphi_1,\varphi_r). \tag{71}$$

To give P_r explicitly, let us temporarily make our polynomials inhomogeneous by setting $y=1$, so identifying A_m with the polynomials in x of degree at most m, and let us write

$$u_k = \varphi_{r+1}(x^k)$$

as a polynomial of degree at most $m-r-1$ in x. Then the equations (70) take the form

$$u_{k+1} - xu_k = v_k$$

for $k=0,\dots,m-1$, where v_k is a given polynomial in x of degree at most $m-r$ and the equations are to be solved for u_k of degree at most $m-r-1$. If a solution exists, then

$$u_m = v_{m-1} + xv_{m-2} + \cdots + x^{m-1}v_0 + x^m u_0.$$

Since all polynomials u_k and v_k here have degree less than m, this equation determines the coefficients of u_0 as linear combinations of the coefficients of the v_k:

$$u_0 = -(x^{-m}v_{m-1} + x^{-m+1}v_{m-2} + \cdots + x^{-1}v_0)_+,$$

where the subscript $+$ means to discard the negative powers of x. Having found u_0, we can express the complete solution, if it exists, by the recurrence

$$u_{k+1} = \mathrm{trunc}_{m-r-1}(v_k + xu_k),$$

where trunc_{m-r-1} is the truncation of the polynomial to the given degree. Whether or not a solution exists, this process defines u_k as a linear function of the v's, and so defines a right inverse P_r for the linear map L_r. In this form, the coefficients of P_r are integers, and this allows us to pass to any ring. These leads to the following version.

Proposition 6.3 *Let R be a Noetherian ring, let $A = R[x, y]$ and let $I \subset A$ be an ideal such that*

- *A/I is a free R-module of rank $N = \frac{1}{2}m(m+1)$;*
- *there is an R-module homomorphism $\varphi\colon A_m \to A^{(m-1)}$ such that $a - \varphi(a) \in I$ for all $a \in A_m$.*

Then I is generated by the elements $a - \varphi(a)$ for $a \in A_m$. Furthermore, if we write

$$\varphi = \varphi_1 + \varphi_2 + \cdots + \varphi_m,$$

with $\varphi_r\colon A_m \to A_{m-r}$, then φ_r for $r \geq 2$ is determined by φ_1 through the iterative solution (71). This establishes a bijection between ideals I satisfying the above two constraints and module homomorphisms $\varphi_1\colon A_m \to A_{m-1}$.

Proof If I satisfies the second condition, the relations $a = \varphi(a) \bmod I$ show that the map $A^{(m-1)} \to A/I$ is surjective. The first of the two conditions tells us that these are free R-modules of equal rank, and it follows that the map is an isomorphism because R is Noetherian. Thus we have a direct sum decomposition $A = I \oplus A^{(m-1)}$. As before, the constraints then lead to the relations (71), which determine φ_r for $r \geq 2$. □

6.2 Syzygies

Proposition 6.3, which determines φ entirely in terms of φ_1, will be applied in Section 6.3 to see that the generators $\overline{W}_h^m$ of the ideal $\bar{\mathcal{J}}_{n,-1}$ can be determined completely in terms of the leading and subleading terms. (The subleading terms are already supplied by Proposition 5.12.) This will provide a complete description of the instanton homology $I(Z_{n,-1})$. First, however, we pursue further our discussion of the Hilbert scheme of points in the plane, to explain that the way in which φ_1 determines φ can be packaged by considering the syzygies of the module A/I. This will lead to quite explicit formulae for the generators.

We return temporarily to the case $A = k[x, y]$ as above, and we take $k = \mathbb{C}$. Fix m again and write $N = \frac{1}{2}m(m+1)$. Let $U \subset \mathcal{H}^N$ be as before (67). An ideal $I \in U$ contains no nonzero polynomials of degree less than m and is generated by $m+1$ elements whose leading terms are a basis for A_m. Choose a basis for A_m so as to identify $A_m = A^{\oplus(m+1)}$, say the monomial basis (66). We then have generators for I in the form

$$g_i = x^{m-i} y^i - \varphi(x^{m-i} y^i).$$

Because A has dimension 2, a resolution of A/I has only one more step, and we therefore have a presentation of the ideal I in the form

$$0 \to A^{\oplus m} \xrightarrow{S} A^{\oplus(m+1)} \xrightarrow{g} I \to 0. \tag{72}$$

Here $g = (g_i)$ is given by the generators (the relations in A/I) and S is the matrix of syzygies.

In the special case that $I = \mathfrak{m}^m$ and $g_i = x^{m-i}y^i$ the syzygy matrix can be taken to be

$$S_0 = \begin{pmatrix} -y & 0 & 0 & \dots & 0 \\ x & -y & 0 & \dots & 0 \\ 0 & x & -y & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -y \\ 0 & 0 & 0 & \dots & x \end{pmatrix}. \tag{73}$$

Lemma 6.4 *For a general $I \in U$, the syzygy matrix S has the form $S = S_0 + S_1$, where S_0 is as above and S_1 is a matrix of scalars (polynomials of degree* 0*).*

Proof Write $g = g(0) + g(1) + \cdots + g(m)$, where $g(r)$ is a vector of homogeneous polynomials of degree $m-r$ and $g(0)$ is the basis of monomials of degree m. (So the entries of $g(r)$ are the polynomials $-\varphi_r(x^{m-i}y^i)$.) Let

$$g^t = g(0) + tg(1) + t^2 g(2) + \cdots,$$

and let I^t be the ideal generated by the entries of g^t. Because the colength of $I = I^1$ is the same as that of I^0, this is a flat family, and the syzygy matrix S^0 for g^0 therefore lifts to a syzygy matrix S^t, whose entries are polynomials in (x, y, t) and which coincides with S^0 at $t = 0$. Because the entries of g^t are homogeneous (of degree m) in (t, x, y), we may assume that S^t is also homogeneous. Since S_0 has homogeneous degree 1, so too does S^t, and it follows that

$$S^t = S_0 + tS_1,$$

where the entries of S_1 have degree 0 in (x, y). □

Note that in the above lemma, the matrix S_1 is entirely determined by the leading term $g(0)$ and the subleading term $g(1)$ (or equivalently by $\varphi_1 \colon A_m \to A_{m-1}$) via the condition

$$g(0) \cdot S_1 + g(1) \cdot S_0 = 0. \tag{74}$$

Quite concretely, taking $g(0)$ to be again the standard monomial basis, taking S_0 as above, and writing the subleading terms $g_i(1)$ as

$$g_i(1) = \sum_{j=0}^{m-1} G_{ij} x^{m-1-j} y^j,$$

then

$$S_1 = \begin{pmatrix} -G_{1,0} & -G_{2,0} & \dots & -G_{m,0} \\ G_{0,0} - G_{1,1} & G_{1,0} - G_{2,1} & \dots & G_{m-1,0} - G_{m,1} \\ \vdots & \vdots & \ddots & \vdots \\ G_{0,m-2} - G_{1,m-1} & G_{1,m-2} - G_{2,m-1} & \dots & G_{m-1,m-2} - G_{m,m-1} \\ G_{0,m-1} & G_{1,m-1} & \dots & G_{m-1,m-1} \end{pmatrix}. \tag{75}$$

Proposition 6.5 *Let $S = S_0 + S_1$ be the syzygy matrix as above, so that S_0 is the matrix of syzygies of the standard monomial ideal $\mathfrak{m}^m$ and S_1 is determined by the subleading terms $g_i(1)$ by (75). Then the generators $g_0, \dots, g_m$ of the ideal I are precisely the $m \times m$ minors of the $(m+1) \times m$ matrix S (ie the determinants of the matrices obtained by deleting a single row of S, with alternating sign).*

Proof Let $h = (h_0, h_1, \dots, h_m)$ be the minors. We have both $h \cdot S = 0$ (by standard properties of determinants) and $g \cdot S = 0$ (by construction), and it follows that $ah = bg$ for some a and b in A, because the rank of the kernel of S^{T} is 1. On the other hand, by inspection, the leading term of h_i is the same as that of g_i, namely $x^{m-i} y^i$. So $h = g$. □

Finally, we can pass from the case of $k[x, y]$ to more general coefficients without difficulty. The next proposition summarizes the situation.

Proposition 6.6 *As in Proposition 6.3, let R be a Noetherian ring, let $A = R[x, y]$ and let $I \subset A$ be an ideal such that*

- *A/I is a free R-module of rank $N = \frac{1}{2}m(m+1)$;*
- *there is an R-module homomorphism $\varphi\colon A_m \to A^{(m-1)}$ such that $a - \varphi(a) \in I$ for all $a \in A_m$.*

Let $(g_0(0), \dots, g_m(0))$ be a basis for $A_m \cong A^{\oplus(m+1)}$ and let

$$g_i = g_i(0) - \varphi(g_i(0)) = g_i(0) + g_i(1) + g_i(2) + \cdots + g_i(m),$$

where $g_i(j)$ is homogeneous of degree $m - j$. Then the elements $(g_0, \dots, g_m)$ are generators of the ideal I. Furthermore, let S_0 be a matrix of syzygies for the leading parts $g_i(0)$, with entries which are homogeneous of degree 1, and let S_1 be the matrix of scalars determined by the subleading parts $g_i(1)$ via equation (74). Then:

(i) *The matrix $S = S_0 + S_1$ is the matrix of syzygies for the generators $(g_0, g_1, \dots, g_m)$ of the ideal I.*

(ii) *If $h_0, \dots, h_m$ are the $m \times m$ minors of the matrix S, then $(h_0, h_1, \dots, h_m)$ is a set of generators for I.*

(iii) *If S_0 is chosen so that its minors are the leading terms $(g_0(0), \dots, g_m(0))$, then the generators g_i for I are equal to the minors h_i of S.*

In this way, the generators g are determined by their leading and subleading terms, $g(0)$ and $g(1)$.

Proof We may take it that $g(0)$ is the standard monomial basis and that S_0 is given (73). The matrix S_1 is then given by (75) where the terms $G_{i,j}$ are the coefficients of the subleading terms $g(1)$. According to Proposition 6.3, the lower terms in the entries of g are expressible as universal polynomials in the coefficients of $g(1)$. On the other hand, the recipe in terms of the minors of S expresses the lower terms of g as polynomials in the coefficients of $g(1)$, at least when R is a field k. The polynomials occurring in the minors have integer coefficients, and must coincide with the polynomials in Proposition 6.3. □

6.3 Equations for the curve D_n

Let $\mathcal{R} = \mathbb{Q}[\tau, \tau^{-1}]$. Let R temporarily denote the ring

$$R = \mathcal{R}[\epsilon]/\langle \epsilon^2 - 1 \rangle.$$

The algebra $\bar{\mathcal{A}}$ in (32) is $R[\alpha, \delta]$ and the instanton homology $I(Z_{n,-1})$ is described as a quotient $\bar{\mathcal{A}}/\bar{\mathcal{J}}_{n,-1}$ in (33). We know that $I(Z_{n,-1})$ is a free $\mathcal{R}$-module of rank $\frac{1}{4}(n^2-1) = m(m+1)$ from Corollary 3.19, and it is a free R-module of rank $\frac{1}{2}m(m+1)$. We know that there are elements $\overline{W}^m_h$ in $\bar{\mathcal{J}}_{n,-1}$ of degree m in (α, δ) having the form

$$\overline{W}^m_h = w(0)_h + \epsilon w(1)_h + \cdots = \bar{w}^m_{n,h} + \epsilon \bar{w}^{m-1}_{n,h'} + \cdots . \tag{76}$$

(see Proposition 5.12). The leading and subleading terms $w(0)$ and $\epsilon w(1)$ are known from Proposition 5.12 and Definition 4.2. We also know that the leading terms $w(0)_h$ are a basis for the m^{th} power of the maximal ideal, $\langle \alpha, \delta \rangle^m$, by Lemma 5.14.

The ideal $\bar{\mathcal{J}}_{n,-1} \subset R[\alpha, \delta]$ therefore satisfies the hypotheses of Propositions 6.3 and 6.6. In the notation of Proposition 6.6, we know φ_1 explicitly, as it is determined by the subleading terms $\epsilon w(1)_h$. We therefore have the following result as a corollary. In this statement, we write $n = 2m+1$ as usual.

Theorem 6.7 *Let $\bar{\mathcal{J}}_{n,-1}$ be the ideal of relations for the instanton homology $I(Z_{n,-1})$ with local coefficients, and let*

$$\overline{W}^m_h = w(0)_h + w(1)_h + \cdots + w(m)_h, \quad \text{with } 0 \le h \le n \text{ and } h = m+1 \bmod 2,$$

be the generators for this ideal, as in (76). There are explicit polynomial formulae which express the coefficients of all the lower terms $w(r)_h$ for $r \ge 2$ in terms of the leading and subleading terms

$$w(0)_h = \bar{w}^m_{n,h} \quad \text{and} \quad w(1)_h = \epsilon \bar{w}^{m-1}_{n,n-h}$$

in Proposition 5.12. If the syzygy matrix

$$S = S_0 + S_1$$

is constructed as in Proposition 6.6, as a matrix whose entries are inhomogeneous linear forms in (α, δ) with coefficients in $R = \mathcal{R}[\epsilon]/\langle \epsilon^2 - 1 \rangle$, then the generators $\overline{W}^m_h$ are the $m \times m$ minors of S. □

To obtain a final form for the generators, we now need to find an explicit formula for the syzygy matrix S, starting from our formulae for $w(0)_h$ and $w(1)_h$. In Section 6.2 above, we illustrated the calculation when the leading terms of the generators were the standard monomial basis in the polynomials in two variables, so that the term S_0 was the standard syzygy matrix (73). The leading terms $w(0)_h$ are not monomials in our case, so we must first write down a suitable matrix of syzygies S_0 for these.

From Proposition 4.7, on setting all δ_i equal to δ to pass from the ring $\mathcal{A}_j$ to $\overline{\mathcal{A}}$, we obtain an expression for $w(0)_h = \overline{w}^m_{n,h}$ as a product of linear factors. It is convenient to remove the combinatorial factor of $1/m!$ and write

$$g(0)_h = m!\,w(0)_h = m!\,\overline{w}^m_{n,h},$$

for which Proposition 4.7 yields the formula

$$\begin{aligned} g(0)_h &= \prod_{\substack{j=-m+1 \\ j=-m+1 \bmod 2}}^{m-1} \big(\alpha + \tfrac{1}{2}(2h-n-2j)\delta\big), \\ &= \big(\alpha + \tfrac{1}{2}(2h-3)\delta\big)\big(\alpha + \tfrac{1}{2}(2h-7)\delta\big)\cdots\big(\alpha + \tfrac{1}{2}(2h-4m+1)\delta\big). \end{aligned}$$

We introduce some abbreviated notation, setting

$$L(k) = (\alpha + k\delta/2) \quad \text{and} \quad P(k,l) = L(k)L(k+4)L(k+8)\cdots L(l).$$

(The latter notation will be used only when $k = l \bmod 4$.) Then we can write

$$g(0)_h = P(2h-4m+1, 2h-3).$$

If we compare $g(0)_h$ to $g(0)_{h+2}$, only the first and last factors in this product differ, so we have a relation

$$-L(2h+1)g(0)_h + L(2h-4m+1)g(0)_{h+2} = 0.$$

That is, for h' in the range $0 \le h' \le n-2$ with $h' = m+1 \bmod 2$, we have

$$\sum_h S_0^{h'h} g(0)_h = 0,$$

where

$$S_0^{h'h} = \begin{cases} -L(2h'+1) & \text{if } h = h', \\ L(2h'-4m+1) & \text{if } h = h'+2, \\ 0 & \text{otherwise.} \end{cases} \tag{77}$$

This is therefore the leading part S_0 of the required syzygy matrix $S = S_0 + S_1$. It is straightforward to verify that the minors of $S_0^{h'h}$ are the terms $g(0)_h$, as required.

We normalize the subleading terms just as we did the leading terms, so that

$$g(1)_h = m!\,w(1)_h = m!\,\epsilon\tau^{n-2h}\,\overline{w}^{m-1}_{n,n-h},$$

from Proposition 5.2. We then have the explicit formulae again from Proposition 4.7 (noting that $|\eta'| = n-h$),

$$\begin{aligned} g(1)_h &= m\epsilon\tau^{n-2h} \prod_{\substack{j=-m+2 \\ j=m \bmod 2}}^{m-2} \big(\alpha + \tfrac{1}{2}(n-2h-2j)\delta\big) \\ &= m\epsilon\tau^{n-2h} P(-2h+5, -2h+4m-3). \end{aligned}$$

To obtain the other term S_1 in the syzygy matrix, we need to solve the following equations for $S_1^{h'h}$:

$$\sum_h S_1^{h'h} g(0)_h + \sum_h S_0^{h'h} g(1)_h = 0,$$

where $h, h' = m+1 \bmod 2$, with $0 \le h \le n$ and $0 \le h' \le n-2$. Using the formulae for $g(0)_h$, $g(1)_h$ and $S_0^{h'h}$, we write this out as

$$\begin{aligned} 0 = \sum_h S_1^{h'h} P(2h-4m+1, 2h-3) \\ - m\epsilon\tau^{n-2h'} L(2h'+1) P(-2h'+5, -2h'+4m-3) \\ + m\epsilon\tau^{n-2h'-4} L(2h'-4m+1) P(-2h'+1, -2h'+4m-7). \end{aligned}$$

The solution $S_1^{h'h}$ consisting of scalars in R is unique, because the terms $g(0)_h$ are a basis for the homogeneous polynomials of degree m in (α, δ).

The last two of the three terms above have at least $m-2$ common linear factors $L(k)$, and have $m-1$ common factors in two edge cases. The $m-2$ factors are the expression

$$Q(h') = P(-2h'+5, -2h'+4m-7).$$

The edge cases are $h' = 0$ (which only occurs when m is odd), and $h' = n-2$ (which occurs only when m is even). In these two edge cases the $m-1$ common factors are respectively,

$$Q_+ = L(1)Q(0) = P(1, 4m-7) \quad \text{and} \quad Q_- = L(-1)Q(n-2) = P(-4m+7, -1).$$

We seek a solution $S_1^{h'h}$ to the above equations in the special form where, for each h', the coefficients $S_1^{h'h}$ are nonzero only for those values of h for which $g(0)_h$ is divisible by $Q(h')$ (respectively Q_+ or Q_- in the edge cases). Excluding the edge cases, there are three such values of h, namely

(78) $$h \in \{n-h'-3, n-h'-1, n-h'+1\}, \quad \text{where } 0 < h' < n-2.$$

In each of the edge cases, there are two such values of h:

(79) $$h \in \begin{cases} \{n-3, n-1\} & \text{if } h' = 0, \\ \{1, 3\} & \text{if } h' = n-2. \end{cases}$$

In the nonedge cases, the equations for the nonzero coefficients $S_1^{h'h}$ then take the general shape

(80) $$S_1^{h', n-h'-3} A + S_1^{h', n-h'-1} B + S_1^{h', n-h'+1} C + D = 0,$$

where A, B and C are the homogeneous quadratic polynomials in (α, δ) given by

$$g(0)_h / Q(h'), \quad \text{where } h \in \{n-h'-3, n-h'-1, n-h'+1\},$$

and D is a quadratic polynomial

$$D = \big(S_0^{h', h'} g(1)_{h'} + S_0^{h', h'+2} g(1)_{h'+2}\big) / Q(h').$$

Explicitly,

$$\begin{aligned}
A &= L(-2h'-3)L(-2h'+1),\\
B &= L(-2h'+1)L(-2h'+4m-3),\\
C &= L(-2h'+4m-3)L(-2h'+4m+1),\\
D &= m\epsilon\tau^{n-2h'}\bigl(-L(2h'+1)L(-2h'+4m-3)+\tau^{-4}L(2h'-4m+1)L(-2h'+1)\bigr).
\end{aligned}$$

The three polynomials A, B and C are independent, and we know there to be a unique solution, which we can now find by equating coefficients of α^2, $\alpha\delta$ and δ^2 in (80). The two edge cases are similar. Thus in the case $h'=0$, the equations for the two unknown coefficients of S_1 take the form

$$S_1^{0,n-3}X + S_1^{0\,n-1}Y = Z, \tag{81}$$

where X, Y and Z are homogeneous linear forms in (α,δ), while in the case $h'=n-2$ we have similar equations

$$S_1^{n-2,1}X' + S_1^{n-2,3}Y' = Z'. \tag{82}$$

Solving the equations (80)–(82) for the coefficients $S_1^{h'h}$ leads to the following answer, valid for all h', whether or not we are in an edge case. We find

$$S_1^{h'h} = \begin{cases} \epsilon\tau^{n-4-2h'}(-n+2+h') & \text{if } h=n-h'-3,\\ \epsilon\tau^{n-4-2h'}(m-h'-1+(m-h')\tau^4) & \text{if } h=n-h'-1,\\ \epsilon\tau^{n-2h'}h' & \text{if } h=n-h'+1,\\ 0 & \text{otherwise,}\end{cases} \tag{83}$$

for all h', h in the range $0\le h\le n$ and $0\le h'\le n-2$ with the parity constraint $h=h'=m+1 \pmod 2$. So we have obtained the desired closed form for the generators of the ideal $\bar{\mathcal{J}}_{n,-1}$ for the instanton homology $I(Z_{n,-1})$:

Theorem 6.8 *Let $S=S_0+S_1$ be an $m\times(m+1)$ with rows indexed by h' and columns indexed by h in the range $0\le h\le n$ and $0\le h'\le n-2$ with the parity constraint $h=h'=m+1 \pmod 2$. Let the entries of S_0 be given by (77) and the entries of S_1 be given by (83), so that the entries of S belong to the ring $\bar{\mathcal{A}}=\mathbb{Q}[\tau,\tau^{-1},\epsilon,\alpha,\delta]/\langle\epsilon^2=1\rangle$. Then the normalized generators $m!\,\overline{W}{}^m_h$ of the ideal $\bar{\mathcal{J}}_{n,-1}$ are given by the $m\times m$ minors of S.*

Remark The matrix the matrix S has $m+1$ different $m\times m$ minors, and explicitly the generators of the ideal can be expressed as

$$m!\,\overline{W}{}^m_h = \pm\det S[h],$$

where $S[h]$ is obtained from S by deleting the column indexed by h. (Recall again that the indexing of the columns is by only those integers h with $h=m+1 \bmod 2$.) The signs alternate as usual. Although there are $m+1$ generators in this description, in fact only two generators suffice, as the following proposition states.

Proposition 6.9 *The ideal $\bar{\mathcal{J}}_{n,-1}$ is generated by the two elements $\overline{W}^m_{m-1}$ and $\overline{W}^m_{m+1}$, or equivalently by the two determinants*

$$G_1(n) = \det S[m-1] \quad \textit{and} \quad G_2(n) = \det S[m+1].$$

Proof It is sufficient to show that the matrix $S[m-1,m+1]$ obtained by deleting both columns $h=m-1$ and $h=m+1$ has full rank $m-1$. To do this, let us examine the $(m-1)\times(m-1)$ matrix T obtained from $S[m-1,m+1]$ by deleting either the first or last row, according as m is odd or even respectively. An inspection of the entries of S reveals first that the entries of T on the contra-diagonal are all units in $\bar{\mathcal{A}}$: they are nonzero integers times powers of τ. Furthermore, a reordering of the rows and columns makes T triangular, with these same units on the diagonal. The determinant of T is therefore nonzero, which shows that $S[m-1,m+1]$ indeed has full rank as desired. □

As illustration, when $m=3$ (ie $n=7$) the two generators $G_1(7)$ and $G_2(7)$ are

$$\tfrac{1}{48}\bigl(8\alpha^3+36\alpha^2\delta+22\alpha\delta^2-21\delta^3+24\epsilon\tau^3\alpha^2-72\epsilon\tau^3\alpha\delta+30\epsilon\tau^3\delta^2-(88\tau^2+16\tau^{-2})\alpha \\ -(52\tau^2+56\tau^{-2})\delta-24\epsilon\tau^5-96\epsilon\tau\bigr)$$

and

$$\tfrac{1}{48}\bigl(8\alpha^3-12\alpha^2\delta-26\alpha\delta^2+15\delta^3+24\epsilon\tau^{-1}\alpha^2+24\epsilon\tau^{-1}\alpha\delta-18\epsilon\tau^{-1}\delta^2-(40\tau^2+64\tau^{-2})\alpha \\ +(68\tau^2-32\tau^{-2})\delta-72\epsilon\tau-48\epsilon\tau^{-3}\bigr).$$

6.4 Relating different values of *n*

Theorem 6.8 provides a complete description of the instanton homology of $Z_{n,-1}$ with local coefficients, but we have not yet presented a full description for the case of Z_n. As preliminary material for this, we describe how the functoriality of instanton homology can be used to obtain relations between the ideal of relations in Z_n for different values of n.

The fact that the ideal $\mathcal{J}_n$ annihilates $I(Z_n)$ leads, via a standard approach, to the interpretation of the elements of $\mathcal{J}_n$ as universal relations that hold for the maps defined by general bifold cobordisms. To spell this out, let W be a homology-oriented bifold cobordism from Z^0 to Z^1, both of which are admissible. We have seen in Section 2.3 that W gives rise to homomorphisms of $\mathcal{R}$-modules

$$I(W,a)\colon I(Z^0)\to I(Z^1)$$

depending linearly on

$$a\in \mathrm{Sym}_*\bigl(H_2(W;\mathbb{Q})\oplus H_0(\Sigma(W);O)\bigr)\otimes\mathcal{R},$$

where O is the orientation bundle of the singular set $\Sigma(W)$ with coefficients $\mathbb{Q}$. Further, given a distinguished 2-dimensional class e we can use marked connections with nonzero w_2 to define maps

$$I(W,a)^e\colon I(Z^0)\to I(Z^1).$$

Using δ_p to denote the generator of the symmetric algebra corresponding the homology class of a point $p \in \Sigma(W)$ with local orientation, let us imitate the definition of $\mathcal{A}_n$ and write

$$\mathcal{A}(W) = \big(\mathrm{Sym}_*\big(H_2(W;\mathbb{Q}) \oplus H_0(\Sigma(W);O)\big) \otimes \mathcal{R}[\epsilon]\big)/\langle \epsilon^2 - 1, \delta_p^2 - \delta_q^2\rangle_{p,q}$$

where the indexing in the ideal runs through all pairs of points p, q in $\Sigma(W)$. We obtain a linear map

$$\Psi\colon \mathcal{A}(W) \to \mathrm{Hom}(I(Z^0), I(Z^1)) \quad \text{by} \quad a_1 + \epsilon a_2 \mapsto I(W, a_1) + I(W, a_2)^e. \tag{84}$$

This construction has been phrased so that, in the special case that W is the product cobordism from Z_n to itself and e is the generator of H_2, the algebra $\mathcal{A}(W)$ coincides with $\mathcal{A}_n$ as defined above, and the map Ψ is the action of the algebra $\mathcal{A}_n$ on the module $I(Z_n)$ via the instanton module structure.

Continuing with the case of a general cobordism W, we suppose now that we have an embedded orbifold sphere $S \subset W$ meeting the singular set in n orbifold points $\{p_1, \dots, p_n\}$. Choose an orientation for S and define local orientations for the singular set in the neighborhood of the n points of intersection in such a way that the intersections are all positive. In this way we obtain elements $\delta_{p_k} \in \mathcal{A}(W)$, where for the class e in the definition of $\mathcal{A}(W)$ we take the fundamental class $[S]$. Let the singular set of W also be oriented globally, and let the operators $\delta^*_{p_k}$ be defined using this global orientation of the singular set. We then have

$$\delta^*_{p_k} = \nu_k \delta_{p_k},$$

where $\nu_k = \pm 1$ according to whether the orientations agree or not.

Let us suppose that the normal bundle of S is trivial so that the boundary of the tubular neighborhood of S is a copy of Z_n. From the definitions, there is a natural map

$$i_*\colon \mathcal{A}_n \to \mathcal{A}(W)$$

arising from the inclusion, which we define so that $i_*(\delta_{p_k}) = \nu_k \delta^*_{p_k}$ for all k, while $i_*(\alpha) = [S] \in H_2(W)$ and $i_*(\epsilon) = (-1)^f \epsilon$, where f is the number of signs ν_k which are -1.

Proposition 6.10 *For an embedded orbifold sphere $S \subset W$ as above, the ideal $\mathcal{J}_n$ lies in the kernel of the map Ψ defined at (84). That is, for $a = a_1 + \epsilon a_2 \in \mathcal{J}_n \subset \mathcal{A}_n$, we have*

$$I(W, i_*(a_1)) + (-1)^f I(W, i_*(a_2))^e = 0.$$

More generally, if b is another class in $\mathcal{A}(W)$ which an be expressed as a polynomial in cycles disjoint from S, then we have

$$I(W, i_*(a_1)b) + (-1)^f I(W, i_*(a_2)b)^e = 0.$$

Proof In its structure, this is a standard argument based on the observation that we can factor the cobordism W as a composite cobordism in which the first factor is the cobordism from Z^0 to $Z^0 \amalg Z_n$. For the disjoint union, we can construct the instanton homology as a tensor product, and then we apply functoriality. See [22] and [32], for example, for similar arguments. The details of the signs, in particular the sign $(-1)^f$ come from Proposition 5.11. □

Our application of Proposition 6.10 is equivalent to [32, Corollary 2.6.8]. (A closely related result appears in [28].) Suppose that

$$n = n' + 2f, \quad \text{where } f \geq 0.$$

Consider an embedding of the orbifold sphere $S = S^2_n$ in the trivial cobordism $W = [0,1] \times Z_{n'}$, representing the generator in homology. This means that S meets the singular locus $[0,1] \times K(Z_{n'})$ geometrically in $n' + 2f$ points, while the algebraic intersection number is n'. There are therefore $2f$ signed intersection points that cancel in pairs. Such a sphere $S \subset [0,1] \times Z_{n'}$ can be constructed by taking the standard generating sphere $S' \subset Z_{n'}$ and introducing $2f$ extra intersection points by doing f "finger moves" to the sphere S'. We take these extra intersection points to be the orbifold points numbered $n'+1, \dots, n'+2f$ in $S \cong S^2_n$, and we suppose that they all lie on the component $[0,1] \times K^{n'} \subset [0,1] \times K(Z_{n'})$. Among these $2f$ points, there are f of them that have negative intersection number, and we can take it that these are the points numbered $n'+f+1, \dots, n'+2f$ in S^2_n. There is a corresponding map

$$i^{n,n'}_* : \mathcal{A}_n \to \mathcal{A}_{n'}, \quad \text{where } n = n' + 2f,$$

and our choice of numbering means that it is given by

$$i^{n,n'}_*(\alpha) = \alpha, \qquad i^{n,n'}_*(\epsilon) = (-1)^f \epsilon \qquad \text{and} \qquad i^{n,n'}_*(\delta_k) = \begin{cases} \delta_k & \text{if } 1 \leq k \leq n', \\ \delta_{n'} & \text{if } n'+1 \leq k \leq n'+f, \\ -\delta_{n'} & \text{if } n'+f+1 \leq k \leq n'+2f. \end{cases}$$

Proposition 6.10 now yields the following.

Corollary 6.11 [32, Corollary 2.6.8] *When $n = n' + 2f$ and $i^{n,n'}_* : \mathcal{A}_n \to \mathcal{A}_{n'}$ is defined as above, we have an inclusion of ideals,*

$$i^{n,n'}_* \mathcal{J}_n \subset \mathcal{J}_{n'}. \qquad \square$$

With a little more work and an examination of the explicit formulae for the leading and subleading terms of the generators of $\mathcal{J}_n$ (Proposition 4.7), we can strengthen the above corollary as follows.

Proposition 6.12 *In the situation of Corollary 6.11 above, we have inclusions*

$$(\tau^4 - 1)^f \mathcal{J}_{n'} \subset i^{n,n'}_* \mathcal{J}_n \subset \mathcal{J}_{n'}.$$

In particular, the ideals $i^{n,n'}_* \mathcal{J}_n$ and $\mathcal{J}_{n'}$ become equal after tensoring with the field of fractions of the ring $\mathcal{R} = \mathbb{Q}[\tau, \tau^{-1}]$.

Proof It suffices to treat the case $f = 1$, so $n' = n-2$. Let $\eta_0 \subset 1, \dots, n-2$, and let $\eta_1, \eta_2 \subset \{1, \dots, n\}$ be respectively the same as η_0 and $\eta_0 \cup \{n-1, n-2\}$. From the explicit formulae, we see

$$i^{n,n-2}_*(w^m_{n,\eta_1}) = i^{n,n-2}_*(w^m_{n,\eta_2}),$$

because $i^{n,n-2}_*(B_{\eta_1}) = i^{n,n-2}_*(B_{\eta_2})$. Similarly

$$i^{n,n-2}_*(w^{m-1}_{n,\eta'_1}) = i^{n,n-2}_*(w^{m-1}_{n,\eta'_2}).$$

We therefore have (using the general shape of the subleading term)

$$\begin{aligned} i_*^{n,n-2}(W^m_{\eta_1} - W^m_{\eta_2}) &= (-1)^f \epsilon(\tau^{n-2h} - \tau^{n-2h-4}) i_*^{n,n-2}(w^{m-1}_{n,\eta_1'}) + \text{lower terms} \\ &= u(\tau^4 - 1) i_*^{n,n-2}(w^{m-1}_{n,\eta_1'}) + \text{lower terms}, \end{aligned}$$

where u is a unit in $\mathbb{Q}[\tau,\tau^{-1}]$. By the previous corollary, these belong to $\mathcal{J}_{n-2}$. It is now enough to show that the elements $i_*^{n,n-2}(w^{m-1}_{n,\eta_1'})$ generate the ideal j_{n-2} of relations in the ordinary cohomology of $\mathrm{Rep}(S^2_{n-2})$, because the statement about instanton homology will follow as before. From the formulae in Proposition 4.7, we see that this is the same as showing that the elements $w^{m-1}_{n-2,\eta_0'}$ generate the ideal j_{n-2}, which has already been established (as the case $n-2$) in Proposition 4.8. □

The homomorphism $i_*^{n,n'}$ does not pass to a homomorphism between the quotient rings $\bar{\mathcal{A}}$. But we can at least compose with the quotient map $\mathcal{A}_{n'} \to \bar{\mathcal{A}}$ to get the following immediate corollary. In the statement of the corollary, we note that the choices of sign in the definition of $i_*^{n,n'}$ are arbitrary and can be replaced by a more general phrasing.

Corollary 6.13 *Let $\nu \in \{\pm 1\}^n$ be any choice of signs. Write $n' = \sum \nu_i$ and assume $n' \geq 1$. Consider the homomorphism $\bar{\iota}_\nu \colon \mathcal{A}_n \to \bar{\mathcal{A}}$ defined by $\bar{\iota}_\nu(\delta_i) = \nu_i \delta$ for all i, and $\bar{\iota}_\nu(\epsilon) = (-1)^{(n-n')/2}\epsilon$. Then we have an inclusions of ideals in $\bar{\mathcal{A}} = \mathcal{R}[\delta,\alpha,\epsilon]/\langle \epsilon^2 - 1\rangle$,*

$$(\tau^4 - 1)^{(n-n')/2} \bar{\mathcal{J}}_{n',-1} \subset \bar{\iota}_\nu(\mathcal{J}_n) \subset \bar{\mathcal{J}}_{n',-1}. \qquad \square$$

We refer to the relations between the ideals in Corollaries 6.11 and 6.13 as "finger-move relations", because of the interpretation of the sphere S as having been obtained from the standard sphere $S' \subset W$ by finger moves.

Remark A second application of Proposition 6.10 will be given in the proof of Proposition 7.1 later in this paper.

6.5 Decomposition of the instanton curve

We are now ready to harness our understanding of $I(Z_{n,-1})$ from Theorem 6.8 to obtain a description of $I(Z_n)$. Write

$$V_n = \operatorname{Spec} \mathbb{Q}[\tau,\tau^{-1},\alpha,\delta_1,\dots,\delta_n,\epsilon].$$

The set of complex-valued points $V_n(\mathbb{C})$ is $\mathbb{C}^\times \times \mathbb{C}^{n+2}$, with τ a coordinate on the first factor. We can describe the $\mathcal{A}_n$-module $I(Z_n)$ geometrically as the coordinate ring of the closed subscheme

$$C_n \subset V_n$$

defined by the vanishing of the elements of the ideal $\mathcal{J}_n$ together with the additional relations that define the algebra $\mathcal{A}_n$, namely the vanishing of $\delta_i^2 - \delta_j^2$ and $\epsilon^2 - 1$. We can write $C_n = \operatorname{Spec}(I(Z_n))$, where $I(Z_n)$ is considered as a quotient ring of the algebra $\mathcal{A}_n$. To describe $I(Z_n)$ as an $\mathcal{A}_n$-module, we can therefore use geometrical language to describe the subscheme C_n. Note that the relation $\epsilon^2 = 1$ means that C_n is contained in the union of the two hyperplanes $\epsilon = 1$ and $\epsilon = -1$, so we may write

$$C_n = C_n^+ \cup C_n^-.$$

In a similar way, let us write

$$\bar{V} = \operatorname{Spec} \mathbb{Q}[\tau, \tau^{-1}, \alpha, \delta, \epsilon],$$

so that the instanton homology group $I(Z_{n,-1})$ defines, (via its ideal of relations $\bar{\mathcal{J}}_{n,-1}$ and the relation $\epsilon^2 = 1$), a subscheme $D_n = \operatorname{Spec}(I(Z_{n,-1}))$, which is a closed subscheme of $\bar{V}$:

$$D_n = D_n^+ \cup D_n^- \subset \bar{V}. \tag{85}$$

We can interpret Corollary 6.13 as describing a relation between the curves C_n for $I(Z_n)$ and D_n for $I(Z_{n,-1})$. First, given any choice of signs $\nu \in \{\pm 1\}^n$, write $n' = \sum \nu_i$, and suppose henceforth that this odd integer n' is positive. Write $f = (n - n')/2$. Define a morphism

$$\bar{\iota}_\nu^* \colon \bar{V} \to V_n$$

by $\delta_i \mapsto \nu_i \delta$ and $\epsilon \mapsto (-1)^f \epsilon$. Write

$$V_{n,\nu} \subset V_n$$

for the image of ι_ν^*. This is the linear subvariety cut out by the linear conditions $\nu_i \delta_i = \nu_j \delta_j$. Their union is the subvariety defined by $\delta_i^2 = \delta_j^2$ for all i, j; so we have

$$C_n \subset \bigcup_\nu V_{n,\nu}.$$

We have an isomorphic copy of the affine scheme $D_{n'}$ as the image of $D_{n'}$ under the embedding ι_ν^*:

$$\iota_\nu^*(D_{n'}) \subset V_{n,\nu}. \tag{86}$$

Proposition 6.14 *The subscheme $C_n \subset V_n$ is the union of the subschemes (86) as ν runs through all choices of sign $\{\pm 1\}^n$ with $n'(\nu) > 0$:*

$$C_n = \bigcup_{\nu;\, n' = n'(\nu) > 0} \bar{\iota}_\nu^*(D_{n'}). \tag{87}$$

The curves $D_{n'}$ are completely known via their defining equations from Theorem 6.8, so the proposition above is a complete characterization of the curve C_n for $I(Z_n)$. In the language of the defining ideals, this proposition is a converse to Corollary 6.13. In other words, we have the following:

Corollary 6.15 *In the notation of Corollary 6.13, the defining ideal $\mathcal{J}_n$ for $I(Z_n)$ can be characterized as*

$$\mathcal{J}_n = \{ w \in \mathcal{A}_n \mid \bar{\iota}_\nu(w) \in \bar{\mathcal{J}}_{n'(\nu),-1} \text{ for all } \nu \}.$$

Thus $I(Z_n)$ is determined as an $\mathcal{A}_n$-module by the finger-move constraints, once $I(Z_{n',-1})$ is known for all odd $n' \le n$.

Proof of Proposition 6.14 Let us write C' for the union on the right-hand side of (87). The inclusion of ideals $\bar{\iota}_\nu(\mathcal{J}_n) \subset \bar{\mathcal{J}}_{n',-1}$ in Corollary 6.13 says that the curve C_n contains C'.

The coordinate ring of the scheme on the left-hand side of (87) is $I(Z_n)$, and if we temporarily write I' for the coordinate ring of the affine scheme C', then the inclusion of schemes means that we have a surjection of rings,

$$I(Z_n) \to I'.$$

We know that $I(Z_n)$ is a free $\mathcal{R}$-module of finite rank, where $\mathcal{R} = \mathbb{Q}[\tau, \tau^{-1}]$. So to prove that the rings are isomorphic, and to complete the proof of the proposition, it will suffice to prove that these two $\mathcal{R}$-modules have the same rank, or in geometrical language,

$$\deg C_n = \deg C',$$

where deg denotes the degree of the projection to the τ coordinate. (The inclusion one way means that we already have $\deg C_n \geq \deg C'$.)

To prove this last equality we note that

$$\deg C_n \leq \sum_{\nu;\, n'(\nu)>0} \deg(C_n \cap V_{n,\nu}), \tag{88}$$

with equality if and only if the schemes $C_n \cap V_{n,\nu}$ for different ν have no common component of positive degree. The two-way inclusions of Corollary 6.13 tell us that $C_n \cap V_{n,\nu}$ and $i_\nu^*(\bar{C}_{n'})$ coincide over the locus where $\tau^4 - 1$ is nonzero. In particular,

$$\deg(C_n \cap V_{n,\nu}) = \deg(i_\nu^*(D_{n'})),$$

and if the schemes on the left have no common component of positive degree for different ν, then the same is true of the schemes on the right. From (88) we therefore obtain

$$\deg C_n \leq \sum_{\nu;\, n'(\nu)>0} \deg D_{n'}, \tag{89}$$

with equality if and only if the schemes on the right-hand side of (87) have no common component of nonzero degree.

In terms of instanton homology, the inequality (88) can be restated as

$$\operatorname{rank}_{\mathcal{R}} I(Z_n) \leq \sum_{\nu;\, n'(\nu)>0} \operatorname{rank}_{\mathcal{R}} I(Z_{n',-1}). \tag{90}$$

On the other hand we can verify directly that we have equality here:

$$\operatorname{rank}_{\mathcal{R}} I(Z_n) = \sum_{\nu;\, n'(\nu)>0} \operatorname{rank}_{\mathcal{R}} I(Z_{n',-1}). \tag{91}$$

Indeed, the right-hand side can be calculated by Corollary 3.19, and is

$$\sum_{f=0}^{(n-1)/2} \binom{n}{f} \cdot \tfrac{1}{4}\big((n-2f)^2 - 1\big).$$

The left-hand side of (91) is twice the rank of the ordinary cohomology of the representation variety $\mathrm{Rep}(S^2_n)$ calculated by Boden [3], and can be expressed as

$$\mathrm{rank}_{\mathcal{R}}\, I(Z_n) = 2^{n-3}(n-1) = \tfrac{1}{8}(F''(1) - F(1)),$$

where $F(t) = (t + t^{-1})^n$. Equality with the right-hand side of (91) can be seen easily from the binomial expansion of $F(t)$.

It follows that the parts making up the union C' on the right-hand side of (87) have no common components of positive degree, and we therefore have, as required,

$$\deg C' = \sum_{\nu} \deg D_{n'(\nu)} = \deg C_n. \qquad \square$$

Remark In the course of the proof, we have seen that C_n has pure dimension 1, and we refer to it as the instanton curve for Z_n. Although it has no embedded points, we have not shown that the curve C_n is reduced: it may perhaps have components with multiplicity larger than 1, but the authors have not seen this arise in calculations.

6.6 Equations for the curve C_n

We now have a geometric description of $I(Z_n)$ as a module, namely as the coordinate ring of an affine curve C_n. The curve C_n is a union of curves each of which is isomorphic to some $D_{n'}$. However, although we have an explicit description of the defining relations for the $D_{n'}$, the resulting description of C_n does not immediately provide explicit generators for the corresponding ideal $\mathcal{J}_n \subset \mathcal{A}_n$. Instead, it describes the ideal $\mathcal{J}_n$ as an intersection of known ideals (expressed essentially in Corollary 6.15).

To practically compute the intersection of the ideals in this particular context, we can leverage what we know about $\mathcal{J}_n$. From Propositions 5.1 and 5.2, we know the ideal $\mathcal{J}_n$ is generated by elements W^m_η which can be written in the form

$$W^m_\eta = w(0) + \epsilon w(1) + w(2) + \epsilon w(3) + \cdots , \tag{92}$$

where $w(i)$ is a homogeneous polynomial of degree $m - i$ in $(\alpha, \delta_1, \dots, \delta_n)$, and furthermore

$$w(0) = w^m_{n,\eta} \quad \text{and} \quad w(1) = w^{m-1}_{n,n-\eta}.$$

Furthermore, the element W^m_η is the unique element of the ideal having leading term $w(0)$. The lower terms in W^m_η are therefore uniquely characterized by the linear constraints of Corollary 6.15, namely that $\bar{\iota}_\nu(W^m_\eta)$ belongs to the known ideal $\bar{\mathcal{J}}_{n'(\nu),-1}$, for all ν. Solving this large linear system provides the generators.

There is an alternative way to package the calculation of W^m_η, which does not explicitly pass through a determination of the ideals $\bar{\mathcal{J}}_{n,-1}$, albeit the same ingredients are used. To set this up, the terms in (92) which are as yet *unknown* are the terms which belong to a lower part of the increasing filtration of $\mathcal{A}_n$, and with this in mind we write

$$L^m_\eta = w(2) + \epsilon w(3) + \cdots ,$$

so that

$$W_h^m = w(0) + \epsilon w(1) + L_\eta^m, \quad \text{where } L_\eta^m \in \mathcal{A}_n^{(m-2)}. \tag{93}$$

There is some symmetry that can be usefully exploited. The braid group B_π for the n-element subset $\pi \subset S^2$ acts on $I(Z_n)$ because of its interpretation as a mapping class group. This action factors through the symmetric group S_π, as one can see from the description of $I(Z_n)$ as a cyclic module for the algebra $\mathcal{A}_n$. Indeed, given a permutation $\sigma \in S_\pi$, we obtain an automorphism $\sigma_* : \mathcal{A}_n \to \mathcal{A}_n$ permuting the generators δ_p and preserving the ideal $\mathcal{J}_n \subset \mathcal{A}_n$, so establishing the automorphism $\sigma_* : I(Z_n) \to I(Z_n)$. From this, we can see that

$$\sigma_*(W_\eta^m) = W_{\sigma(\eta)}^m.$$

In particular, the element $W_\eta^m \in \mathcal{A}_n$ is invariant under the action of group of permutations $S_\eta \times S_{\eta'} \subset S_\pi$. The lower terms L_η^m therefore have the same symmetry. Furthermore, it will be enough if we determine L_η^m for just one subset $\eta \subset \pi$ of each cardinality h satisfying the parity condition (35). Note also that the expression L_η^m is empty unless m is at least 2 (ie n is at least 5).

The proposed recursive procedure for identifying the lower terms L_η^m is to again use Corollary 6.11, which gives us the finger-move relation

$$i_*^{n,n-2}(W_\eta^m) \in \mathcal{J}_{n-2}. \tag{94}$$

We would like to see that, if the ideal $\mathcal{J}_{n-2}$ is already known, then the constraint (94) will be sufficient to determine the lower terms. In line with the remarks above, since either η or η' can be assumed to have at least $m+1$ elements (ie more than half), we will assume that the indices $\{m, m+1, \dots, n\}$ all belong either to η or to η'. In particular this means that W_η^m and its lower terms L_η^m are invariant under the symmetric group S_{m+1} acting by permutation of the variables $\{\delta_m, \delta_{m+1}, \dots, \delta_n\}$. (These indices include the three indices $\{n-2, n-1, n\}$, which are involved in the definition of the finger move $i^{n,n-2}$.)

Lemma 6.16 *Write $n = 2m+1$ and let $L \in \mathcal{A}_n^{(m-2)}$ be an element that is symmetric in the variables $\delta_{m+1}, \dots, \delta_{n-1}, \delta_n$ (ie more than half of the variables). Suppose L satisfies*

$$i_*^{n,n-2}(L) \in \mathcal{J}_{n-2}. \tag{95}$$

Then $L = 0$.

Proof Let σ_k be the k^{th} symmetric polynomial in $\delta_{m+1}, \dots, \delta_n$, and let σ'_k be the symmetric polynomial in $\delta_{m+1}, \dots, \delta_{n-2}$, regarded as elements of $\mathcal{A}_n$ and $\mathcal{A}_{n-2}$ respectively. From Proposition 3.7, we know that $\mathcal{J}_{n-2} \cap \mathcal{A}_{n-2}^{m-2} = 0$, so the hypothesis $i_*^{n,n-2}(L) \in \mathcal{J}_{n-2}$ actually means that $i_*^{n,n-2}(L)$ is zero. We compute what $i_*^{n,n-2}$ does to σ_k, and we find

$$i_*^{n,n-2}(\sigma_k) = \begin{cases} \sigma'_k & \text{if } k = 0, 1, \\ \sigma'_k + \beta\sigma'_{k-2} & \text{if } 2 \le k \le m-1, \\ \beta\sigma'_{k-2} & \text{if } k = m, m+1, \end{cases}$$

where $\beta = -\delta_p^2$ (independent of p).

Because L has degree at most $m-2$, we can write it as

$$L = \sum_{k=0}^{m-2} P_k \sigma_k,$$

where each P_k is an expression in $\mathcal{A}_m$, ie involving only $\delta_1, \dots, \delta_m$. Thus

$$i_*^{n,n-2}(L) = \sum_{k=0}^{m-2} (P_k + \beta P_{k+2}) \sigma'(k),$$

where we set $P_j = 0$ for $j > m-2$. The injectivity of $i_*^{n,n-2}$ is now clear from the upper triangular nature of this linear transformation, because the symmetric functions $\sigma'(k)$ are nonzero in this range. □

The lemma tells us that the finger-move constraint can be used to determine the lower terms L_η^m uniquely. So we obtain a procedure which determines the ideals $\mathcal{J}_n$ recursively for all odd n, as follows.

(i) In the base case $n = 1$, the ideal $\mathcal{J}_1$ is $\langle 1 \rangle$.

(ii) For general $n \geq 3$ (and n odd as always), assume that the ideal $\mathcal{J}_{n'}$ is already known for $n' < n$.

(iii) Write $m = \frac{1}{2}(n-1)$. According to Propositions 5.1 and 5.2, for each η satisfying the parity condition (47), there exists an element $W_\eta^m \in \mathcal{J}_n$ which can be written in the form (93):

$$W_h^m = w(0) + \epsilon w(1) + w(2) + \epsilon w(3) + \cdots = w(0) + \epsilon w(1) + L_\eta^m, \quad \text{where } L_\eta^m \in \mathcal{A}_n^{(m-2)}.$$

The first terms $w(0) + \epsilon w(1)$ are known because $w(0)$ is the Mumford relation and Proposition 5.2 provides the term $w(1)$.

(iv) According to Lemma 6.16, the unknown terms L_η^m in W_η^m are uniquely determined by the finger-move relations (94), which impose linear conditions on the coefficients of L_η^m. Solving these linear equations determines L_η^m and hence determines $W_\eta^m \in \mathcal{A}_n$.

(v) As η runs through the subsets satisfying (47), the elements W_η^m generate the ideal $\mathcal{J}_n \subset \mathcal{A}_n$ according to Proposition 5.1. So we have a known set of generators for $\mathcal{J}_n$. This determines $\mathcal{J}_n$ and completes the inductive step.

7 Further remarks

7.1 Singularities of the instanton curve

When the local coefficient system Γ is replaced by constant coefficients $\mathbb{Q}$, we obtain a description of the instanton homology $I(Z_n; \mathbb{Q})$ which was earlier completely determined by Street [32]. Those results therefore provide a description of the scheme-theoretic intersection of the curve C_n with the hyperplane $\tau = 1$. It is shown in [32] that the simultaneous eigenvalues of the pair of operators (α, δ) on $I(Z_n; \mathbb{Q})$ are of the form (λ, δ), where λ runs through the odd integers in the range $|\lambda| < n$. The multiplicities of the eigenspaces are also computed.

We can apply these results to learn that the curve D_n corresponding to $I(Z_{n,-1};\Gamma)$ intersects the plane $\tau = 1$ in the points x_λ with coordinates

$$(\tau, \alpha, \delta, \epsilon) = (1, \lambda, 0, \pm 1),$$

where λ runs through the same odd integers, and the sign of ϵ is $(-1)^{(\lambda+1)/2}$. We also learn that the intersection multiplicity at x_λ is $\mu_\lambda = \frac{1}{2}(n - |\lambda|)$.

Knowing the intersection multiplicity puts an upper bound on the order of a possible singular point of the curve at x_λ. In particular, it means that D_n is smooth at the points x_λ for the two extreme values of λ, namely $\lambda = \pm(n-2)$, because the intersection multiplicity is 1 at those points.

A little experimentation suggests that equality holds at all the points x_λ where D_n meets $\tau = 1$: that is,

$$\operatorname{ord}(D_n, x_\lambda) = \mu_\lambda = \tfrac{1}{2}(n - |\lambda|). \tag{96}$$

With the understanding that these results have been verified only experimentally for modest values of n, one can describe the singularity of D_n at x_λ in greater detail. First of all, we have seen that the ideal $\bar{\mathcal{J}}_{n,-1}$ which defines D_n has just two generators $G_1(n)$ and $G_2(n)$ (Proposition 6.9), and it follows that the singularity of D_n at x_λ is a local complete intersection. Indeed, each of D_n^+ and D_n^- is cut out as a global complete intersection inside the variety defined by $\epsilon = \pm 1$ and $\tau \neq 0$. Experiment also indicates that the surfaces defined by the vanishing of $G_1(n)$ and $G_2(n)$ are both smooth at x_λ. Indeed, the α-derivative of both is nonzero. By the implicit function theorem, the zero-sets of $G_1(n)$ and $G_2(n)$ are therefore described in a local analytic neighborhood of x_λ by

$$\alpha = \lambda + f_{n,\lambda,1}(\delta, \tau) \quad \text{and} \quad \alpha = \lambda + f_{n,\lambda,2}(\delta, \tau)$$

for two analytic functions $f_{n,\lambda,1}$ and $f_{n,\lambda,2}$. At the singular points (that is, when $|\lambda| < n-2$), the derivatives of both $f_{n,\lambda,1}$ and $f_{n,\lambda,2}$ vanish at $(\delta, \tau) = (0, 1)$. The singular germ (D_n, x_λ) is therefore analytically isomorphic to the germ of the analytic plane singularity

$$g_{n,\lambda}(\delta, \tau) = 0, \quad g_{n,\lambda} = f_{n,\lambda,1} - f_{n,\lambda,2},$$

at $(\delta, \tau) = (0, 1)$.

In computations up to $n = 31$, the function $g_{n,\lambda}$ vanishes to order μ_λ at $(0, 1)$, verifying that μ_λ is indeed the order of the singular point. Furthermore we find

$$g_{n,\lambda}(\delta, \tau) = \text{const.}\big(\delta \pm 2(\tau - 1)\big)^{\mu_\lambda} + O(\delta, \tau - 1)^{\mu_\lambda + 1},$$

where the sign depends on ϵ and λ. This means that the tangent cone to the singular point is the line $\delta \pm 2(\tau - 1) = 0$, with multiplicity μ_λ.

The highest-order singular points on the curve are the points x_λ with $\lambda = \pm 1$, where the order of the singularity is $m = \frac{1}{2}(n-1)$. At these points, the analytic form of the singularity is $x^m = y^{m+1}$, where $x = \delta \pm 2(\tau - 1)$. In particular the singularity is unibranch. The authors have not determined (even experimentally) whether the singularity is unibranch at other singular points. Note, however, that the entire curves $D_n^{\pm}$ are reducible when n is composite (as discussed below) and it follows that the singularities are not unibranch when λ and n have a common factor.

One further experimental observation is that the local form of the surface $G_i(n) = 0$, given by $\alpha = \lambda + f_{n,\lambda,i}(\delta, \tau)$ at x_λ, appears to approach a smooth limit as n increases with λ fixed. Indeed, after scaling by λ, we find that the limit is independent of λ also. That is, there is a convergent power series $F(\delta, \tau)$, independent of n, λ and $i = 1, 2$, such that

$$\lambda + f_{n,\lambda,i}(\delta, \tau) \to \lambda F(\delta, \tau).$$

The difference vanishes at $(0, 1)$ to order $(\delta, \tau - 1)^{O(n)}$. Up to terms of degree 5, the series F is

$$F(\delta, 1+\sigma) = 1 - \tfrac{1}{16}\delta^2 + \tfrac{31}{4}\delta\sigma + \tfrac{1}{4}\sigma^2 - \tfrac{31}{8}\delta\sigma^2 - \tfrac{1}{4}\sigma^3 - \tfrac{5}{1024}\delta^4 + \tfrac{31}{128}\delta^3\sigma + \tfrac{5}{128}\delta^2\sigma^2 + \tfrac{31}{32}\delta\sigma^3 \\ + \tfrac{15}{64}\sigma^4 - \tfrac{31}{256}\delta^3\sigma^2 - \tfrac{5}{128}\delta^2\sigma^3 + \tfrac{31}{64}\delta\sigma^4 - \tfrac{7}{32}\sigma^5 + \cdots .$$

7.2 Reducibility when n is composite

The curves D_n^+ and D_n^- arising as $\mathrm{Spec}(I(Z_{n,-1}))$ are irreducible when n is prime in all cases that the authors have calculated. It seems to be an interesting conjecture whether this holds in general. For composite n, however, the curves D_n^+ and D_n^- are reducible, as the following result implies.

Proposition 7.1 *If n' divides the odd integer n, then the curves D_n^+ and D_n^- contain $\psi(D_{n'}^+)$ and $\psi(D_{n'}^-)$ respectively, where ψ is the map on the ambient space $\overline{V}$ given by*

$$\psi(\tau, \tau^{-1}, \delta, \alpha, \epsilon) = (\tau, \tau^{-1}, \delta, (n/n')\alpha, \epsilon).$$

Proof This is an application of the general principal described by Proposition 6.10. In the context of that proposition, take W to be the product cobordism $[0, 1] \times Z_{n'}$. Write $l = n/n'$. We can embed a sphere $S \hookrightarrow W$ representing l times the generator of $H_2(W)$ and meeting the singular set in ln' points, all with the same orientation. The relevant map Ψ in Proposition 6.10 is then the homomorphism of algebras

$$\Psi_l \colon \mathcal{A}_n \to \mathcal{A}_{n'}$$

which is given (with our standardly named generators, and suitably numbering the intersection points) by

$$\Psi_l(\alpha) = l\alpha, \quad \Psi_l(\delta_k) = \delta_{(k \bmod n')}.$$

The conclusion of Proposition 6.10 is that we have an inclusion of ideals $\Psi_l(\mathcal{J}_n) \subset \mathcal{J}_{n'}$.

Passing to the quotient rings $\overline{\mathcal{A}}$ in which all the δ_k are equal, and using the fact that $\overline{\mathcal{J}}_{n,-1}$ is the image of $\mathcal{J}_n$ in the quotient ring (Proposition 5.13), we obtain an inclusion of ideals $\psi_l(\overline{\mathcal{J}}_{n,-1}) \subset \overline{\mathcal{J}}_{n',-1}$ when $n = ln'$, where ψ_l is algebra homomorphism the with $\psi_l(\alpha) = l\alpha$ and $\psi_l(\delta) = \delta$. Proposition 7.1 is just a restatement of this inclusion of ideals, in the geometrical language of the subschemes that they define. □

7.3 Interpretation as the quantum cohomology ring

For every odd n, the representation variety $M = \mathrm{Rep}(S^2_n)$ is naturally a smooth symplectic manifold, by a standard construction [11]. If n points in $\mathbb{CP}^1$ are chosen, then M becomes also a smooth complex-algebraic variety of dimension $n-3$, as a consequence of its interpretation as a moduli space of stable parabolic bundles. With the symplectic form, it is a Kähler manifold, and the cohomology class of the Kähler form is a negative multiple of the canonical class. The latter assertion is the statement of "monotonicity" for the symplectic structure. It can be deduced as a particularly simple case from [20], for example, or it can be deduced from the fact that there is only one class in H^2 which is invariant under the "flip" symmetries [32]. This is therefore a Fano variety. (A concrete description is discussed in [4].)

The quantum cohomology ring of such a Fano variety is defined using a deformation of the usual triple intersection product. Given cycles A, B, C, the quantum intersection product is a scalar which is a weighted count of isolated pseudoholomorphic curves $u\colon \mathbb{CP}^1 \to M$, with the constraint that u maps three marked points to A, B and C. For our purposes, the weight will be of the form $\tau^{[u]\cdot T}$ for a suitable 2-dimensional cohomology class $T = 2\sum \delta_i$. This leads to a quantum cohomology ring $\mathrm{QH}(M)$ which is a module over the ring of Laurent polynomials $\mathcal{R}$. In the spirit of results from [27] and [7], one should expect that the $\epsilon = 1$ component of $I(Z_n)$ is isomorphic to $\mathrm{QH}(M)$ as an algebra.

The special case $n = 5$ in particular is discussed in [31], where the symplectic manifold M is the blow-up of $\mathbb{CP}^2$ at five points, and the quantum cup-product is computed. Also relevant from [30; 31] is Seidel's long exact sequence [31, Proposition 3.5]. In the special case that M is $\mathbb{CP}^2 \mathbin{\#} 5\overline{\mathbb{CP}}^2$, this long exact sequence essentially recovers the skein exact sequence in the proof of Proposition 3.18, involving the orbifold $X_{5,4}$ from Figure 3, restricted to the $+1$ eigenspace of ϵ. The orbifold $X_{5,4}$ plays the role of $H_*(S^2;\Lambda)$ in [31, Proposition 3.5]. Seidel's exact sequence is generalized by Wehrheim and Woodward in [34, Theorem 6.12], motivated by the application to skein triangles, and the generalization is relevant to the case of the skein triangle involving $X_{n,n-1}$ for larger n.

7.4 General local coefficients

As an alternative to the local coefficient system Γ for $I(Z_n)$, there is a larger local coefficient system Γ_n that can be used. Rather than being a system of rank-1 modules over $\mathcal{R} = \mathbb{Q}[\tau^{-1}, \tau]$, the ground ring for Γ_n is the ring of finite Laurent series in n distinct variables $\tau_1, \dots, \tau_n$ attached to the n components of the singular set of Z_n:

$$\mathcal{R}_n = \mathbb{Q}[\tau_1, \tau_1^{-1}, \dots, \tau_n, \tau_n^{-1}].$$

The instanton homology $I(Z_n; \Gamma_n)$ is then a module over the ring

$$\mathcal{R}_n[\delta_1, \dots, \delta_n, \alpha, \epsilon].$$

It is no longer true that $\delta_i^2 = \delta_j^2$; instead we have

$$\delta_i^2 - \tau_i^2 - \tau_i^{-2} = \delta_j^2 - \tau_j^2 - \tau_j^{-2} \quad \text{for all } i, j.$$

It should be possible to compute $I(Z_n;\Gamma_n)$ by adapting the ideas of this paper. As the simplest example, our two generators for the relations in $I(Z_{3,-1})$, where all δ_i and all τ_i are equal, were

$$\alpha+\tfrac{3}{2}\delta+\epsilon\tau^3 \quad \text{and} \quad \alpha-\tfrac{1}{2}\delta+\epsilon\tau^{-1}.$$

For $I(Z_3;\Gamma_3)$ the corresponding relations are

$$\alpha+\tfrac{1}{2}(\delta_1+\delta_2+\delta_3)+\epsilon\tau_1\tau_2\tau_3 \quad \text{and} \quad \alpha+\tfrac{1}{2}(\delta_1-\delta_2-\delta_3)+\epsilon\tau_1\tau_2^{-1}\tau_3^{-1},$$

together with cyclic rotations of the second one. The instanton homology $I(Z_3;\Gamma_3)$ is a free $\mathcal{R}_3$-module of rank 2.

There is an additional symmetry present when using Γ_n, which comes from the flip relation. So the ideal of generators is invariant under the symmetry which changes the sign of δ_i and δ_j for any two distinct indices while changing τ_i and τ_j to τ_i^{-1} and τ_j^{-1}. In the example of $I(Z_3;\Gamma_3)$ there are four generators corresponding to the four subsets $\eta\subset\{1,2,3\}$ of even parity, and the corresponding relations are all obtained from the first one (corresponding to $\eta=\varnothing$) by applying flips. For larger n, the leading and subleading terms follow the same pattern. So the adaptation of Proposition 5.2 to the case of Γ_n has the same leading term while the factor of τ^{n-2h} in front of the subleading term is replaced by

$$\prod_{i\notin\eta}\tau_i\prod_{i\in\eta}\tau_i^{-1}.$$

7.5 Instanton homology for torus knots

As mentioned in the introduction, a motivation for this paper comes from wishing to calculate variants of framed instanton homology for torus knots. In [24], concordance invariants of knots were defined using a version of framed instanton homology $I^\sharp$. In that paper, for a knot $K\subset Y$, the framed instanton homology is defined using the connected sum $(Y,K)\,\#\,(S^3,\Theta)$, where Θ is a theta-graph in S^3. A local coefficient system is used in [24], where the ground ring is the Laurent polynomials in three variables τ_i corresponding to the three edges of Θ. Because of the phenomenon of bubbling in codimension 2 which arises from the vertices of Θ, it was necessary in [24] to use a ring of characteristic 2.

It is possible instead to work in characteristic zero by abandoning the pair (S^3,Θ) and using the pair Z_3 instead (as described just above). The local coefficient system comes from Γ_3. Because $I(Z_3;\Gamma_3)$ has rank 2, one should take just the $+1$ eigenspace of ϵ to obtain a rank-1 module. Thus one can define $I^\sharp(Z;\Gamma_3)$ for general bifolds Z as being $I(Z\,\#\,Z_3;\Gamma_3)_+$. The connected sum is of the 3-manifolds, not a connected sum of pairs. But a connected sum of pairs can be used instead to define a reduced version $I^\natural(Z;\Gamma_3)$.

A variant of the connected sum theorem from [5] allows one to pass to $I^\natural(Z_{n,-1};\Gamma_3)$ starting from the calculation of $I(Z_{n,-1})$ in this paper. Using the surgery exact triangle for instanton homology, one can therefore take the calculation of $I(Z_{n,-1})$ as a first step towards understanding the reduced instanton homology with local coefficients for torus knots in S^3. The authors hope to return to this in a future paper.

7.6 Universal relations

The relations in the instanton homology of Z_n and $Z_{n,-1}$ give rise to universal relations for general admissible bifolds (Y,K) containing spheres. The following is an illustration.

Proposition 7.2 *Let (Y,K) be a bifold and suppose that the singular set K is a knot meeting an embedded sphere $S \subset Y$ transversely with odd geometric intersection number n and algebraic intersection number n'. Orient the sphere and K so that $0 < n' \le n$. Let α be the operator on $I(Y,K)$ corresponding the sphere S and let δ be the operator arising from a point on K. Let ϵ^* be the involution on $I(Y,K)$ arising from S, and let $\epsilon = (-1)^{(n-n')/2}\epsilon^*$. Let $*$ denote the automorphism of the algebra $\bar{\mathcal{A}}_n$ determined by $*\colon \epsilon \mapsto \epsilon^*$. Then the elements of the ideal*

$$*(\tau^4-1)^{(n-n')/2}\bar{\mathcal{J}}_{n',-1} \subset \mathcal{R}[\delta,\alpha,\epsilon]/\langle \epsilon^2-1\rangle$$

annihilate $I(Y,K)$.

Proof Let $\delta_1,\dots,\delta_n$ be the operators corresponding the intersection points of K with S, all oriented with the normal orientation to S. From an application of the general principle of Proposition 6.10, the instanton homology $I(Y,K)$ is annihilated by the ideal $\mathcal{J}_n$ in the algebra $\mathcal{A}_n$. On the other hand, because K is a knot, all the operators δ_i are equal up to sign, so the action of the algebra $\mathcal{A}_n$ factors through the quotient $\bar{\mathcal{A}} = \mathcal{R}[\delta,\alpha,\epsilon]/\langle\epsilon^2-1\rangle$ in which we set $\delta_i = \pm\delta$ according to the sign of the corresponding intersection point of K with S. From Corollaries 6.11 and 6.13 the image of $\mathcal{J}_n$ in the quotient contains the ideal described in the proposition. □

As a simplest example, if K is a knot in $Y = S^1\times S^2$ which has geometric intersection 3 and algebraic intersection 1 with S^2, then $I(Y,K)$ is a torsion $\mathcal{R}$-module annihilated by τ^4-1. In general, the proposition provides a lower bound on the geometric intersection number of K and S^2.

Corollary 7.3 *Let Y contain an oriented 2-sphere S, and let $K \subset Y$ be a knot having odd algebraic intersection number $n' > 0$ with S. Then a lower bound for the transverse geometric intersection number $K\cap S$ for any knot isotopic to K is $n'+2f$, where*

$$f = \min\{F \ge 0 \mid (\tau^4-1)^F G_i(n') \text{ annihilates } I(Y,K) \text{ for } i=1,2\}.$$

Here $G_1(n')$ and $G_2(n')$ are the two generators in Proposition 6.9. □

In light of the results from [36] concerning higher-genus orbifolds, it is possible that the bound $n'+2f$ defined in the corollary is not particularly strong. It may be that $n'+2f$ is a lower bound for n_g+2g, where n_g is the geometric intersection number with a surface S_g of genus g homologous to S. It is easy to visualize examples where n_1+2 is much smaller than n_0, for example.

In the case that $n = n'$ in Proposition 7.2 (ie when algebraic and geometric intersection numbers are equal), the $\bar{\mathcal{A}}$-module $I(Y,K)$ is annihilated by the defining ideal of the curve D_n. This means that we can interpret $I(Y,K)$ as a coherent sheaf on D_n.

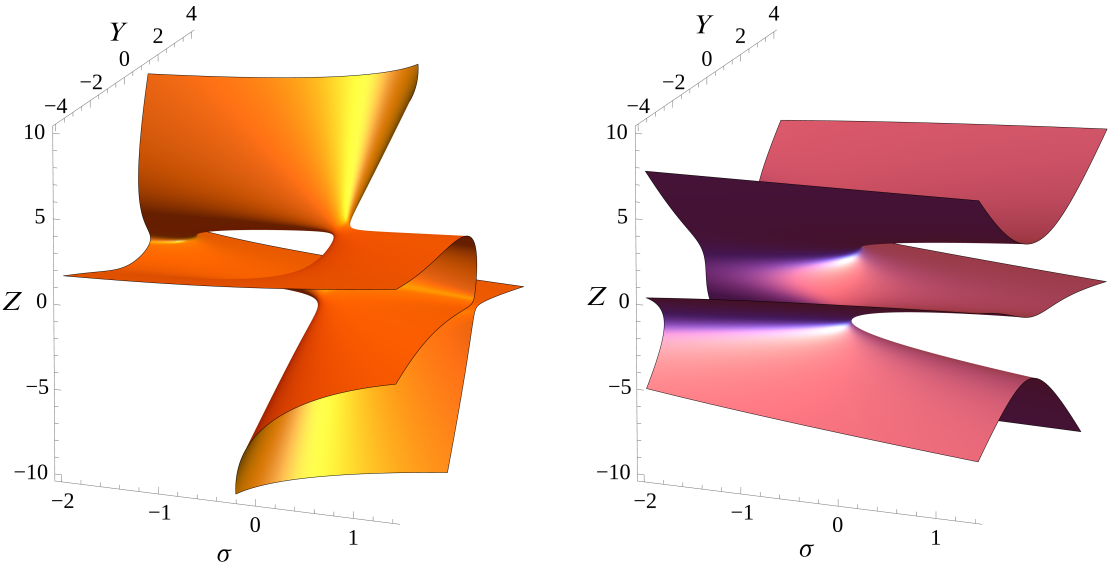


Figure 4: The real loci defined by the vanishing of the generators $G_1(n)$, left, and $G_2(n)$, right, for $n=7$ in the coordinates (σ, Y, Z). Only the part with $\epsilon = 1$ is shown. The part with $\epsilon = -1$ is obtained by changing the sign of Y and Z. These are smooth affine cubic surfaces.

7.7 The degrees of the relations

The two generators $G_1(n), G_2(n)$ for the ideal of relations for $I(Z_{n,-1})$ both have total degree $m = \frac{1}{2}(n-1)$ in (α, δ) but larger degree in τ. However, a substitution simplifies the polynomials a little: if we substitute

$$Z = \tau\alpha \quad \text{and} \quad Y = \tau\delta$$

then, after clearing unnecessary powers of τ from the denominator, we obtain a polynomial in Z, Y and τ^4. Writing $\sigma = \tau^4$, the total degree of the generators $G_i(n)$ in (σ, Z, Y) is m. The real loci defined by the vanishing of these two polynomials in (σ, Y, Z) are shown in Figure 4 for $n = 7$.

References

[1] **M F Atiyah**, **R Bott**, *The Yang–Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. A 308 (1983) 523–615 MR Zbl

[2] **I Biswas**, **N Raghavendra**, *Canonical generators of the cohomology of moduli of parabolic bundles on curves*, Math. Ann. 306 (1996) 1–14 MR Zbl

[3] **H U Boden**, *Representations of orbifold groups and parabolic bundles*, PhD thesis, Brandeis University (1991) Available at `https://www.proquest.com/docview/303936590`

[4] **C Casagrande**, *Rank 2 quasiparabolic vector bundles on $\mathbb{P}^1$ and the variety of linear subspaces contained in two odd-dimensional quadrics*, Math. Z. 280 (2015) 981–988 MR Zbl

[5] **A Daemi**, **C Scaduto**, *Equivariant aspects of singular instanton Floer homology*, preprint (2019) arXiv 1912.08982

[6] **S K Donaldson**, *The orientation of Yang–Mills moduli spaces and 4-manifold topology*, J. Differential Geom. 26 (1987) 397–428 MR Zbl

[7] **S Dostoglou**, **D A Salamon**, *Self-dual instantons and holomorphic curves*, Ann. of Math. 139 (1994) 581–640 MR Zbl

[8] **R Earl**, **F Kirwan**, *Complete sets of relations in the cohomology rings of moduli spaces of holomorphic bundles and parabolic bundles over a Riemann surface*, Proc. Lond. Math. Soc. 89 (2004) 570–622 MR Zbl

[9] **D Eisenbud**, *The geometry of syzygies: a second course in commutative algebra and algebraic geometry*, Grad. Texts in Math. 229, Springer (2005) MR Zbl

[10] **R Fintushel**, **R J Stern**, *Instanton homology of Seifert fibred homology three spheres*, Proc. Lond. Math. Soc. 61 (1990) 109–137 MR Zbl

[11] **W M Goldman**, *The symplectic nature of fundamental groups of surfaces*, Adv. Math. 54 (1984) 200–225 MR Zbl

[12] **Y-H Kiem**, *On the redundancy of the Mumford relations*, J. Algebraic Geom. 9 (2000) 395–402 MR Zbl

[13] **P A Kirk**, **E P Klassen**, *Representation spaces of Seifert fibered homology spheres*, Topology 30 (1991) 77–95 MR Zbl

[14] **F Kirwan**, *The cohomology rings of moduli spaces of bundles over Riemann surfaces*, J. Amer. Math. Soc. 5 (1992) 853–906 MR Zbl

[15] **U Koschorke**, *Infinite-dimensional K-theory and characteristic classes of Fredholm bundle maps*, from "Global analysis" (S-s Chern, S Smale, editors), Proc. Sympos. Pure Math. 15, Amer. Math. Soc., Providence, RI (1970) 95–133 MR Zbl

[16] **P B Kronheimer**, *An obstruction to removing intersection points in immersed surfaces*, Topology 36 (1997) 931–962 MR Zbl

[17] **P B Kronheimer**, **T S Mrowka**, *Gauge theory for embedded surfaces, I*, Topology 32 (1993) 773–826 MR Zbl

[18] **P B Kronheimer**, **T S Mrowka**, *Gauge theory for embedded surfaces, II*, Topology 34 (1995) 37–97 MR Zbl

[19] **P B Kronheimer**, **T S Mrowka**, *Khovanov homology is an unknot-detector*, Publ. Math. Inst. Hautes Études Sci. 113 (2011) 97–208 MR Zbl

[20] **P B Kronheimer**, **T S Mrowka**, *Knot homology groups from instantons*, J. Topol. 4 (2011) 835–918 MR Zbl

[21] **P B Kronheimer**, **T S Mrowka**, *A deformation of instanton homology for webs*, Geom. Topol. 23 (2019) 1491–1547 MR Zbl

[22] **P B Kronheimer**, **T S Mrowka**, *Tait colorings, and an instanton homology for webs and foams*, J. Eur. Math. Soc. 21 (2019) 55–119 MR Zbl

[23] **P B Kronheimer**, **T S Mrowka**, *Instantons and Bar-Natan homology*, Compos. Math. 157 (2021) 484–528 MR Zbl

[24] **P B Kronheimer**, **T S Mrowka**, *Instantons and some concordance invariants of knots*, J. Lond. Math. Soc. 104 (2021) 541–571 MR Zbl

[25] **V B Mehta**, **C S Seshadri**, *Moduli of vector bundles on curves with parabolic structures*, Math. Ann. 248 (1980) 205–239 MR Zbl

[26] **E Miller**, **B Sturmfels**, *Combinatorial commutative algebra*, Grad. Texts in Math. 227, Springer (2005) MR Zbl

[27] **V Muñoz**, *Quantum cohomology of the moduli space of stable bundles over a Riemann surface*, Duke Math. J. 98 (1999) 525–540 MR Zbl

[28] **V Muñoz**, *Ring structure of the Floer cohomology of* $\Sigma \times \mathbb{S}^1$, Topology 38 (1999) 517–528 MR Zbl

[29] **I R Porteous**, *Simple singularities of maps*, from "Proceedings of Liverpool Singularities: symposium I" (C T C Wall, editor), Lecture Notes in Math. 192, Springer (1971) 286–307 MR Zbl

[30] **P Seidel**, *A long exact sequence for symplectic Floer cohomology*, Topology 42 (2003) 1003–1063 MR Zbl

[31] **P Seidel**, *Lectures on four-dimensional Dehn twists*, from "Symplectic 4-manifolds and algebraic surfaces" (F Catanese, G Tian, editors), Lecture Notes in Math. 1938, Springer (2008) 231–267 MR Zbl

[32] **E J Street**, *Towards an instanton Floer homology for tangles*, PhD thesis, Harvard University (2012) Available at `https://www.proquest.com/docview/1027897859`

[33] **M Thaddeus**, *Conformal field theory and the cohomology of the moduli space of stable bundles*, J. Differential Geom. 35 (1992) 131–149 MR Zbl

[34] **K Wehrheim**, **C T Woodward**, *Exact triangle for fibered Dehn twists*, Res. Math. Sci. 3 (2016) art. id. 17 MR Zbl

[35] **J Weitsman**, *Geometry of the intersection ring of the moduli space of flat connections and the conjectures of Newstead and Witten*, Topology 37 (1998) 115–132 MR Zbl

[36] **Y Xie**, **B Zhang**, *Instanton Floer homology for sutured manifolds with tangles*, preprint (2019) arXiv 1907.00547

[37] **D Zagier**, *On the cohomology of moduli spaces of rank two vector bundles over curves*, from "The moduli space of curves" (R Dijkgraaf, C Faber, G van der Geer, editors), Progr. Math. 129, Birkhäuser, Boston, MA (1995) 533–563 MR Zbl

Department of Mathematics, Harvard University
Cambridge, MA, United States

Department of Mathematics, Massachusetts Institute of Technology
Cambridge, MA, United States

`kronheim@math.harvard.edu`, `mrowka@math.mit.edu`

Proposed: Simon Donaldson
Seconded: Martin R Bridson, Ian Agol

Received: 29 November 2022
Revised: 2 November 2023

msp

Geometry & Topology **29**:4 (2025) 2047–2103
DOI: 10.2140/gt.2025.29.2047
Published: 27 June 2025

Holonomic Poisson geometry of Hilbert schemes

Mykola Matviichuk
Brent Pym
Travis Schedler

We undertake a detailed study of the geometry of Bottacin's Poisson structures on Hilbert schemes of points in Poisson surfaces, ie smooth complex surfaces equipped with an effective anticanonical divisor. We focus on three themes that, while logically independent, are linked by the interplay between (characteristic) symplectic leaves and deformation theory. Firstly, we construct the symplectic groupoids of the Hilbert schemes and develop the classification of their symplectic leaves, using the methods of derived symplectic geometry. Secondly, we establish local normal forms for the Poisson brackets, and combine them with a toric degeneration argument to verify that Hilbert schemes satisfy our recent conjecture characterizing holonomic Poisson manifolds in terms of the geometry of the modular vector field. Finally, using constructible sheaf methods, we compute the space of first-order Poisson deformations when the anticanonical divisor is reduced and has only quasihomogeneous singularities. (The latter is automatic if the surface is projective.) Along the way, we find a tight connection between the Poisson geometry of the Hilbert schemes and the finite-dimensional Lie algebras of affine transformations, which is mediated by syzygies. In particular, we find that the Hilbert scheme has a natural subvariety that serves as a global counterpart of the nilpotent cone, and we prove that the Lie algebras of affine transformations have holonomic dual spaces — the first such series of Lie algebras to be discovered.

53D17; 14A30, 14B12, 14C05

1 Introduction

Let X be a Poisson surface, ie a smooth connected complex variety or analytic space of dimension two, equipped with a holomorphic Poisson bivector (a section of the anticanonical line bundle). In [5], Bottacin constructed a natural Poisson structure on the Hilbert scheme $X^{[n]}$ parametrizing zero-dimensional subschemes of X, generalizing the celebrated symplectic structure on the Hilbert schemes of K3 and abelian surfaces due to Beauville [2, Section 6] and Mukai [28]. Concretely, this Poisson manifold is obtained by resolving the singularities of the symmetric power $X^{(n)} = X^n/S_n$, but it also has a natural interpretation as a moduli space. Our goal in this paper is to develop some of the beautiful geometry of Bottacin's Poisson structures, focusing on the following three themes:

(1) symplectic leaves and symplectic groupoids,

(2) local normal forms and holonomicity,

(3) deformation theory.

In so doing, we encounter connections with a range of topics, from combinatorial linear algebra and Lie theory, to D-modules and derived algebraic geometry.

As we shall see, the three themes listed above are closely intertwined on a conceptual level. However, the results we establish for each of them (summarized in Sections 1.1 through 1.3 below, respectively) are based on different techniques and are essentially logically independent. Thus, for instance, while we make use of the formalism of derived geometry and shifted symplectic structures for the first theme, and this informs our intuition throughout the paper, the results in the second theme are proven using classical techniques of Poisson geometry and toric degeneration, and those in the third theme use constructible sheaves.

1.1 Symplectic leaves and the Hilbert groupoid

Bottacin's description of $X^{[n]}$ as a moduli space gives rise to an explicit description of the Poisson bivector as a pairing on Ext groups, which when combined with a deformation-theoretic argument, leads to a modular description of its symplectic leaves. Namely, the vanishing locus of the Poisson structure on X is a curve $D \subset X$, and two elements $Z_1, Z_2 \in X^{[n]}$ lie on the same symplectic leaf if and only if the restrictions of their ideal sheaves to D are isomorphic as $\mathcal{O}_D$-modules; see eg Pym and Rains [36] and Rains [38; 39], where such results are established in the more general context of moduli spaces of perfect complexes on noncommutative surfaces, although for Hilbert schemes we obtain this by different means below, as a direct consequence of Theorem 1.1.

It is natural to ask for a geometric (rather than module-theoretic) interpretation of the leaves. For instance, one can check that the map sending an element $Z \in X^{[n]}$ to the intersection $Z \cap D$ is invariant under Hamiltonian flow, and hence if two elements Z_1, Z_2 lie on the same symplectic leaf, we must have $Z_1 \cap D = Z_2 \cap D$. Over the open set of $X^{[n]}$ corresponding to collections of n distinct points of X, this characterizes the symplectic leaves completely, but the situation in general is not so simple. Indeed, even for a smooth curve D, there exist pairs $Z_1, Z_2 \in X^{[n]}$ that have the same scheme-theoretic intersection with D, but lie on different symplectic leaves.

The reason for this subtlety is that the intersection $\mathsf{Z} \cap \mathsf{D}$, if nonempty, is never transverse. Hence, it is natural to consider the derived intersection $\mathsf{Z} \cap^h \mathsf{D}$, which enhances the naive algebra of functions $\mathcal{O}_{\mathsf{Z}\cap\mathsf{D}} \cong \mathcal{O}_{\mathsf{Z}} \otimes_{\mathcal{O}_{\mathsf{X}}} \mathcal{O}_{\mathsf{D}}$ to the dg algebra $\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}} := \mathcal{O}_{\mathsf{Z}} \otimes^L_{\mathcal{O}_{\mathsf{X}}} \mathcal{O}_{\mathsf{D}}$, whose cohomology gives the Tor modules $\mathcal{T}or^{\mathcal{O}_{\mathsf{X}}}_{\bullet}(\mathcal{O}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}})$. An important feature is that the derived intersection $\mathsf{Z} \cap^h \mathsf{D}$ can have nontrivial automorphisms that preserve the embedding $\mathsf{Z} \cap^h \mathsf{D} \hookrightarrow \mathsf{D}$, so that the notion of the equality $\mathsf{Z}_1 \cap \mathsf{D} = \mathsf{Z}_2 \cap \mathsf{D}$ of ordinary scheme-theoretic intersections is enhanced to a whole space of equivalences $\mathsf{Z}_1 \cap^h \mathsf{D} \sim \mathsf{Z}_2 \cap^h \mathsf{D}$ of derived subschemes of D, which we call D-*equivalences* of the pair $(\mathsf{Z}_1, \mathsf{Z}_2)$; these turn out to be the same as determinant-one $\mathcal{O}_{\mathsf{D}}$-module isomorphisms between the restrictions of their ideal sheaves to D. These equivalences can be composed, giving a groupoid $(\mathsf{X}, \mathsf{D})^{[n]}$ whose objects are elements of $\mathsf{X}^{[n]}$ and whose morphisms are homotopy classes of D-equivalences. We call this groupoid the *Hilbert groupoid of the pair* (X, D), and develop its structure in Sections 2 and 3 below. We summarize these results as follows:

Theorem 1.1 *Let* X *be a smooth complex surface*, *and let* $\mathsf{D} \subset \mathsf{X}$ *be any curve* (*which may be singular or even nonreduced*). *Then* $(\mathsf{X}, \mathsf{D})^{[n]}$ *naturally has the structure of a smooth groupoid* (*ie a complex Lie groupoid*), *which is birationally equivalent to the product* $\mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$. *If*, *in addition*, *the curve* D *is anticanonical*, *then a choice of Poisson structure on* X *whose vanishing locus is* D *endows* $(\mathsf{X}, \mathsf{D})^{[n]}$ *with a natural symplectic structure*, *making it into a symplectic groupoid that integrates Bottacin's Poisson structure on* $\mathsf{X}^{[n]}$. *In particular*, *the symplectic leaves are the connected components of the* D*-equivalence classes*, *and these are locally closed subvarieties of* $\mathsf{X}^{[n]}$.

We remark that although the output of the construction is an ordinary smooth variety, the proof we present takes full advantage of the derived geometry toolbox, particularly foundational results on moduli stacks of sheaves and shifted symplectic structures from Brav and Dyckerhoff [6], Calaque [9], Pantev, Toën, Vaquié and Vezzosi [34], Toën and Vaquié [46] and Schürg, Toën and Vezzosi [43]. Such an approach naturally (and rather easily) constructs $(\mathsf{X}, \mathsf{D})^{[n]}$ as a derived stack; our main new addition is a representability statement, which shows that this derived stack is in fact a smooth variety. Combining this with normal forms for sheaves on smooth and nodal curves, we obtain a classification of symplectic leaves in the case that D has only nodal singularities. For smooth curves, this problem was also treated in [38, Section 11.2]; we expand on the results in [38], emphasizing connections with the combinatorics of Young diagrams and monomial ideals.

1.2 Local models and holonomicity

In Section 4 we turn to our second theme, namely the construction of explicit local models for the Poisson manifold $\mathsf{X}^{[n]}$ and its symplectic groupoid.

The guiding principle here is that the derived intersection $\mathsf{Z} \cap^h \mathsf{D}$ can be described in purely classical terms using the Hilbert–Burch resolution of the structure sheaf $\mathcal{O}_{\mathsf{Z}}$, which expresses the ideal defining Z locally as the maximal minors of a $(k+1) \times k$ matrix with entries in $\mathcal{O}_{\mathsf{X}}$ (the "syzygy matrix"). As explained by Ellingsrud and Strømme [12] one can construct natural coordinates on $\mathsf{X}^{[n]}$ by choosing suitable normal forms for the syzygy matrices. We use these coordinates to establish a tight connection between the local structure of the Poisson manifolds $\mathsf{X}^{[n]}$ and the Lie algebras $\mathfrak{aff}_k(\mathbb{C}) \cong \mathfrak{gl}_k(\mathbb{C}) \ltimes \mathbb{C}^k$ of affine transformations of $\mathbb{C}^k$, for $k \geq 1$. Indeed, by passing to open subsets and taking products, one can reduce the problem of understanding the local structure of the Hilbert schemes to the case of open balls around the elements $\mathsf{Z} = p(k) \in \mathsf{X}^{[n]}$ that correspond to the k^{th} order neighbourhoods of points $p \in \mathsf{X}$, where $k \geq 1$, for which we establish the following description.

Theorem 1.2 *A choice of log Darboux coordinates on* X *at* p *induces an analytic Poisson isomorphism from a neighbourhood of* $p(k)$ *in* $\mathsf{X}^{[k(k+1)/2]}$ *to a neighbourhood of the origin in* $\mathfrak{aff}_k(\mathbb{C})^\vee$, *where the latter is equipped with the following Poisson structure*:

- *a constant symplectic structure with explicit Darboux coordinates*, *if* $p \in \mathsf{X} \setminus \mathsf{D}$,
- *the Lie–Poisson structure*, *if* p *is a smooth point of* D, *or*
- *a Lie-compatible quadratic Poisson structure*, *if* p *is a node of* D.

As a consequence, when D is smooth, the symplectic groupoid $\mathfrak{aff}_k(\mathbb{C})^\vee \rtimes \mathsf{Aff}_k(\mathbb{C})$ of $\mathfrak{aff}_k(\mathbb{C})^\vee$ locally integrates the Poisson structure; in fact, we show that this a local model for the groupoid $(\mathsf{X}, \mathsf{D})^{[n]}$ constructed in Theorem 1.1, when $n = \frac{1}{2}k(k+1)$. Another interesting aspect of this case is that we can identify a natural subvariety in $\mathsf{X}^{[k(k+1)/2]}$ that globalizes the embedding $\mathfrak{gl}_k(\mathbb{C})^\vee \subseteq \mathfrak{aff}_k(\mathbb{C})^\vee$, namely the locus W of schemes whose intersection with D has length k. The resulting map $\mathsf{W} \to \mathrm{Sym}^k \mathsf{D}$ locally encodes the universal Poisson deformation of the nilpotent cone, giving a link to Springer theory — in particular, the fibre over $k \cdot p$ is a global version of the nilpotent cone; see Remark 4.9.

Of the three cases in Theorem 1.2, the third one is the most subtle. The quadratic Poisson structure in question seems to be new, and has the remarkable property that it admits a canonical isotrivial toric degeneration. The construction of this degeneration involves a careful analysis of the combinatorics of the weights of a natural torus action on $\mathfrak{aff}_k(\mathbb{C})$, which we recast visually as a sort of "game of dominoes"; see Section 4.6.

As an application of our local models, we show that the Hilbert schemes give new examples of Poisson manifolds satisfying a natural and subtle nondegeneracy condition, called holonomicity. (This was the problem that motivated the broader investigation we present in this paper.) The notion of holonomicity was introduced by the second- and third-named authors in [37], motivated by deformation theory; it is roughly equivalent to the statement that for each point in the Poisson manifold, the space of derived Poisson deformations of its germ is finite-dimensional. This notion was further developed and applied in our joint work [27], where we emphasized the importance of a certain collection of symplectic leaves, which we call "characteristic": these are the symplectic leaves that are preserved by the flow of the modular vector field. In particular, while the original definition of holonomicity involves $\mathcal{D}$-modules, we conjectured that it is equivalent to a simple geometric criterion:

Conjecture 1.3 [27] *A Poisson manifold is holonomic if and only if its germ at any point has only finitely many characteristic symplectic leaves.*

In the present work, we show that the symplectic leaf through a point $\mathsf{Z} \in \mathsf{X}^{[n]}$ is characteristic if and only if the corresponding derived intersection $\mathsf{Z} \cap^h \mathsf{D}$ is preserved by the flow of the modular vector field on D. Once again, this can be turned into a concrete local computation in examples without relying on derived techniques. Combining such concrete calculations with the results in [27; 37], the aforementioned toric degenerations, and some subtle combinatorial linear algebra, we establish our conjecture in (almost) complete generality for Hilbert schemes:

Theorem 1.4 *For a Poisson surface* X *whose vanishing locus is the anticanonical divisor* $\mathsf{D} \subset \mathsf{X}$, *the following statements are equivalent*:

1. *For every* $n \geq 0$, *the induced Poisson structure on the Hilbert scheme* $\mathsf{X}^{[n]}$ *is holonomic.*
2. *For every* $n \geq 0$, *the germ of* $\mathsf{X}^{[n]}$ *at any point has only finitely many characteristic symplectic leaves.*
3. *The only singularities of* D *are nodes.*

Note that the order of the quantifiers in statements (1) and (2) is important: for $n < 6$ there exist some exceptions for which $\mathsf{X}^{[n]}$ has finitely many characteristic leaves, but D has nonnodal singularities. However, we can treat most of these cases directly, leaving only one case left to determine whether Conjecture 1.3 holds for all Hilbert schemes, namely the case when $n = 5$ and D has only A_2-singularities (cusps).

On the other hand, combining Theorem 1.4 with Theorem 1.2, we obtain the first (and so far only known) infinite series of holonomic Lie algebras:

Corollary 1.5 *The Lie–Poisson structure on* $\mathfrak{aff}_k(\mathbb{C})^\vee$ *is holonomic for any* $k \geq 1$, *and in particular*, *its Poisson cohomology is finite-dimensional.*

It would be interesting to know if there are other series of Lie algebras with holonomic duals. Note that such Lie algebras must have an open coadjoint orbit, ie they are Frobenius. However, many of the Frobenius Lie algebras known to us are not holonomic.

1.3 Deformation theory

We close the paper in Section 5 by discussing the Poisson deformations of $\mathsf{X}^{[n]}$. Such deformations are governed by the Poisson cohomology $\mathsf{H}^\bullet_{\sigma^{[n]}}(\mathsf{X}^{[n]})$, which in turn is controlled by the geometry of the open symplectic leaf and the codimension-two characteristic leaves. We compute the Poisson cohomology in low degrees under the additional assumption that D has only quasihomogeneous singularities (which includes all cases where X is projective and D is reduced). This extends earlier work of Ran [40], who treated the case where D is smooth. In particular, we determine space of first-order deformations (the second Poisson cohomology):

Theorem 1.6 *Assume that* X *is connected, and that the anticanonical divisor* D *is reduced and has only quasihomogeneous singularities. Then the second Poisson cohomology of* $\mathsf{X}^{[n]}$*, is given, for all* $n \geq 2$*, by*

$$\mathsf{H}^2_{\sigma^{[n]}}(\mathsf{X}^{[n]}) \cong \mathsf{H}^2((\mathsf{X} \setminus \mathsf{D})^{[n]}; \mathbb{C}) \oplus \mathsf{H}^0(\mathcal{K}_\mathsf{X}^\vee|_{\mathsf{D}_{\mathrm{sing}}}) \cong \mathsf{H}^2_\sigma(\mathsf{X}) \oplus \wedge^2 \mathsf{H}^1(\mathsf{X} \setminus \mathsf{D}; \mathbb{C}) \oplus \mathbb{C} \cdot [\mathsf{E}],$$

where $\mathcal{K}_\mathsf{X}^\vee$ *is the anticanonical bundle and* $[\mathsf{E}]$ *is the Poisson Chern class of the exceptional divisor of the Hilbert–Chow morphism.*

This decomposition of the deformation space has the following interpretation. The summand $\mathsf{H}^2_\sigma(\mathsf{X})$ corresponds to deformations of $\mathsf{X}^{[n]}$ induced by applying Bottacin's construction to a deformation of X itself. Meanwhile the summand $\wedge^2 \mathsf{H}^1(\mathsf{X} \setminus \mathsf{D}; \mathbb{C})$ gives deformations in which the symplectic form on the open symplectic leaf $(\mathsf{X} \setminus \mathsf{D})^{[n]}$ is deformed by the pullback of a closed two-form along the Albanese map $(\mathsf{X} \setminus \mathsf{D})^{[n]} \to \mathsf{H}^1(\mathsf{X} \setminus \mathsf{D}; \mathbb{C})^\vee / \mathsf{H}_1(\mathsf{X} \setminus \mathsf{D}; \mathbb{Z})$. Finally, the deformations in the direction $\mathbb{C} \cdot [\mathsf{E}]$ should correspond to Hilbert schemes of *noncommutative* surfaces given by deformation quantization of (X, σ) as in [38; 39]. In the case $\mathsf{X} = \mathbb{P}^2$ and D is an elliptic curve, this interpretation was established by Hitchin [21], who proved that the resulting deformations are exactly the Hilbert schemes of Sklyanin algebras introduced by Nevins and Stafford [32].

Acknowledgements We thank Marco Gualtieri, Jacques Hurtubise, Anton Izosimov, Eric Rains and Pavel Safronov for helpful conversations. Pym is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada; a Research Support for New Academics grant from the Fonds de recherche du Québec (Nature et technologies); and a Faculty Startup Grant from McGill University. Matviichuk is supported by a Chapman Fellowship, and thanks Jacques Hurtubise, Niky Kamran and CIRGET for partially funding his postdoctoral position at McGill, during which most of his contribution to this paper took place. Schedler thanks the Max Planck Institute for Mathematics in Bonn and the Hausdorff Institute of Mathematics in Bonn for excellent working conditions during preliminary stages of work on this project. The authors thank the reviewer for their helpful comments, which helped improve the exposition of certain parts of the manuscript.

2 Hilbert groupoids of curves in surfaces

2.1 Hilbert schemes of surfaces

Throughout this paper, by a $\mathbb{C}$*-scheme* we mean a separated scheme of finite type over $\mathbb{C}$ or a complex analytic space. By a *smooth (complex) surface*, we mean a smooth $\mathbb{C}$-scheme of dimension two.

Let X be a smooth complex surface and suppose that $n \in \mathbb{Z}_{>0}$ is a positive integer. We denote by $\mathsf{X}^{[n]}$ the Hilbert scheme (or Douady space) parametrizing zero-dimensional subschemes of X of total length n. There are many equivalent presentations of this space, each of which will be useful in what follows:

- As a functor of points: if S is a $\mathbb{C}$-scheme, the set of maps from S to $\mathsf{X}^{[n]}$ is given by

  $$\mathsf{X}^{[n]}(\mathsf{S}) := \{\text{subschemes of } \mathsf{S}\times\mathsf{X} \text{ that are finite and flat of length } n \text{ over } \mathsf{S}\}.$$

  It is a theorem of Fogarty [13] that this functor is representable by a smooth scheme.
- As a resolution of singularities: let $\mathsf{X}^{(n)}$ be the n^{th} symmetric power of X, ie the quotient of X^n by the permutations of the factors; it is a singular variety when $n \geq 2$. There is a natural map $\mathsf{X}^{[n]} \to \mathsf{X}^{(n)}$, called the Hilbert–Chow morphism, which sends a length-n subscheme $\mathsf{Z} \subset \mathsf{X}$ to its support (the collection of points in Z, counted with multiplicity). This map is a resolution of singularities that is crepant [2, proof of Proposition 5] and strictly semismall [22, Theorem 2.12].
- As a moduli space of sheaves: a point $\mathsf{Z} \in \mathsf{X}^{[n]}$ is equivalent to the data of its ideal sheaf $\mathcal{I}_\mathsf{Z} \subset \mathcal{O}_\mathsf{X}$. We may view $\mathcal{I}_\mathsf{Z}$ as a perfect complex of coherent sheaves on X. This complex is canonically isomorphic to $\mathcal{O}_\mathsf{X}$ away from Z, and this induces a canonical trivialization of the determinant $\det \mathcal{I}_\mathsf{Z}$ away from Z, which extends to all of X since X is smooth and Z has codimension two. In this way, we identify $\mathsf{X}^{[n]}$ with an open substack of the derived stack $\mathsf{Perf}_0(\mathsf{X})$ of perfect complexes with trivialized determinant [43]. Note that the tangent complex of $\mathsf{X}^{[n]}$ at Z is then given by

  $$\mathbb{T}_\mathsf{Z}\mathsf{X}^{[n]} \cong \mathbb{T}_{\mathcal{I}_\mathsf{Z}}\mathsf{Perf}_0(\mathsf{X}) \cong \mathsf{R}\Gamma(\mathcal{R}\mathcal{E}nd(\mathcal{I}_\mathsf{Z})_0)[1],$$

  where $\mathcal{R}\mathcal{E}nd(\mathcal{I}_\mathsf{Z})_0$ is the homotopy fibre of the Illusie trace map $\mathcal{R}\mathcal{E}nd(\mathcal{I}_\mathsf{Z}) \to \mathcal{O}_\mathsf{X}$. The cohomology of this complex is concentrated in degree zero, giving the tangent space of $\mathsf{X}^{[n]}$ at Z as

(2-1) $$\mathsf{T}_\mathsf{Z}\mathsf{X}^{[n]} \cong \mathsf{H}^1(\mathsf{R}\Gamma(\mathcal{R}\mathcal{E}nd(\mathcal{I}_\mathsf{Z})_0)) =: \mathsf{Ext}^1_\mathsf{X}(\mathcal{I}_\mathsf{Z}, \mathcal{I}_\mathsf{Z})_0 \cong \mathsf{Hom}_\mathsf{X}(\mathcal{I}_\mathsf{Z}, \mathcal{O}_\mathsf{Z}).$$

2.2 Derived intersection of subschemes with curves

If X is a smooth surface, a *curve in* X is an effective divisor $\mathsf{D} \subset \mathsf{X}$; unless otherwise specified, we allow the possibility that D is singular, or even nonreduced.

Suppose that $i : \mathsf{D} \hookrightarrow \mathsf{X}$ is a curve. If $\mathsf{Z} \subset \mathsf{X}$ is a length-n subscheme, then the scheme-theoretic intersection $\mathsf{Z} \cap \mathsf{D} \subset \mathsf{D}$ is a closed subscheme of D of length at most n, with structure sheaf $\mathcal{O}_{\mathsf{Z}\cap\mathsf{D}} \cong \mathcal{O}_\mathsf{Z} \otimes_{\mathcal{O}_\mathsf{X}} \mathcal{O}_\mathsf{D}$. The *derived intersection* $\mathsf{Z} \cap^h \mathsf{D}$, is the derived subscheme of D obtained as the spectrum of the sheaf of commutative dg $\mathcal{O}_\mathsf{D}$-algebras $\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}} := \mathcal{O}_\mathsf{Z} \otimes^L_{\mathcal{O}_\mathsf{X}} \mathcal{O}_\mathsf{D}$, where $\otimes^L$ denotes the derived tensor product. In other words, we consider the dg algebra $Li^*\mathcal{O}_\mathsf{Z}$, where

$$Li^* : \mathsf{Perf}(\mathsf{X}) \to \mathsf{Perf}(\mathsf{D}), \quad Li^*\mathcal{E} := i^{-1}\mathcal{E} \overset{L}{\underset{i^{-1}\mathcal{O}_\mathsf{X}}{\otimes}} \mathcal{O}_\mathsf{D},$$

is the derived restriction of perfect complexes. Resolving $\mathcal{O}_\mathsf{D}$ by the Koszul complex

$$\mathcal{O}_\mathsf{D} \cong (\mathcal{O}_\mathsf{X}(-\mathsf{D}) \to \mathcal{O}_\mathsf{X})$$

we have concretely that

(2-2) $$\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}} \cong (\mathcal{O}_\mathsf{X}(-\mathsf{D})|_\mathsf{Z} \to \mathcal{O}_\mathsf{Z}),$$

with $\mathcal{O}_\mathsf{Z}$ placed in cohomological degree zero.

Remark 2.1 The number of points of intersection, counted with multiplicity, is equal to

$$\#(\mathsf{Z} \cap \mathsf{D}) := \dim_{\mathbb{C}} \mathsf{H}^0(\mathcal{O}_{\mathsf{Z}\cap\mathsf{D}}) \tag{2-3}$$

by definition. For the derived intersection, it is natural to consider the virtual count of points, defined by the Euler characteristic of $\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}}$, rather than its zeroth cohomology. Using the Koszul resolution (2-2), we find

$$\#_{\mathrm{vir}}(\mathsf{Z} \overset{h}{\cap} \mathsf{D}) := \chi(\mathsf{H}^\bullet(\mathcal{O}_{\mathsf{Z}\cap^h D})) = \dim_{\mathbb{C}} \mathsf{H}^0(\mathcal{O}_{\mathsf{Z}}) - \dim_{\mathbb{C}} \mathsf{H}^0(\mathcal{O}_{\mathsf{X}}(-\mathsf{D})|_{\mathsf{Z}}) = 0$$

for all Z, since Z is zero-dimensional and $\mathcal{O}_{\mathsf{X}}(-\mathsf{D})|_{\mathsf{Z}}$ is an invertible $\mathcal{O}_{\mathsf{Z}}$-module, in accordance with the general principle that derived intersection numbers are deformation-invariant. ◊

We now determine when two elements $\mathsf{Z}_1, \mathsf{Z}_2 \in \mathsf{X}^{[n]}$ have equivalent derived intersection with D. Note that since the ideal sheaf $\mathcal{I}_{\mathsf{Z}}$ is torsion-free, we have $\mathcal{T}or_j(\mathcal{I}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}}) = 0$ for $j > 0$, so the natural map $Li^*\mathcal{I}_{\mathsf{Z}} \to i^*\mathcal{I}_{\mathsf{Z}}$ comparing the derived and ordinary pullbacks is a quasi-isomorphism. Hence we obtain an exact triangle

$$i^*\mathcal{I}_{\mathsf{Z}} \to \mathcal{O}_{\mathsf{D}} \to Li^*\mathcal{O}_{\mathsf{Z}} \tag{2-4}$$

of perfect complexes on D, corresponding to the exact sequence

$$0 \to \mathcal{T}or_1^{\mathcal{O}_{\mathsf{X}}}(\mathcal{O}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}}) \to i^*\mathcal{I}_{\mathsf{Z}} \to \mathcal{O}_{\mathsf{D}} \to \mathcal{O}_{\mathsf{Z}\cap\mathsf{D}} \to 0 \tag{2-5}$$

of coherent sheaves. Note that $\mathcal{T}or_1(\mathcal{O}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}})$ is a torsion module on D. Since $\mathcal{O}_{\mathsf{D}}$ is torsion-free (by definition), the image

$$i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}} := \mathrm{image}(i^*\mathcal{I}_{\mathsf{Z}} \to \mathcal{O}_{\mathsf{D}}) \cong i^*\mathcal{I}_{\mathsf{Z}}/\mathcal{T}or_1^{\mathcal{O}_{\mathsf{X}}}(\mathcal{O}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}})$$

is the maximal torsion-free quotient of $i^*\mathcal{I}_{\mathsf{Z}}$. We then have the following.

Lemma 2.2 *For any* $(\mathsf{Z}_1, \mathsf{Z}_2) \in \mathsf{X}^{[n]}$, *the following spaces are canonically homotopy equivalent*:

1. *The space of equivalences* $\mathsf{Z}_1 \cap^h \mathsf{D} \xrightarrow{\sim} \mathsf{Z}_2 \cap^h \mathsf{D}$ *of derived subschemes of* D.
2. *The space of quasi-isomorphisms* $Li^*\mathcal{O}_{\mathsf{Z}_1} \xrightarrow{\sim} Li^*\mathcal{O}_{\mathsf{Z}_2}$ *of commutative differential graded* $\mathcal{O}_{\mathsf{D}}$*-algebras.*
3. *The space of quasi-isomorphisms* $Li^*\mathcal{O}_{\mathsf{Z}_1} \xrightarrow{\sim} Li^*\mathcal{O}_{\mathsf{Z}_2}$ *of perfect complexes of* $\mathcal{O}_{\mathsf{D}}$*-modules, commuting up to coherent homotopy with the natural maps* $Li^*\mathcal{O}_{\mathsf{Z}_1} \leftarrow \mathcal{O}_{\mathsf{D}} \to Li^*\mathcal{O}_{\mathsf{Z}_2}$.
4. *The space of equivalences* $i^*\mathcal{I}_{\mathsf{Z}_1} \xrightarrow{\sim} i^*\mathcal{I}_{\mathsf{Z}_2}$ *in* $\mathsf{Perf}_0(\mathsf{D})$.
5. *The discrete set of isomorphisms* $i^*\mathcal{I}_{\mathsf{Z}_1} \xrightarrow{\sim} i^*\mathcal{I}_{\mathsf{Z}_2}$ *commuting with the natural maps* $i^*\mathcal{I}_{\mathsf{Z}_1} \to \mathcal{O}_{\mathsf{D}} \leftarrow i^*\mathcal{I}_{\mathsf{Z}_2}$.
6. *The discrete set of isomorphisms* $i^*\mathcal{I}_{\mathsf{Z}_1} \xrightarrow{\sim} i^*\mathcal{I}_{\mathsf{Z}_2}$ *that induce the identity map on the maximal torsion-free quotients* $i^*\mathcal{I}_{\mathsf{Z}_j}^{\mathrm{tf}} \subset \mathcal{O}_{\mathsf{D}}$.

In particular, all of these spaces are equivalent to their discrete sets of connected components.

Proof The spaces in (1) and (2) are equivalent by definition of the derived intersection.

For the equivalence of (2) and (3), choose an arbitrary cofibrant resolution $\mathcal{E}^\bullet \to \mathcal{O}_{\mathsf{Z}_1}$ of $\mathcal{O}_{\mathsf{Z}_1}$ as a sheaf of $\mathcal{O}_{\mathsf{X}}$-dg algebras concentrated in degrees ≤ 0, with degree zero term $\mathcal{E}^0 = \mathcal{O}_{\mathsf{X}}$. The dg $\mathcal{O}_{\mathsf{D}}$-algebra $i^*\mathcal{E} \cong \mathcal{E}^\bullet \otimes_{\mathcal{O}_{\mathsf{X}}} \mathcal{O}_{\mathsf{D}}$ is then also cofibrant. Meanwhile, we may resolve $\mathcal{O}_{\mathsf{D}}$ by the Koszul complex $\mathcal{B}_{\mathsf{D}} := (\mathcal{O}_{\mathsf{X}}(-\mathsf{D}) \to \mathcal{O}_{\mathsf{X}})$. The space of equivalences $Li^*\mathcal{O}_{\mathsf{Z}_1} \cong Li^*\mathcal{O}_{\mathsf{Z}_2}$ of dg $\mathcal{O}_{\mathsf{D}}$-algebras is then equivalent to the space of dg $\mathcal{O}_{\mathsf{D}}$-algebra maps $i^*\mathcal{E}^\bullet \to i^*(\mathcal{O}_{\mathsf{Z}_2} \otimes \mathcal{B}_{\mathsf{D}})$. By degree considerations, such a dg algebra map is equivalent to a map of complexes whose degree zero part is the natural projection $\mathcal{O}_{\mathsf{D}} \to \mathcal{O}_{\mathsf{Z}_2}$, as desired. Note further that the full space of equivalences is modelled by the simplicial set of dg algebra maps

$$i^*\mathcal{E} \to i^*(\mathcal{O}_{\mathsf{Z}_2} \otimes \mathcal{B}_{\mathsf{D}}) \otimes \Omega^\bullet_\Delta, \tag{2-6}$$

where $\Omega^\bullet_\Delta$ is the simplicial dg algebra of polynomial differential forms on simplices, and the degree-zero part is the canonical map $\mathcal{O}_{\mathsf{D}} \to \mathcal{O}_{\mathsf{Z}_2}$ (constant in the simplex direction). Similar considerations now show that the sets of higher simplices of spaces (2) and (3) are the same, and agree with the set of 0-simplices, so that these spaces are equivalent and both discrete. Namely, examining the degrees and using the compatibility with the differentials, we see that if a map (2-6) has degree-zero part the map $\mathcal{O}_{\mathsf{D}} \to \mathcal{O}_{\mathsf{Z}_2}$, it must take values in the subspace $i^*(\mathcal{O}_{\mathsf{Z}_2} \otimes \mathcal{B}_{\mathsf{D}}) \otimes \Omega^{0,\mathrm{cl}}_\Delta \cong i^*(\mathcal{O}_{\mathsf{Z}_2} \otimes \mathcal{B}_{\mathsf{D}})$, so that it reduces to a 0-simplex, as desired.

The equivalence of (3) and (4) follows from the exact triangle (2-4), using the fact that the trivialization of the determinant of $\mathcal{I}_{\mathsf{Z}}$ is defined by the inclusion $\mathcal{I}_{\mathsf{Z}} \hookrightarrow \mathcal{O}_{\mathsf{X}}$. Note that it is also clear that the space (4) is discrete: since $i^*\mathcal{I}_{\mathsf{Z}_1}, i^*\mathcal{I}_{\mathsf{Z}_2}$ are sheaves concentrated in degree zero, the complex $\mathsf{RHom}(i^*\mathcal{I}_{\mathsf{Z}_1}, i^*\mathcal{I}_{\mathsf{Z}_2})$ is concentrated in nonnegative degrees, so there are no nontrivial homotopies between morphisms, and similarly for the determinants.

Finally, (5) and (6) are clearly equivalent by definition of $i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}$, and they are equivalent to (4) because for an arbitrary morphism $\phi\colon \mathcal{I}_{\mathsf{Z}_1} \to \mathcal{I}_{\mathsf{Z}_2}$, the diagram

$$\begin{array}{ccc} \mathcal{I}_{\mathsf{Z}_1} & \xrightarrow{\ \phi\ } & \mathcal{I}_{\mathsf{Z}_2} \\ \downarrow & & \downarrow \\ \mathcal{O}_{\mathsf{D}} & \xrightarrow{\ \det\phi\ } & \mathcal{O}_{\mathsf{D}} \end{array}$$

is commutative. □

Definition 2.3 We refer to any of the equivalent discrete sets in Lemma 2.2 as the *set of* D*-equivalences from* Z_1 *to* Z_2. We say that Z_1 and Z_2 are D*-equivalent* if there exists a D-equivalence from Z_1 to Z_2.

Evidently D-equivalence defines an equivalence relation on $\mathsf{X}^{[n]}$. If we are just interested in the equivalence classes, an a priori weaker notion will suffice:

Lemma 2.4 *Two elements $\mathsf{Z}_1, \mathsf{Z}_2 \in \mathsf{X}^{[n]}$ are D-equivalent if and only if the $\mathcal{O}_{\mathsf{D}}$-modules $i^*\mathcal{I}_{\mathsf{Z}_1}$ and $i^*\mathcal{I}_{\mathsf{Z}_2}$ are isomorphic.*

Proof By part (4) of Lemma 2.2, it suffices to show that $i^*\mathcal{I}_{\mathsf{Z}_1}$ and $i^*\mathcal{I}_{\mathsf{Z}_2}$ are isomorphic as $\mathcal{O}_{\mathsf{D}}$-modules if and only if they are isomorphic by a map with trivial determinant. But if $\phi\colon i^*\mathcal{I}_{\mathsf{Z}_1} \to i^*\mathcal{I}_{\mathsf{Z}_2}$ is any isomorphism, then its determinant gives an isomorphism

$$\mathcal{O}_{\mathsf{D}} \cong \det \mathcal{I}_{\mathsf{Z}_1} \xrightarrow{\sim} \det \mathcal{I}_{\mathsf{Z}_2} \cong \mathcal{O}_{\mathsf{D}},$$

or equivalently an invertible element $g := \det\phi \in \mathsf{H}^0(\mathcal{O}_{\mathsf{D}}^\times)$. Since the rank of an ideal sheaf is equal to one, the determinant is a linear functional on endomorphisms and we have

$$\det(g^{-1}\phi) = g^{-1}\det(\phi) = \det(\phi)^{-1}\det(\phi) = 1.$$

Therefore $g^{-1}\phi\colon \mathcal{I}_{\mathsf{Z}_1} \to \mathcal{I}_{\mathsf{Z}_2}$ is an isomorphism with trivial determinant, as desired. □

Remark 2.5 For the rest of the paper, we work almost exclusively with the spaces 4–6 in Lemma 2.2; in those cases, as explained in the proof, the discreteness is immediate because the objects involved are sheaves concentrated in degree zero, so there is no room for higher homotopies. ◊

Remark 2.6 Lemma 2.2 holds in much more general circumstances, with essentially the same proof. Namely, statements (1), (2), (3), (5) and (6) are equivalent for any pair of subschemes $\mathsf{Z}_1, \mathsf{Z}_2$ in a scheme X with an effective Cartier divisor D. For the equivalence with (4), we need to speak about trivialization of the determinant being determined away from codimension two subsets, and for this we need to assume in addition that X is S_2 (eg smooth or more generally normal) and that the subschemes $\mathsf{Z}_1, \mathsf{Z}_2 \subset \mathsf{X}$ have codimension at least two. ◊

Remark 2.7 In the proof, the discreteness of the space (1) is deduced from the equivalence with the other spaces. We can also argue directly using standard results on mapping spaces for derived schemes, as follows.

Let $\mathsf{U} = \mathsf{Z}_1 \cap^h \mathsf{D}$ and let $\mathsf{V} = \mathsf{Z}_2 \cap^h \mathsf{D}$. By definition, space (1) is the space of equivalences $\mathsf{U} \xrightarrow{\sim} \mathsf{V}$ commuting with the inclusions into D, ie it is the homotopy fibre of the map $\mathsf{Maps}(\mathsf{U},\mathsf{V}) \to \mathsf{Maps}(\mathsf{U},\mathsf{D})$ induced by composition with the inclusion $\mathsf{V} \hookrightarrow \mathsf{D}$. Now observe that

$$\mathsf{Maps}(\mathsf{U},\mathsf{V}) \cong \mathsf{Maps}\Big(\mathsf{U}, \mathsf{Z}_j \overset{h}{\underset{\mathsf{X}}{\times}} \mathsf{D}\Big) \cong \mathsf{Maps}(\mathsf{U},\mathsf{Z}_2) \overset{h}{\underset{\mathsf{Maps}(\mathsf{U},\mathsf{X})}{\times}} \mathsf{Maps}(\mathsf{U},\mathsf{D});$$

hence the homotopy fibre in question is equivalent to the homotopy fibre W of the map $\mathsf{Maps}(\mathsf{U},\mathsf{Z}_2) \to \mathsf{Maps}(\mathsf{U},\mathsf{X})$. It thus suffices to show that the homotopy groups $\pi_i(\mathsf{W})$ are trivial for $i > 0$. By the long exact sequence of homotopy, it suffices to show that

(a) $\pi_i(\mathsf{Maps}(\mathsf{U},\mathsf{Z}_2)) = \pi_i(\mathsf{Maps}(\mathsf{U},\mathsf{X})) = 0$ for $i > 1$, and

(b) the map $\pi_1(\mathsf{Maps}(\mathsf{U},\mathsf{Z}_2)) \to \pi_1(\mathsf{Maps}(\mathsf{U},\mathsf{X}))$ is injective.

For this, we use that U is affine with $\pi_i(\mathcal{O}(\mathsf{U})) := \mathsf{H}^{-i}(\mathcal{O}(\mathsf{U})) = 0$ for $i > 1$, and that Z_2 and X are ordinary schemes. Moreover, without loss of generality, the latter are affine, hence (-1)-geometric as derived stacks. Thus (a) follows immediately from [46, Corollary 2.2.4.6]. Furthermore from the proof of [loc. cit.] using the Postnikov system for U, one sees that the map on fundamental groups in (b) is identified with the canonical map

$$\mathsf{Hom}(\mathbb{L}_{\mathsf{Z}_2}, \mathcal{F}) \to \mathsf{Hom}(\mathbb{L}_{\mathsf{X}}, \mathcal{F}), \tag{2-7}$$

where $\mathcal{F} := \mathsf{H}^{-1}(\mathcal{O}_{\mathsf{U}})$. Since Z_2 and X are schemes, their cotangent complexes are concentrated in nonpositive degrees, with zeroth cohomology the Kähler differentials Ω^1. Hence (2-7) is, in turn, identified with the map $\mathsf{Hom}(\Omega^1_{\mathsf{Z}_2}, \mathcal{F}) \to \mathsf{Hom}(\Omega^1_{\mathsf{X}}, \mathcal{F})$ induced by the quotient $\Omega^1_{\mathsf{X}} \to \Omega^1_{\mathsf{Z}_2}$, giving the desired injectivity. ◊

2.3 The Hilbert groupoid

Evidently D-equivalences between points of $\mathsf{X}^{[n]}$ can be composed, so that they form the arrows of a groupoid whose objects are the points of $\mathsf{X}^{[n]}$. We will now show that this defines a smooth groupoid scheme (or in the analytic context, a Lie groupoid).

More precisely, let

$$(\mathsf{X},\mathsf{D})^{[n]} := \mathsf{X}^{[n]} \underset{\mathsf{Perf}_0(\mathsf{D})}{\times} \mathsf{X}^{[n]} \subset \mathsf{Perf}_0(\mathsf{X}) \underset{\mathsf{Perf}_0(\mathsf{D})}{\times} \mathsf{Perf}_0(\mathsf{X})$$

be the derived stack defined as the homotopy fibre product of $\mathsf{X}^{[n]}$ with itself along the map sending a subscheme Z to $i^*\mathcal{I}_{\mathsf{Z}}$, where $\mathcal{I}_{\mathsf{Z}}$ is the ideal sheaf of Z equipped with the canonical trivialization of $\det \mathcal{I}_{\mathsf{Z}}$. By definition of the homotopy fibre product, points of $(\mathsf{X},\mathsf{D})^{[n]}$ consist of triples $(\mathsf{Z}_1, \mathsf{Z}_2, \lambda)$, where $\mathsf{Z}_1, \mathsf{Z}_2$ are points of $\mathsf{X}^{[n]}$, and $\lambda\colon \mathsf{Z}_1 \cap^h \mathsf{D} \xrightarrow{\sim} \mathsf{Z}_2 \cap^h \mathsf{D}$ is a D-equivalence. Composing the D-equivalences gives $(\mathsf{X},\mathsf{D})^{[n]}$ the structure of a groupoid in the category of derived stacks.

Definition 2.8 The groupoid $(\mathsf{X},\mathsf{D})^{[n]}$ over $\mathsf{X}^{[n]}$ is the n^{th} *Hilbert groupoid of* (X,D).

In fact, this object is a classical smooth variety:

Theorem 2.9 *The derived stack* $(\mathsf{X},\mathsf{D})^{[n]}$ *is a smooth classical scheme. The projections* $s, t\colon (\mathsf{X},\mathsf{D})^{[n]} \to \mathsf{X}^{[n]}$ *are smooth morphisms, so that* $(\mathsf{X},\mathsf{D})^{[n]}$ *defines a smooth groupoid scheme over* $\mathsf{X}^{[n]}$ *whose orbits are the* D*-equivalence classes. Moreover, the induced map* $(s,t)\colon (\mathsf{X},\mathsf{D})^{[n]} \to \mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$ *restricts to an isomorphism over the Zariski open set* $(\mathsf{X}\setminus\mathsf{D})^{[n]} \times (\mathsf{X}\setminus\mathsf{D})^{[n]}$*, and is therefore birational.*

Proof To prove representability, we will use the interpretation of D-equivalences from part (5) of Lemma 2.2: the fibre of $(\mathsf{X},\mathsf{D})^{[n]}$ over a point $(\mathsf{Z}_1,\mathsf{Z}_2) \in \mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$ is the set of isomorphisms $\mathcal{I}_{\mathsf{Z}_1} \otimes \mathcal{O}_{\mathsf{D}} \to \mathcal{I}_{\mathsf{Z}_2} \otimes \mathcal{O}_{\mathsf{D}}$ commuting with the natural maps to $\mathcal{O}_{\mathsf{D}}$. Recall that there is a universal ideal sheaf $\mathcal{I}$ on $\mathsf{X}^{[n]} \times \mathsf{X}$ whose restriction to any slice $\{\mathsf{Z}\} \times \mathsf{X} \cong \mathsf{X}$ is the ideal defining Z. Let $\widetilde{\mathsf{X}} := \mathsf{X}^{[n]} \times \mathsf{X}^{[n]} \times \mathsf{X}$ and let $\mathcal{I}_1, \mathcal{I}_2 \subset \mathcal{O}_{\widetilde{\mathsf{X}}}$ be the pullbacks of $\mathcal{I}$ along the two projections to $\mathsf{X}^{[n]} \times \mathsf{X}$. Let $\widetilde{\mathsf{D}} := \mathsf{X}^{[n]} \times \mathsf{X}^{[n]} \times \mathsf{D} \subset \widetilde{\mathsf{X}}$. Finally, let $p\colon \widetilde{\mathsf{X}} \to \mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$ be the projection. Observe that composition with the map $\mathcal{I}_2 \to \mathcal{O}_{\widetilde{\mathsf{X}}} \to \mathcal{O}_{\widetilde{\mathsf{D}}}$ gives a natural map $p_*\mathcal{H}om(\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}, \mathcal{I}_2 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}) \to p_*\mathcal{H}om(\mathcal{I}_1, \mathcal{O}_{\widetilde{\mathsf{D}}})$ of quasicoherent sheaves on $\mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$, and that the morphisms $\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}} \to \mathcal{I}_2 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}$ compatible with the maps to $\mathcal{O}_{\widetilde{\mathsf{D}}}$ form a torsor for the quasicoherent sheaf

$$\mathcal{E} := \ker\bigl(p_*\mathcal{H}om(\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}, \mathcal{I}_2 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}) \to p_*\mathcal{H}om(\mathcal{I}_1, \mathcal{O}_{\widetilde{\mathsf{D}}})\bigr).$$

We claim that $\mathcal{E}$ is, in fact, a coherent sheaf. Indeed, applying the exact sequence (2-5) to the universal scheme $\mathcal{Z}_2 \subset \widetilde{\mathsf{X}}$ given by the vanishing of $\mathcal{I}_2$, we see that $\mathcal{E}$ is identified with the image of the canonical map

$$p_*\mathcal{H}om(\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}, \mathcal{T}or_1(\mathcal{O}_{\mathcal{Z}_2}, \mathcal{O}_{\widetilde{\mathsf{D}}})) \to p_*\mathcal{H}om(\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}, \mathcal{I}_2 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}).$$

Since $\mathcal{T}or_1(\mathcal{O}_{\mathcal{Z}_2}, \mathcal{O}_{\widetilde{\mathsf{D}}})$ is supported on $\mathcal{Z}_2$, and the latter is proper over $\mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$, it follows that $p_*\mathcal{H}om(\mathcal{I}_1 \otimes \mathcal{O}_{\widetilde{\mathsf{D}}}, \mathcal{T}or_1(\mathcal{O}_{\mathcal{Z}_2}, \mathcal{O}_{\widetilde{\mathsf{D}}}))$ is coherent, and hence so is $\mathcal{E}$, as claimed. It now follows that the total space $\mathsf{Tot}(\mathcal{E})$ is a scheme that is relatively affine over $\mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$, and hence so is the torsor in question. The invertible morphisms then form an open subscheme of the torsor, as desired.

To establish the smoothness of s, t, it is now enough to check that their relative tangent complexes $\mathbb{T}s, \mathbb{T}t$ at any closed point of $(\mathsf{X}, \mathsf{D})^{[n]}$ are concentrated in degree zero; note that this statement is independent from the details of the explicit construction of the scheme above, and is indeed proved most easily from the abstract definition as a derived stack. We will prove the smoothness for s; the argument for t is identical.

We have the commutative diagram of tangent complexes

$$\begin{array}{ccccc}
\mathbb{T}s & \longrightarrow & \mathbb{T}(\mathsf{X}, \mathsf{D})^{[n]} & \longrightarrow & s^*\mathbb{T}\mathsf{X}^{[n]} \\
\downarrow & & \downarrow & & \| \\
t^*\mathbb{T}\mathsf{X}^{[n]} & \longrightarrow & s^*\mathbb{T}\mathsf{X}^{[n]} \oplus t^*\mathbb{T}\mathsf{X}^{[n]} & \longrightarrow & s^*\mathbb{T}\mathsf{X}^{[n]} \\
\downarrow & & \downarrow & & \downarrow \\
(Li^*)^*\mathbb{T}\mathsf{Perf}_0(\mathsf{D}) & = & (Li^*)^*\mathbb{T}\mathsf{Perf}_0(\mathsf{D}) & \longrightarrow & 0
\end{array}$$

where $\mathbb{T}s$ denotes the relative tangent complex of the map $s\colon (\mathsf{X}, \mathsf{D})^{[n]} \to \mathsf{X}^{[n]}$. Since homotopy fibre products of stacks give homotopy fibre sequences of tangent complexes, all three rows and the right two columns are fibre sequences, from which we deduce that

$$\mathbb{T}s \cong \mathsf{fibre}\bigl(t^*\mathbb{T}\mathsf{X}^{[n]} \to (Li^*)^*\mathbb{T}\mathsf{Perf}_0(\mathsf{D})\bigr). \tag{2-8}$$

Suppose that $(\mathsf{Z}_1, \mathsf{Z}_2, \lambda) \in (\mathsf{X}, \mathsf{D})^{[n]}$ is a closed point. Let $\mathcal{I} = \mathcal{I}_{\mathsf{Z}_2}$ be the ideal defining Z_2. Identifying the tangent complexes of $\mathsf{X}^{[n]} \subset \mathsf{Perf}_0(\mathsf{X})$ and $\mathsf{Perf}_0(\mathsf{D})$ in (2-8) with the complexes of traceless derived endomorphisms of $\mathcal{I}$ and $Li^*\mathcal{I}$ respectively, we have

$$\begin{aligned}
\mathbb{T}_{(\mathsf{Z}_1, \mathsf{Z}_2, \lambda)}s &\cong \mathsf{fibre}\bigl(\mathsf{R}\Gamma(\mathcal{R}\mathcal{E}nd_{\mathsf{X}}(\mathcal{I})_0) \to \mathsf{R}\Gamma(\mathcal{R}\mathcal{E}nd_{\mathsf{D}}(Li^*\mathcal{I}))_0\bigr)[1] \\
&\cong \mathsf{R}\Gamma(\mathcal{R}\mathcal{E}nd_{\mathsf{X}}(\mathcal{I})_0(-\mathsf{D}))[1],
\end{aligned}$$

where the second equivalence is induced by tensoring the complex $\mathcal{REnd}_{\mathsf{X}}(\mathcal{I})_0$ with the exact sequence for the quotient $\mathcal{O}_{\mathsf{D}} \cong \mathcal{O}_{\mathsf{X}}/\mathcal{O}_{\mathsf{X}}(-\mathsf{D})$. As explained in [5, Corollary 2.6], we have

$$\mathcal{REnd}_{\mathsf{X}}(\mathcal{I})_0[1] \cong \mathcal{H}om_{\mathsf{X}}(\mathcal{I}, \mathcal{O}_{\mathsf{Z}_2}),$$

so that

$$\mathcal{REnd}_{\mathsf{X}}(\mathcal{I})_0(-\mathsf{D})[1] \cong \mathcal{H}om_{\mathsf{X}}(\mathcal{I}, \mathcal{O}_{\mathsf{Z}_2}) \otimes_{\mathcal{O}_{\mathsf{X}}} \mathcal{O}_{\mathsf{X}}(-\mathsf{D})$$

is a skyscraper sheaf concentrated on the zero-dimensional scheme Z_2. Therefore

$$\begin{aligned}\mathbb{T}_{(\mathsf{Z}_1,\mathsf{Z}_2,\lambda)}s &\cong \mathsf{R}\Gamma(\mathcal{H}om(\mathcal{I}, \mathcal{O}_{\mathsf{Z}_2}) \otimes_{\mathcal{O}_{\mathsf{X}}} \mathcal{O}_{\mathsf{X}}(-\mathsf{D})) \\ &\cong \mathsf{Hom}_{\mathsf{X}}(\mathcal{I}, \mathcal{O}_{\mathsf{Z}_2}(-\mathsf{D}))\end{aligned}$$

is concentrated in degree zero, as desired. □

Example 2.10 The construction of the Hilbert groupoid is already interesting in the case when D is smooth and $n = 1$, where it recovers a known construction of Lie groupoids associated with curves in surfaces [11; 18], or more precisely a groupoid integrating the Lie algebroid $\mathcal{T}_{\mathsf{X}}(-\mathsf{D})$ of vector fields vanishing on D, via blowing up. Namely, a length-one subscheme is simply a point of X with the reduced scheme structure, so that $\mathsf{X}^{[1]} = \mathsf{X}$, and hence by Theorem 2.9, we have a natural birational map

$$(s,t)\colon (\mathsf{X},\mathsf{D})^{[1]} \to \mathsf{X}\times\mathsf{X}.$$

A pair of points in $\mathsf{X}^{[1]}$ are D-equivalent if and only if one of two things occurs: either they both lie in $\mathsf{X}\setminus\mathsf{D}$, or they are the same point of D. It follows that the isotropy groups of points $p \in \mathsf{D}$ are two-dimensional, and so the preimage of the diagonal embedding of D under the map (s,t) is a hypersurface in $(\mathsf{X},\mathsf{D})^{[1]}$. Hence (s,t) factors through the blowup of $\mathsf{X}\times\mathsf{X}$ along the diagonal copy of D. One can check (eg using Theorem 4.7 below) that this identifies $(\mathsf{X},\mathsf{D})^{[1]}$ with the complement of the strict transforms of $\mathsf{D}\times\mathsf{X}$ and $\mathsf{X}\times\mathsf{D}$ in the blowup. It would be interesting to give a similarly explicit description of the birational map $(\mathsf{X},\mathsf{D})^{[n]} \to \mathsf{X}\times\mathsf{X}$ when $n > 1$ or D is singular. ◊

2.4 Stabilizer groups

Definition 2.11 We denote by $\mathsf{G}_{\mathsf{Z},\mathsf{D}}$ the stabilizer group algebra of $\mathsf{Z} \in \mathsf{X}^{[n]}$ in the groupoid $(\mathsf{X},\mathsf{D})^{[n]}$, ie the group of homotopy classes of infinitesimal self D-equivalences of Z, and by $\mathfrak{g}_{\mathsf{Z},\mathsf{D}}$ its Lie algebra.

By Lemma 2.2, $\mathsf{G}_{\mathsf{Z},\mathsf{D}}$ can be identified with the group of automorphisms of $i^*\mathcal{I}_{\mathsf{Z}}$ that act as the identity on the maximal torsion-free quotient $i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}$. Correspondingly, $\mathfrak{g}_{\mathsf{Z},\mathsf{D}}$ is identified with the endomorphisms of $i^*\mathcal{I}_{\mathsf{Z}}$ that act by zero on $i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}$. Since the maximal torsion subsheaf

$$\tau_{\mathsf{Z},\mathsf{D}} := \mathcal{T}or_1^{\mathcal{O}_{\mathsf{X}}}(\mathcal{O}_{\mathsf{Z}}, \mathcal{O}_{\mathsf{D}}) = \mathcal{H}^{-1}(\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}})$$

of $i^*\mathcal{I}_{\mathsf{Z}}$ is preserved by any endomorphism, we immediately deduce the following:

Lemma 2.12 *There is a canonical exact sequence of groups*

$$0 \to \mathsf{Hom}_{\mathsf{D}}(i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}, \tau_{\mathsf{Z},\mathsf{D}}) \to \mathsf{G}_{\mathsf{Z},\mathsf{D}} \to \mathsf{Aut}_{\mathsf{D}}(\tau_{\mathsf{Z},\mathsf{D}}),$$

where the group law on $\mathsf{Hom}_{\mathsf{D}}(i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}, \tau_{\mathsf{Z},\mathsf{D}})$ *is addition. Correspondingly we have an exact sequence of Lie algebras*

$$0 \to \mathsf{Hom}_{\mathsf{D}}(i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}, \tau_{\mathsf{Z},\mathsf{D}}) \to \mathfrak{g}_{\mathsf{Z},\mathsf{D}} \to \mathsf{End}_{\mathsf{D}}(\tau_{\mathsf{Z},\mathsf{D}}),$$

where $\mathsf{Hom}_{\mathsf{D}}(i^*\mathcal{I}_{\mathsf{Z}}^{\mathrm{tf}}, \tau_{\mathsf{Z},\mathsf{D}})$ *is abelian.*

2.5 Syzygy matrices and orbit data

We now turn to the classification of the orbits of $(\mathsf{X},\mathsf{D})^{[n]}$, ie the enumeration of the D-equivalence classes. In light of Lemma 2.4, this is equivalent to the classification of $\mathcal{O}_{\mathsf{D}}$-modules $\mathcal{F}$ such that $\mathcal{F} \cong i^*\mathcal{I}_{\mathsf{Z}}$ for some $\mathsf{Z} \in \mathsf{X}^{[n]}$.

In order to characterize such modules, note that since X is smooth of dimension two the stalk of $\mathcal{I}_{\mathsf{Z}}$ at any point $p \in \mathsf{X}$ is an $\mathcal{O}_{\mathsf{X},p}$-module of projective dimension at most two; indeed, if $p \notin \mathsf{Z}$ then $\mathcal{I}_{\mathsf{Z},p} = \mathcal{O}_{\mathsf{X},p}$, while if $p \in \mathsf{Z}$, the classical Hilbert–Burch theorem states that the minimal resolution has the form

$$0 \to \mathcal{O}_{\mathsf{X},p}^{\oplus k} \xrightarrow{S} \mathcal{O}_{\mathsf{X}}^{\oplus(k+1)} \xrightarrow{M(S)} \mathcal{I}_{\mathsf{Z},p} \to 0, \tag{2-9}$$

where S is a $(k+1)\times k$-matrix of elements of the maximal ideal $\mathfrak{m}_{\mathsf{X},p} < \mathcal{O}_{\mathsf{X},p}$, called the *syzygy matrix*, and $M(S)$ is the row vector formed from the maximal minors of S with alternating signs, ie

$$M(S) = \begin{pmatrix}\det(S_0) & -\det(S_1) & \cdots & (-1)^k \det(S_k)\end{pmatrix},$$

where S_j is the $k \times k$ submatrix obtained by removing the $(j+1)$st row from S. The matrix S is determined by Z and p up to multiplication on the left and right by elements in $\mathsf{GL}_{k+1}(\mathcal{O}_{\mathsf{X},p})$ and $\mathsf{GL}_k(\mathcal{O}_{\mathsf{X},p})$, respectively.

Applying the functor Li^* to (2-9), we obtain a minimal resolution

$$0 \to \mathcal{O}_{\mathsf{D},p}^{\oplus k} \xrightarrow{S|_{\mathsf{D}}} \mathcal{O}_{\mathsf{D},p}^{\oplus(k+1)} \xrightarrow{M(S|_{\mathsf{D}})} i^*\mathcal{I}_{\mathsf{Z},p} \to 0 \tag{2-10}$$

for any $p \in \mathsf{D}$, so that the D-equivalence class of Z near p is controlled by the equivalence class of the matrix $S|_{\mathsf{D}}$. For instance, $S|_{\mathsf{D}}$ can be used to construct an explicit model for $\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D}}$ as a dg $\mathcal{O}_{\mathsf{D}}$-algebra. Namely, by extending (2-9) one step to the right we obtain a minimal resolution of $\mathcal{O}_{\mathsf{Z},p}$, which carries a canonical dg algebra structure due to Herzog [20] (see also [3, Section 6]), whose structure constants are determined by S. Tensoring the resolution with $\mathcal{O}_{\mathsf{D},p}$, we obtain an explicit model for $\mathcal{O}_{\mathsf{Z}\cap^h\mathsf{D},p}$ as an $\mathcal{O}_{\mathsf{D},p}$-algebra.

This motivates the following definition. Note that in this definition, there is a unique local orbit datum of size $k = 0$, corresponding to the case in which the point $p \in \mathsf{D}$ does not lie in Z and $S|_{\mathsf{D}}$ is the zero map $0 \to \mathcal{O}_{\mathsf{D},p}$, which we think of as a matrix of size 1×0.

Definition 2.13 A *local orbit datum supported at* $p \in \mathsf{D}$ *of size* $k \geq 0$ is an element of the double quotient

$$\mathsf{GL}_{k+1}(\mathcal{O}_{\mathsf{D},p}) \backslash \mathfrak{m}_{\mathsf{D},p}^{(k+1)\times k} / \mathsf{GL}_k(\mathcal{O}_{\mathsf{D},p}), \tag{2-11}$$

which can be realized as the restriction to D of the syzygy matrix of a zero-dimensional subscheme $\mathsf{Z} \subset \mathsf{X}$. We denote the set of local orbit data of size k by $\mathsf{OrbDat}_k(\mathsf{D}, p)$, and set

$$\mathsf{OrbDat}(\mathsf{D}, p) := \bigsqcup_{k \geq 0} \mathsf{OrbDat}_k(\mathsf{D}, p).$$

A *(global) orbit datum* is an assignment $\phi\colon p \mapsto \phi(p)$ of a local orbit datum $\phi(p) \in \mathsf{OrbDat}_k(\mathsf{D}, p)$ to every point $p \in \mathsf{D}$ such that there are only finitely many p with $\phi(p)$ of positive size.

Note that every global orbit datum can be realized by a zero-dimensional subscheme, simply by taking the disjoint union of the zero-dimensional subschemes at each p where $\phi(p)$ has positive size.

If $\mathsf{Z} \in \mathsf{X}^{[n]}$, we denote by

$$[\mathsf{Z}]_p \in \mathsf{OrbDat}(\mathsf{D}, p)$$

the image of the syzygy matrix of Z, which is well-defined. If ϕ is a global orbit datum, we define a subset

$$\mathsf{X}_\phi^{[n]} := \{\mathsf{Z} \in \mathsf{X}^{[n]} \mid [\mathsf{Z}]_p = \phi(p) \text{ for all } p \in \mathsf{D}\}.$$

Then clearly $\mathsf{X}_\phi^{[n]}$, if nonempty, is a D-equivalence class, ie an orbit of $(\mathsf{X}, \mathsf{D})^{[n]}$, and every D-equivalence class arises in this way.

2.6 Classification of orbits

2.6.1 Smooth points Let $p \in \mathsf{D}$ be a smooth point of D and suppose that $\mathsf{Z} \in \mathsf{X}^{[n]}$. Then the ring $\mathcal{O}_{\mathsf{D},p}$ of analytic germs at p is a principal ideal domain, isomorphic to the ring $\mathbb{C}\{x\}$ of convergent power series in a coordinate x centred at p. Then since $i^*\mathcal{I}_{\mathsf{Z},p}$ is a finitely generated module whose free part has rank one, it must have the form

$$i^*\mathcal{I}_{\mathsf{Z},p} \cong \mathcal{O}_{\mathsf{D},p} \oplus \bigoplus_{j=1}^{k} \mathcal{O}_{\mathsf{D},p}/\mathfrak{m}_{\mathsf{D},p}^{\mu_j} \cong \mathbb{C}\{x\} \oplus \bigoplus_{j=1}^{k} \mathbb{C}[x]/x^{\mu_j} \tag{2-12}$$

for a unique collection $\mu = (\mu_1 \geq \mu_2 \geq \cdots)$ of positive integers. Correspondingly, the syzygy matrix can be put in the Smith normal form

$$S|_{\mathsf{D}} \sim S_\mu(x) := \begin{pmatrix} x^{\mu_1} & 0 & 0 & \dots & 0 \\ 0 & x^{\mu_2} & 0 & \dots & 0 \\ 0 & 0 & x^{\mu_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & x^{\mu_k} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{2-13}$$

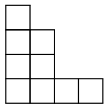
Recall the following definition.

Definition 2.14 A *Young diagram* (of *length* k) is a decreasing sequence $\mu = (\mu_1 \geq \cdots \geq \mu_k)$ of positive integers. Its *size* is its sum:

$$|\mu| = \mu_1 + \cdots + \mu_k.$$

By convention, we set $\mu_j = 0$ for $j > k$.

We depict a Young diagram λ by placing λ_j boxes in the j^{th} row, starting the numbering of rows from 1, for example

$(4, 2, 2, 1) \quad \leftrightarrow$ 

is a Young diagram of length four and size nine.

Rephrasing the above, we have the following characterization of the local orbit data at p:

Lemma 2.15 *If $p \in \mathsf{D}$ is a smooth point, the map sending a Young diagram μ to the equivalence class of the corresponding Smith normal form matrix gives a canonical bijection*

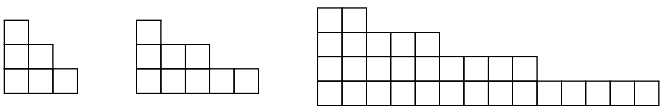
$$\mathsf{OrbDat}_k(\mathsf{D}, p) \cong \{\text{Young diagrams of length } k\}.$$

In particular, if D is smooth, then the global orbit data are in bijection with functions from D to the set of Young diagrams of arbitrary size.

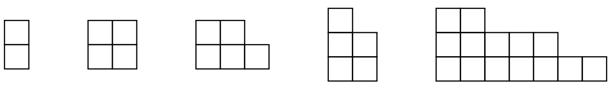
We now describe the basic characteristics of the orbit datum corresponding to a Young diagram μ. In order to do so, we introduce the following auxiliary notion:

Definition 2.16 We say that a Young diagram $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots)$ is *horizontally convex* if $\lambda_i - \lambda_{i+1} \geq \lambda_{i+1} - \lambda_{i+2}$ for every $i > 0$.

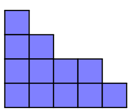
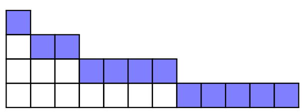
For instance, the following Young diagrams are horizontally convex:



whereas the following are *not* horizontally convex:



For a Young diagram μ, let $\mathrm{hc}(\mu)$ be the unique Young diagram λ such that $\lambda_j - \lambda_{j+1} = \mu_j$ for all $j > 0$, or equivalently $\lambda_j = \sum_{l \geq j} \mu_l$. Then $\mathrm{hc}(\mu)$ is horizontally convex, and evidently every horizontally convex diagram can be obtained in this way. For example,

(2-14) $\mu =$  $\mapsto \quad \mathrm{hc}(\mu) =$ 

Note that the number of boxes of $\mathrm{hc}(\mu)$ is given by

$$|\mathrm{hc}(\mu)| = \sum_{j\geq 1} j\mu_j, \tag{2-15}$$

which will be useful below.

Given a Young diagram μ and a smooth point $p \in \mathsf{D}$, we define a zero-dimensional subscheme

$$\mathsf{Z}_p^\mu \subset \mathsf{X}$$

supported at p, as follows. First, choose coordinates (x, y) centred at p, such that D is given locally by the equation $y = 0$, and let $\lambda = \mathrm{hc}(\mu)$. Then set

$$\mathsf{Z}_p^\mu := \text{the vanishing locus of the monomials } x^j y^l \text{ for } j \geq \lambda_{l+1}. \tag{2-16}$$

Using the horizontal convexity of λ, it is straightforward to verify that this definition is independent of the choice of coordinates (x, y).

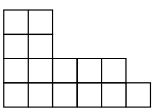
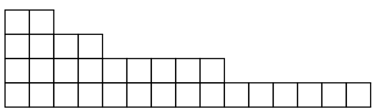
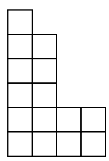
Evidently the algebra $\mathcal{O}_{\mathsf{Z}_p^\mu} \cong \mathbb{C}[x, y]/(x^j y^l)_{j<\lambda_{l+1}}$ has a basis given by the monomials $x^j y^l$, where $j < \lambda_{l+1}$ for $l = 0, \ldots, k-1$. The following is therefore immediate:

Lemma 2.17 *The algebra $\mathcal{O}_{\mathsf{Z}_p^\mu}$ has a basis indexed by the boxes of the Young diagram* $\mathrm{hc}(\mu)$*. In particular, the length of the scheme Z_p^μ is given by*

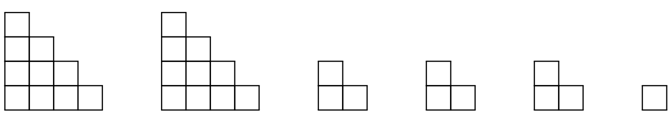
$$\#\mathsf{Z}_p^\mu = |\mathrm{hc}(\mu)|.$$

Remark 2.18 Z_p^μ can be defined in a coordinate-free fashion as follows. Let μ^T be the transpose of μ, defined by $\mu_j^T = \#\{l > 0 \mid \mu_l \geq j\}$, and let $k^T = \mu_1$ be its length. For a collection $p_1, \ldots, p_{k^T}$ of pairwise distinct points in D, consider the subscheme given by the disjoint union of $(\mu_j^T)^{\text{th}}$ order neighbourhoods of p_j for $1 \leq j \leq k^T$. Taking the limit as the points $p_1, \ldots, p_{k^T} \to p$ we obtain the subscheme Z_p^μ.

For example, consider the Young diagrams

$\mu :=$  $\lambda := \mathrm{hc}(\mu) =$  $\mu^T =$ 

Then Z_p^μ is obtained as the limit of the schemes defined by the intersection of the ideals $\mathfrak{m}_{\mathsf{X},p_1}^4$, $\mathfrak{m}_{\mathsf{X},p_2}^4$, $\mathfrak{m}_{\mathsf{X},p_3}^2$, $\mathfrak{m}_{\mathsf{X},p_4}^2$, $\mathfrak{m}_{\mathsf{X},p_5}^2$, $\mathfrak{m}_{\mathsf{X},p_6}$ when all points $p_i \in \mathsf{D}$ tend to p while remaining distinct. Pictorially, this corresponds to the fact that the Young diagram λ can be obtained by taking the Young diagrams



and piling their boxes onto one another horizontally. ◊

The importance of the subscheme Z_p^μ is that it gives a canonical representative of the corresponding orbit datum:

Proposition 2.19 *Let μ be a Young diagram. Then $[\mathsf{Z}_p^\mu] = \mu \in \mathsf{OrbDat}(\mathsf{D}, p)$.*

Proof Let $\mathsf{Z} = \mathsf{Z}_p^\mu$. In light of the above, $i^*\mathcal{I}_{\mathsf{Z},p}$ is determined up to isomorphism by its torsion subsheaf, which is isomorphic as an $\mathcal{O}_\mathsf{D}$-module to $\mathcal{T}or_1(\mathcal{O}_\mathsf{Z}, \mathcal{O}_\mathsf{D})$. In local coordinates (x, y) such that D is the vanishing locus of y, we can use the Koszul resolution of $\mathcal{O}_\mathsf{D}$ to compute

$$\mathcal{T}or_1(\mathcal{O}_\mathsf{Z}, \mathcal{O}_\mathsf{D}) \cong \ker(\mathcal{O}_\mathsf{Z} \xrightarrow{y\cdot} \mathcal{O}_\mathsf{Z}).$$

Using the basis of $\mathcal{O}_\mathsf{Z}$ indexed by the boxes of $\lambda := \mathrm{hc}(\mu)$ as above, the operator y corresponds to a translation upwards by one step. Therefore $\mathcal{T}or_1(\mathcal{O}_\mathsf{Z}, \mathcal{O}_\mathsf{D})$ has a $\mathbb{C}$-basis indexed by the boxes of λ that sit at the top of the columns, and these are in bijection with the boxes of μ, as illustrated in (2-14). Moreover, x restricts to a coordinate on D, identifying $\mathcal{O}_{\mathsf{D},p} \cong \mathbb{C}\{x\}$. The action of x on the basis corresponds to translation rightwards by one step in the Young diagram, and in this way we deduce that $\mathcal{T}or_1(\mathcal{O}_\mathsf{X}, \mathcal{O}_\mathsf{D})$ is isomorphic to the direct sum of the modules $\mathbb{C}\{x\}/(x^{\mu_j})$ for $j = 1, \dots, k$, as desired. □

Proposition 2.20 *If $\mathsf{Z} \in \mathsf{X}^{[n]}$ is any element whose orbit datum at p is the Young diagram μ, then Z_p^μ is a subscheme of Z. Moreover, the orbit of Z contains as a Zariski open dense subset the locus of elements of the form $\mathsf{Z}' \sqcup \mathsf{Z}_p^\mu$ such that $p \notin \mathsf{Z}'$.*

Proof Choose local coordinates (x, y) centred at p for which $\mathsf{D} = \{y = 0\}$. By assumption, the syzygy matrix for Z is equivalent to one of the form

$$S_\mathsf{Z} = S_\mu(x) + yM = \sum_{j=1}^{k} x^{\mu_j} E_j + yM,$$

where E_j is whose (j, j)-entry is 1 and whose other entries are zero, M is in $\mathcal{O}_{\mathsf{X},p}^{(k+1)\times k}$ and $S_\mu(x)$ is the Smith normal form matrix (2-13). Moreover Z is given locally by the vanishing of the maximal minors of S_Z. Using the skew-multilinearity of the minors of a matrix, we deduce that each maximal minor is a sum of terms the form $x^{\mu_{j_1}} \cdots x^{\mu_{j_l}} y^{k-l} \det(M_{j_1,\dots,j_l})$ for a $(k-l) \times (k-l)$ submatrix $M_{j_1,\dots,j_l}$ of M, where the indices $j_l, \dots, j_l$ are pairwise distinct. Since $\mu_k \le \mu_{k-1} \le \cdots$, each such term is divisible by $x^{\mu_k + \cdots + \mu_{k-l+1}} y^{k-l} = x^{\mathrm{hc}(\mu)_{l-1}} y^{k-l}$ and hence $\mathcal{I}_\mathsf{Z}$ is contained in the ideal defining Z_p^μ, or equivalently $\mathsf{Z}_p^\mu \subset \mathsf{Z}$, as desired.

For the second statement, first note that the Zariski open property follows by upper semicontinuity of the multiplicity of the point p; this upper semicontinuity follows from the continuity of the Hilbert–Chow morphism, and the fact that multiplicity is clearly upper semicontinuous on the base. So we only need to prove that the given locus is dense. The problem localizes around the points of Z, so we may assume without loss of generality that $\mathsf{X} = \mathbb{C}^2$; in this case D is anticanonical so by Corollary 3.5 below, the orbits are disjoint unions of symplectic leaves of one of Bottacin's Poisson structures on $\mathsf{X}^{[n]}$. Furthermore, we may assume that Z is supported at a point $p \in \mathsf{D}$, with $\mathsf{Z} \neq \mathsf{Z}_p^\mu$. Hence $\#\mathsf{Z} > \#\mathsf{Z}_p^\mu = |\mathrm{hc}(\mu)|$, so by Lemma 2.17 above and Lemma 2.22 below, the orbit of Z has positive dimension. Since the monomial ideals give isolated points in $\mathsf{X}^{[n]}$, we may assume in addition that the ideal defining Z is not generated by monomials. Then the vector field on $\mathsf{X}^{[n]}$ induced by the Euler vector field $E = \sum_{i=1}^n y_i \partial_{y_i}$ generating the standard $\mathbb{C}^*$ action in coordinates is nonzero at Z; see eg [29, Proposition 7.4]. Pick a locally defined function $g \in \mathcal{O}_{\mathsf{X}^{[n]}}$ such that $E(g) = 1$. Then we have $\{g, h\} = 1$, where h is the locally defined function on $\mathsf{X}^{[n]}$ sending Z to the sum of the x-coordinates of its support. Therefore, the Hamiltonian vector field of g pushes Z away from the locus where $\sum_{i=1}^n x_i = 0$. In particular, it pushes Z away from the set of elements of $(\mathbb{C}^2)^{[n]}$ supported at p. Since the flow cannot change the intersection with D, we deduce that Z is equivalent to a scheme of the form $\mathsf{Z}' \sqcup \mathsf{Z}''$, where $\mathsf{Z}' \neq \varnothing$ is disjoint from p and $\mathsf{Z}'' \supset \mathsf{Z}_p^\mu$. The result follows by downward induction on the length of Z. □

Remark 2.21 In the above proof, we cite Corollary 3.5 below. There is no circular logic because the above proposition is not used after the present subsection. Moreover Corollary 3.5 (and all of Section 3.2) do not require anything from the present Section 2.6. ◊

In light of (2-12) we have a (noncanonical) splitting $i^*\mathcal{I}_{\mathsf{Z},p} \cong i^*\mathcal{I}^{\mathrm{tf}}_{\mathsf{Z},p} \oplus \tau_{\mathsf{Z},\mathsf{D}}$, where

$$i^*\mathcal{I}^{\mathrm{tf}}_{\mathsf{Z},p} \cong \mathcal{O}_{\mathsf{D},p} \cong \mathbb{C}\{x\} \quad \text{and} \quad \tau_{\mathsf{Z},\mathsf{D},p} \cong \bigoplus_{j=1}^k \mathcal{O}_{\mathsf{D},p}/\mathfrak{m}_{\mathsf{D},p}^{\mu_j} \cong \bigoplus_{j=1}^k \mathbb{C}[x]/(x^{\mu_j})$$

are the torsion-free and torsion parts, respectively. It follows immediately that the natural map $\mathsf{G}_{\mathsf{Z},\mathsf{D}} \to \mathsf{Aut}_{\mathsf{D}}(\tau_{\mathsf{Z},\mathsf{D}})$ is a split surjection, with kernel $\mathsf{Hom}(i^*\mathcal{I}^{\mathrm{tf}}_{\mathsf{Z}}, \tau_{\mathsf{Z},\mathsf{D}}) \cong \mathsf{H}^0(\tau_{\mathsf{Z},\mathsf{D}})$, giving the following description of the stabilizer group:

Lemma 2.22 *Suppose $\mathsf{Z} \cap \mathsf{D}$ is supported in the smooth locus of D. Then the stabilizer group of Z in $(\mathsf{X},\mathsf{D})^{[n]}$ is given by*

$$\mathsf{G}_{\mathsf{Z},\mathsf{D}} \cong \mathsf{Aut}_{\mathsf{D}}(\tau_{\mathsf{Z},\mathsf{D}}) \ltimes \mathsf{H}^0(\tau_{\mathsf{Z},\mathsf{D}}),$$

where $\mathsf{H}^0(\tau_{\mathsf{Z},\mathsf{D}})$ is viewed as an abelian group under addition. In particular, the orbit of Z has codimension

$$\operatorname{codim}((\mathsf{X},\mathsf{D})^{[n]} \cdot \mathsf{Z}) = 2 \sum_{p \in \mathsf{Z}} |\mathrm{hc}(\mu(p))|,$$

where $\mu(p)$ is the Young diagram representing the orbit datum $[\mathsf{Z}]_p$ for all $p \in \mathsf{D}$.

Proof It remains to establish the formula for the codimension of the orbit, or equivalently the dimension of the Lie algebra $\mathfrak{g}_{\mathsf{Z},\mathsf{D}}$ of $\mathsf{G}_{\mathsf{Z},\mathsf{D}}$. But the latter is the sum of the dimensions of the vector spaces

$$\mathsf{H}^0(\tau_{\mathsf{Z},\mathsf{D}}) \cong \bigoplus_{j=1}^k \mathbb{C}[x]/(x^{\mu_j}) \quad \text{and} \quad \mathsf{End}_{\mathsf{D}}(\tau_{\mathsf{Z},\mathsf{D}}) \cong \bigoplus_{j,l=1}^k \mathsf{Hom}_{\mathbb{C}[x]}(\mathbb{C}[x]/(x^{\mu_j}), \mathbb{C}[x]/(x^{\mu_l})).$$

Note that $\dim \mathbb{C}[x]/(x^{\mu_j}) = \mu_j$, while $\dim \mathsf{Hom}_{\mathbb{C}[x]}(\mathbb{C}[x]/(x^{\mu_j}), \mathbb{C}[x]/(x^{\mu_l})) = \min\{\mu_j, \mu_l\}$. Therefore,

$$\dim \mathsf{G}_{\mathsf{G},\mathsf{Z}} = \sum_{j\geq 1} \mu_j + \sum_{j,l\geq 1} \min\{\mu_j,\mu_l\} = \sum_{j\geq 1}\mu_j + \sum_{j\geq 1}(2j-1)\mu_j = 2\sum_{j\geq 1} j\mu_j = 2|\mathrm{hc}(\mu)|,$$

by (2-15). □

Combined with Lemma 2.17, we deduce the following:

Corollary 2.23 *The zero-dimensional orbits of* $(\mathsf{X},\mathsf{D})^{[n]}$ *are exactly the elements of the form* $\mathsf{Z} = \bigsqcup_{p\in\mathsf{D}} \mathsf{Z}_p^{\mu(p)}$, *such that* $\sum_{p\in\mathsf{D}} |\mathrm{hc}(\mu(p))| = n$.

Combining these results, we have obtained the following description of the orbits in the case where D is smooth:

Theorem 2.24 *Suppose that* $\mathsf{D}\subset\mathsf{X}$ *is smooth. Then the orbits of* $(\mathsf{X},\mathsf{D})^{[n]}$ *are in bijection with functions* μ *from* D *to the set of all Young diagrams, such that*

$$\sum_{p\in\mathsf{D}} |\mathrm{hc}(\mu(p))| \leq n.$$

Moreover, for any such μ, *the set*

$$\left\{ \mathsf{Z}' \sqcup \bigsqcup_{p\in\mathsf{Z}} \mathsf{Z}_p^{\mu(p)} \;\middle|\; \mathsf{Z}'\subset \mathsf{X}\setminus\mathsf{D} \right\} \cong (\mathsf{X}\setminus\mathsf{D})^{[n-\sum_p |\mathrm{hc}(\mu(p))|]}$$

is open and dense in the corresponding orbit $\mathsf{X}_\mu^{[n]}$. *In particular, if* X *is connected, then all orbits are connected.*

Remark 2.25 In addition, one can show that the orbit $\mathsf{X}_\mu^{[n]}$ lies in the closure of the orbit $\mathsf{X}_{\tilde\mu}^{[n]}$ if and only if $\tilde\mu \geq \mu$ in the dominance order on Young diagrams. However the proof uses more refined information about the local structure of $\mathsf{X}^{[n]}$, so we postpone it; see Proposition 4.10. ◊

2.6.2 Nodes We now briefly describe the classification of the orbit data at a nodal singularity $p\in\mathsf{D}$. Choose coordinates (x,y) centred at p, so that $\mathsf{D}=\{xy=0\}$, giving an isomorphism

$$\mathcal{O}_{\mathsf{D},p} \cong \mathbb{C}\{x,y\}/(xy).$$

The classification of finitely generated modules over this ring is known: the torsion modules were classified in [14] (see also [4]), and the classification of all finitely generated modules over $\mathbb{C}[x,y]/(xy)$ was done in [25; 26], based on earlier works [30; 31]. In particular, every finitely generated module decomposes as a direct sum of indecomposable ones, and this decomposition can be determined algorithmically [24]. Furthermore, the minimal resolutions of the indecomposables are presented in [8, Appendix A]. Of these, only two families have projective dimension at most one, namely those listed as (1) and (3) on pages 224–225 of [8]:

- A "continuous series" of torsion modules, expressed as the cokernels of the square matrices

$$M(\boldsymbol{d},k,u) := \begin{pmatrix} x^{\nu_1} I_k & 0 & 0 & \cdots & y^{\mu_l} J_k(u) \\ y^{\mu_1} I_k & x^{\nu_2} I_k & 0 & \cdots & 0 \\ 0 & y^{\mu_2} I_k & x^{\nu_3} I_k & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y^{\mu_{l-1}} I_k & x^{\nu_l} I_k \end{pmatrix}$$

  of size $k \cdot l$, where $\boldsymbol{d} = (\nu_1,\mu_1),(\nu_2,\mu_2),\dots,(\nu_l,\mu_l)$ is a nonperiodic sequence of pairs of positive integers, I_k denotes the identity matrix of size k and $J_k(u)$ is a maximal Jordan block of size k with eigenvalue $u \neq 0$. We include here the degenerate case $l = 1$, in which case $M(\boldsymbol{d},k,u) = x^{\nu_1} I_k - y^{\mu_1} J_k(u)$.

- A "discrete series" of rank-one modules, expressed as the cokernels of the $(k+1)\times k$ matrices

$$\mathscr{D}(\boldsymbol{d}) := \begin{pmatrix} x^{\nu_1} & 0 & 0 & \cdots & 0 \\ y^{\mu_1} & x^{\nu_2} & 0 & \cdots & 0 \\ 0 & y^{\mu_2} & x^{\nu_3} & \cdots & 0 \\ 0 & 0 & y^{\mu_3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & x^{\nu_k} \\ 0 & 0 & 0 & \cdots & y^{\mu_k} \end{pmatrix},$$

  where $\boldsymbol{d} = (\mu_1,\nu_1),\dots,(\mu_k,\nu_k)$ is a sequence of pairs of positive integers. We include here the degenerate case $k = 0$, which corresponds to the trivial module.

Since for any $\mathsf{Z} \in \mathsf{X}^{[n]}$, the module $i^*\mathcal{I}_{\mathsf{Z},p}$ has rank one and projective dimension at most one, we have the following classification of the orbit data:

Proposition 2.26 *If p is a node, then each local orbit datum is presented as the direct sum of a single matrix of the form $\mathscr{D}(\boldsymbol{d})$ and at most finitely many matrices of the form $M(\boldsymbol{d},k,u)$. This presentation is unique up to permutation of the summands.*

Remark 2.27 By choosing suitable lifts of the direct sums of the matrices $\mathscr{D}(\boldsymbol{d})$ and $M(\boldsymbol{d},k,u)$ to matrices valued in $\mathcal{O}_{\mathsf{X},p}$, one can show that all direct sums as in Proposition 2.26 are realized as orbit data for sufficiently large n. It would be interesting to know the minimal n for which a given direct sum can occur. ◊

Remark 2.28 It is straightforward to check that the indecomposable rank-one modules in the discrete series above are determined up to isomorphism by their torsion submodules, which have infinite projective dimension and form the discrete series described in item (2) on page 225 of [8]. Hence, if the singularities of D are at worst nodal, the orbit type of an element $\mathsf{Z} \in \mathsf{X}^{[n]}$ is completely determined by the isomorphism class of the torsion submodule $\tau_{\mathsf{Z},\mathsf{D}}$ of $\mathcal{I}_{\mathsf{Z}}$. ◊

3 Symplectic and Poisson structures

3.1 Bottacin's Poisson structure

From now on, consider the particular case in which the divisor $\mathsf{D} \subset \mathsf{X}$ is anticanonical, and fix a nonzero section

$$\sigma \in \mathsf{H}^0(\mathcal{K}_\mathsf{X}^\vee)$$

vanishing on D. Thus σ is a Poisson structure on X having the open set $\mathsf{X} \setminus \mathsf{D}$ as a two-dimensional symplectic leaf, and the points of D as zero-dimensional symplectic leaves. In [5], Bottacin shows that for every $n \geq 0$, the induced Poisson structure $\sigma^{(n)}$ on the symmetric power $\mathsf{X}^{(n)}$ lifts along the Hilbert–Chow morphism to a Poisson structure on the Hilbert scheme $\mathsf{X}^{[n]}$, and that the corresponding bivector

$$\sigma^{[n]} \in \mathsf{H}^0(\wedge^2 \mathcal{T}_{\mathsf{X}^{[n]}})$$

has the following description: using the various identifications (2-1) of the tangent space at $\mathsf{Z} \in \mathsf{X}^{[n]}$, and the corresponding Serre dual descriptions of the cotangent space, the anchor map $(\sigma^{[n]})^\sharp : \mathsf{T}_\mathsf{Z}^*\mathsf{X}^{[n]} \to \mathsf{T}_\mathsf{Z}\mathsf{X}^{[n]}$ is determined by the following commutative diagram:

$$\begin{array}{ccccc} \mathsf{T}^*_\mathsf{Z}\mathsf{X}^{[n]} & \xrightarrow{\sim} & \mathsf{Ext}^1_\mathsf{X}(\mathcal{I}_\mathsf{Z},\mathcal{I}_\mathsf{Z}\otimes\mathcal{K}_\mathsf{X})_0 & \xrightarrow{\sim} & \mathsf{Hom}_\mathsf{X}(\mathcal{I}_\mathsf{Z},\mathcal{O}_\mathsf{Z}\otimes\mathcal{K}_\mathsf{X}) \\ \big\downarrow{\scriptstyle(\sigma^{[n]})^\sharp} & & \big\downarrow{\scriptstyle\sigma} & & \big\downarrow{\scriptstyle\sigma} \\ \mathsf{T}_\mathsf{Z}\mathsf{X}^{[n]} & \xleftarrow[\sim]{} & \mathsf{Ext}_1^\mathsf{X}(\mathcal{I}_\mathsf{Z},\mathcal{I}_\mathsf{Z})_0 & \xleftarrow[\sim]{} & \mathsf{Hom}_\mathsf{X}(\mathcal{I}_\mathsf{Z},\mathcal{O}_\mathsf{Z}) \end{array}$$

3.2 Symplectic groupoid and symplectic leaves

We now sketch an alternative approach to the construction of Bottacin's Poisson structure $\sigma^{[n]}$, based on the formalism of shifted symplectic structures from [34], and the now standard correspondence between Poisson structures, symplectic groupoids and shifted Lagrangian morphisms. This conceptual approach has the advantage of immediately yielding a description of the symplectic groupoid and symplectic leaves of $\sigma^{[n]}$, and is indeed how we arrived at the definition of the groupoid $(\mathsf{X},\mathsf{D})^{[n]}$. However, a posteriori one can construct the symplectic structure on the groupoid without invoking the theory of shifted symplectic structures; see Remark 3.4 below.

We will assume here that X is algebraic, since the cited references make this assumption. However, the constructions are effectively local in X, so that the result holds also in the nonalgebraic setting with essentially the same arguments.

First, we observe that if $\mathsf{D} \subset \mathsf{X}$ is an anticanonical divisor, the map $Li^* : \mathsf{Perf}(\mathsf{X}) \to \mathsf{Perf}(\mathsf{D})$ of derived stacks carries a one-shifted Lagrangian structure in a neighbourhood of any perfect complex on X with proper support. If X itself is proper, this can be obtained by viewing the anticanonical section σ as a relative orientation on the inclusion $\mathsf{D} \to \mathsf{X}$, in the sense of [9], and combining the main result of [9] with the 2-shifted symplectic structure on the stack $\mathsf{Perf}(-)$ of perfect complexes from [34]. More generally, one can invoke the results of [6]. We may then pull this Lagrangian structure back along the square

$$\begin{array}{ccc} \mathsf{Perf}_0(\mathsf{X}) & \longrightarrow & \mathsf{Perf}(\mathsf{X}) \\ \downarrow & & \downarrow \\ \mathsf{Perf}_0(\mathsf{D}) & \longrightarrow & \mathsf{Perf}(\mathsf{D}) \end{array}$$

thus obtaining a canonical isotropic structure on the map $\mathsf{Perf}_0(\mathsf{X}) \to \mathsf{Perf}_0(\mathsf{D})$ over the locus of complexes with proper support.

Lemma 3.1 *The induced isotropic structure on* $\mathsf{Perf}_0(\mathsf{X}) \to \mathsf{Perf}_0(\mathsf{D})$ *is nondegenerate*, *ie a Lagrangian structure.*

Since the construction of Lagrangian structures on mapping stacks is functorial for maps of the shifted symplectic structures on the target, the lemma is an immediate consequence of the definitions and the following basic fact:

Proposition 3.2 *The pullback of the* 2*-shifted symplectic structure on* $\mathsf{Perf}(-)$ *along the canonical map* $\mathsf{Perf}_0(-) \to \mathsf{Perf}(-)$ *is nondegenerate*, *and thus defines a* 2*-shifted symplectic structure on* $\mathsf{Perf}_0(-)$.

Proof Let $\mathsf{Pic}(-) = B\mathbb{G}_\mathrm{m}$ be the Picard stack. Then by definition

$$\mathsf{Perf}_0(-) \cong \mathsf{Perf}(-) \underset{\mathsf{Pic}(-)}{\times} *$$

is the fibre of the derived determinant map $\det\colon \mathsf{Perf}(-) \to \mathsf{Pic}(-)$ from [43] at the point $* \to \mathsf{Pic}(-)$ corresponding to the universal trivial line bundle. Let $x\colon \mathsf{Spec}(A) \to \mathsf{Perf}_0(-)$ be the point classifying a perfect complex $\mathcal{E}$ on $\mathsf{Spec}(A)$ with trivialized determinant. By [43, Proposition 3.2], we have a homotopy fibre sequence

$$x^*\mathbb{T}\mathsf{Perf}_0(-) \to \mathcal{REnd}_A(\mathcal{E}) \xrightarrow{\mathrm{tr}} A$$

of complexes on X, where tr is the Illusie trace map, so that

$$x^*\mathbb{T}\mathsf{Perf}_0(-) \cong \mathcal{REnd}_A(\mathcal{E})_0 := \mathrm{fibre}(\mathcal{REnd}_A(\mathcal{E}) \xrightarrow{\mathrm{tr}} A)$$

is the complex of "traceless" derived endomorphisms. Meanwhile, as constructed in [34, Section 2.3], the two-form underlying the 2-shifted symplectic structure on $\mathsf{Perf}(-)$ is given by the trace pairing

$$\omega\colon \mathcal{REnd}_A(\mathcal{E}) \overset{L}{\underset{A}{\otimes}} \mathcal{REnd}_A(\mathcal{E}) \to \mathcal{REnd}_A(\mathcal{E}) \xrightarrow{\mathrm{tr}} A.$$

Hence the induced two-form on $\mathsf{Perf}_0(-)$ is given by the restriction of the pairing ω to the traceless endomorphisms $\mathcal{REnd}_A(\mathcal{E})_0$. But by definition, every element of $\mathsf{Perf}(A)$ is equivalent to a strictly perfect one, ie a dg module $\mathcal{E}$ whose underlying module is projective of finite type, ie a graded vector bundle. Hence the statement reduces to the classical nondegeneracy of the trace pairing on traceless matrices. □

Restricting to the open substack $\mathsf{X}^{[n]} \subset \mathsf{Perf}_0(\mathsf{X})$, we deduce that the map $\mathsf{X}^{[n]} \to \mathsf{Perf}_0(\mathsf{D})$ is one-shifted Lagrangian, so that our groupoid

$$(\mathsf{X},\mathsf{D})^{[n]} \cong \mathsf{X}^{[n]} \underset{\mathsf{Perf}_0(\mathsf{D})}{\times} \mathsf{X}^{[n]}$$

carries a zero-shifted symplectic structure by the Lagrangian intersection theorem [34, Theorem 2.9]. But by Theorem 2.9, the derived stack $(\mathsf{X},\mathsf{D})^{[n]}$ is a smooth classical scheme, so by [34, pages 297–298] this 0-shifted symplectic structure is an ordinary symplectic structure in the classical sense. Moreover, by definition, this symplectic structure on $(\mathsf{X},\mathsf{D})^{[n]}$ is the difference of the pullbacks of the isotropic structure on $\mathsf{X}^{[n]}$ along the two maps $s,t\colon (\mathsf{X},\mathsf{D})^{[n]} \to \mathsf{X}^{[n]}$. It is therefore automatically multiplicative, so that $(\mathsf{X},\mathsf{D})^{[n]}$ has the structure of a symplectic groupoid. As for any symplectic groupoid, the embedding of the identity arrows $\mathsf{X}^{[n]} \hookrightarrow (\mathsf{X},\mathsf{D})^{[n]}$ is Lagrangian, giving an identification $\mathbb{L}\mathsf{X}^{[n]}|_{\mathsf{Z}} \cong \mathbb{T}s|_{(\mathsf{Z},\mathsf{Z},\mathrm{id})}$. The induced Poisson bivector on $\mathsf{X}^{[n]}$ is the composition of this isomorphism with the differential $\mathbb{T}s|_{(\mathsf{Z},\mathsf{Z},\mathrm{id})} \to \mathbb{T}\mathsf{X}^{[n]}|_{\mathsf{Z}}$ of the target map. Considering the definition of the symplectic structure in terms of Serre duality and using the identifications of the (relative) tangent spaces from the proof of Theorem 2.9, we deduce that this Poisson structure is exactly the one defined by Bottacin.

In summary, we have the following:

Theorem 3.3 *Let (X,σ) be a Poisson surface with degeneracy curve $\mathsf{D} = \sigma^{-1}(0) \subset \mathsf{X}$. Then $(\mathsf{X},\mathsf{D})^{[n]}$ carries a canonical symplectic structure, making it a symplectic groupoid over $\mathsf{X}^{[n]}$ that integrates Bottacin's Poisson structure $\sigma^{[n]}$ on $\mathsf{X}^{[n]}$.*

Remark 3.4 The symplectic structure on $(\mathsf{X},\mathsf{D})^{[n]}$ constructed above is the unique two-form for which the map $(s,t)\colon (\mathsf{X},\mathsf{D})^{[n]} \to \mathsf{X}^{[n]} \times \mathsf{X}^{[n]}$ is a symplectomorphism over the open dense set $(\mathsf{X}\setminus\mathsf{D})^{[n]} \times (\mathsf{X}\setminus\mathsf{D})^{[n]}$, where the latter is equipped with the symplectic form $\Omega := \mathrm{pr}_1^*\,\omega - \mathrm{pr}_2^*\,\omega$ obtained by pulling back the symplectic form ω on $\mathsf{X}\setminus\mathsf{D}^{[n]}$ along the two projections $\mathrm{pr}_1,\mathrm{pr}_2$. Thus, to prove that $(\mathsf{X},\mathsf{D})^{[n]}$ is symplectic, one simply needs to check that the pullback of Ω to $(\mathsf{X},\mathsf{D})^{[n]}$ has no poles, and similarly that the pullback of Ω^n has no zeroes.

These properties can be verified directly in at least two ways, giving a construction of the symplectic form that does not rely on the theory of shifted symplectic structures. The first way is to explicitly write down the induced nondegenerate pairing on the tangent space at any point of $(\mathsf{X},\mathsf{D})^{[n]}$ using Grothendieck duality for coherent sheaves, and see directly that it agrees with the pullback of Ω. The second way begins with the observation that the desired properties are local and can be checked in codimension one. Thus, using Weinstein's splitting theorem, one can reduce the problem to the case in which $n=1$, and the reduced curve underlying D is smooth. In this case, the desired properties can be obtained by direct calculation using the explicit description of $(\mathsf{X},\mathsf{D})^{[1]} \to \mathsf{X}\times\mathsf{X}$ in terms of blowups from Example 2.10, in the spirit of [18, Section 2.4], which treats the reduced case. ◊

Since the connected components of the orbits of any symplectic groupoid are exactly the symplectic leaves of the corresponding Poisson structure, and moreover the orbits of any smooth algebraic groupoid are locally closed algebraic subvarieties, Theorem 3.3 has the following immediate consequences.

Corollary 3.5 *The symplectic leaves of* $(\mathsf{X}^{[n]}, \sigma^{[n]})$ *are the connected components of the* D*-equivalence classes. In particular, they are locally closed in the Zariski topology when* (X, σ) *is algebraic, and if* D *is smooth they are in bijection with Young-diagram-valued functions on* D *as in Theorem 2.24.*

Note that the stabilizer groups in $(\mathsf{X}, \mathsf{D})^{[n]}$ are exactly the self D-equivalences, and meanwhile, as for any symplectic groupoid, the Lie algebras of these groups are exactly the conormal Lie algebras of the Poisson structure (ie the linearization of the Poisson structure in the direction transverse to the symplectic leaves) so we have the following:

Corollary 3.6 *If* $\mathsf{Z} \in \mathsf{X}^{[n]}$ *is any point, then its conormal Lie algebra*

$$\ker\big((\sigma^{[n]})^\sharp \colon \mathsf{T}^*_\mathsf{Z}\mathsf{X}^{[n]} \to \mathsf{T}_\mathsf{Z}\mathsf{X}^{[n]}\big)$$

is canonically identified with the Lie algebra $\mathfrak{g}_{\mathsf{Z},\mathsf{D}}$ *of infinitesimal self-*D*-equivalences defined in Definition 2.11 and described in Lemma 2.12.*

3.3 Characteristic leaves

Of particular interest to us are the symplectic leaves that are "characteristic" in the sense of [27]; let us recall the definition. First, recall from [7; 35; 47] the notion of the modular vector field of a Poisson structure σ on a manifold W with respect to a volume form μ: it is the derivation of $\mathcal{O}_\mathsf{W}$ sending a function f to the μ-divergence of the Hamiltonian vector field of f. The modular vector field is an infinitesimal symmetry of the Poisson structure, and is independent of μ up to the addition of a Hamiltonian vector field, so that its projection to the normal space of any symplectic leaf is unambiguously defined. We then have the following.

Definition 3.7 [27] A symplectic leaf of a Poisson manifold is *characteristic* if it is preserved by the flow of the modular vector field, ie the projection of the modular vector field to the normal bundle of the leaf is identically zero.

Now suppose that (X, σ) is a Poisson surface, and let $\mathsf{D} \subset \mathsf{X}$ be the anticanonical divisor on which σ vanishes. Then the restriction of the modular vector field to D is independent of the choice of volume form, giving a canonically defined global vector field

$$\zeta \in \mathsf{H}^0(\mathcal{T}_\mathsf{D}).$$

The characteristic leaves in the Hilbert scheme then have the following description.

Proposition 3.8 *The symplectic leaf of* $\sigma^{[n]}$ *through a point* $\mathsf{Z} \in \mathsf{X}^{[n]}$ *is characteristic if and only if the derived subscheme* $\mathsf{Z} \cap^h \mathsf{D} \subset \mathsf{D}$ *is preserved (up to equivalence) by the flow of* ζ*.*

Proof The problem is local in X around Z, so we may assume without loss of generality that X admits a global volume form μ and let $\tilde{\zeta}$ be the corresponding modular vector field, so that $\tilde{\zeta}|_\mathsf{D} = \zeta$ by definition of ζ. Consider the product X^n and let $p_i : \mathsf{X}^n \to \mathsf{X}$ be the i^{th} projection. For a tensor ξ

on X let $\xi^n = p_1^*\xi + \cdots + p_n^*\xi$ denote its symmetric lift to X^n. Then σ^n, μ^n and $\widetilde{\zeta}^n$ are, respectively, a Poisson structure, a volume form, and the corresponding modular vector field on X^n. These are invariant under permutation, and hence descend to the symmetric power $\mathsf{X}^{(n)}$, giving a Poisson structure $\sigma^{(n)}$, a trivialization $\mu^{(n)}$ of the canonical sheaf, and a vector field $\widetilde{\zeta}^{(n)}$, which agrees with the modular vector field over the smooth locus of $\mathsf{X}^{(n)}$. We claim that these structures lift to corresponding structures on the Hilbert scheme under the Hilbert–Chow morphism $\mathsf{X}^{[n]} \to \mathsf{X}^{(n)}$. Indeed, $\sigma^{(n)}$ lifts to Bottacin's structure $\sigma^{[n]}$ by definition, and $\mu^{(n)}$ lifts to a volume form $\mu^{[n]}$ since the Hilbert–Chow morphism is crepant. We may then form the modular vector field $\widetilde{\zeta}^{(n)}$ of $\sigma^{[n]}$ with respect to $\mu^{[n]}$, which must agree with $\widetilde{\zeta}^{(n)}$ over the smooth locus of $\mathsf{X}^{(n)}$, and hence it must be a lift of $\widetilde{\zeta}^{(n)}$ (which is necessarily unique). It follows that the modular flow on $\mathsf{X}^{[n]}$ is given by pulling back subschemes $\mathsf{Z} \subset \mathsf{X}$ along the flow of $\widetilde{\zeta}$ on X. The result now follows immediately, since $\widetilde{\zeta}|_{\mathsf{D}} = \zeta$. □

Remark 3.9 Let us sketch an interpretation of this result in terms of derived symplectic geometry. We first remark, following a discussion with P Safronov, that the modular vector field of a Poisson manifold can be thought of as the "Hamiltonian vector field of the first Chern class of the leaf space", in the following sense. If Y is a one-shifted symplectic stack, then the symplectic form induces an isomorphism $\mathbb{T}_{\mathsf{Y}} \cong \mathbb{L}_{\mathsf{Y}}[1]$. Under the induced isomorphism $\mathsf{H}^1(\mathbb{L}_{\mathsf{Y}}) \cong \mathsf{H}^0(\mathbb{L}_{\mathsf{Y}}[1]) \cong \mathsf{H}^0(\mathbb{T}_{\mathsf{Y}})$, the first Chern class $c_1(\mathsf{Y}) \in \mathsf{H}^1(\mathbb{L}_{\mathsf{Y}})$ corresponds to a canonical vector field $\zeta \in \mathsf{H}^0(\mathbb{T}_{\mathsf{Y}})$ (its "Hamiltonian vector field"). If $p\colon \mathsf{W} \to \mathsf{Y}$ is a Lagrangian morphism, then W inherits a 0-shifted Poisson structure whose symplectic leaves are the fibres of p, and the vector field ζ agrees up to a constant factor with the projection of the modular vector field to the leaf space. In particular, the characteristic symplectic leaves in W correspond to fixed points of ζ in Y.

In the case at hand, we have $\mathsf{Y} = \mathsf{Perf}_0(\mathsf{D})$, and $\mathbb{T}_{\mathsf{Y}}$ is the complex of traceless derived endomorphisms of the universal sheaf. One can then compute the first Chern class of Y using the Grothendieck–Riemann–Roch theorem, and show that the corresponding vector field ζ generates the one-parameter group of automorphisms of Y given by pulling back complexes along the modular vector field on the curve D as above, recovering Proposition 3.8. ◊

Corollary 3.10 *Let X be a Poisson surface with anticanonical divisor D. If $\mathsf{Z} \in \mathsf{X}^{[n]}$ lies on a characteristic symplectic leaf, then $\mathsf{Z} \cap \mathsf{D}$ is contained in the singular locus of D.*

Proof The modular vector field is nonzero at any smooth point of D. Hence if the support of $\mathsf{Z} \cap \mathsf{D}$ contains a smooth point, it will not be fixed by the flow of the modular vector field. □

Lemma 3.11 *If X is connected, the codimension-two characteristic leaves are exactly the subvarieties of the form $\{\mathsf{Z} \in \mathsf{X}^{[n]} \mid \mathsf{Z} \cap \mathsf{D} = \{p\}$ as schemes$\}$, where p is a singular point of D.*

Proof If Z lies on a characteristic leaf of codimension two, then by Corollary 3.10, $\mathsf{Z} \cap \mathsf{D}$ must be contained in the singular locus of D. If $\#(\mathsf{Z} \cap \mathsf{D}) > 1$, then the symplectic leaf through Z has codimension greater than two, so can assume that $\mathsf{Z} \cap \mathsf{D} = \{p\}$ for some singular point $p \in \mathsf{D}$. Since $p \in \mathsf{D}$ is a singular point, it follows that Z itself is reduced at p, and hence $\mathsf{Z} = \{p\} \sqcup \mathsf{Z}'$, where $\mathsf{Z}' \subset \mathsf{X} \setminus \mathsf{D}$. Hence the leaf is exactly an embedded copy of the symplectic variety $(\mathsf{X} \setminus \mathsf{D})^{[n-1]}$. The transverse germ is isomorphic to the germ of X at p, and hence any such subvariety is a characteristic leaf. □

Proposition 3.12 *If* D *has only nodal singularities, then each germ of* $\mathsf{X}^{[n]}$ *has only finitely many characteristic symplectic leaves.*

Proof The problem is local and invariant under rescaling the Poisson structure by a constant, so by the log Darboux theorem it is enough to treat the case in which $\mathsf{X} = \mathbb{C}^2$ and $\sigma = xy\,\partial_x \wedge \partial_y$ (eg see [1]), in which case the modular vector field is given by $\zeta = y\partial_y - x\partial_x$ and its flow integrates to the $\mathbb{C}^*$-action on X given by $t \cdot (x, y) = (tx, t^{-1}y)$. It is straightforward to verify that this action preserves the equivalence class of the matrices $\mathscr{D}(\boldsymbol{d})$, and changes the equivalence classes of the matrices $M(\boldsymbol{d}, k, u)$ by a suitable rescaling of u. Hence according to Proposition 2.26, the orbit data for $\mathsf{X}^{[n]}$ that are fixed by the modular flow are exactly those given by a single matrix from the series $\mathscr{D}(\boldsymbol{d})$, of which only finitely many can occur in $\mathsf{X}^{[n]}$. Hence there are only finitely many characteristic leaves globally in $\mathsf{X}^{[n]}$. But the leaves are algebraic subvarieties (locally closed in the Zariski topology), so the analytic germ at every point can have only finitely many connected components (bounded by the number of irreducible components in the analytic or formal germ of the leaf closure), as desired. □

For $n = 1$, the characteristic leaves are given by the open leaf $\mathsf{X} \setminus \mathsf{D}$ and the singular points of D, and hence are locally finite as long as D is reduced. However if D has nonnodal singularities and n is sufficiently large, one finds the existence of infinitely many characteristic symplectic leaves in $\mathsf{X}^{[n]}$. Recall that a point $p \in \mathsf{D}$ is a *double point* if the multiplicity of D at p is exactly two, ie the defining equation vanishes to order exactly two, in which case it is an A_k singularity for some $k \geq 1$, meaning that there exists coordinates such that D is given locally by the equation

$$y^2 - x^{k+1} = 0.$$

We have the following:

Proposition 3.13 *Let* X *be a Poisson surface with anticanonical divisor* D. *Then the following statements hold*:

(1) $\mathsf{X}^{[2]}$ *has finitely many characteristic symplectic leaves if and only if the only singularities of* D *are double points.*

(2) *If* D *has a nonnodal singularity, then* $\mathsf{X}^{[n]}$ *has infinitely many characteristic leaves for every* $n \geq n(\mathsf{D})$, *where*

$$n(\mathsf{D}) := \begin{cases} 2 & \text{if } \mathsf{D} \text{ has a triple point}, \\ 3 & \text{if } \mathsf{D} \text{ has a singularity of type } A_k \text{ for some } k \geq 3, \\ 6 & \text{if } \mathsf{D} \text{ has an } A_2 \text{ singularity}. \end{cases}$$

Proof Suppose $\mathsf{Z} \in \mathsf{X}^{[2]}$. If Z is reduced, then $\mathsf{Z} = \{p_1, p_2\}$ is a pair of distinct points and $\mathsf{X}^{[2]}$ is locally isomorphic to a neighbourhood of (p_1, p_2) in the product $\mathsf{X} \times \mathsf{X}$, and hence there are only finitely many characteristic leaves in a neighbourhood of Z.

If Z is not reduced, then it is the image of an embedding $\mathsf{Spec}(\mathbb{C}[\epsilon]/\epsilon^2) \to \mathsf{X}$ and is therefore classified by a pair (p, L) where $p \in \mathsf{X}$ and $\mathsf{L} < \mathsf{T}_p\mathsf{X}$ is a line. By Corollary 3.10 we need only consider the case in which p is a singular point of D, in which case Z is contained entirely in D, so that σ vanishes at Z. Let us choose local coordinates x, y on X centred at p so that the bivector on X has the form $\sigma = f(x, y)\, \partial_x \wedge \partial_y$. The modular vector field is then given in terms of the partial derivatives f_x, f_y by the formula

$$\zeta = f_x \partial_y - f_y \partial_x$$

and hence the linearization of ζ at p is given by the matrix

$$\begin{pmatrix} -f_{xy}(0) & -f_{yy}(0) \\ f_{xx}(0) & f_{xy}(0) \end{pmatrix}$$

If f vanishes to order three or more, then this matrix is zero. Hence every line $\mathsf{L} < \mathsf{T}_p\mathsf{X}$ is preserved by the flow of ζ, and we obtain infinitely many characteristic leaves in $\mathsf{X}^{[2]}$, and hence also in $\mathsf{X}^{[n]}$ for $n > 2$. On the other hand, if f vanishes to order two, this matrix is nonzero. Since it is traceless, it preserves at most two lines in $\mathsf{T}_p\mathsf{X}$ and so the set of characteristic leaves with support at p is finite.

Now suppose that D has an A_k-singularity at $p \in \mathsf{X}$. Then we may choose coordinates (x, y) centred at p such that $f(x, y) = h(x, y)(y^2 - x^{k+1})$ and $h(0, 0) \neq 0$. Then the modular vector field of π is given by

$$\zeta = h(x, y)\big(-2y\partial_x - (k+1)x^k\partial_y\big) + (y^2 - x^{k+1})\big(h_x(x, y)\partial_y - h_y(x, y)\partial_x\big).$$

Suppose first that $k \geq 3$, and consider the ideal $\mathcal{I}_a := (y + ax^2, x^3)$, where $a \in \mathbb{C}$. Then each $\mathcal{I}_a$ defines a point in $\mathsf{X}^{[3]}$ and contains the ideal $\mathcal{I}_{\mathsf{D}} = (y^2 - x^{k+1})$ defining D. Therefore, multiplication by f annihilates $\mathrm{Hom}(\mathcal{I}_a, \mathcal{O}_{\mathsf{X}}/\mathcal{I}_a)$, and so $\sigma^{[3]}$ vanishes at $\mathcal{I}_a$. Moreover, the modular vector field ζ sends each generator of $\mathcal{I}_a$ to another element of $\mathcal{I}_a$. Therefore, the vector field $\zeta^{[3]}$ on $\mathsf{X}^{[3]}$ vanishes at $\mathcal{I}_a$, so that the singleton $\{\mathcal{I}_a\}$ is a characteristic leaf for all $a \in \mathbb{C}$.

Finally, assume $k = 2$. Then we can assume $h(x, y) = 1$ in the expression for σ above by a theorem of Arnol'd [1]. For $a \in \mathbb{C}$, consider the ideal $\mathcal{J}_a = (y^2 + ax^3, x^2y, xy^2)$, which defines an element of $\mathsf{X}^{[6]}$. We claim that multiplication by $f = y^2 - x^3$ annihilates each element $\varphi \in \mathrm{Hom}(\mathcal{J}_a, \mathcal{O}_{\mathsf{X}}/\mathcal{J}_a)$ for each $a \in \mathbb{C}$. Denoting the generators by $g_1 = y^2 + ax^3$, $g_2 = x^2y$, $g_3 = xy^2$, we have

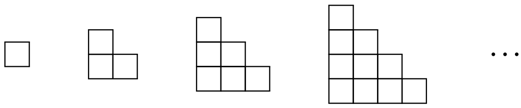
$$y\varphi(g_2) - x\varphi(g_3) = 0 \quad \text{and} \quad -xy\varphi(g_1) + ax^2\varphi(g_2) + y\varphi(g_3) = 0.$$

It is easy to deduce from these equations that the image of any such φ lies inside $(x, y)\mathcal{O}_{\mathsf{X}}/\mathcal{J}_a$, which is clearly annihilated by multiplication by f. Therefore, the Poisson tensor $\sigma^{[6]}$ vanishes at $\mathcal{J}_a$. Moreover, the vector field $\zeta = -2y\partial_x - 3x^2\partial_y$ sends each generator of $\mathcal{J}_a$ to another element of $\mathcal{J}_a$. Therefore, $\zeta^{[6]}$ vanishes at $\mathcal{J}_a$, so that the singleton $\{\mathcal{J}_a\}$ is a characteristic symplectic leaf for each $a \in \mathbb{C}$. □

4 Local models and holonomicity

4.1 Coordinates on the Hilbert scheme

There are two well-known methods for constructing coordinate charts on the Hilbert schemes, indexed by Young diagrams: the Haiman coordinates [19], and the Ellingsrud–Strømme coordinates [12]. In this section we use these coordinate systems to describe the local behaviour of Bottacin's Poisson structures. Of particular relevance to us are the charts associated to the "triangular" Young diagrams



which are defined as follows.

Choose a point $p \in \mathsf{X}$ and an integer $k > 0$, and let $p(k)$ be the k^{th} order neighbourhood of p, ie the vanishing locus of the ideal $\mathfrak{m}_p^{k+1} < \mathcal{O}_{\mathsf{X}}$ where $\mathfrak{m}_p$ is the maximal ideal corresponding to p. In local coordinates (x, y) on an open set U centred at p, the scheme $p(k)$ is expressed as the vanishing locus of the monomials $x^j y^l$ where $j + l \geq k$, so that $\mathcal{O}_{p(k),p}$ has a basis given by the monomials $x^j y^l$ with $j + l < k$. The latter are in bijection with the boxes of the triangular Young diagram

$$\triangle := (k > k-1 > k-2 > \cdots)$$

so that

$$p(k) \in \mathsf{X}^{[n]},$$

where

$$n := |\triangle| = \tfrac{1}{2}k(k+1)$$

is the size of $\triangle$. The point $p(k) \in \mathsf{X}^{[n]}$ then has an open neighbourhood $\mathsf{U}_{\triangle} \subset \mathsf{U}^{[n]} \subset \mathsf{X}^{[n]}$ given by

$$\mathsf{U}_{\triangle} := \{\mathsf{Z} \in \mathsf{U}^{[n]} \mid \text{the monomials } (x^j y^l)|_{\mathsf{Z}} \text{ for } j + l < k \text{ form a basis for } \mathsf{H}^0(\mathcal{O}_{\mathsf{Z}})\} \subset \mathsf{X}^{[n]},$$

and the two coordinate systems on these charts are defined as follows.

- **Ellingsrud–Strømme coordinates** If $E \in \mathbb{C}^{(k+1)\times k}$ is a $(k+1) \times k$ matrix, let

(4-1)
$$S_E(x,y) := E + \begin{pmatrix} -x & 0 & 0 & \cdots & 0 \\ y & -x & 0 & \cdots & 0 \\ 0 & y & -x & \cdots & 0 \\ 0 & 0 & y & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & -x \\ 0 & 0 & 0 & \cdots & y \end{pmatrix} = E - x\begin{pmatrix} I_k \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ I_k \end{pmatrix},$$

  and define a subscheme $\mathsf{Z}(E) \subset \mathsf{X}$ by the formula

$$\mathsf{Z}(E) := \text{vanishing locus of the maximal minors of } S_E(x,y).$$

Then the map $E \mapsto \mathsf{Z}(E)$ identifies a neighbourhood of the origin in $\mathbb{C}^{(k+1)\times k}$ with $\mathsf{U}_{\triangleright}$. Note that $S_E(x,y)$ is a Hilbert–Burch syzygy matrix for $\mathsf{Z}(E)$. The functions

$$E_i^j : \mathsf{U}_{\triangleright} \to \mathbb{C} \quad \text{for } 0 \le i \le k-1 \text{ and } 0 \le j \le k$$

sending $\mathsf{Z}(E)$ to the (j,i)-entry of the corresponding matrix E are the Ellingsrud–Strømme coordinates on $\mathsf{U}_{\triangleright}$. Note that here we start numbering of rows and columns from 0.

- **Haiman coordinates** For every $\mathsf{Z} \in \mathsf{U}_{\triangleright}$, the ideal $\mathcal{I}_{\mathsf{Z}}$ is generated by functions of the form

$$f_{\mathsf{Z},j} = x^j y^{k-j} - \sum_{a+b \le k-1} c_{ab}^j(\mathsf{Z}) x^a y^b \quad \text{for } 0 \le j \le k, \tag{4-2}$$

where c_{ab}^j are certain regular functions on $\mathsf{U}_{\triangleright}$ called Haiman's functions. Haiman's coordinates are then given by the collection

$$C_i^j := c_{i,k-1-i}^j \quad \text{for } 0 \le i \le k-1 \text{ and } 0 \le j \le k.$$

These coordinate systems are related by the linear transformations

$$E_i^j = C_j^{i+1} - C_{j-1}^i \qquad \text{and} \qquad C_i^j = \begin{cases} E_{j-1}^i + E_{j-2}^{i-1} + \cdots + E_{j-i-1}^0 & \text{if } j \ge i+1, \\ -E_j^{i+1} - E_{j+1}^{i+2} - \cdots - E_{j+k-i-1}^k & \text{if } j \le i, \end{cases}$$

where we use the convention that $C_{-1}^j = C_k^j = 0$ for any j. One way to explicitly verify the second transformation is by expanding the maximal minors of $S_E(x,y)$, considering only the linear term in E (equivalently, total degree $k-1$ in x, y).

Remark 4.1 Each set of coordinates E_i^j and C_i^j splits into two halves: $j \le i$ and $j \ge i+1$. In the formula above, the E-coordinates in one half are expressed in terms of the C-coordinates in the other half, and vice versa. ◊

Remark 4.2 In the particular case where $\mathsf{X} = \mathbb{C}^2$ and (x,y) are the standard coordinates, these charts are defined on the whole space of $(k+1)\times k$-matrices, giving open embeddings $\mathbb{C}^{(k+1)\times k} \hookrightarrow (\mathbb{C}^2)^{[n]}$ for all $k > 0$. ◊

An important feature of these coordinate charts is that their construction is $\mathbb{C}^\times$-equivariant, in the following sense. Define an action of $\mathbb{C}^\times$ on the germ of X at p by dilation of the coordinates we chose:

$$u \cdot (x,y) = (ux, uy) \quad \text{for } u \in \mathbb{C}^\times.$$

By functoriality, we have an induced action on the germ of $\mathsf{X}^{[n]}$ at $p(k)$.

Lemma 4.3 *The Ellingsrud–Strømme and Haiman coordinates have weight one with respect to the induced $\mathbb{C}^\times$-action, ie*

$$u \cdot E_i^j = uE_i^j \quad \text{and} \quad u \cdot C_i^j = uC_i^j$$

for all indices i, j and all $u \in \mathbb{C}^\times$.

Proof Since the two coordinate systems are related by a linear transformation, it suffices to prove the statement for one of them; we do so for the Ellingsrud–Strømme coordinates. For this, note that

$$S_E(ux,uy) = E - ux\begin{pmatrix} I_k \\ 0 \end{pmatrix} + uy\begin{pmatrix} 0 \\ I_k \end{pmatrix} = u\left(u^{-1}E - x\begin{pmatrix} I_k \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ I_k \end{pmatrix}\right) = uS_{u^{-1}E}(x,y).$$

Hence the ideals generated by the minors of the matrices $S_E(ux,uy)$ and $S_{u^{-1}E}(x,y)$ are the same, which implies that the Ellingsrud–Strømme coordinates scale linearly, as desired. □

4.2 Expression for the Poisson structure

Our interest in the triangular charts stems, in part, from the fact that they exhibit all possible local behaviours of Bottacin's Poisson structures. Assume as before that σ is a nonzero Poisson structure on a surface X, vanishing on the anticanonical divisor D. Recall from [27] that two germs W_1 and W_2 of Poisson manifolds are *stably equivalent* if there exist symplectic germs S_1 and S_2 such that $\mathsf{W}_1 \times \mathsf{S}_1 \cong \mathsf{W}_2 \times \mathsf{S}_2$. Then we have the following:

Lemma 4.4 *Let (X,σ) be a Poisson surface as above and suppose $\mathsf{Z} \in \mathsf{X}^{[m]}$ for some $m \geq 0$. Then the germ of $(\mathsf{X}^{[m]},\sigma^{[m]})$ at Z is stably equivalent to a product of germs of points in the triangular charts of Hilbert schemes $\mathsf{X}^{[k(k+1)/2]}$ for $k \geq 0$.*

Proof Any element $\mathsf{Z} \in \mathsf{X}^{[m]}$ has an analytic neighbourhood consisting of products of neighbourhoods of its connected components, so we may assume without loss of generality that Z is supported at a single point $p \in \mathsf{D}$. It suffices to show that there exists an integer $k > 0$ and a subscheme $\mathsf{Z}' \subset \mathsf{X} \setminus \mathsf{D}$ such that $\mathsf{Z} \sqcup \mathsf{Z}'$ lies in the triangular chart $\mathsf{U}_{\triangleright} \subset \mathsf{X}^{[k(k+1)/2]}$, since in this case the germ at Z' is symplectic and hence the germs at Z and at $\mathsf{Z} \sqcup \mathsf{Z}'$ are stably equivalent.

To see this choose local coordinates on an open set U centred at p and choose k large enough that $\mathsf{H}^0(\mathcal{O}_\mathsf{Z})$ is spanned by the images of the monomials $A = \{x^j y^l \mid j+l < k\} \subset \mathsf{H}^0(\mathcal{O}_\mathsf{U})$. Now choose a collection of distinct points $p_1, p_2, \ldots \in \mathsf{U} \setminus (\mathsf{U} \cap \mathsf{D})$ such that the evaluation maps $\mathrm{ev}_{p_j} : \mathsf{H}^0(\mathcal{O}_\mathsf{U}) \to \mathbb{C}$ form a basis for $\mathrm{span}(A)^\vee$. After reordering, we can assume by linear independence that $\mathrm{ev}_{p_1}, \ldots, \mathrm{ev}_{p_{k-m}}$ restrict to a basis of $\bigl(\ker(\mathrm{span}(A) \to \mathsf{H}^0(\mathcal{O}_\mathsf{Z}))\bigr)^\vee = \mathrm{coker}\bigl(\mathsf{H}^0(\mathcal{O}_\mathsf{Z})^\vee \to \mathrm{span}(A)^\vee\bigr)$, and then $\mathsf{Z}' = \{p_1, \ldots, p_{k-m}\}$ is the desired subscheme. □

An explicit expression for the Poisson bracket can be obtained via the following algorithm:

(1) First obtain a formula in the case of the Poisson structure $\sigma_0 := \partial_x \wedge \partial_y$; we do so in Section 4.3 below.

(2) Define the *recursion operators* J_x, J_y to be the endomorphisms of the tangent bundle given by the action of x, y on the tangent spaces under the identification

$$\mathsf{T}_\mathsf{Z}\mathsf{U}_{\triangleright} \cong \mathsf{Hom}_\mathsf{U}(\mathcal{I}_\mathsf{Z}, \mathcal{O}_\mathsf{Z}).$$

Note that these operators have weight one, which implies they are linear in Ellingsrud–Strømme or Haiman coordinates. Explicitly, they are given as follows (we postpone their derivation to Section 4.4):

$$
\begin{aligned}
J_x \cdot \partial_{C_i^j} &= \sum_{b=0}^{k} E_i^b \partial_{E_{j-1}^b} - \sum_{a=0}^{k-1} E_a^j \partial_{E_a^{i+1}}, \\
J_y \cdot \partial_{C_i^j} &= \sum_{b=0}^{k} E_i^b \partial_{E_j^b} - \sum_{a=0}^{k-1} E_a^j \partial_{E_a^i}.
\end{aligned}
\tag{4-3}
$$

(3) Now given an arbitrary Poisson structure

$$\sigma = f(x,y)\partial_x \wedge \partial_y = f(x,y)\sigma_0$$

on U, it follows from Bottacin's formula that $\sigma^{[n]}$ is the bivector corresponding to the linear map $(\sigma^{[n]})^\sharp \colon \mathsf{T}^*\mathsf{U}_{\triangle} \to \mathsf{T}\mathsf{U}_{\triangle}$ given by the formula

$$(\sigma^{[n]})^\sharp = f(J_x, J_y)(\sigma_0^{[n]})^\sharp.$$

Note that the pair $(\sigma^{[n]}, \sigma_0^{[n]})$ is a nondegenerate bi-Hamiltonian structure and $f(J_x, J_y)$ is its associated recursion operator in the sense of integrable systems; this is the source of our name for J_x and J_y.

In the following sections we describe the local structure in the cases in which D has at worst nodal singularities; by the Darboux theorem and its generalizations, this corresponds to the case in which σ is homogeneous of nonpositive weight in a suitable chart (x, y).

4.3 Darboux coordinates

Suppose that $p \in \mathsf{X} \setminus \mathsf{D}$. Then σ is nondegenerate at p, so by the Darboux theorem there exists coordinates (x, y) centred at p such that

$$\sigma = \partial_x \wedge \partial_y.$$

In particular σ has weight -2, ie is constant, and hence the corresponding Poisson structure $\sigma^{[n]}$ is induced by a constant symplectic structure $\omega^{[n]}$ in the chart $\mathsf{U}_{\triangle}$. Natural Darboux coordinates for $\omega^{[n]}$ are then given by the following.

Proposition 4.5 *The Ellingsrud–Strømme and Haiman coordinates on $\mathsf{U}_{\triangle}$ are symplectically dual, in the sense that*

$$\{E_i^j, C_{i'}^{j'}\} = \begin{cases} 1 & \text{if } i = i' \text{ and } j = j', \\ 0 & \text{otherwise.} \end{cases}$$

Hence E_i^j and C_i^j for $j \le i$ form Darboux coordinates, ie

$$\omega^{[n]} = \sum_{0 \le j \le i \le k-1} dC_i^j \wedge dE_i^j.$$

Proof $\mathsf{U}_{\triangle}$ has an open dense set consisting of reduced schemes Z lying in U. Near such an element we have Darboux coordinates $x_1, y_1, \dots, x_n, y_n$ indicating the x and y coordinates of the support the subscheme Z' near Z. Thus

$$\omega^{[n]} = \sum_{i=1}^{n} dy_i \wedge dx_i = \mathrm{Tr}(dM_y \wedge dM_x),$$

where M_x and M_y are the matrices of functions representing the operators of multiplication by x and y on $\mathsf{H}^0(\mathcal{O}_{\mathsf{Z}})$ in the basis of indicator functions of the points of Z. By Lemma 4.6 below, the right-hand side may in fact by computed using the same formula with respect to *any* basis for $\mathsf{H}^0(\mathcal{O}_{\mathsf{Z}})$. In particular, in $\mathsf{U}_{\triangle}$ we may choose the basis of monomials $e_{ij} := x^i y^j$, with $i + j \le k-1$, and compute the actions x and y using Haiman's defining relations (4-2) for $\mathcal{O}_{\mathsf{Z}}$:

$$x \cdot e_{ij} = \begin{cases} e_{i+1,j} & \text{if } i+j \le k-2, \\ \sum_{a+b\le k-1} c_{ab}^k e_{ab} & \text{if } i+j = k-1, \end{cases}$$
$$y \cdot e_{ij} = \begin{cases} e_{i,j+1} & \text{if } i+j \le k-2, \\ \sum_{a+b\le k-1} c_{ab}^k e_{ab} & \text{if } i+j = k-1, \end{cases}$$

from which we deduce that

$$\omega^{[n]} = \mathrm{Tr}(dM_y \wedge dM_x) = \sum_{i+j, a+b \le k-1} d\langle y \cdot e_{ij}, e_{ab}^* \rangle \wedge d\langle x \cdot e_{ab}, e_{ij}^* \rangle = \sum_{i,a=0}^{k-1} dC_a^i \wedge dC_i^{a+1}.$$

Therefore

$$\omega^{[n]}(\partial_{C_i^j}, -) = dC_j^{i+1} - dC_{j-1}^i = dE_i^j, \tag{4-4}$$

so that the Hamiltonian vector field of E_i^j is $\partial_{C_i^j}$, as claimed.

To prove the claim about the Darboux coordinates, note that due to Remark 4.1 we have $\{E_i^j, E_{i'}^{j'}\} = \{C_i^j, C_{i'}^{j'}\} = 0$ for all $j \le i$, $j' \le i'$. □

Lemma 4.6 *Let A and B be a pair of commuting square matrices with functional entries. Then the two-form $\mathrm{Tr}(dA \wedge dB)$ is invariant under simultaneous conjugation of A and B via an invertible matrix with functional entries.*

Proof We learned this from the preprint version of [45, proof of Proposition C.1]. Since the published version does not contain the proof, we include it here for completeness. Let $\widetilde{A} = gAg^{-1}$, $\widetilde{B} = gBg^{-1}$ for some invertible matrix of functions g. Then we have

$$\mathrm{Tr}(d\widetilde{A} \wedge d\widetilde{B}) = \mathrm{Tr}(dA \wedge dB) - \mathrm{Tr}\, d([A, B]\, g^{-1} dg).$$

which gives the result. □

4.4 Recursion operators

In this subsection, we derive the formulae for the recursion operators (4-3) in the triangular chart of the Hilbert scheme of $\mathsf{X}=\mathbb{C}^2$. Note that since x and y have weight one with respect to the natural $\mathbb{C}^\times$-action, the operators J_x and J_y also have weight one, and in particular they send the basis vector fields $\partial_{C_i^j}$ in Haiman coordinates to linear vector fields. By applying the automorphism that interchanges x and y (which has the form $E_i^j \mapsto -E_{k-i-1}^{k-j}$ in Ellingsrud–Strømme coordinates) the calculation of J_x reduces to that of J_y, so we will just explain how to do the latter.

The tangent vector $\partial_{C_i^j}$, with $0\le i\le k-1$ and $0\le j\le k$, corresponds to an element $\varphi_i^j\in\mathsf{Hom}(\mathcal{I},\mathcal{O}_{\mathsf{Z}})$ of the form

$$\varphi_i^j(f_r)=\sum_{a+b\le k-1}\gamma_{abi}^{rj}\,x^a y^b \quad \bmod \mathcal{I},$$

where $f_r=f_{\mathsf{Z},r}$ for $r=0,1,\dots,k$ are the generators (4-2) of $\mathcal{I}$. Note that each f_r is a homogeneous polynomial of degree k in variables x, y and the Haiman coordinates C_i^j. Therefore, each γ_{abi}^{rj} is a homogeneous polynomial of degree $k-1-a-b$ in Haiman coordinates. By definition of the Haiman coordinates, we have

$$\varphi_i^j(f_r)=\delta_{jr}x^i y^{k-1-i}+\sum_{a=0}^{k-2}\theta_{ai}^{rj}\,x^a y^{k-2-a} \quad \bmod V_3, \tag{4-5}$$

where δ_{jr} is the Kronecker delta symbol, θ_{ai}^{rj} are linear functions in Haiman coordinates, and for $l\ge 0$ we denote by

$$V_l:=\mathcal{I}\oplus\bigoplus_{c+d\le k-l}\mathbb{C}\cdot x^c y^d\subset\mathbb{C}[x,y]$$

the linear subspace generated by $\mathcal{I}$ and the monomials of degree at most $k-l$.

By using the equalities $\varphi_i^j(xf_r-yf_{r+1})=x\varphi_i^j(f_r)-y\varphi_i^j(f_{r+1})$ for $r=0,1,\dots,k-1$, and looking the terms modulo V_2, we obtain the following linear equations on γ:

$$\theta_{ai}^{r+1,j}-\theta_{a-1,i}^{rj}=\delta_{rj}C_a^{i+1}-\delta_{r+1,j}C_a^i-\delta_{ai}(C_j^{r+1}-C_{j-1}^r)\quad \text{for } 0\le i,r\le k-1 \text{ and } 0\le j,a\le k.$$

Here and below we adopt the convention that if an index of C or θ falls outside the allowed range, then we declare the value of such C or θ to be zero.

Finally, to compute $J_y\cdot\partial_{C_i^j}$, we need to calculate $y\varphi_i^j$. By (4-5), one has

$$\begin{aligned} y\varphi_i^j(f_b)=\delta_{jb}x^i y^{k-i}+\sum_{a=0}^{k-2}\theta_{ai}^{bj}\,x^a y^{k-1-a}&=\delta_{bj}\sum_{a=0}^{k-1}C_a^i x^a y^{k-1-a}+\sum_{a=0}^{k-2}\theta_{ai}^{bj}x^a y^{k-1-a}\\ &=\sum_{a=0}^{k-1}(\delta_{bj}C_a^i+\theta_{ai}^{bj})x^a y^{k-1-a}\quad \bmod V_2.\end{aligned}$$

This implies that

$$J_y \cdot \partial_{C_i^j} = \sum_{b=0}^{k} \sum_{a=0}^{k-1} (\delta_{bj} C_a^i + \theta_{ai}^{bj}) \frac{\partial}{\partial C_a^b},$$

and therefore

$$\begin{aligned}
\langle J_y \cdot \partial_{C_i^j}, dE_a^b \rangle &= \left\langle \sum_{\beta=0}^{k} \sum_{\alpha=0}^{k-1} (\delta_{\beta j} C_\alpha^i + \theta_{\alpha i}^{\beta j}) \frac{\partial}{\partial C_\alpha^\beta}, dC_b^{a+1} - dC_{b-1}^a \right\rangle \\
&= \delta_{a+1,j} C_b^i - \delta_{aj} C_{b-1}^i + \theta_{b,i}^{a+1,j} - \theta_{b-1,i}^{a,j} \\
&= \delta_{a+1,j} C_b^i - \delta_{aj} C_{b-1}^i + \delta_{aj} C_b^{i+1} - \delta_{a+1,j} C_b^i - \delta_{bi} (C_j^{a+1} - C_{j-1}^a) \\
&= \delta_{aj} (C_b^{i+1} - C_{b-1}^i) - \delta_{bi} (C_j^{a+1} - C_{j-1}^a) \\
&= \delta_{aj} E_i^b - \delta_{bi} E_a^j,
\end{aligned}$$

as desired.

4.5 Smooth curves and the Lie algebra $\mathfrak{aff}_k(\mathbb{C})$

Now suppose that $p \in \mathsf{D}$ is a smooth point. Then by the log Darboux theorem there exist coordinates (x, y) centred at p such that D is given by $y = 0$ and

$$\sigma = y \partial_x \wedge \partial_y$$

In particular σ is linear in these coordinates, thus $\sigma^{[n]}$ is linear in the corresponding Haiman/Ellingsrud–Strømme coordinates, so that it corresponds dually to a Lie algebra structure on the space of $(k+1) \times k$ matrices.

We claim that this is precisely the Lie algebra $\mathfrak{aff}_k(\mathbb{C}) \cong \mathfrak{gl}_k(\mathbb{C}) \ltimes \mathbb{C}^k$ of the group $\mathsf{Aff}_k(\mathbb{C}) \cong \mathsf{GL}_k(\mathbb{C}) \ltimes \mathbb{C}^k$ of affine transformations of $\mathbb{C}^k$. To see this, we use the natural embedding

$$\psi : \mathsf{Aff}_k(\mathbb{C}) \hookrightarrow \mathsf{GL}_{k+1}(\mathbb{C}), \quad \psi(g, v) = \begin{pmatrix} g & v \\ 0 & 1 \end{pmatrix},$$

where $g \in \mathsf{GL}_k(\mathbb{C})$ and $v \in \mathbb{C}^k$, which identifies $\mathfrak{aff}_k(\mathbb{C})$ with $\mathbb{C}^{(k+1) \times k}$. Identifying the duals via the trace pairing, the coadjoint action of $\mathsf{Aff}_k(\mathbb{C})$ on the vector space $\mathfrak{aff}_k(\mathbb{C})^\vee \cong \mathbb{C}^{(k+1) \times k}$ is given by the formula

$$(g, v) \cdot E := \psi(g, v) E g^{-1}.$$

Now if $E \in \mathbb{C}^{(k+1) \times k}$, the syzygy matrix of the scheme $\mathsf{Z}(E)$ is the matrix $S_E(x, y)$ from (4-1), and hence $i^* \mathcal{I}_{\mathsf{Z}(E)}$ is presented as the cokernel of the matrix

$$S_E(x, y)|_{\mathsf{D}} = S_E(x, 0) = E - x \begin{pmatrix} I_k \\ 0 \end{pmatrix},$$

and direct computation shows that

$$\psi(g,v)S_E(x,0)g^{-1} = S_{(g,v)\cdot E}(x,0),$$

so that (g,v) induces a D-equivalence from $\mathsf{Z}(E)$ to $\mathsf{Z}((g,v)\cdot E)$. This defines a map of groupoids between $\mathsf{Aff}_k(\mathbb{C}) \ltimes \mathfrak{aff}_k(\mathbb{C})^\vee$ and $(\mathsf{X},\mathsf{D})^{[n]}$ over the open sets corresponding to $\mathsf{U}_\triangleright$, giving a local model for the symplectic groupoid $(\mathsf{X},\mathsf{D})^{[n]}$:

Theorem 4.7 *The map $\mathsf{E} \to \mathsf{Z}(E)$ gives a Poisson isomorphism from an open set in $\mathfrak{aff}_k(\mathbb{C})^\vee$ to $\mathsf{U}_\triangleright$, and the corresponding map of groupoids is a symplectomorphism onto its image.*

Proof We first prove that the map $E \mapsto \mathsf{Z}(E)$ is Poisson. For this we compute the Poisson bivector on $\mathsf{U}_\triangleright$ following the algorithm of Section 4.2. From the formulae in Proposition 4.5 and (4-3) we have $\{E_i^j, E_a^b\} = \langle J_y \cdot \partial_{C_i^j}, dE_a^b\rangle = \delta_{aj}E_i^b - \delta_{bi}E_a^j$, which is exactly the formula for the bracket on $\mathfrak{aff}_k(\mathbb{C})^\vee$ in the given basis, as desired. Hence the map $E \mapsto \mathsf{Z}(E)$ is Poisson.

This implies immediately that the map of groupoids is compatible with symplectic structures, and hence it is étale onto its image. It remains to check that it is injective. That the map is injective on objects of the groupoid is simply the statement that $\mathsf{U}_\triangleright$ is a chart, so we only need to show injectivity on the target (or source) fibres. Thus, it suffices to fix $E \in \mathfrak{aff}_k(\mathbb{C})^\vee$ and show that we obtain an injective map from $\mathsf{Aff}_k(\mathbb{C}) \times \{E\}$ to D-equivalences. For this, we note that any element $(g,v) \in \mathsf{Aff}_k(\mathbb{C})$ can be recovered from the corresponding D-equivalence, where the latter is viewed as an automorphism of the first two terms of the resolution (2-9). Indeed, evaluating the aforementioned complex at p and considering the endomorphism of the second term, $\mathcal{O}_{\{p\}}^{\oplus(k+1)}$, we obtain a $(k+1)\times(k+1)$-matrix lying in $\mathsf{Aff}_k(\mathbb{C})$. This evaluation is well-defined modulo isomorphisms of D-equivalences, ie, depends only on the homotopy class of the map of cochain complexes, because the nontrivial differential in the two-term complex becomes zero after restriction to p. □

Corollary 4.8 *For $\mathsf{X} = \mathsf{U} = \mathbb{C}^2$ and $\sigma = y\partial_x \wedge \partial_y$, we have $\mathsf{U}_\triangleright \cong \mathfrak{aff}_k(\mathbb{C})^\vee$ as Poisson manifolds. Under this identification*:

(1) *The symplectic leaves in $\mathsf{U}_\triangleright$ are identified with the coadjoint orbits of $\mathfrak{aff}_k(\mathbb{C})$.*

(2) *The linear subspace $\mathfrak{gl}_k(\mathbb{C})^\vee \subset \mathfrak{aff}_k(\mathbb{C})^\vee$ corresponds to the set $\mathcal{A}_k$ of elements $\mathsf{Z} \in \mathsf{U}_\triangleright$ such that $\#(\mathsf{Z} \cap \mathsf{D}) = k$.*

(3) *The nilpotent cone in $\mathfrak{gl}_k(\mathbb{C})^\vee$ is identified with the set of elements $\mathsf{Z} \in \mathsf{U}_\triangleright$ such that $\mathcal{I}_{\mathsf{Z}\cap\mathsf{D}} = (x^k, y)$.*

Proof It remains to establish points (2) and (3) above. For this, note that the subspace $\mathfrak{gl}_k(\mathbb{C})^\vee$ is cut out by the k equations $E_i^k = 0$ for $i = 0,1,\dots,k-1$. If $E_i^k \neq 0$ for some $0 \le i \le k-1$, then the $(k-i)^{\text{th}}$ minor of $S_E(x,0)$ is a polynomial whose highest-degree term is $\pm E_i^k x^{k-1}$. Since this function vanishes on $\mathsf{Z}\cap\mathsf{D}$, we deduce that $\#\mathsf{Z}\cap\mathsf{D} < k$. Conversely, if $E_i^k = 0$ for all i, then all maximal minors of $S_E(x,0)$ are zero, except the last one, which is equal to the characteristic polynomial of the square matrix $\widetilde{E} := (E_i^j)_{i,j=0}^{k-1}$, which has degree exactly k. Hence $\#(\mathsf{Z}\cap\mathsf{D}) = k$, which establishes statement (2). Statement (3) follows immediately since $\widetilde{E}$ lies in the nilpotent cone if and only if its characteristic polynomial is x^k. □

Remark 4.9 More generally, if (X,σ) is a Poisson surface with smooth zero divisor D, then for any $k \le n$ one can define a locally closed Poisson submanifold $\mathsf{W} \subset \mathsf{X}^{[n]}$ as the locus of elements $\mathsf{Z} \in \mathsf{X}^{[n]}$ such that $\#(\mathsf{Z}\cap\mathsf{D}) = k$, and one has a natural projection $\mathsf{W} \to \mathsf{D}^{(k)}$, whose fibres are Poisson subvarieties of $\mathsf{X}^{[n]}$. Away from the big diagonal in $\mathsf{D}^{(k)}$, the fibres are smooth and symplectic, but in general they can be singular. Corollary 4.8 implies that the singularities are products of slices of nilpotent orbits in $\mathfrak{sl}_m(\mathbb{C})$ for $m \le k$. These singularities are all symplectic. Near a most singular fibre $(\mathsf{Z}\cap\mathsf{D}) = k\cdot p$ with $n \ge \frac{1}{2}k(k+1)$, the family restricts to the universal Poisson deformation of the nilpotent cone (times a parameter giving the centre of mass of Z, which makes sense in local coordinates). We expect it to have a unique simultaneous resolution $\widetilde{\mathsf{W}} \to \mathsf{W}$, giving a global analogue of the Grothendieck–Springer resolution, which restricted to the fibre over $k\cdot p$ is a global analogue of the Springer resolution. It would be interesting to further study these objects. For example, are all Poisson deformations of $\mathsf{W}, \widetilde{\mathsf{W}}$ given by deformations of the Hilbert scheme? And are all Poisson deformations of the fibre over $k\cdot p \in D$ given by deforming the Hilbert scheme and moving the point in $\mathsf{D}^{(k)}$ (at least infinitesimally)? ◊

Using this result, we obtain a complete characterization of the closure relation between symplectic leaves as a condition on the corresponding Young diagrams. Note that the necessity of this condition was proven by Rains in [38, Section 11.2], but our proof of sufficiency uses our local normal form in an essential way.

Proposition 4.10 *Let (X,σ) be a Poisson surface with a smooth zero divisor D, and let $\mu, \widetilde{\mu}$ be global orbit data as in Lemma 2.15. Then the closure of the symplectic leaf $\mathsf{X}^{[n]}_{\mu}$ contains $\mathsf{X}^{[n]}_{\widetilde{\mu}}$ if and only if for every $p \in \mathsf{D}$, we have the inclusion $\mathrm{hc}(\mu(p)) \subseteq \mathrm{hc}(\widetilde{\mu}(p))$, or equivalently, $\mu(p) \le \widetilde{\mu}(p)$ in the dominance order on Young diagrams.*

Proof The second equivalence is a simple consequence of the definitions. Let us prove the first equivalence.

Let us start by remarking for any ideal $\mathsf{Z} \in \mathsf{X}^{[m]}$, with $m < n$, the set

$$\{\mathsf{Z}' \in \mathsf{X}^{[n]} \mid \mathsf{Z}' \supseteq \mathsf{Z}\}$$

is a closed subvariety. By Theorem 2.24, every element $\mathsf{Z} \in \mathsf{X}^{[n]}_{\mu}$ satisfies $\mathsf{Z} \supset \bigsqcup_{p\in\mathsf{D}} \mathsf{Z}_p^{\mu(p)}$. Therefore, if $\mathsf{X}^{[n]}_{\widetilde{\mu}}$ lies in the closure of $\mathsf{X}^{[n]}_{\mu}$, then we have $\mathsf{Z}_p^{\widetilde{\mu}(p)} \supseteq \mathsf{Z}_p^{\mu(p)}$ for each $p \in \mathsf{D}$.

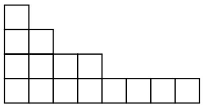
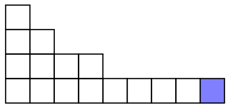
For the converse, suppose that $\mu \subseteq \widetilde{\mu}$. We wish to show that $\mathsf{X}^{[n]}_{\widetilde{\mu}} \subset \overline{\mathsf{X}^{[n]}_{\mu}}$. It suffices to prove this result in the complex topology, since it is finer.

It is enough to consider the case when both μ and $\widetilde{\mu}$ consist of one nontrivial Young diagram each, concentrated at the same point $p \in \mathsf{D}$. By abuse of notation, we denote these diagrams simply by μ and $\widetilde{\mu}$.

We can also, without loss of generality, take $\mathsf{X} = \mathbb{C}^2$ equipped with the Poisson bivector, $\sigma = y\,\partial_x \wedge \partial_y$, and $p = (0,0)$. Furthermore, note that by Theorem 2.24 it suffices to prove the closure relation for the open dense sets consisting of subschemes of the form $\mathsf{Z}_p^{\mu(p)} \sqcup \mathsf{Z}'$, where $\mathsf{Z}' \subset \mathsf{X} \setminus \mathsf{D}$ is reduced. Such an element has an analytic neighbourhood of the form $\mathsf{U}^{[k]} \times \mathsf{V}^{[n-k]}$, where U is an analytic neighbourhood of p and V is an analytic neighbourhood of Z' such that $\overline{\mathsf{V}} \subset \mathsf{X} \setminus \mathsf{D}$ and $\overline{\mathsf{U}} \cap \overline{\mathsf{V}} = \varnothing$. Hence by adding or removing points from Z', we see that it is enough to prove the statement for any single value of $n \geq |\mathrm{hc}(\widetilde{\mu})|$.

Let $\lambda = \mathrm{hc}(\mu)$ and $\widetilde{\lambda} = \mathrm{hc}(\widetilde{\mu})$. We consider two special cases for the pair $(\lambda, \widetilde{\lambda})$ below, and then explain how the general case follows from these.

Case 1 ($\lambda_1 < \widetilde{\lambda}_1$, but $\lambda_i = \widetilde{\lambda}_i$, for $i > 1$) In this case, by induction on the difference $\widetilde{\lambda}_1 - \lambda_1$, we may assume without loss of generality that $\lambda_1 + 1 = \widetilde{\lambda}_1$. In other words, the Young diagram $\widetilde{\lambda}$ is obtained from λ by adding one box to the first row. For instance:

$\lambda =$  $\widetilde{\lambda} =$ 

We may assume without loss of generality that $n = |\lambda| + 1 = |\widetilde{\lambda}|$. Then $\mathsf{X}_{\widetilde{\mu}}^{[n]}$ consists of only one point given by $\mathsf{Z}_p^{\widetilde{\mu}}$, whereas $\mathsf{X}_{\mu}^{[n]}$ is two-dimensional and its generic point is of the form $\mathsf{Z}_p^{\mu} \sqcup \{(x_1, y_1)\}$, $y_1 \neq 0$. The idea is to send the point (x_1, y_1) to the origin along a curve that is tangent to the divisor $\{y = 0\}$ to a high enough order, so that the limiting ideal will be $\mathsf{Z}_p^{\widetilde{\mu}}$. Here is the calculation that makes this heuristic precise.

Let us write

$$\mathcal{I}_{\mathsf{Z}_p^{\mu}} = (x^{a_j} y^j,\ j = 0, 1, \ldots, \ell),$$

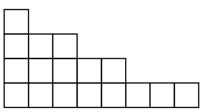
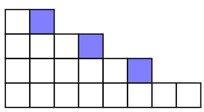
where $a_j = \lambda_{j+1}$ for $j = 0, 1, \ldots, \ell$ and $\ell = \lambda_1^T$. Let $x_1 = \varepsilon$, and $y_1 = \varepsilon^N$, where $\varepsilon \in \mathbb{C}^*$, and N is a large positive integer to be determined in a moment. Then

$$\begin{aligned}\mathcal{I}_{\mathsf{Z}_p^{\mu} \sqcup \{(x_1,y_1)\}} &= (x^{a_j} y^j,\ j = 0, 1, \ldots, \ell) \cap (x - \varepsilon, y - \varepsilon^N) \\ &= (x^{a_0+1} - \varepsilon x^{a_0}, x^{a_j} y^j - \varepsilon^{b_j} x^{a_0},\ j = 1, 2, \ldots, \ell),\end{aligned}$$

where $b_j = jN + a_j - a_0$ for $j = 1, 2, \ldots, \ell$. Now choose N so that $b_j > 0$ for all j. Then

$$\mathsf{Z}_p^{\mu} \sqcup \{(x_1, y_1)\} \to \mathsf{Z}_p^{\widetilde{\mu}} \quad \text{as } \varepsilon \to 0.$$

Case 2 ($\lambda_1 = \widetilde{\lambda}_1$) In this case, we cannot assume that $|\widetilde{\lambda}| = |\lambda| + 1$, because we must maintain the condition that the Young diagrams involved are horizontally convex. For instance, in the following example there are no intermediate horizontally convex diagrams between λ and $\widetilde{\lambda}$, even though $|\widetilde{\lambda}| = |\lambda| + 3$:

$\lambda =$  $\widetilde{\lambda} =$ 

We may assume without loss of generality that $n = \frac{1}{2}k(k+1)$, where k is the common value of λ_1 and $\widetilde{\lambda}_1$. It is then enough to show that

$$\mathsf{X}^{[n]}_{\widetilde{\mu}} \cap \mathcal{U}_{\triangle} \subset \overline{\mathsf{X}^{[n]}_{\mu}} \cap \mathcal{U}_{\triangle},$$

where $\mathcal{U}_{\triangle}$ is the triangular chart in $(\mathbb{C}^2)^{[n]}$. But under the isomorphism $\mathcal{A}_k \cong \mathfrak{gl}_k(\mathbb{C})^\vee$, described in Corollary 4.8(2), the symplectic leaves $\mathsf{X}^{[n]}_{\mu} \cap \mathcal{U}_{\triangle}$ and $\mathsf{X}^{[n]}_{\widetilde{\mu}} \cap \mathcal{U}_{\triangle}$ correspond to the conjugacy classes of nilpotent Jordan type μ and $\widetilde{\mu}$, respectively. Recall that the condition $\lambda \subseteq \widetilde{\lambda}$ implies that $\mu \leq \widetilde{\mu}$ in the dominance order. Hence the result follows from the classical Gerstenhaber–Hesselink theorem; see eg [33].

General case Note that for any pair $\lambda \subseteq \widetilde{\lambda}$, we can find an intermediate diagram $\widehat{\lambda}$ such that $\lambda \subseteq \widehat{\lambda}$ falls into Case 1 and $\widehat{\lambda} \subseteq \widetilde{\lambda}$ falls into Case 2. This completes the proof. □

4.6 Nodal points and toric degenerations

Now suppose that $p \in \mathsf{D}$ is a nodal singularity. Then by [1], there exist coordinates (x, y) such that, after rescaling σ by a nonzero constant, we have

$$\sigma = xy\,\partial_x \wedge \partial_y.$$

Applying the algorithm from Section 4.2 to compute the bracket $\sigma^{[n]} = J_x J_y \sigma_0^{[n]}$, we find after a straightforward but tedious calculation that

$$\begin{aligned}\{E_i^j, E_a^b\} = {} & \delta_{a\geq j}\sum_{p=0}^{a} E_i^{p+j-a}E_p^b - \sum_{p=i}^{a} E_{a+i-p}^j E_p^b \\ & - \delta_{a+1\leq j}\sum_{p=a+1}^{k-1} E_i^{p+j-a}E_p^b + \sum_{p=a+1}^{i-1} E_{a+i-p}^j E_p^b + \sum_{q=0}^{\min(j,b-1)} E_i^{b+j-q}E_a^q \\ & - \delta_{b\leq i}\sum_{q=0}^{b-1} E_{q+i-b}^j E_a^q - \sum_{q=\max(j+1,b)}^{k} E_i^{b+j-q}E_a^q + \delta_{b-1\geq i}\sum_{q=b}^{k} E_{q+i-b}^j E_a^q.\end{aligned} \tag{4-6}$$

A remarkable feature of this quadratic Poisson structure is that it admits a canonical toric degeneration, as we now explain.

4.6.1 The torus-invariant part Consider the action of the torus $(\mathbb{C}^\times)^{2n} = (\mathbb{C}^\times)^{k(k+1)}$ on the germ of $\mathsf{U}_{\triangle}$ by independent dilation of the Ellingsrud–Strømme coordinates. The induced action on the space of germs of bivectors preserves the space of quadratic bivectors, and hence we may project $\sigma^{[n]}$ onto its torus-invariant part, which is also quadratic.

Definition 4.11 We denote by $\sigma_{\Delta}^{[n]}$ the torus-invariant part of $\sigma^{[n]}$.

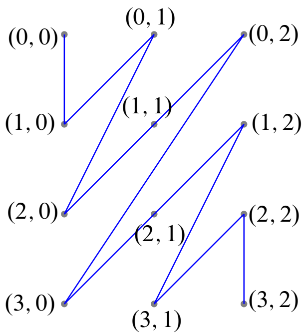


Figure 1: Ordering the indices of a matrix.

Explicitly, we have

$$\sigma_\Delta^{[n]} = \sum_{i,j,a,b} \Pi_{ia}^{jb} \, E_i^j E_a^b \partial_{E_i^j} \wedge \partial_{E_a^b},$$

where the coefficients are given by

$$\Pi_{ia}^{jb} = \delta_{a\geq j} - \delta_{a\geq i} - \delta_{bj} \, \delta_{a\geq i+1} + \delta_{b\geq j+1} \, \delta_{ai} - \delta_{b\geq j+1} + \delta_{b\geq i+1}. \tag{4-7}$$

This bivector is generically nondegenerate, and degenerates along the union of the coordinate hyperplanes. The inverse log symplectic form is then given by

$$\omega_\Delta^{[n]} = \sum_{i,j,a,b} B_{ia}^{jb} \frac{\mathrm{d}E_i^j}{E_i^j} \wedge \frac{\mathrm{d}E_a^b}{E_a^b},$$

where

$$B_{ia}^{jb} = -\delta_{a+b,i+j} \, \delta_{a\geq i+1} - \delta_{a+b,i+j+1} \, \delta_{a\leq i} + \delta_{a+b,i+j} \, \delta_{a\leq i-1} + \delta_{a+b,i+j-1} \, \delta_{a\geq i} \tag{4-8}$$

is the "biresidue" of ${\omega_\Delta}^{[n]}$ along the intersection of hyperplanes $E_i^j = E_a^b = 0$ in $\mathbb{C}^{k(k+1)}$. To understand more clearly the structure of the biresidues, it is helpful to order the coordinates according to their position in the syzygy matrix as follows: first order by the sum of the indices, and then order by the column index to break ties, ie

$$(0,0) \prec (1,0) \prec (0,1) \prec (2,0) \prec (1,1) \prec (0,2) \prec \cdots. \tag{4-9}$$

For instance, the case $k = 3$ is illustrated in Figure 1. With this ordering the biresidues B_{ia}^{jb} form a skew-symmetric matrix B of size $m := k(k+1)$ with the following property, which will be useful in our study of holonomicity in Section 4.7:

Definition 4.12 A skew-symmetric $m \times m$ matrix $(b_{\alpha,\beta})_{\alpha,\beta=1}^m$ is *cyclically monotone* if, perhaps after multiplying all of its entries by the same nonzero constant, we have

$$b_{\alpha,\alpha+1} \geq b_{\alpha,\alpha+2} \geq \cdots \geq b_{\alpha,m} \geq b_{\alpha,1} \geq b_{\alpha,2} \geq \cdots \geq b_{\alpha,\alpha-1}$$

for each $1 \leq \alpha \leq n$.

For instance, for $k = 2$, we have

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

4.6.2 Toric degeneration: a game of dominoes We will prove the following:

Theorem 4.13 *There exists a rank-one subtorus* $\mathsf{G} \cong \mathbb{C}^\times \hookrightarrow (\mathbb{C}^\times)^{(k+1)\times k}$ *such that*

$$\lim_{g \in \mathsf{G}, g \to 0} g \cdot \sigma^{[n]} = \sigma_\Delta^{[n]}.$$

Remark 4.14 In light of this result, we can view $\sigma^{[n]}$ as a one-parameter deformation of the toric log symplectic structure $\sigma_\Delta^{[n]}$. As explained in [27], such deformations are obtained by smoothing out the nodal singularities along pairwise intersections of hyperplanes. The combinatorics of this process can be encoded in a "smoothing diagram" where we draw a vertex corresponding to each hyperplane, a coloured edge joining two vertices when the corresponding intersection is smoothed, and decorating triangle with angles that indicate the order to which the smoothing degenerates along triple intersections. In the case at hand, the hyperplanes are given by the equations $E_i^j = 0$ and are in bijection with the positions in a $(k+1) \times k$ matrix; the corresponding smoothing diagram is shown in Figure 2. ◊

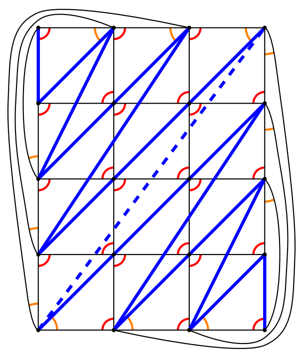


Figure 2: Smoothing diagram encoding the deformation from $\sigma_\Delta^{[n]}$ to $\sigma^{[n]}$ for a nodal curve in the triangular chart (with $k = 4$). The $k(k+1)$ hyperplanes on which $\sigma_\Delta^{[n]}$ degenerates merge into two irreducible components corresponding to the two components of the node $xy = 0$; accordingly, the corresponding vertices are joined by two collections of solid blue lines. The solid dashed line corresponds to an additional possible deformation, which is induced by smoothing $xy = 0$ to $xy = \epsilon \neq 0$ in X, so that the divisor becomes irreducible.

We will prove Theorem 4.13 by analyzing the decomposition of the space of quadratic bivectors into weight spaces for the torus action. Namely, the theorem is equivalent to the statement that $\sigma^{[n]} - \sigma^{[n]}_{\Delta}$ is a sum of weight vectors of the torus action, whose weights with respect to the subtorus G are strictly positive. Note that since the torus action on $\mathsf{U}_{\triangle}$ is defined by rescaling matrix entries, it is natural to depict the weights of the action as matrices of the same size with integer coefficients. With this convention, the weight of a monomial bivector of the form

$$E_i^j E_k^l \partial_{E_a^b} \wedge \partial_{E_c^d} \tag{4-10}$$

is obtained by starting with the zero matrix, adding 1 to the positions (i, j) and (k, l) in turn, and then subtracting 1 from the positions (a, b) and (c, d) in turn.

In the Poisson bracket (4-6), only certain weights can appear. We will describe them in terms of the following objects.

Definition 4.15 A $(k+1) \times k$ matrix is a *domino* if it has exactly two nonzero entries, one of which is $+1$ and the other of which is -1, and they lie either on the same row or on the same column. We refer to the position of the entries $+1$ and -1 as the *head* and *tail* of the domino, respectively. We say that a domino is oriented to the *north* (resp. *south*, *east* or *west*) if its head is above (resp. below, right of, or left of) its tail. See Figure 3.

This *size* of a domino is the distance between its head and tail, and the *valuation* is defined as follows:

- If the domino is oriented north or south, its valuation is the number of vertical translations required to move it so that its uppermost nonzero entry lies in the top column.
- If the domino is oriented east or west, its valuation is the number of horizontal translations required to move it so that its leftmost nonzero entry lies in the leftmost column, plus $\frac{1}{2}$.

Definition 4.16 A *rectangular weight* is a $(k+1) \times k$ matrix given by a sum of two dominoes with opposite orientations, whose heads and tails form the vertices of a rectangle, with a head in the top left corner as in Figure 4.

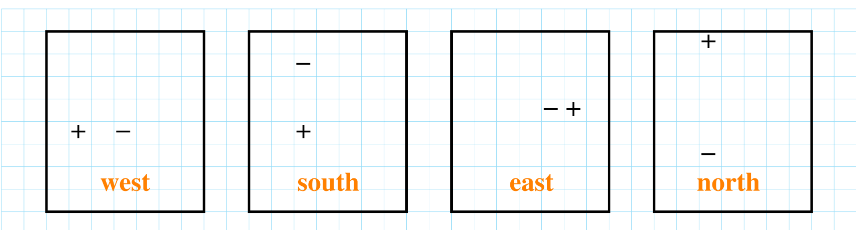


Figure 3: Four examples of dominoes, one from each possible orientation. Reading from left to right, the lengths are 2, 3, 1 and 5, and the valuations are $1 + \frac{1}{2}$, 1, $4 + \frac{1}{2}$ and 0, respectively.

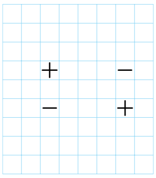


Figure 4: A rectangular weight.

Definition 4.17 An ordered pair of dominoes is *admissible* if both dominoes have the same length, the first domino is directed west or south, the second is directed east or north, and the valuation of the first is greater than the valuation of the second.

By direct inspection of the formula (4-6), we have the following:

Lemma 4.18 *The weight of every monomial appearing in* $\sigma^{[n]}$ *is either rectangular, or the sum of an admissible pair of dominoes.*

We will reduce such weights to sums of the following elementary ones, which correspond to the weights of the first-order deformations of $\sigma_{\Delta}^{[n]}$ that smooth a given pairwise intersection of divisor components (ie to the smoothable strata in the sense of [27]).

Definition 4.19 A *smoothable weight* is a $(k+1)\times k$ matrix given by a sum of two dominoes of length one placed in one of the following configurations:

- **Type I** The two dominoes are adjacent, forming a square-shaped rectangular weight.
- **Type IIa** A south–east admissible pair of dominoes concentrated in the leftmost column and top row such that the tail of the first lies on the same diagonal as the head of the second.
- **Type IIb** A west–north admissible pair of dominoes concentrated in the bottom row and rightmost column such that the head of the first lies on the same diagonal as the tail of the second.

Such matrices are depicted in Figure 5. We remark that for each smoothable weight of type IIa or IIb, the valuation of the southwestern domino is only $\frac{1}{2}$ higher than the valuation of the northeastern domino, which is the minimum possible for the pair to be admissible.

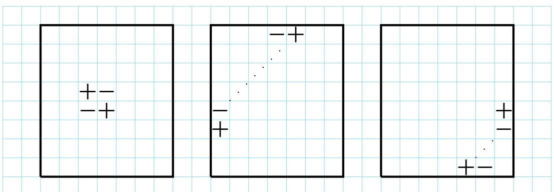


Figure 5: Smoothable weights; from left to right, the types are I, IIa and IIb.

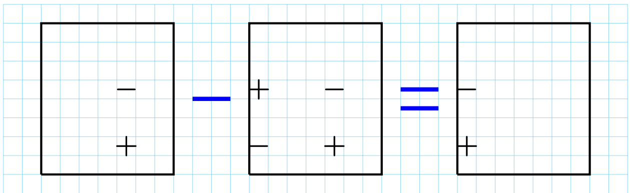


Figure 6: Pushing a southern domino to the leftmost column.

Lemma 4.20 *The set of smoothable weights is linearly independent.*

Proof Consider the square matrix of size $k(k+1)$ formed from the biresidues B_{ip}^{jq} (4-8) of the toric log symplectic form with indices ordered as in (4-9). Then every smoothable weight can be expressed as the difference of a unique pair of consecutive rows of B_{ip}^{jq}. But this matrix is invertible, and hence the differences of its consecutive rows are linearly independent. □

Lemma 4.21 *Any rectangular weight is a sum of smoothable weights of Type* 1 *with nonnegative integer coefficients.*

Proof Consider the sum of all smoothable weights of Type I that are contained inside the rectangle; this sum telescopes, so that only the vertices of the rectangle remain. □

Proposition 4.22 *The sum of any admissible pair of dominoes is a linear combination of smoothable quadratic weights with nonnegative integer coefficients.*

Proof The proof is a sort of game, in which we translate dominoes in the plane by certain admissible operations, which correspond mathematically to subtracting collections of smoothable weights. We will show that by repeated application of such moves, every admissible pair may be reduced to zero.

Firstly, by subtracting a rectangular weight, we can push any southern domino all the way to the leftmost column, as illustrated in Figure 6. Therefore, whenever we have an admissible pair of dominoes, one of which is southern, we can assume without loss of generality that the southern domino is located in the leftmost column. Likewise, we can assume that every northern domino in an admissible pair is located in the rightmost column, every eastern domino in the top row, and every western domino in the bottom row (just as the dominoes in the smoothable weights of type II).

Secondly, we claim that by subtracting several rectangles and smoothable weights of types IIa and IIb, we can move any southern domino in the leftmost column one step to the north. For this, we proceed in three steps. In the first step, we subtract some smoothable weights of type IIa to turn the domino to the west, as in Figure 7. In the second step, we subtract a rectangular weight to push the obtained western domino to the bottom row as in Figure 8. Finally, in the third step, we subtract some smoothable weights of type IIb to turn the western domino back to the south, as in Figure 9. Because the matrix has one more row than it has columns, the domino is now one step northwards from where it started, as desired.

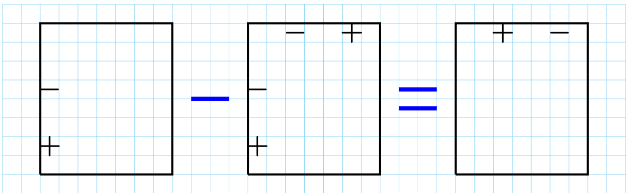


Figure 7: Pushing a southern domino to the north, step 1.

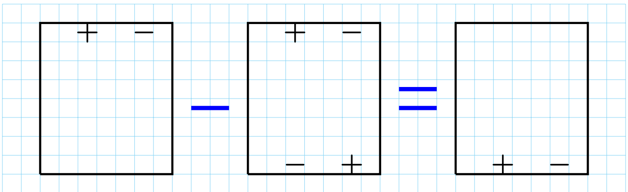


Figure 8: Pushing a southern domino to the north, step 2.

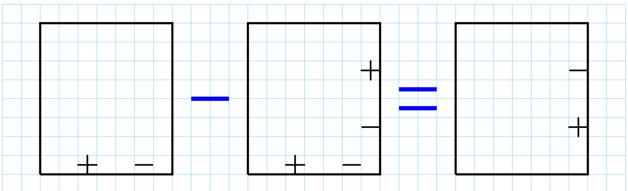


Figure 9: Pushing an eastern domino to the west, step 3.

Using these operations, we are free to push any southern domino in an admissible pair to the north, without changing the second domino in the pair, as long as the pair stays admissible. Similarly, we may push any western domino to the east, any northern domino to the south and any eastern domino to the west.

Now by definition, there are four types of admissible pairs of dominoes: south–east, south–north, west–north, and west–east. Note that by subtracting smoothable weights of type II as in Figure 7, we can convert any admissible south–north pair into an admissible west–north pair, and similarly we may convert any admissible west–east pair into an admissible south–east pair, so it suffices to treat the south–east and west–north cases.

To this end, consider a south–east pair and assume without loss of generality that the southern domino is in the left column, while the eastern one is in the top row. Pushing the southern one as far as possible to the north, we can reduce its valuation to the minimum value for which the pair remains admissible. In this way we obtain an admissible pair that is evidently a telescoping sum of smoothable weights of type IIa. Similarly, pushing a west–north pair to the bottom row and right column, and moving the northern domino south to maximize its valuation, we obtain a sum of smoothable weights of type IIb. □

We can now complete the construction of the toric degeneration:

Proof of Theorem 4.13 By Lemma 4.20 there exists a coweight w of the torus $(\mathbb{C}^*)^{(k+1)k}$ that pairs with each smoothable weight to give a positive integer. By construction, the corresponding one-parameter subgroup $\mathsf{G} \cong \mathbb{C}^\times$ then acts with positive weight on every smoothable weight space, and hence by Lemma 4.21 and Proposition 4.22 acts with positive weight on the weights spaces corresponding to rectangular weights or admissible pairs of dominoes. □

4.7 Holonomicity

Recall from [37] that every holomorphic Poisson manifold (W, σ) has an associated *characteristic variety*

$$\mathrm{Char}(\sigma) \subset \mathsf{T}^*\mathsf{W}$$

defined as the singular support of the complex of $\mathcal{D}_{\mathsf{W}}$-modules associated to the deformation complex of (W, σ). We say that (W, σ) is *holonomic* if its characteristic variety is Lagrangian. This is a local property that depends only on the stable equivalence classes of the germs of σ. It was shown in [37] that if (W, σ) is holonomic then every point has a neighbourhood with only finitely many characteristic symplectic leaves and that the converse holds when $\dim \mathsf{W} = 2$. In [27], we conjectured that the converse holds in all dimensions, and we proved this when (W, σ) is generically symplectic and degenerates along a normal crossing divisor. In this section, we give further evidence for the conjecture.

Considering the invariance under stable equivalence and the results in Section 3.3, there are three cases one must consider in order to verify the conjecture for all Hilbert schemes: either

(a) D has at worst nodal singularities and n is arbitrary,

(b) D has only double points and $n = 2$, or

(c) D has only A_2 singularities and $n = 5$.

We shall treat the cases (a) and (b); we do not know whether the conjecture holds in case (c) but have no reason to doubt it. Case (a) is the subject of the following theorem, whose proof will occupy most of this section. Note that it implies that $\mathsf{X}^{[n]}$ is holonomic if D is smooth; a more direct proof is possible in this case, but we omit it. The smooth case allows us to immediately treat case (b) as a corollary.

Theorem 4.23 *For a Poisson surface X whose vanishing locus is the anticanonical divisor $\mathsf{D} \subset \mathsf{X}$, the following statements are equivalent*:

(1) *For every $n \geq 0$, the induced Poisson structure on the Hilbert scheme $\mathsf{X}^{[n]}$ is holonomic.*

(2) *For every $n \geq 0$, the germ of $\mathsf{X}^{[n]}$ at any point has only finitely many characteristic symplectic leaves.*

(3) *The only singularities of D are nodes.*

Corollary 4.24 *Let X be a Poisson surface with reduced anticanonical divisor $\mathsf{D} \subset \mathsf{X}$. Then $\mathsf{X}^{[2]}$ is holonomic if and only if every singular point of D is a double point.*

Proof of Corollary 4.24 Let $\mathsf{V} \subset \mathsf{X}^{[2]}$ be the set of elements set-theoretically supported at a single singular point of D. Then V is the disjoint union of several copies of $\mathbb{P}^1$, one for each singular point, and $\mathsf{X}^{[2]} \setminus \mathsf{V}$ is holonomic by Theorem 4.23. It therefore suffices to show that $\mathrm{Char}(\sigma^{[2]}) \cap (\mathsf{T}^*\mathsf{X}^{[2]})|_{\mathsf{V}}$ has pure dimension four. But since all singular points are double points, Proposition 3.8 implies that the zeros of the modular vector field in V are isolated, so it defines a function on the smooth five-dimension variety $\mathsf{T}^*\mathsf{X}^{[2]}|_{\mathsf{V}}$ that is nonzero on every connected component. Its vanishing locus contains $\mathrm{Char}(\sigma^{[2]}) \cap \mathsf{T}^*\mathsf{X}^{[2]}|_{\mathsf{V}}$ by [37, Theorem 3.4, part 1], and hence the latter has dimension four, as desired. □

We now proceed with the proof of Theorem 4.23. The equivalence of statements (2) and (3) was proven in Section 3.3, so in light of the above it remains only to prove that if D is nodal, then $\mathsf{X}^{[n]}$ is holonomic for all n. The problem is local and invariant under stable equivalence and taking products, so by Lemma 4.4, it suffices to prove these statements for the triangular chart. Moreover, in light of Lemma 4.25 below and Theorem 4.13, the problem reduces to proving the statement for the toric degeneration $\sigma_\Delta^{[n]}$:

Lemma 4.25 *Let π_t for $t \in \mathbb{C}$ be a polynomial family of quadratic Poisson tensors on $\mathbb{C}^{2n}$ such that π_0 is holonomic. Then π_t is holonomic for all but finitely many $t \neq 0$.*

Proof Let $\mathsf{V}_t \subset \mathsf{T}\mathbb{C}^{2n} \cong \mathbb{C}^{4n}$ be the characteristic variety of the complex of $\mathcal{D}$-modules $\mathcal{M}^\bullet_{\pi_t}$, which is thus invariant under the action of $\mathbb{C}^\times$ by dilation of the fibres. Additionally, since all Poisson structures π_t are quadratic, each V_t is also invariant under the dilation on the base $\mathbb{C}^{2n}$. Combining these two symmetries, we get that all V_t are invariant under the uniform dilation of all directions $\mathbb{C}^{4n}$. In other words, the projectivization $\mathbb{P}(\mathsf{V}_t) \subset \mathbb{P}^{4n-1}$ is well-defined for each t. Since π_t is assumed holonomic, we have $\dim_{\mathbb{C}} \mathbb{P}(\mathsf{V}_t) = 2n-1$. Then by semicontinuity of dimension for proper maps, we have that $\dim_{\mathbb{C}} \mathbb{P}(\mathsf{V}_t) \leq 2n-1$ for t in a Zariski neighbourhood of the origin, as desired. □

The toric degeneration $\sigma_\Delta^{[n]}$ is a log symplectic manifold with normal crossings boundary, so the results of our earlier work [27] apply. In particular, by condition (4) of [27, Theorem 1.5], and by the cyclic monotonicity property of the biresidues of $\sigma_\Delta^{[n]}$ (Definition 4.12), it suffices to prove the following linear-algebraic statement:

Lemma 4.26 *Let B be a cyclically monotone matrix of odd size, with only zeros and ones above the diagonal. Then the row span of B does not contain the constant vector $(1, 1, \dots, 1) \in \mathbb{C}^m$.*

Although the statement is elementary, we did not manage to find an easy direct proof. We will instead prove a more general statement, that is amenable to an inductive argument. For a tuple $J = (J_1, \dots, J_m)$ of nonempty open intervals $J_1 = (c_1, d_1), \dots, J_m = (c_m, d_m) \subset \mathbb{R}$, such that all the endpoints of J_i are distinct, let

$$\ell(J) := (d_1 - c_1, d_2 - c_2, \dots, d_m - c_m)$$

be the vector of their lengths. Let us say that J_α *overlaps* J_β *on the left* (resp. *right*) if $c_\alpha < c_\beta < d_\alpha < d_\beta$ (resp. $c_\beta < c_\alpha < d_\beta < d_\alpha$). Define an $m \times m$ skew-symmetric matrix $B(J)$ by the formula

$$B(J)_{\alpha\beta} = \begin{cases} 1 & \text{if } J_\alpha \text{ overlaps } J_\beta \text{ on the left,} \\ -1 & \text{if } J_\alpha \text{ overlaps } J_\beta \text{ on the right,} \\ 0 & \text{otherwise.} \end{cases} \tag{4-11}$$

Lemma 4.27 *Every matrix B as in Lemma 4.26 is of the form $B = B(J)$ for some collection J of intervals of lengths $\ell(J) = (1, \dots, 1)$.*

Proof For $1 \le \beta \le m$, let k_β be the row index of the uppermost 1 in the βth column of B, with the convention that if there are no ones in the β^{th} column, we define $k_\beta = \beta$. Set $J_1 = (0, 1)$, and choose intervals $J_2, \dots, J_m$ of length one inductively by choosing the left endpoint $c_\beta > c_{\beta-1} + 1$ if $k_\beta = \beta$ and $c_\beta \in (\max\{c_{\beta-1}, c_{k_\beta - 1} + 1\}, c_{k_\beta} + 1)$ if $k_\beta < \beta$. Then $\ell(J) = (1, \dots, 1)$ and $B(J) = B$, as desired. □

Proposition 4.28 *Let $J = (J_1, \dots, J_m)$ be a tuple of intervals, where m is odd and the endpoints of J_i are distinct. Then the vector $\ell(J) \in \mathbb{R}^m$ of their lengths does not lie in the row span of $B(J)$.*

Proof We proceed by induction on the odd integer m. The base case $m = 1$ is obvious, so assume the proposition holds for some odd integer $m - 2$; we will prove that it holds for m.

Let view $B(J)$ as a bivector $B(J) \in \wedge^2 \mathbb{R}^m$ so that we have

$$B(J) := \sum_{1 \le \alpha < \beta \le m} B(J)_{\alpha,\beta} e_\alpha \wedge e_\beta \quad \text{and} \quad \ell(J) = \sum_{\alpha=1}^{m} \ell(J)_\alpha e_\alpha,$$

where $e_1, e_2, \dots$ is the standard basis of $\mathbb{R}^m$.

By permuting the indices, we may assume without loss of generality that the sequence of left endpoints of the intervals is strictly increasing. Then if $\alpha < \beta$, the interval J_α can never overlap J_β on the right, and hence $B(J)_{\alpha,\beta} \in \{0, 1\}$ for $\alpha < \beta$.

If no interval J_γ with $\gamma < m$ overlaps J_m, then $B(J)_{\alpha,m} = 0$ for all α, and since $\ell(J)_m > 0$ we conclude that $\ell(J)$ is not in the row span of $B(J)$, as desired. Hence we may assume without loss of generality that there exists a maximal index $\gamma < m$ such that J_γ overlaps J_m on the left. Define a new basis $\tilde{e}_1, \dots, \tilde{e}_m$ for $\mathbb{R}^m$ by the formula

$$\tilde{e}_\alpha := \begin{cases} e_\gamma - \sum_{\beta \ne \gamma, m} B(J)_{m,\beta} \cdot e_\beta & \text{if } \alpha = \gamma, \\ e_m + \sum_{\beta \ne \gamma, m} B(J)_{\gamma,\beta} \cdot e_\beta & \text{if } \alpha = m, \\ e_\alpha & \text{otherwise.} \end{cases}$$

It is tedious but straightforward to check that in this new basis, we have

$$B(J) = \tilde{e}_\gamma \wedge \tilde{e}_m + \sum_{\alpha,\beta,\gamma,m \text{ distinct};\, \alpha<\beta} B(\tilde{J})_{\alpha,\beta} \tilde{e}_\alpha \wedge \tilde{e}_\beta \quad \text{and} \quad \ell = a_1 \tilde{e}_\gamma + a_2 \tilde{e}_m + \sum_{\alpha \ne \gamma, m} \ell(\tilde{J})_\alpha \tilde{e}_\alpha,$$

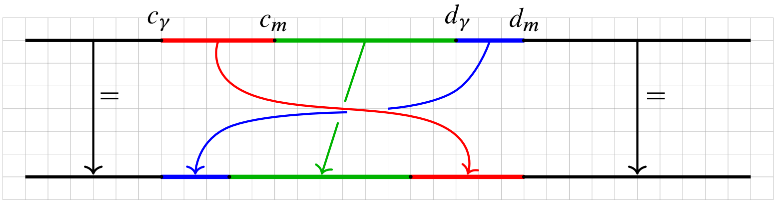


Figure 10: The interval exchange transformation T.

where $a_1, a_2 \in \mathbb{Z}$, and $\widetilde{J} = (T \cdot J_1, \ldots, \widehat{T \cdot J_\gamma}, \ldots, T \cdot J_{m-1})$ is the collection of $m-2$ intervals defined as follows. First, define the "interval exchange transformation" to be the bijection $T : \mathbb{R} \to \mathbb{R}$ which translates the intervals between the endpoints of $J_\gamma = (c_\gamma, d_\gamma)$ and $J_m = (c_m, d_m)$ as shown in Figure 10, and acts as the identity on the rest of the real line. Then $T \cdot J_\alpha = (T(c_\alpha), T(d_\alpha))$ is the interval spanned by the T-images of the endpoints of J_α. By induction, $\ell(\widetilde{J})$ is not in the image of $B(\widetilde{J})$, and hence $\ell(J)$ is not in the image of $B(J)$, as desired. □

5 Deformation theory

In this section we address some aspects of the deformation theory of Bottacin's Poisson structures. Recall that if (W, σ) is a holomorphic Poisson manifold, then the sheaf $\mathscr{X}^\bullet_{\mathsf{W}} = \wedge^\bullet \mathcal{T}_{\mathsf{W}}$ of polyvector fields comes equipped with a differential $\mathrm{d}_\sigma := [\sigma, -]$. The hypercohomology $\mathsf{H}^\bullet_\sigma(\mathsf{W})$ of this complex (called the Poisson cohomology) controls deformations of W as a holomorphic Poisson manifold, or more generally, as a generalized complex manifold.

We shall compute the Poisson cohomology of $\mathsf{X}^{[n]}$ in low degrees, and in particular determine the deformation space, under the assumption that D is reduced, and has only quasihomogeneous singularities (ie is locally the zero set of a quasihomogeneous polynomial). As remarked in [37, Section 4.4], the quasihomogeneity is automatic if X is projective, thanks to the classification in [23, Section 7], or by analysis of the minimal models and their blowups.

The basic technique, developed in [27; 37], is to decompose the complex $\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}$ (in the derived category) into pieces supported on characteristic symplectic leaves; the only pieces that contribute to the cohomology in low degree are the ones supported on leaves of small codimension. We start with the surface case $n = 1$, and then bootstrap from there to treat the general case.

5.1 Surface case

Let us briefly recall the description of the Poisson cohomology in the case $n = 1$ from [37, Section 4.4] and [15]. Let us denote by $\mathsf{U} := \mathsf{X} \setminus \mathsf{D}$ the open symplectic leaf, and by

$$j : \mathsf{U} \hookrightarrow \mathsf{X}$$

its open embedding. Since σ is symplectic over U, we have a canonical isomorphism

$$j^*(\mathscr{X}^\bullet_{\mathsf{X}}, \mathrm{d}_\sigma) \cong (\Omega^\bullet_{\mathsf{U}}, \mathrm{d}) \cong \mathbb{C}_{\mathsf{U}}$$

in the constructible derived category of U, which gives, by adjunction, a natural map

$$(\mathscr{X}^\bullet_{\mathsf{X}}, \mathrm{d}_\sigma) \to Rj_*\mathbb{C}_{\mathsf{U}} \tag{5-1}$$

in the constructible derived category of X.

On the other hand, let $\mathsf{D}_{\mathrm{sing}} \subset \mathsf{D}$ denote the scheme-theoretic singular locus of D, cut out locally by the vanishing of a defining equation for D and its partial derivatives, and let

$$i : \mathsf{D}_{\mathrm{sing}} \hookrightarrow \mathsf{X}$$

be the corresponding closed embedding. Its image is the union of all codimension-two characteristic symplectic leaves in X. One easily checks that the restriction of bivectors on X to $\mathsf{D}_{\mathrm{sing}}$ annihilates the image of d_σ, and hence there is a canonical map

$$(\mathscr{X}^\bullet_{\mathsf{X}}, \mathrm{d}_\sigma) \to i_*i^*\mathcal{K}^\vee_{\mathsf{X}}[-2], \tag{5-2}$$

where the right-hand side is a skyscraper sheaf supported on $\mathsf{D}_{\mathrm{sing}}$, viewed as a complex concentrated in degree two with trivial differential. Its stalk at a point $p \in \mathsf{D}_{\mathrm{sing}}$ is canonically identified with the tangent cohomology $\mathsf{H}^1(\mathbb{T}_{\mathsf{D},p})$, ie the space of smoothings of the singularities of D at p.

The maps (5-1) and (5-2) determine the Poisson cohomology as follows:

Theorem 5.1 [37, Theorem 4.7] *If D is reduced with only quasihomogeneous singularities, then the canonical map*

$$(\mathscr{X}^\bullet_{\mathsf{X}}, \mathrm{d}_\sigma) \to Rj_*\mathbb{C}_{\mathsf{U}} \oplus i_*i^*\mathcal{K}^\vee_{\mathsf{X}}[-2]$$

is a quasi-isomorphism. In particular,

$$\mathsf{H}^\bullet_\sigma(\mathsf{X}) \cong \mathsf{H}^\bullet(\mathsf{U}; \mathbb{C}) \oplus \mathsf{H}^0(\mathsf{D}_{\mathrm{sing}}, i^*\mathcal{K}^\vee_{\mathsf{X}})[-2].$$

as graded vector spaces.

5.2 Higher-dimensional case

We now use the results in the previous section to treat the case $n > 1$. Since we are interested in cohomology in degrees up to 2, we need to understand the contributions made by characteristic leaves of codimension ≤ 2. According to Lemma 3.11, these are given by the open symplectic leaf $\mathsf{U}^{[n]} \subset \mathsf{X}^{[n]}$, and the elements $\mathsf{Z} \subset \mathsf{X}^{[n]}$ whose intersection with D is a single singular point of D. The latter give several copies $\mathsf{U}^{[n-1]}$, one for each singular point. We treat the contributions from these leaves as follows.

- **The open leaf** We have the natural open embedding

$$j^{[n]} \colon \mathsf{U}^{[n]} \hookrightarrow \mathsf{X}^{[n]},$$

identifying $\mathsf{U}^{[n]}$ with the open symplectic leaf in $\mathsf{X}^{[n]}$. This gives a natural map

$$(\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}, \mathrm{d}_{\sigma^{[n]}}) \to Rj_* \mathbb{C}_{\mathsf{U}^{[n]}}. \tag{5-3}$$

- **The codimension-two leaves** We have a Poisson rational map

$$\mathsf{X}^{[n-1]} \times \mathsf{X} \dashrightarrow \mathsf{X}^{[n]}$$

given by $(\mathsf{Z}, p) \mapsto \mathsf{Z} \sqcup \{p\}$ whenever $p \notin \mathsf{Z}$. Since U and $\mathsf{D}_{\mathrm{sing}}$ are disjoint by definition, this map restricts to an embedding

$$i^{[n]} \colon \mathsf{U}^{[n-1]} \times \mathsf{D}_{\mathrm{sing}} \hookrightarrow \mathsf{X}^{[n]}$$

with trivial normal bundle. By Lemma 3.11, the image of this embedding is the union of the codimension-two characteristic symplectic leaves of $\mathsf{X}^{[n]}$. Since $\mathsf{U}^{[n-1]}$ is symplectic, we have by Künneth decomposition and Theorem 5.1 that

$$\begin{aligned}(i^{[n]})^*(\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}, \mathrm{d}_{\sigma^{[n]}}) &\cong \mathbb{C}_{\mathsf{U}^{[n-1]}} \boxtimes i^*(Rj_*\mathbb{C}_{\mathsf{U}} \oplus \mathcal{K}^\vee_{\mathsf{X}}[-2]) \\ &\cong (i^{[n]})^* Rj^{[n]}_* \mathbb{C}_{\mathsf{U}^{[n]}} \oplus (\mathbb{C}_{\mathsf{U}^{[n-1]}} \boxtimes i^* \mathcal{K}^\vee_{\mathsf{X}}[-2]),\end{aligned}$$

where $\boxtimes$ denotes the external tensor product of sheaves on the product $\mathsf{U}^{[n-1]} \times \mathsf{D}_{\mathrm{sing}}$. Projecting onto the second summand, we obtain a natural map

$$(\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}, \mathrm{d}_{\sigma^{[n]}}) \to i^{[n]}_*(\mathbb{C}_{\mathsf{U}^{[n-1]}} \boxtimes i^* \mathcal{K}^\vee_{\mathsf{X}})[-2]. \tag{5-4}$$

Combining (5-3) and (5-4), we obtain a canonical map

$$\phi \colon (\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}, \mathrm{d}_{\sigma^{[n]}}) \to Rj_*\mathbb{C}_{\mathsf{U}^{[n]}} \oplus i^{[n]}_*(\mathbb{C}_{\mathsf{U}^{[n-1]}} \boxtimes i^* \mathcal{K}^\vee_{\mathsf{X}})[-2] \tag{5-5}$$

that encapsulates the contributions to $\mathscr{X}^\bullet_{\mathsf{X}^{[n]}}$ made by characteristic leaves of codimension ≤ 2.

Theorem 5.2 *If D is reduced and has only quasihomogeneous singularities, then the induced map*

$$\mathsf{H}^j(\phi) \colon \mathsf{H}^j_{\sigma^{[n]}}(\mathsf{X}^{[n]}) \to \mathsf{H}^j(\mathsf{U}^{[n]}; \mathbb{C}) \oplus (\mathsf{H}^{j-2}(\mathsf{U}^{[n-1]}; \mathbb{C}) \otimes \mathsf{H}^0(i^* \mathcal{K}^\vee_{\mathsf{X}}))$$

is an isomorphism for $j \leq 2$.

Before giving the proof, which will occupy the rest of this section, let us note that the cohomology $\mathsf{H}^\bullet(\mathsf{U}^{[n]}; \mathbb{C})$ can be computed from $\mathsf{H}^\bullet(\mathsf{U}; \mathbb{C})$ using the Göttsche–Soergel formula [16]. In particular, applying the theorem to the degree-two hypercohomology, we obtain the following.

Corollary 5.3 *If* D *is reduced and has only quasihomogeneous singularities, then the space of first-order deformations of* $(\mathsf{X}^{[n]}, \sigma^{[n]})$ *is given, for all* $n \geq 2$, *by*

$$\begin{aligned}\mathsf{H}^2_{\sigma^{[n]}}(\mathsf{X}^{[n]}) &\cong \mathsf{H}^2(\mathsf{U}^{[n]};\mathbb{C}) \oplus \mathsf{H}^0(i^*\mathcal{K}_\mathsf{X}^\vee)\\ &\cong (\mathsf{H}^2(\mathsf{U};\mathbb{C}) \oplus \wedge^2\mathsf{H}^1(\mathsf{U};\mathbb{C}) \oplus \mathbb{C}\cdot[\mathsf{E}]) \oplus \mathsf{H}^0(i^*\mathcal{K}_\mathsf{X}^\vee)\\ &\cong \mathsf{H}^2_\sigma(\mathsf{X}) \oplus \wedge^2\mathsf{H}^1(\mathsf{U};\mathbb{C}) \oplus \mathbb{C}\cdot[\mathsf{E}],\end{aligned}$$

where $[\mathsf{E}] \in \mathsf{H}^2(\mathsf{U}^{[n]};\mathbb{C})$ *is the first Chern class of the exceptional divisor of the Hilbert–Chow morphism.*

Before proving Theorem 5.2, we summarize the strategy as follows. The key observation is that, by a generalization of Hartogs' principle, the Poisson cohomology in degree $\leq m$ is unchanged if we discard a subset of codimension $\geq m+2$, and the restriction map is injective to the complement of a subset of codimension $\geq m+1$. Thus, it is enough to show that

(1) the isomorphism holds outside a subset of codimension three, and

(2) ϕ is surjective in degree two.

For (1), the open subset we use is the locus of elements $\mathsf{Z} \in \mathsf{X}^{[n]}$ for which the intersection $\mathsf{Z} \cap \mathsf{D}$ consists of at most two reduced smooth points of D, or one reduced singular point of D. Both conditions are codimension at most two, so it is a natural subset to consider. Moreover, the case where Z does not touch the singular locus is normal crossings so that it is handled by the results of [27], and the case of a single singular point is immediate. For (2), we will explicitly exhibit the two summands of the codomain of ϕ as the image of elements of ϕ.

Proceeding with the proof, let

$$\mathsf{D}^{[n]} := \mathsf{X}^{[n]} \setminus \mathsf{U}^{[n]} \cong \{\mathsf{Z} \in \mathsf{X}^{[n]} \mid \mathsf{Z} \cap \mathsf{D} \neq \varnothing\}$$

be the degeneracy divisor of the Poisson structure on $\mathsf{X}^{[n]}$. (Note that despite the notation, it is *not* isomorphic to the Hilbert scheme of D.) Let $\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]})$ be the sheaf of logarithmic differential forms in the sense of K Saito [42]. Then, as for any log symplectic manifold, the natural map $\sigma^\sharp\colon \Omega^1_\mathsf{X} \to \mathcal{T}_\mathsf{X}$ induces a commutative triangle

$$\begin{array}{ccc} & (\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}), \mathrm{d}) & \\ \swarrow & & \searrow \\ (\mathscr{X}^\bullet_\mathsf{X}, \mathrm{d}_\sigma) & \longrightarrow & Rj_*\mathbb{C}_\mathsf{U} \end{array} \tag{5-6}$$

where all maps are quasi-isomorphisms away from the singular locus $\mathsf{D}^{[n]}_{\mathrm{sing}}$ of $\mathsf{D}^{[n]}$; see [37, page 695].

For our purposes, it suffices to understand these complexes along the generic part of the singular locus. To this end, define locally closed subvarieties $\mathsf{S}_1, \mathsf{S}_2 \subset \mathsf{D}^{[n]}_{\mathrm{sing}}$ by

$$\mathsf{S}_1 := \{\mathsf{Z} \in \mathsf{X}^{[n]} \mid \mathsf{Z} \cap \mathsf{D} \text{ consists of two reduced points}\} \quad \text{and} \quad \mathsf{S}_2 := i^{[n]}(\mathsf{U}^{[n-1]} \times \mathsf{D}_{\mathrm{sing}}).$$

Evidently S_1 and S_2 are disjoint. Moreover, we have the following.

Lemma 5.4 *The subset*

$$\mathsf{W} := \mathsf{X}^{[n]} \setminus (\mathsf{U}^{[n]} \sqcup \mathsf{D}^{[n]}_{\mathrm{reg}} \sqcup \mathsf{S}_1 \sqcup \mathsf{S}_2)$$

is a closed subvariety of codimension three.

Proof If $\mathsf{Z} \in \mathsf{W}$ then Z contains at least one of the following configurations: three distinct points of D; a singular point and a smooth point; or a nonreduced point tangent to D. Each of these conditions defines a locally closed subvariety of $\mathsf{X}^{[n]}$ that maps birationally onto its image in $\mathsf{X}^{(n)}$, which clearly has codimension three. □

Lemma 5.5 *The morphism ϕ from (5-5) restricts to a quasi-isomorphism over $\mathsf{X}^{[n]} \setminus \mathsf{W}$.*

Proof It is already a quasi-isomorphism away from $\mathsf{D}^{[n]}_{\mathrm{sing}}$, and is furthermore a quasi-isomorphism over the image of $i^{[n]}$, so it suffices to show that it is an isomorphism over S_1. Since S_1 is disjoint from the support of $i^{[n]}_*(\mathbb{C}_{\mathsf{U}^{[n-1]}} \boxtimes i^*\mathcal{K}^\vee_{\mathsf{X}})$, this is equivalent to showing that the map (5-3) is a quasi-isomorphism there. But in a neighbourhood of this locus, $\mathsf{D}^{[n]}$ has only normal crossings singularities, and the only characteristic leaf is the open one. Hence the result follows from our computation of the Poisson cohomology in the normal crossings case [27]; see also [41]. □

Since W has codimension three and $\mathscr{X}^\bullet_{\mathsf{X}}$ is locally free and concentrated in nonnegative degree, the restriction map from Poisson cohomology of $\mathsf{X}^{[n]}$ to that of $\mathsf{X} \setminus \mathsf{W}^{[n]}$ is an isomorphism in degrees 0 and 1, and injective in degree 2, by a standard local cohomology argument (a generalization of Hartogs' principle). Since $\mathsf{U}^{[n]}$ and $i^{[n]}(\mathsf{U}^{[n-1]} \times \mathsf{D}_{\mathrm{sing}})$ are disjoint from W we have the following.

Corollary 5.6 *The map $\mathsf{H}^j(\phi)$ on hypercohomology is an isomorphism in degree* 0 *and* 1, *and injective in degree* 2.

Hence to prove Theorem 5.2, it remains to show that the map $\mathsf{H}^2(\phi)$ is surjective. To this end, we require a lemma:

Lemma 5.7 *If D has only quasihomogeneous singularities, then the natural map*

$$\mathsf{H}^j(\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]})) \to \mathsf{H}^j(\mathsf{U}^{[n]};\mathbb{C})$$

is surjective for $j \le 2$.

Proof The argument is a de Rham analogue of the approach of Göttsche and Soergel [16]. We have the natural maps

$$\begin{array}{ccccc} \mathsf{U}^{[n]} & \longrightarrow & \mathsf{U}^{(n)} & \longleftarrow & \mathsf{U}^n \\ \downarrow{\scriptstyle j^{[n]}} & & \downarrow{\scriptstyle j^{(n)}} & & \downarrow{\scriptstyle j^n} \\ \mathsf{X}^{[n]} & \xrightarrow{r} & \mathsf{X}^{(n)} & \xleftarrow{q} & \mathsf{X}^n \end{array}$$

and corresponding divisors $\mathsf{D}^{[n]}$, $\mathsf{D}^{(n)}$ and D^n; for instance, D^n is the set of n-tuples $(p_1, \dots, p_n) \in \mathsf{X}^n$ such that $p_i \in \mathsf{D}$ for some i.

Since $\mathsf{X}^{(n)} = \mathsf{X}^n/\mathsf{S}_n$ is a finite quotient, it has only rational klt singularities, and hence by [44, Lemma 1.8] and [17, Theorem 1.4] we have

$$r_*\Omega^\bullet_{\mathsf{X}^{[n]}} = (q_*\Omega^\bullet_{\mathsf{X}^n})^{\mathsf{S}_n},$$

where on the right-hand side we have taken invariants with respect to the symmetric group action on X^n. An identical statement holds replacing X everywhere with U. Note that since the exceptional divisor $\mathsf{E} \subset \mathsf{X}^{[n]}$ is not contained in $\mathsf{D}^{[n]}$, a form $\omega \in \Omega^\bullet_{\mathsf{U}^{[n]}}$ extends to a logarithmic form on $\mathsf{X}^{[n]}$ if and only if it extends to a logarithmic form on $\mathsf{X}^{[n]} \setminus \mathsf{E}$ that has no poles on E. Since the maps r and q are étale away from $r(\mathsf{E})$ and relate the corresponding divisors, we conclude that

$$\begin{aligned}(r_*\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}), \mathrm{d}) &\cong (q_*\Omega^\bullet_{\mathsf{X}^n}(\log \mathsf{D}^n)^{\mathsf{S}_n}, \mathrm{d}) \\ &\cong (q_* R(j^n_*)\mathbb{C}_{\mathsf{U}^n})^{\mathsf{S}_n} \\ &\cong Rj^{(n)}_*\mathbb{C}_{\mathsf{U}^{(n)}},\end{aligned}$$

where the second isomorphism is a consequence of the Künneth decomposition and the results of [10] (using that D is quasihomogeneous).

Passing to hypercohomology, we obtain a commutative diagram

$$\begin{array}{ccccc} \mathsf{H}^\bullet(r_*\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}), \mathrm{d}) & \longrightarrow & \mathsf{H}^\bullet(\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}), \mathrm{d}) & \longleftarrow & \mathsf{H}^\bullet(\Omega^\bullet_{\mathsf{X}^{[n]}}) \\ \downarrow \sim & & \downarrow & & \downarrow \sim \\ \mathsf{H}^\bullet(\mathsf{U}^n;\mathbb{C})^{\mathsf{S}_n} \cong \mathsf{H}^\bullet(\mathsf{U}^{(n)};\mathbb{C}) & \longrightarrow & \mathsf{H}^\bullet(\mathsf{U}^{[n]};\mathbb{C}) & \xleftarrow{(j^{[n]})^*} & \mathsf{H}^\bullet(\mathsf{X}^{[n]};\mathbb{C}) \end{array}$$

where the bottom left arrow is the inclusion given by the codimension-zero contribution to the Göttsche–Soergel formula. Said map is an isomorphism in degrees zero and one, and in degree two it is an injection with complement spanned by the class of the exceptional divisor $\mathsf{E} \subset \mathsf{U}^{[n]}$. Since the latter extends canonically to X, it follows that the map $\mathsf{H}^\bullet(\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}), \mathrm{d}) \to \mathsf{H}^\bullet(\mathsf{U}^{[n]};\mathbb{C})$ is surjective in degree two. □

Now let us observe that by applying Bottacin's construction in families, any Poisson deformation of (X, σ) induces a Poisson deformation of $(\mathsf{X}^{[n]}, \sigma^{[n]})$. Under this correspondence, deformations of the singularities of $\mathsf{D} \subset \mathsf{X}$ induce identical deformations of the singularities of $\mathsf{D}^{[n]} \subset \mathsf{X}^{[n]}$ transverse to the codimension-two characteristic leaves. Applied to first-order deformations, this means that we have a canonical morphism $\mathsf{H}^2_\sigma(\mathsf{X}) \to \mathsf{H}^2_{\sigma^{[n]}}(\mathsf{X}^{[n]})$ such that the composition

$$\mathsf{H}^0(i^*\mathcal{K}^\vee_{\mathsf{X}}) \hookrightarrow \mathsf{H}^2_{\sigma^{[n]}}(\mathsf{X}^{[n]}) \xrightarrow{\phi} \mathsf{H}^2(\mathsf{U}^{[n]};\mathbb{C}) \oplus \mathsf{H}^0(i^*\mathcal{K}^\vee_{\mathsf{X}}) \to \mathsf{H}^0(i^*\mathcal{K}^\vee_{\mathsf{X}})$$

is the identity map. Furthermore, the image of $\mathsf{H}^0(i^*\mathcal{K}^\vee_{\mathsf{X}})$ in $\mathsf{H}^2_{\sigma^{[n]}}(\mathsf{X}^{[n]})$ is complementary to the image of $\mathsf{H}^2(\Omega^\bullet_{\mathsf{X}^{[n]}}(\log \mathsf{D}^{[n]}))$, because the latter can only produce deformations in which the divisor $\mathsf{D}^{[n]}$ deforms locally trivially. It follows that $\mathsf{H}^2(\phi)$ is surjective, as desired. This completes the proof of Theorem 5.2.

References

[1] **V I Arnold**, *Remarks on Poisson structures on a plane and on other powers of volume elements*, Trudy Sem. Petrovsk. (1987) 37–46 MR Zbl In Russian; translated in J. Soviet Math. 47 (1989) 2509–2516

[2] **A Beauville**, *Variétés Kähleriennes dont la première classe de Chern est nulle*, J. Differential Geom. 18 (1983) 755–782 MR Zbl

[3] **K A Beck**, **S Sather-Wagstaff**, *A somewhat gentle introduction to differential graded commutative algebra*, from "Connections between algebra, combinatorics, and geometry" (S M Cooper, S Sather-Wagstaff, editors), Springer Proc. Math. Stat. 76, Springer (2014) 3–99 MR Zbl

[4] **V M Bondarenko**, **T G Gerasimova**, **V V Sergeichuk**, *Pairs of mutually annihilating operators*, Linear Algebra Appl. 430 (2009) 86–105 MR Zbl

[5] **F Bottacin**, *Poisson structures on Hilbert schemes of points of a surface and integrable systems*, Manuscripta Math. 97 (1998) 517–527 MR Zbl

[6] **C Brav**, **T Dyckerhoff**, *Relative Calabi–Yau structures, II: Shifted Lagrangians in the moduli of objects*, Selecta Math. 27 (2021) art. id. 63 MR Zbl

[7] **J L Brylinski**, **G Zuckerman**, *The outer derivation of a complex Poisson manifold*, J. Reine Angew. Math. 506 (1999) 181–189 MR Zbl

[8] **I Burban**, **Y Drozd**, *Coherent sheaves on rational curves with simple double points and transversal intersections*, Duke Math. J. 121 (2004) 189–229 MR Zbl

[9] **D Calaque**, *Lagrangian structures on mapping stacks and semi-classical TFTs*, from "Stacks and categories in geometry, topology, and algebra" (T Pantev, C Simpson, B Toën, M Vaquié, G Vezzosi, editors), Contemp. Math. 643, Amer. Math. Soc., Providence, RI (2015) 1–23 MR Zbl

[10] **F J Castro-Jiménez**, **L Narváez-Macarro**, **D Mond**, *Cohomology of the complement of a free divisor*, Trans. Amer. Math. Soc. 348 (1996) 3037–3049 MR Zbl

[11] **C Debord**, *Holonomy groupoids of singular foliations*, J. Differential Geom. 58 (2001) 467–500 MR Zbl

[12] **G Ellingsrud**, **S A Strømme**, *On a cell decomposition of the Hilbert scheme of points in the plane*, Invent. Math. 91 (1988) 365–370 MR Zbl

[13] **J Fogarty**, *Algebraic families on an algebraic surface*, Amer. J. Math. 90 (1968) 511–521 MR Zbl

[14] **I M Gelfand**, **V A Ponomarev**, *Indecomposable representations of the Lorentz group*, Uspehi Mat. Nauk 23 (1968) 3–60 MR Zbl In Russian; translated in Russian Math. Surv. 23 (1968) 1–58

[15] **R Goto**, *Unobstructed deformations of generalized complex structures induced by C^∞ logarithmic symplectic structures and logarithmic Poisson structures*, from "Geometry and topology of manifolds" (A Futaki, R Miyaoka, Z Tang, W Zhang, editors), Springer Proc. Math. Stat. 154, Springer (2016) 159–183 MR Zbl

[16] **L Göttsche**, **W Soergel**, *Perverse sheaves and the cohomology of Hilbert schemes of smooth algebraic surfaces*, Math. Ann. 296 (1993) 235–245 MR Zbl

[17] **D Greb**, **S Kebekus**, **S J Kovács**, **T Peternell**, *Differential forms on log canonical spaces*, Publ. Math. Inst. Hautes Études Sci. 114 (2011) 87–169 MR Zbl

[18] **M Gualtieri**, **S Li**, *Symplectic groupoids of log symplectic manifolds*, Int. Math. Res. Not. 2014 (2014) 3022–3074 MR Zbl

[19] **M Haiman**, *t,q-Catalan numbers and the Hilbert scheme*, Discrete Math. 193 (1998) 201–224 MR Zbl

[20] **J Herzog**, *Komplexe, Auflösungen und Dualität in der lokalen Algebra*, Habilitationsschrift, Universität Regensburg (1974)

[21] **N Hitchin**, *Deformations of holomorphic Poisson manifolds*, Mosc. Math. J. 12 (2012) 567–591 MR Zbl

[22] **A A Iarrobino**, *Punctual Hilbert schemes*, Mem. Amer. Math. Soc. 188, Amer. Math. Soc., Providence, RI (1977) MR Zbl

[23] **C Ingalls**, *Quantizable orders over surfaces*, J. Algebra 207 (1998) 616–656 MR Zbl

[24] **R C Laubenbacher**, **B Sturmfels**, *A normal form algorithm for modules over $k[x,y]/\langle xy\rangle$*, J. Algebra 184 (1996) 1001–1024 MR Zbl

[25] **L S Levy**, *$\mathbb{Z}G_n$-modules, G_n cyclic of square-free order n*, J. Algebra 93 (1985) 354–375 MR

[26] **L S Levy**, *Modules over Dedekind-like rings*, J. Algebra 93 (1985) 1–116 MR Zbl

[27] **M Matviichuk**, **B Pym**, **T Schedler**, *A local Torelli theorem for log symplectic manifolds*, preprint (2020) arXiv 2010.08692

[28] **S Mukai**, *Symplectic structure of the moduli space of sheaves on an abelian or $K3$ surface*, Invent. Math. 77 (1984) 101–116 MR Zbl

[29] **H Nakajima**, *Lectures on Hilbert schemes of points on surfaces*, Univ. Lect. Ser. 18, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[30] **L A Nazarova**, **A V Roĭter**, *Finitely generated modules over a dyad of two local Dedekind rings, and finite groups which possess an abelian normal divisor of index p*, Izv. Akad. Nauk SSSR Ser. Mat. 33 (1969) 65–89 MR In Russian; translated in Math. USSR-Izv. 3 (1969) 65–86

[31] **L A Nazarova**, **A V Roĭter**, **V V Sergeichuk**, **V M Bondarenko**, *Application of modules over a dyad for the classification of finite p-groups possessing an abelian subgroup of index p and of pairs of mutually annihilating operators*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 28 (1972) 69–92 MR Zbl In Russian; translated in J. Soviet Math. 3 (1975) 636–653

[32] **T A Nevins**, **J T Stafford**, *Sklyanin algebras and Hilbert schemes of points*, Adv. Math. 210 (2007) 405–478 MR Zbl

[33] **J O'Halloran**, *A simple proof of the Gerstenhaber–Hesselink theorem for nilpotent matrices*, Comm. Algebra 15 (1987) 2017–2023 MR Zbl

[34] **T Pantev**, **B Toën**, **M Vaquié**, **G Vezzosi**, *Shifted symplectic structures*, Publ. Math. Inst. Hautes Études Sci. 117 (2013) 271–328 MR Zbl

[35] **A Polishchuk**, *Algebraic geometry of Poisson brackets*, J. Math. Sci. (NY) 84 (1997) 1413–1444 MR Zbl

[36] **B Pym**, **E Rains**, *Shifted Poisson structures, noncommutative surfaces and stacks of Calabi–Yau categories*, in preparation

[37] **B Pym**, **T Schedler**, *Holonomic Poisson manifolds and deformations of elliptic algebras*, from "Geometry and physics, II" (J E Andersen, A Dancer, O García-Prada, editors), Oxford Univ. Press (2018) 681–703 MR Zbl

[38] **E M Rains**, *The noncommutative geometry of elliptic difference equations*, preprint (2016) arXiv 1607.08876

[39] **E M Rains**, *The birational geometry of noncommutative surfaces*, preprint (2019) arXiv 1907.11301

[40] **Z Ran**, *Incidence stratifications on Hilbert schemes of smooth surfaces, and an application to Poisson structures*, Int. J. Math. 27 (2016) art. id. 1650006 MR Zbl

[41] **Z Ran**, *Deformations of holomorphic pseudo-symplectic Poisson manifolds*, Adv. Math. 304 (2017) 1156–1175 MR Zbl

[42] **K Saito**, *Theory of logarithmic differential forms and logarithmic vector fields*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 27 (1980) 265–291 MR Zbl

[43] **T Schürg**, **B Toën**, **G Vezzosi**, *Derived algebraic geometry, determinants of perfect complexes, and applications to obstruction theories for maps and complexes*, J. Reine Angew. Math. 702 (2015) 1–40 MR Zbl

[44] **J H M Steenbrink**, *Mixed Hodge structure on the vanishing cohomology*, from "Real and complex singularities" (P Holm, editor), Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands (1977) 525–563 MR Zbl

[45] **A Thomas**, *Generalized punctual Hilbert schemes and $\mathfrak{g}$-complex structures*, Int. J. Math. 33 (2022) art. id. 2250004 MR Zbl See also preprint (2019) arXiv 1910.08504

[46] **B Toën**, **M Vaquié**, *Moduli of objects in dg-categories*, Ann. Sci. École Norm. Sup. 40 (2007) 387–444 MR Zbl

[47] **A Weinstein**, *The modular automorphism group of a Poisson manifold*, J. Geom. Phys. 23 (1997) 379–394 MR Zbl

MM, TS: *Department of Mathematics, Imperial College London*
London, United Kingdom

BP: *Department of Mathematics and Statistics, McGill University*
Montreal, QC, Canada

m.matviichuk@imperial.ac.uk, brent.pym@mcgill.ca, t.schedler@imperial.ac.uk

Proposed: Jim Bryan Received: 25 January 2023
Seconded: Sándor J Kovács, Arend Bayer Revised: 13 May 2024

msp

Geometry & Topology **29**:4 (2025) 2105–2173
DOI: 10.2140/gt.2025.29.2105
Published: 27 June 2025

Mutations and faces of the Thurston norm ball dynamically represented by multiple distinct flows

ANNA PARLAK

A pseudo-Anosov flow on a hyperbolic 3-manifold dynamically represents a top-dimensional face F of the Thurston norm ball if the cone on F is dual to the cone spanned by the homology classes of closed orbits of the flow. Fried showed that for every fibered face of the Thurston norm ball there is a unique, up to isotopy and reparametrization, flow which dynamically represents the face. Using veering triangulations we have found that there are nonfibered faces of the Thurston norm ball which are dynamically represented by multiple topologically inequivalent flows. This raises the question of how distinct flows representing the same face are related.

We define combinatorial mutations of veering triangulations along surfaces that they carry. We give sufficient and necessary conditions for the mutant triangulation to be veering. After appropriate Dehn filling, these veering mutations correspond to transforming one 3-manifold M with a pseudo-Anosov flow transverse to an embedded surface S into another 3-manifold admitting a pseudo-Anosov flow transverse to a surface homeomorphic to S. We show that a nonfibered face of the Thurston norm ball can be dynamically represented by two distinct flows that differ by a veering mutation. Furthermore, one of the discussed pairs of homeomorphic veering mutants can be used to construct counterexamples to the classification theorem of Anosov flows on Bonatti–Langevin manifolds published in the 90s.

57K30, 57Q15; 37D20, 57K32

1 Introduction

Let M be a compact oriented 3-manifold M whose interior admits a complete hyperbolic structure. The *Thurston norm* on $H_2(M, \partial M; \mathbb{R})$ measures the minimal topological complexity of surfaces that represent a homology class; see Thurston [50]. It has been intensively studied in various different contexts. It is related to finite depth foliations (see Gabai [18]), the Alexander polynomial (see Dunfield [9] and McMullen [35]), the L^2-torsion function (see Friedl and Lück [16]), Floer homology (see Ozsváth and Szabó [43]) and many other aspects of 3-dimensional topology. Here we focus on the connection between the Thurston norm on $H_2(M, \partial M; \mathbb{R})$ and nonsingular flows on M. Originally this connection was drawn by Fried [13; 14] and Mosher [41; 42]. The topic has reemerged recently in work of Landry [28; 29; 30] and Landry, Minsky and Taylor [31; 32], where they relate the Thurston norm with veering triangulations.

Since the unit norm ball $\mathbb{B}_{\mathrm{Th}}$ of the Thurston norm is a compact polytope [50, Theorem 2], we can speak about its *faces*. Thurston proved that all ways in which M fibers over the circle are encoded by finitely many, potentially zero, top-dimensional faces of $\mathbb{B}_{\mathrm{Th}}$, called *fibered faces* [50, Theorem 3]. The first known connection between pseudo-Anosov flows and the Thurston norm concerned only these faces. Assuming that M is closed, Fried proved that associated to a fibered face $\mathtt{F}$ there is a unique, up to isotopy and reparametrization, pseudo-Anosov flow Ψ on M with the property that a class $\eta \in H_2(M; \mathbb{Z})$ can be represented by a cross-section to Ψ if and only if η is in the interior of the cone $\mathbb{R}_+ \cdot \mathtt{F}$ [14, Theorem 7]. Mosher extended Fried's result by showing that $\eta \in H_2(M; \mathbb{Z})$ can be represented by a surface that is *almost transverse* to Ψ if and only if η is in $\mathbb{R}_+ \cdot \mathtt{F}$ [40, Theorem 1.4]. Results of Fried and Mosher are stated for closed manifolds, but they can be generalized to the case of flows on 3-manifolds with toroidal boundary whose interior admits a complete hyperbolic structure; see [30, Theorem 3.5]. In this case, the relevant flows are obtained from pseudo-Anosov flows by *blowing-up* finitely many closed orbits into toroidal boundary components; see Mosher [42, Section 3.2] and Bonatti and Iakovoglou [4, Section 3.6].

The relationship between pseudo-Anosov flows and the Thurston norm extends beyond the fibered case. In this more general setup, we consider flows which do not admit cross-sections. Such flows are called *noncircular*. Given a potentially noncircular flow Ψ on M denote, by $\mathcal{C}(\Psi)$ the cone in $H_2(M, \partial M; \mathbb{R})$ spanned by the homology classes whose algebraic intersection with the homology classes of closed orbits of Ψ is nonnegative. Following Mosher [41], we say that Ψ *dynamically represents* a (not necessarily fibered, not necessarily top-dimensional) face $\mathtt{F}$ of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ if $\mathcal{C}(\Psi) = \mathbb{R}_+ \cdot \mathtt{F}$ and $\mathbb{R}_+ \cdot \mathtt{F}$ is the maximal cone in $H_2(M, \partial M; \mathbb{R})$ in which the Thurston norm agrees with the minus Euler class of the normal plane bundle to Ψ. From the results of Fried and Mosher mentioned in the last paragraph, it follows that every fibered face is dynamically represented by a flow which is unique up to isotopy and reparametrization. In the nonfibered case, Mosher found sufficient conditions on a noncircular flow to dynamically represent a face of the Thurston norm ball [41, Theorem 2.7] and showed that there are noncircular flows representing nonfibered faces [41, Section 4]. However, the question of whether, for every nonfibered face $\mathtt{F}$ of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$, there is a (blown-up) pseudo-Anosov flow Ψ which dynamically represents $\mathtt{F}$ remains open.

We answer two closely related questions. First, if there is a flow which dynamically represents a nonfibered face, is this flow necessarily unique, up to isotopy and reparametrization? In Section 5 we give explicit examples of flows which represent the same nonfibered face of the Thurston norm ball but are not even topologically equivalent, thus showing that the answer to this question is negative; see Theorem 5.2. These examples have been found using *veering triangulations*, a combinatorial tool to study pseudo-Anosov flows. We refer the reader to Section 2.3 for an outline of the connection between veering triangulations and pseudo-Anosov flows.

Once we know that a nonfibered face can be dynamically represented by two topologically inequivalent flows we may ask how the two distinct flows which dynamically represent the same face are related. Veering triangulations can be helpful in solving this problem as well. The veering census (see Giannopolous, Schleimer and Segerman [20]) and other computational tools to study triangulations [7; 8; 47] can be used to find many examples of veering triangulations that *combinatorially represent* the same face of the Thurston norm ball. At the beginning of Section 4 and in Section 4.2 we briefly outline what the search for appropriate examples boils down to. Since veering triangulations are finite objects that satisfy very restrictive conditions, comparing two veering triangulations is easier than comparing their underlying flows. An analysis of certain examples of veering triangulations which combinatorially represent the same face of the Thurston norm ball led us to define *combinatorial mutations of veering triangulations* along surfaces that they carry. Our main goal is to carefully study these operations and demonstrate that in special cases they can yield distinct flows representing the same face of the Thurston norm ball.

1.1 Combinatorial mutations of veering triangulations

A veering triangulation $\mathcal{V}$ of a 3-manifold M is determined by three pieces of combinatorial data: an ideal triangulation $\mathcal{T}$, a taut structure α on $\mathcal{T}$ and a smoothing of the dual spine of $\mathcal{T}$ into a branched surface $\mathcal{B}$ with certain properties; see Definitions 2.1 and 2.5. Associated to $(\mathcal{T}, \alpha)$ there is a finite system of *branch equations* such that if w is a nonzero nonnegative integral solution to this system then w determines a surface S_w which is *carried* by $(\mathcal{T}, \alpha)$; see Section 2.1.3. We also say that S_w is carried by $\mathcal{V}$. We will restrict our attention to those carried surfaces whose boundary components lying in the same boundary component of M have the same orientation. We say that such surfaces are *properly carried*. We characterize surfaces properly carried by a veering triangulation in Corollary 2.14. If S_w is properly carried then the result of cutting M along a surface S_w^ϵ properly embedded in M and homotopic to S_w is a *sutured manifold*; see Section 3.3. We denote it by $M|S_w^\epsilon$.

The surface S_w^ϵ is naturally equipped with an ideal triangulation $\mathcal{Q}_{\mathcal{V},w}$ induced from $\mathcal{V}$. Let $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ be the group of orientation-preserving combinatorial automorphisms of $\mathcal{Q}_{\mathcal{V},w}$. Associated to $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ there always is a *mutant manifold* M^φ, obtained from the sutured manifold $M|S_w^\epsilon$ by identifying the two copies of S_w^ϵ in its boundary via φ. Our goal is to mimic this construction in the combinatorial setup of triangulations. Unfortunately, it is not as straightforward as it may sound. The main difficulty is the fact that S_w is often not embedded. Thus we may view cutting $\mathcal{T}$ along S_w as equivalent to cutting it along a certain branched surface F_w which fully carries S_w; see Section 3.4. This in turn causes the problem of not being able to use φ directly to reglue the top boundary F_w^+ of $\mathcal{T}|F_w$ to its bottom boundary F_w^-. In Section 3.5 we define a *regluing map* $r(\varphi)\colon F_w^+ \to F_w^-$ determined by φ and use it to define a *mutant* *triangulation* $\mathcal{T}^\varphi$. Without further assumptions on φ, not only can this triangulation fail to be veering, but it also might not be a triangulation of M^φ. We deal with these issues in Sections 3.6 and 3.7.

Studying mutations has a long history, particularly in knot theory; see for instance Dunfield, Garoufalidis, Shumakovitch and Thistlethwaite [10], Kirk and Livingston [26], Millichap [37] and Morton and Traczyk [39]. Mutant knots share many properties, and much work on mutations concentrates on establishing which knot invariants distinguish mutants. Another thread in the theory is finding sufficient conditions on a surface S and its homeomorphism φ so that M and M^φ share some property. For instance, in [48, Theorem 4.4] Ruberman considered mutations of hyperbolic 3-manifolds and found sufficient conditions for the mutant manifold M^φ to be hyperbolic and have the same hyperbolic volume as M. Our goal to find conditions under which $\mathcal{T}^\varphi$ is a veering triangulation of M^φ fits into this second framework.

1.2 Properties of the mutant triangulation

To analyze the homeomorphism type of the manifold underlying $\mathcal{T}^\varphi$, we introduce the notion of *edge product disks*, a special type of product disks in the sutured manifold $M|S_w^\epsilon$; see Section 3.4. Then we define what it means for $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ to *misalign edge product disks*; see Definition 3.8. Using this we prove:

Theorem 3.10 *The mutant triangulation $\mathcal{T}^\varphi$ is an ideal triangulation of M^φ if and only if φ misaligns edge product disks.*

To find sufficient conditions for the mutant triangulation to be veering, we first need to ensure that it admits a taut structure. It turns out that for this it also suffices to assume that φ misaligns edge product disks. However, we prove a slightly stronger result:

Proposition 3.16 *The mutant triangulation $\mathcal{T}^\varphi$ admits a taut structure if and only if every vertical annulus or Möbius band in M^φ lies in a prismatic region of M^φ.*

The backward direction of Proposition 3.16 is proved by explicitly constructing a taut structure α^φ on $\mathcal{T}^\varphi$ from the taut structure α on $\mathcal{T}$. We say that $(\mathcal{T}, \alpha)$ and $(\mathcal{T}^\varphi, \alpha^\varphi)$ are *taut mutants*.

Intuitively, the condition that appears in the above proposition means that φ might align edge product disks, but it does so in a way which is not visible from the perspective of $\mathcal{T}^\varphi$; see Lemma 3.15. Nonetheless, in light of Theorem 3.10 it is convenient to assume that φ misalign edge product disks, so that we deal only with triangulations of M^φ. This assumption is further justified by the fact that in Proposition 3.17 we prove that when φ aligns edge product disks M^φ cannot admit a veering triangulation.

To obtain sufficient conditions on the taut triangulation $(\mathcal{T}^\varphi, \alpha^\varphi)$ to admit a veering structure, we make use of the branched surface $\mathcal{B}$ defining the veering structure on $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$. This branched surface intersects the 2-skeleton of $\mathcal{T}$ in a train track; see Figure 6. Therefore any surface S_w carried by $\mathcal{V}$ inherits a train track $\tau_{\mathcal{V},w}$ which is dual to its ideal triangulation $\mathcal{Q}_{\mathcal{V},w}$. By $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ we denote the subgroup of $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ consisting of orientation-preserving combinatorial automorphisms of $\mathcal{Q}_{\mathcal{V},w}$ which preserve $\tau_{\mathcal{V},w}$.

Theorem 3.19 *Let S_w be a surface properly carried by a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ of M. Suppose that $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ misaligns edge product disks. If additionally $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ then $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure.*

Under the assumptions of this theorem, the branched surface $\mathcal{B}$ dual to $\mathcal{T}$ mutates into a branched surface $\mathcal{B}^\varphi$ that is dual to $\mathcal{T}^\varphi$ and satisfies Definition 2.5. We say that $\mathcal{V}^\varphi = (\mathcal{T}^\varphi, \alpha^\varphi, \mathcal{B}^\varphi)$ is obtained from $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ by a *veering mutation* or that $\mathcal{V}^\varphi$ and $\mathcal{V}$ are *veering mutants*.

Observe that Theorem 3.19 gives a sufficient condition for a taut mutant $(\mathcal{T}^\varphi, \alpha^\varphi)$ to be veering. It is, however, possible that $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure even when $\varphi \notin \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$. This can happen whenever, after cutting $\mathcal{T}$ along F_w, the cut triangulation $\mathcal{T}|F_w$ admits a veering structure $\mathcal{B}^*|F_w$ which mutates into a branched surface that is dual to $\mathcal{T}^\varphi$ and satisfies Definition 2.5. If $\mathcal{B}^*|F_w \neq \mathcal{B}|F_w$ we do not consider such triangulations to be veering mutants. This construction can be used to prove a generalization of Theorem 3.19 giving a sufficient and necessary conditions on a taut mutant of a veering triangulation to be veering.

Theorem 3.21 *Let S_w be a surface properly carried by a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ of M. Suppose that $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ misaligns edge product disks. The taut triangulation $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure if and only if there is a veering structure $\mathcal{B}^*|F_w$ on $(\mathcal{T}|F_w, \alpha|F_w)$ such that the isomorphism $\varphi\colon \mathcal{Q}^+_{\mathcal{V},w} \to \mathcal{Q}^-_{\mathcal{V},w}$ sends $\tau^{*\,+}_{\mathcal{V},w}$ to $\tau^{*\,-}_{\mathcal{V},w}$.*

In Section 3.8 we give an example of a pair of veering triangulations which are taut mutants but not veering mutants. This proves that the generalization appearing in Theorem 3.21 is not just theoretical, but actually arises in practice. In the same subsection we also define a *veering mutation with insertion*, a certain generalization of a veering mutation where the related triangulations have different numbers of tetrahedra.

1.3 Homeomorphic veering mutants

In Section 4 we analyze a few examples of homeomorphic veering mutants. Apart from illustrating our constructions, we use these examples to establish the following facts connecting veering mutations and faces of the Thurston norm ball:

Fact 4.2 (Veering mutations and faces of the Thurston norm ball)

(1) *There are nonfibered faces of the Thurston norm ball that can be represented by two combinatorially nonisomorphic veering mutants.*

(2) *A veering mutation along a surface representing a class lying at the boundary of the cone on a fibered face may yield a veering triangulation representing a nonfibered face of the Thurston norm ball of the mutant manifold.*

Analyzing two veering mutants of the complement of the 10^3_{12} link leads to the following discovery:

Fact 4.7 *The complement of the* 10^3_{12} *link admits two fibrations over the circle such that*:

- *The fiber is a genus-two surface with four punctures.*
- *The monodromy of one fibration is obtained from the monodromy of the other fibration by post-composing it with an involution. In particular, the stretch factors of the monodromies are equal.*
- *The monodromies are not conjugate in the mapping class group of a genus-two surface with four punctures.*

The last part of Fact 4.7 follows from the observation that the Euler classes of the two fibrations lie in different orbits under the action of $\mathrm{Homeo}(M)$ on $H^2(M, \partial M; \mathbb{R})$. Examples of such fibrations of the same manifold were known previously; see for instance McMullen and Taubes [36, Theorem 1.2]. What is new here is that we get fiber bundles which are not isomorphic, even though both their total spaces and fibers are homeomorphic, and the stretch factors of their monodromies are the same.

1.4 Multiple distinct flows dynamically representing the same face of the Thurston norm ball

Given a veering triangulation $\mathcal{V}$ of M, it is possible to construct a transitive pseudo-Anosov flow Ψ on a closed Dehn filling N of M, provided that a certain natural condition on the Dehn filling slopes is satisfied; see Agol and Tsang [1, Theorem 5.1], stated here as Theorem 2.20. Let Ψ° be the blown-up flow on M. If $\mathcal{V}$ combinatorially represents a face $\mathtt{F}$ of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ then Ψ° dynamically represents $\mathtt{F}$; see Landry, Minsky and Taylor [31, Theorem 6.1], stated here as Theorem 2.29. Under additional assumptions on the Dehn filling slopes, there is also a face $\mathtt{F}_N$ of the Thurston norm ball in $H_2(N; \mathbb{R})$ such that $\mathbb{R}_+ \cdot \mathtt{F}_N = \mathcal{C}(\Psi)$; see Landry [29, Theorem A], stated here as Theorem 2.30. These results are the main ingredients to prove the following theorem:

Theorem 5.2 *There are nonfibered faces of the Thurston norm ball that can be dynamically represented by two topologically inequivalent flows.*

In the case of manifolds with nonempty boundary, we show that a nonfibered face can be dynamically represented by two topologically inequivalent blown-up Anosov flows constructed from a pair of homeomorphic veering mutants. Unfortunately, the corresponding Anosov flows on the Dehn-filled manifold cannot be used to prove the theorem in the closed case because the manifold is toroidal. For this reason, we refer to a different pair of veering triangulations which represent the same face of the Thurston norm ball and, after appropriate Dehn filling, yield transitive pseudo-Anosov flows on a hyperbolic 3-manifold. These veering triangulations are not related by a veering mutation; see Fact 4.6. In particular, it is worth emphasizing that not all pairs of veering triangulations combinatorially representing the same face of the Thurston norm ball are related by a veering mutation or even a veering mutation with insertion.

Remark Although Anosov flows underlying homeomorphic veering mutants $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ discussed in Section 4.1 cannot be used to prove Theorem 5.2 in the closed case, they have another interesting feature. The closed manifold N obtained by Dehn filling $M \cong M^{\varrho\sigma}$ along the boundary of the mutating surface is a graph manifold constructed from the orientable circle bundle over a 2-holed $\mathbb{R}P^2$ by identifying its two toroidal boundary components. In [2] Barbot calls such manifolds *BL-manifolds*. The Anosov flows Ψ and $\Psi^{\varrho\sigma}$ on N built from $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$, respectively, are counterexamples to the claim, which appears as [2, Theorem B(2)], that all non-$\mathbb{R}$-covered Anosov flows on a fixed BL-manifold are topologically equivalent; see Remark 5.3.

1.5 Polynomial invariants of veering triangulations representing the same face of the Thurston norm ball

In [32] Landry, Minsky, and Taylor introduced two polynomial invariants of veering triangulations: the *taut polynomial* and the *veering polynomial*. They proved that the taut polynomial generalizes the *Teichmüller polynomial*, an invariant of a fibered face of the Thurston norm ball defined by McMullen in [34], to faces of the Thurston norm ball that are combinatorially represented by veering triangulations [32, Theorem 7.1]. In Table 3 we list the taut and veering polynomials of veering triangulations representing the same face of the Thurston norm ball that we discussed in Section 4. We deduce that in the nonfibered case the taut and veering polynomials are not invariants of faces of the Thurston norm ball combinatorially represented by veering triangulations.

Fact 6.1 *There are nonfibered faces of the Thurston norm ball that can be combinatorially represented by two distinct veering triangulations with different taut polynomials, and different veering polynomials.*

1.6 Further questions

In Section 7 we speculate about what happens on the level of flows when we perform a veering mutation. We also ask a few questions concerning veering mutations, faces of the Thurston norm ball dynamically represented by multiple distinct flows, connections between this work and a recent result of Barthelmé, Frankel and Mann [3] characterizing topologically inequivalent pseudo-Anosov flows on a fixed manifold, and hyperbolic volumes of veering mutants.

Acknowledgements

I am grateful to Michael Landry and Chi Cheuk Tsang for answering my questions about their work connecting veering triangulations and pseudo-Anosov flows. I thank Saul Schleimer for discussions on veering triangulations analyzed in Section 4.1. I thank Thomas Barthelmé and Chi Cheuk Tsang for discussions that led to Remark 5.3. Finally, I thank the referee for their valuable comments and suggestions.

This project was partially supported by the Simons Investigator Award 409745 of Vladimir Marković.

2 Veering triangulations and pseudo-Anosov flows

Let M be a compact oriented 3-manifold. By an *ideal triangulation* of M we mean an expression of $M - \partial M$ as a collection of finitely many ideal tetrahedra with triangular faces identified in pairs by homeomorphisms which send vertices to vertices. Links of ideal vertices of the triangulation correspond to boundary components of M.

Let $\mathcal{T}$ be a finite ideal triangulation of M. Every triangular face of $\mathcal{T}$ has two *embeddings* into two, not necessarily distinct, tetrahedra. Every edge of $\mathcal{T}$ has finitely many embeddings into tetrahedra of $\mathcal{T}$ and the same number of embeddings into faces of $\mathcal{T}$. By *edges of a triangle/tetrahedron* or *triangles of a tetrahedron* we mean embeddings of these ideal simplices into the boundary of a higher-dimensional ideal simplex. Similarly, by *triangles/tetrahedra attached to an edge* we mean triangles/tetrahedra in which the edge is embedded, together with this embedding. Observe that triangles/tetrahedra attached to an edge can be circularly ordered, and hence we can speak about consecutive triangles/tetrahedra attached to an edge.

Every ideal triangulation $\mathcal{T}$ determines a 2-dimensional complex $\mathcal{D}$ dual to $\mathcal{T}$, called the *dual spine* of $\mathcal{T}$. For every tetrahedron t of $\mathcal{T}$ there is a vertex $v = v(t)$ of $\mathcal{D}$. If tetrahedra t_1 and t_2 of $\mathcal{T}$ admit faces f_1 and f_2, respectively, which are identified in $\mathcal{T}$, then in $\mathcal{D}$ there is an edge joining their dual vertices v_1 and v_2. Finally, each edge e of $\mathcal{T}$ gives a 2-cell of $\mathcal{D}$ which is glued along the edges of $\mathcal{D}$ that are dual to the consecutive triangles attached to e. Since there are no higher-dimensional cells in $\mathcal{D}$, and 0- and 1-cells have special names, we will often refer to the 2-cells of $\mathcal{D}$ as just "cells".

Translating between properties of an ideal triangulation and properties of its dual spine is straightforward. Throughout the paper we freely alternate between these two perspectives depending on which one is more useful in a given context.

2.1 Taut triangulations

In [27, Introduction] Lackenby introduced *taut ideal triangulations* of 3-manifolds. Using the duality between an ideal triangulation and its dual spine we define tautness of an ideal triangulation in terms of properties of its dual spine.

Definition 2.1 A *taut structure* α on an ideal triangulation $\mathcal{T}$ is a choice of orientations on the edges of its dual spine $\mathcal{D}$ such that

(1) every vertex v of $\mathcal{D}$ has two incoming edges and two outgoing edges,

(2) every cell s of $\mathcal{D}$ has exactly one vertex b_s such that the two edges of s adjacent to v both point out of b_s,

(3) every cell s of $\mathcal{D}$ has exactly one vertex t_s such that the two edges of s adjacent to v both point into t_s.

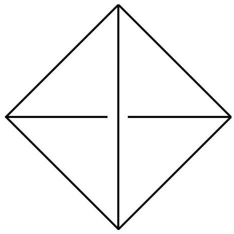


Figure 1: The taut tetrahedron.

A *taut triangulation* is a pair $(\mathcal{T}, \alpha)$, where $\mathcal{T}$ is an ideal triangulation, and α is a taut structure on $\mathcal{T}$. If $(\mathcal{T}, \alpha)$ is taut then for every cell s of the dual spine of $\mathcal{T}$ the vertex from Definition 2.1(2) is called the *bottom vertex* of s, and the vertex from Definition 2.1(3) is called the *top vertex* of s.

Remark 2.2 Taut triangulations are often called *transverse taut triangulations*; see eg [12; 45; 46].

Intuitively, tautness of an ideal triangulation gives an upwards direction which is consistent throughout the whole triangulation. Under the duality, orientations on the edges of $\mathcal{D}$ translate into coorientations on the faces of $\mathcal{T}$. If $(\mathcal{T}, \alpha)$ is taut then, by Definition 2.1(1), every tetrahedron t of $\mathcal{T}$ has two faces whose coorientations point into t, and two faces whose coorientations point out of t. We call the pair of faces whose coorientations point out of t the *top faces* of t and the pair of faces whose coorientations point into t the *bottom faces* of t. We also define the *top diagonal* of t to be the common edge of the two top faces of t and the *bottom diagonal* of t to be the common edge of the two bottom faces of t. By Definition 2.1(2), every edge of $\mathcal{T}$ is embedded as the top diagonal in precisely one tetrahedron of $\mathcal{T}$. Similarly, Definition 2.1(3) implies that every edge of $\mathcal{T}$ is embedded as the bottom diagonal in precisely one tetrahedron of $\mathcal{T}$. We encode a taut structure on a tetrahedron by drawing it as a quadrilateral with two diagonals — one on top of the other; see Figure 1. Then the convention is that coorientations on all faces point towards the reader. In other words, we view the tetrahedron *from above*.

2.1.1 The horizontal branched surface Let e be an edge of a taut triangulation $(\mathcal{T}, \alpha)$. To every embedding $\epsilon(e)$ of e into a tetrahedron t of $\mathcal{T}$ we assign a dihedral angle 0 or π in the following way. If $\epsilon(e)$ is either the top or the bottom diagonal of t we assign to $\epsilon(e)$ the angle π. Otherwise we assign to $\epsilon(e)$ the angle 0. This equips the 2-skeleton of $\mathcal{T}$ with a structure of a branched surface with branch locus equal to the 1-skeleton of $\mathcal{T}$; see Figure 2. We call it the *horizontal branched surface* associated to $\mathcal{T}$, and denote it by $\mathcal{H}$.

Let $-\alpha$ denote the taut structure on $\mathcal{T}$ obtained by reversing orientations of all edges of the dual spine $\mathcal{D}$ of $\mathcal{T}$. Taut triangulations $(\mathcal{T}, \alpha)$ and $(\mathcal{T}, -\alpha)$ determine the same dihedral angles between consecutive faces attached to edges of $\mathcal{T}$ and thus the same horizontal branched surface. We call $(\mathcal{T}, \pm\alpha)$ a *taut angle structure* on $\mathcal{T}$. This term will appear in Section 2.2.3.

2.1.2 The boundary track Recall that boundary components of M correspond to links of vertices of $\mathcal{T}$. An ideal vertex of tetrahedron t of $\mathcal{T}$ meets three faces of t. Thus an ideal triangulation $\mathcal{T}$ of M

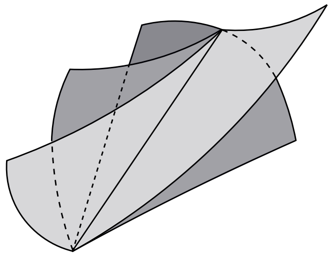


Figure 2: The horizontal branched surface associated to a taut triangulation.

determines a triangulation $\partial\mathcal{T}$ of ∂M. If $\mathcal{T}$ is additionally taut, the smoothing of the 2-skeleton $\mathcal{T}^{(2)}$ into the horizontal branched surface determines a smoothing of $\partial\mathcal{T}$ into a train track. We call this train track the *boundary track* of $\mathcal{T}$ and denote it by β. If an ideal vertex of t meets faces f_1, f_2 and f_3 of t then exactly one pair (f_i, f_j), for $i \neq j$, is adjacent either along the top or along the bottom diagonal of t. In the construction of the horizontal branched surface of $\mathcal{T}$ we assign to such a pair the dihedral angle π, and to the remaining pairs we assign the dihedral angle 0. Thus every complementary region of β is a bigon. This has important implications for the topology of ∂M. If τ is a train track in a surface S without boundary then the Euler characteristic of S is equal to half the sum of indices of all complementary regions of τ in S, where the index of a complementary region C is the quantity

$$\operatorname{index}(C) = 2\chi(C) - \#\text{cusps in } \partial C.$$

It follows that any surface admitting a bigon train track has zero Euler characteristic. Among closed orientable surfaces only the torus satisfies this condition. Below we state this observation as a lemma. We will refer to it in the proof of Proposition 3.16 to show that in some situations the mutant triangulation does not admit a taut structure.

Lemma 2.3 *Suppose that an oriented* 3*-manifold* M *admits a taut ideal triangulation. Then the boundary of* M *is nonempty and consists of tori.* □

2.1.3 Surfaces carried by a taut triangulation Let $(\mathcal{T}, \alpha)$ be a taut triangulation of an oriented 3-manifold M, with the set T of tetrahedra, the set F of faces and the set E of edges. Recall from Section 2.1.1 that α determines a branched surface structure on $\mathcal{T}^{(2)}$, which we call the horizontal branched surface and denote by $\mathcal{H}$; see Figure 2. The 1-skeleton of $\mathcal{T}$ is the branch locus of $\mathcal{H}$. Thus each (oriented) edge e of $(\mathcal{T}, \alpha)$ determines a *branch equation* defined as follows. Let $f_1, f_2, \dots, f_k$ be triangles attached to e on the left side, ordered from the bottom to the top. Let $f_1', f_2', \dots, f_l'$ be triangles attached to e on the right side, also ordered from the bottom to the top. Then the branch equation determined by e is given by

(2.4)
$$f_1 + f_2 + \cdots + f_k = f_1' + f_2' + \cdots + f_l'.$$

Let $w = (w_f)_{f \in F}$ be a nonzero nonnegative integral solution to the system of branch equations of $(\mathcal{T}, \alpha)$. We call the number w_f the *weight* of f and w a *weight system* on $(\mathcal{T}, \alpha)$. Using weights of triangles

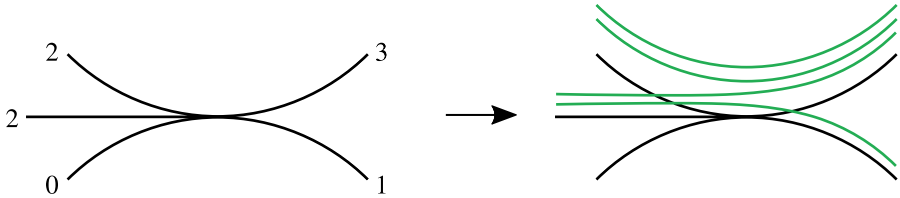


Figure 3: From a solution to branch equations to an embedded surface.

we can define the weight w_e of an edge $e \in E$ as the sum of weights of faces attached to e on one of its sides. For an edge satisfying the branch equation (2.4) we have

$$w_e = \sum_{i=1}^{k} w_{f_i} = \sum_{j=1}^{l} w_{f'_j}.$$

Equip the triangles of $\mathcal{T}$ with an orientation determined by their coorientation via the right-hand rule. Then the (relative) 2-chain

$$S_w = \sum_{f \in F} w_f \, f$$

is a 2-cycle giving an oriented surface properly immersed in M. It is embedded if and only if $w_e \le 1$ for every $e \in E$. Let $x \in E \cup F$. If $w_x > 1$ then multiple copies of x are pinched together. Pulling these overlapping regions of S_w slightly apart yields an oriented surface S_w^ϵ which is properly embedded in M; see Figure 3. We say that S_w^ϵ is *carried* by $(\mathcal{T}, \alpha)$. If additionally for every boundary component T of M all connected components of $S_w^\epsilon \cap T$ have the same orientation, we say that S_w^ϵ (or S_w) is *properly carried* by $(\mathcal{T}, \alpha)$.

More generally, we say that a surface S properly embedded in M is *carried* by $(\mathcal{T}, \alpha)$ if there exists a nonzero nonnegative integral solution $w = (w_f)_{f \in F}$ to the system of branch equations of $(\mathcal{T}, \alpha)$ such that S is homotopic to the relative 2-cycle S_w. If S_w is properly carried then we say that S is properly carried. Note that the same properly embedded surface S can be carried by $(\mathcal{T}, \alpha)$ in multiple different ways. When we write S_w or S_w^ϵ we always mean a surface in a fixed carried position corresponding to the weight system w.

If there exists a strictly positive integral solution w to the system of branch equations of $(\mathcal{T}, \alpha)$, we say that $(\mathcal{T}, \alpha)$ is *layered*. If there exists a nonnegative nonzero integral solution, but no strictly positive integral solution, then we say that $(\mathcal{T}, \alpha)$ is *measurable*. If there is no nonnegative nonzero solution to the system of branch equations of $(\mathcal{T}, \alpha)$ then we say that $(\mathcal{T}, \alpha)$ is *nonmeasurable*.

2.1.4 The Euler class of a taut triangulation A taut triangulation $(\mathcal{T}, \alpha)$ of M determines the *Euler class* $\chi_{(\mathcal{T},\alpha)}$, an element of $H^2(M, \partial M; \mathbb{R})$ which satisfies $\chi_{(\mathcal{T},\alpha)}([S]) = \chi(S)$ for every surface S carried by $(\mathcal{T}, \alpha)$; see [27, page 390; 32, Section 5.2]. It can be defined as follows. Let Γ be the 1-skeleton of the dual spine of $(\mathcal{T}, \alpha)$. By Definition 2.1(1), every vertex v of Γ has two incoming edges and two outgoing

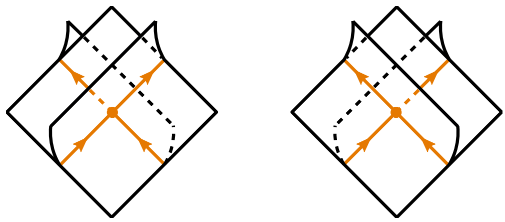


Figure 4: The dual spine of a veering triangulation can be smoothed into a branched surface which around every vertex looks like one of the above configurations. Orientation on the edges of the branch locus is relevant.

edges. Therefore Γ forms a 1-cycle. Let $[\Gamma]$ be the associated homology class in $H_1(M;\mathbb{R})$. The *Euler class* $\chi_{(\mathcal{T},\alpha)}$ is defined by the equality

$$\chi_{(\mathcal{T},\alpha)}(\cdot) = -\tfrac{1}{2}\langle[\Gamma],\cdot\rangle,$$

where $\langle\cdot,\cdot\rangle$ denotes the algebraic intersection pairing.

2.2 Veering triangulations

Taut triangulations are abundant in 3-manifolds [27, Theorem 1]. In contrast, veering triangulations, a subclass of taut triangulations defined below, are very rare. It is conjectured that any hyperbolic 3-manifold with toroidal boundary admits only finitely many, potentially zero.

Definition 2.5 A *veering structure* on a taut ideal triangulation is a smoothing of its dual spine into a branched surface $\mathcal{B}$ which locally around every vertex looks as in either of the pictures in Figure 4.

A *veering triangulation* is a taut ideal triangulation with a veering structure. We call the branched surface $\mathcal{B}$ from Definition 2.5 the *stable branched surface* of a veering triangulation. We emphasize that its branch locus is oriented by the taut structure on the triangulation.

The stable branched surface $\mathcal{B}$ of a veering triangulation can be transversely orientable or not. If $\mathcal{B}$ is transversely orientable, we say that $\mathcal{V}$ is *edge-orientable*. Otherwise we say that $\mathcal{V}$ is not edge-orientable. We refer the reader to [45] for more information about edge-orientability and how it affects certain polynomial invariants of veering triangulations.

Remark 2.6 In [51] a branched surface which locally around every vertex looks like in Figure 4, and has only solid tori or torus shells as complementary regions, is called a *veering branched surface*. However, the author of [51] orients the branch locus of this branched surface in the opposite direction.

Definition 2.5 implies that a veering triangulation is determined by three pieces of combinatorial data:

(1) an ideal triangulation $\mathcal{T}$,

(2) a taut structure α; see Definition 2.1,

(3) a veering structure $\mathcal{B}$; see Definition 2.5.

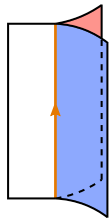


Figure 5: One of the sectors on the two-sheeted side veers to the right, the other veers to the left.

For brevity, we typically denote a veering triangulation by a calligraphic letter $\mathcal{V}$, potentially with some sub- or superscript, by which we mean $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$.

Definition 2.5 is "dual" to the now-standard definition of a veering triangulation that requires the existence of a certain coloring on the edges of triangulation; see [12, Definition 5.1]. Below we explain how to translate between the two definitions. When viewing the dual spine of a veering triangulation as a branched surface, we call its 2-cells *sectors*. Every edge d of $\mathcal{B}$ is adjacent to three sectors of $\mathcal{B}$. The structure of a branched surface on $\mathcal{B}$ determines the *one-sheeted side* of d and the *two-sheeted side* of d; see Figure 5. We say that a sector s adjacent to d is *large* relative to d if it is on the one-sheeted side of d. Otherwise we say that s is *small* relative to d. Thus two out of three sectors adjacent to d are small relative to d. Since d is oriented, and the manifold underlying $\mathcal{V}$ is oriented, we can detect in which direction (right/left) each of these small sectors veers. One of them veers to the right of d, and the other to the left of d. These directions are marked in Figure 5.

Lemma 2.7 *Let s be a sector of the stable branched surface of a veering triangulation. Let d_1 and d_2 be two consecutive edges of s. Let v be the common vertex of d_1 and d_2.*

(1) *If the orientations of d_1 and d_2 both point into v, then s is large relative to both d_1 and d_2.*
(2) *If the orientations of d_1 and d_2 both point out of v, then s is small relative to both d_1 and d_2, and if it veers right (respectively, left) of d_1, then it veers right (respectively, left) of d_2.*
(3) *If the orientation of d_1 points into v and the orientation of d_2 points out of v, then either s is small relative to d_1 and large relative to d_2, or s is small relative to both d_1 and d_2, in which case if it veers right (respectively, left) of d_1, then it veers right (respectively, left) of d_2.*

In particular, s has at least four edges.

Proof The statement of this lemma is a verbalization of the local picture of $\mathcal{B}$ presented in Figure 4. □

Lemma 2.7 says that if s veers to the right (respectively, left) of d then for every other edge d' of s such that s is small relative to d', s veers to the right (respectively, left) of d'. Since, by Lemma 2.7(2), s is small relative to at least two of its edges, the *veering direction* of s is well defined. We can therefore assign colors, red and blue, to the sectors of $\mathcal{B}$ so that right-veering sectors are colored blue and left-veering sectors are colored red; see Figure 5. We call them the *veering colors* on $\mathcal{B}$. Dually, we obtain a coloring on the edges of $\mathcal{V}$.

Corollary 2.8 *Let $\mathcal{V}$ be a veering triangulation. The veering colors on sectors of the stable branched surface of $\mathcal{V}$ determine colors on edges of $\mathcal{V}$ such that for every tetrahedron t of $\mathcal{V}$ the following two conditions hold*:

- *Let e_0, e_1 and e_2 be edges of a top face of t, ordered counterclockwise as viewed from above and so that e_0 is the top diagonal of t. Then e_1 is red and e_2 is blue.*
- *Let e_0, e_1 and e_2 be edges of a bottom face of t, ordered counterclockwise as viewed from above and so that e_0 is the bottom diagonal of t. Then e_1 is blue and e_2 is red.* □

The conditions from Corollary 2.8 are exactly the veeringness conditions that appear in [12, Definition 5.1]. Therefore if a triangulation is veering in the sense of Definition 2.5 then it is veering in the sense of [12, Definition 5.1]. The converse also holds. This can be seen by observing that colors on edges of a veering tetrahedron t that satisfy the conditions listed in Corollary 2.8 determine how to smooth the dual spine of t into a branched surface which locally around its vertices looks like the one presented in Figure 4: blue edges are dual to right veering sectors, and red edges are dual to left veering sectors; see Figure 5.

Remark 2.9 The dual spine of a veering triangulation $\mathcal{V}$ can be smoothed into another branched surface, called the *unstable branched surface* of $\mathcal{V}$. We denote it by $\mathcal{B}^u$; see [12, Section 6.1]. It is also encoded by the colors on edges of $\mathcal{V}$. If t is a tetrahedron of $\mathcal{V}$ whose bottom diagonal is blue (respectively, red) then $\mathcal{B}^u_t = \mathcal{B}^u \cap t$ is obtained from Figure 4, left (respectively, Figure 4, right), by rotating it by π in the plane of the page and then reversing orientations of all edges in the branch locus.

Remark 2.10 If M admits a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ then it also admits a veering triangulation $-\mathcal{V} = (\mathcal{T}, -\alpha, -\mathcal{B}^u)$, where $-\alpha$ is obtained from α by reversing orientations of all edges of the dual spine of $\mathcal{T}$, and $-\mathcal{B}^u$ is the unstable branched surface of $\mathcal{V}$ with orientation on the branch locus given by $-\alpha$.

In Proposition 3.17 we will use the following crucial fact about veering triangulations:

Theorem 2.11 (Hodgson, Rubinstein, Segerman and Tillmann [23, Theorem 1.5]) *Suppose that M is a compact oriented 3-manifold that admits a veering triangulation. Then the interior of M admits a complete hyperbolic metric.* □

2.2.1 The stable train track of a veering triangulation Let $\mathcal{B}$ be the stable branched surface of a veering triangulation $\mathcal{V}$. For every face f of $\mathcal{V}$ the intersection of $\mathcal{B}$ with f is a train track with one switch v_f in the interior of f and three branches, each joining v_f with the midpoint of an edge of f. The union of all these train tracks in faces of $\mathcal{V}$ gives a train track in the horizontal branched surface of $\mathcal{V}$. We call it the *stable train track* of $\mathcal{V}$, and denote it by τ. A picture of τ restricted to the faces of one veering tetrahedron is presented in Figure 6.

Let $\tau_f = \tau \cap f$. It is a trivalent train track with one large branch and two small branches. We say that an edge e of f is the *large edge* of f if it is dual to the large branch of τ_f. Otherwise we say that e is a small edge of f. The key property of the stable train track is stated in the following lemma:

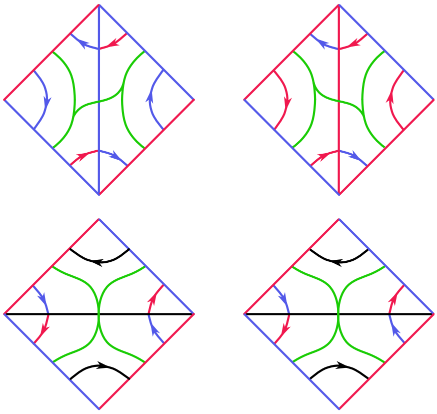


Figure 6: Left column: a veering tetrahedron with blue top diagonal. Right column: a veering tetrahedron with red top diagonal. Top faces are presented in the top row, bottom faces in the bottom row. The stable train track is in green. Oriented arcs around vertices correspond to branches of the boundary track.

Lemma 2.12 *Let f be a top face of a tetrahedron t of a veering triangulation $\mathcal{V}$. Then the large edge of f is identified with the bottom diagonal of the tetrahedron immediately above f.*

Proof Let t' be the tetrahedron immediately above f. There is a bottom face f' of t' such that the large edge of f is identified with an edge e' of f'. The picture of the stable train track in the bottom faces of a veering tetrahedron (Figure 6) indicates that e' must be the bottom diagonal of t'. □

The stable branched surface and the stable train track carry the same combinatorial information. However, it is beneficial to have both these perspectives on the same object, as they have different applications in this paper. In Section 3.7 we use the stable train track to define a certain subgroup of the group of orientation-preserving combinatorial automorphisms of a surface carried by a veering triangulation. The whole branched surface is more natural to use in the proof of Theorem 3.19, which says that under certain conditions a mutant of a veering triangulation is veering.

2.2.2 The boundary track of a veering triangulation Let $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ be a veering triangulation of a 3-manifold M. In Section 2.1.2 we defined the boundary track β of $(\mathcal{T}, \alpha)$. A (bigon) complementary region r of β has three switches, corresponding to three edges of some tetrahedron t meeting at a vertex of that tetrahedron. A switch in the boundary of r is smooth if and only if it is an endpoint of either the top or the bottom diagonal of t. We say that r is *ascending* if its smooth switch is an endpoint of the top diagonal of t. Otherwise we say that r is *descending*.

Now that we have a veering structure on the triangulation as well, we can color the switches and branches of the boundary track using the colors of edges determined by Corollary 2.8 and the following rules. If a switch s of β is an endpoint of an edge e we color s by the color of e. If a branch b of β corresponds to an arc around a vertex v of a triangle f we color b by the color of the edge of f opposite to v; see Figure 6.

This coloring on the boundary track of a veering triangulation was first introduced by Futer and Guéritaud in [17, Section 2]. We orient b using the coorientation on f and the right-hand rule. With this orientation, if S_w is a surface carried by $\mathcal{V}$, as defined in Section 2.1.3, then ∂S_w is an oriented smooth 1-cycle in β.

If b connects two switches of the same color (necessarily of a different color than b; see Corollary 2.8), we say that b is a *ladderpole branch*. In Figure 6 ladderpole branches are shown as arcs connecting a diagonal edge to a nondiagonal edge of the same color, or vice versa. Since every edge of $\mathcal{V}$ is the top diagonal of exactly one tetrahedron and the bottom diagonal of exactly one tetrahedron, every switch of β is adjacent to exactly two ladderpole branches. It follows that the union of all ladderpole branches is a disjoint union of simple closed curves on ∂M. We call these curves the *ladderpole curves* of $\mathcal{V}$. We say that a ladderpole curve l is red (respectively, blue) if it consists of red (respectively, blue) ladderpole branches. Each boundary component of M contains an even number of ladderpole curves which alternate in color.

Let S_w^ϵ be a surface properly embedded in M obtained by slightly pulling apart the overlapping regions of S_w. Recall from Section 2.1.3 that S_w is not properly carried if there is a boundary component T of M such that two connected components $S_w^\epsilon \cap T$ have opposite orientations. In the following lemma we characterize the boundary slopes of such surfaces.

Lemma 2.13 *A surface S_w carried by a veering triangulation of M is not properly carried if and only if there is a boundary component T of M such that ∂S_w runs both along a blue ladderpole curve of T and along a red ladderpole curve of T.*

Proof Using Figure 6 we can make the following observations:

(1) Let l_b be a blue ladderpole branch. Let r_b^a and r_b^d be the ascending and the descending complementary region of β adjacent to l_b, respectively. Then l_b is oriented from the smooth switch of r_b^a to the smooth switch of r_b^d. The remaining branches of r_b^a are oriented towards l_b, while the remaining branches of r_b^d are oriented away from l_b.

(2) Let l_r be a red ladderpole branch. Let r_r^a and r_r^d be the ascending and the descending complementary region of β adjacent to l_r, respectively. Then l_r is oriented from the smooth switch of r_r^d to the smooth switch of r_r^a. The remaining branches of r_r^a are oriented away from l_r, while the remaining branches of r_r^d are oriented towards l_b.

It follows that two smooth parallel cycles in β contained in the same boundary component of M have opposite orientations if and only if one of them runs along a red ladderpole curve and the other runs along a blue ladderpole curve. □

Observe that a surface S_w carried by $\mathcal{V}$ inherits from $\mathcal{V}$ both an ideal triangulation and the stable train track. We denote these structures in S_w by $\mathcal{Q}_{\mathcal{V},w}$ and $\tau_{\mathcal{V},w}$, respectively. It turns out that one can recognize whether S_w is properly carried by analyzing the complementary regions of $\tau_{\mathcal{V},w}$. If a boundary component of S_w runs along a blue ladderpole slope then the corresponding vertex of $\mathcal{Q}_{\mathcal{V},w}$ is contained in a cusp-free complementary region r of $\tau_{\mathcal{V},w}$; see Figure 6. Moreover, the boundary of r crosses only red edges. We will say that r is red. Analogously, if a boundary component of S_w runs along a red ladderpole slope then the corresponding vertex of $\mathcal{Q}_{\mathcal{V},w}$ is contained in a cusp-free complementary region of $\tau_{\mathcal{V},w}$ whose boundary crosses only blue edges. We will say that such complementary regions are blue. Using this terminology, Lemma 2.13 implies the following characterization of carried surfaces which are not properly carried:

Corollary 2.14 *Let S_w be a surface carried by a veering triangulation $\mathcal{V}$ of M. The surface S_w is properly carried by $\mathcal{V}$ if and only if any two cusp-free complementary regions of $\tau_{\mathcal{V},w}$ that have different color intersect different boundary components of M.* □

2.2.3 The veering census Data on veering triangulations of orientable 3-manifolds consisting of up to 16 tetrahedra is available in the veering census [20]. A veering triangulation in the census is described by a string of the form

(2.15) `[isoSig]_[taut angle structure].`

The first part of this string is the isomorphism signature of the triangulation. It identifies a triangulation uniquely up to combinatorial isomorphism [6, Section 3]. The second part of the string records a taut angle structure, that is a taut structure up to reversing the orientation of all dual edges. A string of the form (2.15) is called a *taut signature* and we use it whenever we refer to any particular veering triangulation from the veering census.

The following lemma has been well known since the development of the veering census. It explains why there is at most one veering triangulation with a fixed underlying taut ideal triangulation. We include its proof here because we will use it in Section 3.8.

Lemma 2.16 *Suppose that $(\mathcal{T},\alpha)$ is a taut triangulation. If $(\mathcal{T},\alpha)$ admits a veering structure then this structure is unique.*

Proof Suppose that there are two veering triangulations $\mathcal{V}=(\mathcal{T},\alpha,\mathcal{B})$ and $\mathcal{V}'=(\mathcal{T},\alpha,\mathcal{B}')$. If they are distinct then there is a tetrahedron t of $\mathcal{T}$ such that $\mathcal{B}_t=\mathcal{B}\cap t$ and $\mathcal{B}'_t=\mathcal{B}'\cap t$ are different. Let f be a top face of t. Let τ_f and τ'_f be the stable train tracks in f determined by $\mathcal{B}$ and $\mathcal{B}'$, respectively. Definition 2.5 and the assumption that $\mathcal{B}_t\neq\mathcal{B}'_t$ imply that $\tau_f\neq\tau'_f$. In particular, there is an edge e_1 of f which is dual to the large branch of τ_f and a distinct edge e_2 of f, of a different color than e_1, which is dual to the large branch of τ'_f. Applying Lemma 2.12 to $\mathcal{V}$ yields that e_1 is identified with the bottom diagonal of the tetrahedron immediately above f, and applying it to $\mathcal{V}'$ yields that e_2 is identified with the bottom diagonal of the tetrahedron immediately above f. Since e_1 and e_2 cannot be identified in $\mathcal{T}$, this is a contradiction to the assumption that the taut ideal triangulations underlying $\mathcal{V}$ and $\mathcal{V}'$ are the same. □

2.3 The connection with pseudo-Anosov flows

Recall from the introduction that veering triangulations are combinatorial tools to study pseudo-Anosov flows. In this subsection we will make this statement more precise.

Definition 2.17 A continuous flow $\Psi\colon N \times \mathbb{R} \to N$ on a closed 3-manifold N is *pseudo-Anosov* if there are 2-dimensional singular foliations $\mathcal{F}^s$ and $\mathcal{F}^u$ on N with the following properties:

- $\mathcal{F}^s$ and $\mathcal{F}^u$ intersect along the flow lines of Ψ.
- Ψ admits finitely many (potentially zero) isolated closed orbits $\ell_1, \dots, \ell_k$ such that for $i = 1, 2, \dots, k$ in a sufficiently small tubular neighborhood of ℓ_i the foliation $\mathcal{F}^{s/u}$ is isotopic to the mapping torus of the $(2\pi m_i / p_i)$-rotation of the p_i-pronged foliation of a disk (with the prong singularity in the center), for some $p_i \geq 3$ and $m_i \in \mathbb{Z}$. These orbits are called the *singular orbits* of Ψ.
- Away from the singular orbits of Ψ, foliations $\mathcal{F}^s$ and $\mathcal{F}^u$ are nonsingular and transverse to each other.
- Two flow lines contained in the same leaf of $\mathcal{F}^s$ are forward asymptotic, and two flow lines contained in the same leaf of $\mathcal{F}^u$ are backward asymptotic.

If Ψ has a dense orbit then we say that Ψ is *transitive*. If Ψ does not have any singular orbits then it is called an *Anosov flow*. We will not pay much attention to the parametrization of a flow. In fact, we will consider two flows which are topologically equivalent to be the same; see definition below.

Definition 2.18 Two flows Ψ and Ψ' on N are *topologically equivalent* (or *orbit equivalent*) if there is a homeomorphism $h\colon N \to N$ which takes oriented orbits of Ψ to oriented orbits of Ψ'.

In the literature there is also a notion of a smooth pseudo-Anosov flow; see [1, Definition 5.8]. Recently Shannon proved that any continuous transitive Anosov flow is topologically equivalent to a smooth Anosov flow [49, Section 5]; see Definition 2.18. His methods generalize to transitive continuous pseudo-Anosov flows [1, Theorem 5.10]. Thus up to topological equivalence we may assume that our flows are smooth. The reason we prefer the above definition is that it focuses on topological properties of pseudo-Anosov flows which are crucial for our purposes — namely, the existence of foliations $\mathcal{F}^s$ and $\mathcal{F}^u$ with prescribed behavior. These foliations are called the *stable* and *unstable foliations* of the flow, respectively. If ℓ is a p-pronged orbit of Ψ, where $p \geq 2$ and $p = 2$ corresponds to a nonsingular orbit, the compact core of $N - \ell$ has a boundary torus T_ℓ which meets the singular leaves of $\mathcal{F}^s$ along p parallel simple closed curves, and similarly for the singular leaves of $\mathcal{F}^u$. We call all these $2p$ parallel curves the *prong curves* of Ψ in T_ℓ. Splitting $\mathcal{F}^s$ and $\mathcal{F}^u$ open along their singular leaves yields a pair of laminations $\mathcal{L}^s$ and $\mathcal{L}^u$ in N, called the *stable* and *unstable laminations* of Ψ. Leaves of $\mathcal{L}^{s/u}$ are open annuli, open Möbius bands or planes. Möbius bands appear if and only if $\mathcal{L}^{s/u}$ is not transversely orientable.

The simplest examples of pseudo-Anosov flows are the suspension flows of pseudo-Anosov homeomorphisms of closed surfaces on their mapping tori. Their stable/unstable foliations are formed by the mapping tori of the stable/unstable foliations of the monodromy of the fibration. Suspension flows have an additional property: an embedded surface which intersects every flow line with positive sign. Such a surface is called a *cross-section* to the flow. Any flow which admits a cross-section is called *circular*. From the suspension flow of a pseudo-Anosov homeomorphism one can construct infinitely many other pseudo-Anosov flows via the Goodman–Fried surgery [15; 21]. In particular, Goodman–Fried surgery can be used to construct noncircular pseudo-Anosov flows, including pseudo-Anosov flows on nonfibered 3-manifolds.

Let Ψ be a pseudo-Anosov flow on a closed 3-manifold N. Fix a finite nonempty collection Λ of closed orbits of Ψ which includes all singular orbits of Ψ. To be able to construct a veering triangulation of $N-\Lambda$ encoding Ψ a technical condition, called *no perfect fits relative to* Λ, has to be satisfied. We refer the reader to [1, Definition 5.12] for a precise definition of this term. Here we will work combinatorially with veering triangulations, and deduce appropriate statements concerning flows using the following two theorems.

Theorem 2.19 (Agol and Guéritaud, unpublished) *Let Ψ be a pseudo-Anosov flow on a closed 3-manifold N. Suppose that Λ is a finite nonempty collection of closed orbits of Ψ which includes all singular orbits of Ψ and such that Ψ has no perfect fits relative to Λ. Then $N-\Lambda$ admits a veering triangulation $\mathcal{V}$ such that the stable (respectively, unstable) branched surface of $\mathcal{V}$, when embedded in N via the inclusion $(N-\Lambda) \hookrightarrow N$, fully carries the stable (respectively, unstable) lamination of Ψ. The ladderpole curves of $\mathcal{V}$ are homotopic to the prong curves of Ψ in $\partial(N-\Lambda)$.* □

If $\mathcal{V}$ arises from Ψ via the Agol–Guéritaud construction we will say that $\mathcal{V}$ *encodes* Ψ. Tsang proved that every pseudo-Anosov flow Ψ is without perfect fits relative to some collection Λ of closed orbits of Ψ which contains all singular orbits and one additional orbit [53, Proposition 2.7]. Thus every pseudo-Anosov flow can be encoded by some veering triangulation. The situation is the cleanest when Ψ is not an Anosov flow and does not have perfect fits relative to its collection $\mathrm{Sing}(\Psi)$ of singular orbits. Then the veering triangulation of $N-\mathrm{Sing}(\Psi)$ obtained via the Agol–Guéritaud construction can be considered to be canonical for Ψ. In the remaining cases, there might be many choices for an additional orbit to be added to the set Λ, and thus no canonical veering triangulations encoding the flow. The proof of Theorem 2.19 appears in [31, Section 4]. The theorem is stated there only for the case when Ψ does not have perfect fits relative to $\mathrm{Sing}(\Psi)$, but the proof applies equally well to the case when the set Λ only properly contains $\mathrm{Sing}(\Psi)$.

Another theorem connecting veering triangulations and pseudo-Anosov flows says that one can go also in the other direction: use veering triangulations to construct pseudo-Anosov flows.

Theorem 2.20 (Agol and Tsang [1, Theorem 5.1]) *Let $\mathcal{V}$ be a veering triangulation of a 3-manifold M. Suppose that M has k boundary components $T_1, \dots, T_k$. Let l_i be the collection of blue ladderpole curves of $\mathcal{V}$ on T_i, and let s_i be a connected simple closed curve on T_i. If $p_i = |\langle \ell_i, s_i\rangle|$ is greater than one for every $i = 1, 2, \dots, k$, then the Dehn filled manifold $M(s_1, \dots, s_k)$ admits a transitive pseudo-Anosov flow Ψ with the following properties*:

- *Ψ is without perfect fits relative to a collection $\Lambda = \{\ell_1, \dots, \ell_k\}$ of closed orbits isotopic to the cores of the filling solid tori.*

- *The orbit ℓ_i is p_i-pronged for $i = 1, 2, \dots, k$.*
- *The stable* (*respectively*, *unstable*) *lamination of* Ψ *is fully carried by the stable* (*respectively*, *unstable*) *branched surface of* $\mathcal{V}$. □

The fact that the stable lamination of Ψ is fully carried by the stable branched surface of the veering triangulation is not explicitly stated in [1, Theorem 5.1] but it follows from [1, Proposition 5.13]. (Note that the authors call the stable branched surface from Definition 2.5 the unstable branched surface, and orient the edges of its branch locus in the opposite direction.) If $\mathcal{V}$ is a veering triangulation of M and Ψ is a pseudo-Anosov flow on some closed Dehn filling N of M constructed by the Agol–Tsang construction we will say that Ψ is *built from* $\mathcal{V}$. The fact that Agol and Guéritaud's construction and Agol and Tsang's construction are each other's inverses appears in [52, Theorem 2.1]; see also the program of Schleimer and Segerman outlined in [12, Section 1.2].

Let $\mathcal{V}$ be a veering triangulation of M encoding a pseudo-Anosov flow Ψ on some closed Dehn filling N of M. If we view M as a cusped 3-manifold $N - \Lambda$, then it is naturally equipped with a flow $\Psi_{N-\Lambda}$, the restriction of Ψ to the complement of Λ. The flow $\Psi_{N-\Lambda}$ is not pseudo-Anosov: orbits of Ψ that are asymptotic to an element of Λ become orbits of $\Psi_{N-\Lambda}$ that escape to infinity. Alternatively, we can view M as a manifold obtained from N by *blowing up* elements of Λ into toroidal boundary components. There is a notion of the *blown-up flow* Ψ° on M [42, Section 3.2]. See also [4, Section 3.6] for the construction of Ψ° in the context of smooth flows. The orbits of Ψ that are asymptotic to an element of Λ become orbits of Ψ° that are asymptotic to a prong curve on some boundary component of M (which is an orbit of Ψ°). By splitting the stable/unstable foliations of Ψ open along the leaves through each orbit of Λ we obtain a pair of laminations in M. We will call these laminations the stable/unstable laminations of Ψ°, respectively. Theorem 2.20 immediately implies the following statement:

Corollary 2.21 *Suppose that a pseudo-Anosov flow Ψ is built from a veering triangulation $\mathcal{V}$ of M. The following statements are equivalent.*

- *$\mathcal{V}$ is edge-orientable.*
- *The stable lamination of Ψ is transversely orientable.*
- *The stable lamination of Ψ° is transversely orientable.* □

Given a flow Ψ on an oriented 3-manifold N and an oriented surface S properly embedded in N, we will say that S is *transverse* to Ψ if S is transverse to the orbits of Ψ and the orientation on $TS \oplus T\Psi$ agrees with the orientation of M. Another useful connection between veering triangulations and pseudo-Anosov flows says that if a veering triangulation $\mathcal{V}$ is built from a pseudo-Anosov flow Ψ then all surfaces carried by $\mathcal{V}$ are transverse to the blown-up flow Ψ°.

Theorem 2.22 (Landry, Minsky and Taylor [31, Theorem 5.1]) *Suppose that $\mathcal{V}$ is a veering triangulation of M encoding a pseudo-Anosov flow Ψ on some closed Dehn filling N of M. Then any surface carried by $\mathcal{V}$ is transverse to the blown-up flow Ψ° on M.* □

2.4 Veering triangulations and the Thurston norm

Given a compact oriented 3-manifold M, Thurston defined a seminorm $\|\cdot\|_{\mathrm{Th}}$ on $H_2(M, \partial M; \mathbb{R})$ as follows. Every integral class $\eta \in H_2(M, \partial M; \mathbb{Z})$ can be represented by a properly embedded surface $S \subset M$ [50, Lemma 1]. If S is connected, we set

$$\chi_-(S) = \max\{0, -\chi(S)\},$$

where $\chi(S)$ denotes the Euler characteristic of S. Otherwise, denote by $S_1, S_2, \dots, S_k$ connected components of S and set

$$\chi_-(S) = \sum_{i=1}^{k} \chi_-(S_i).$$

We define a quantity $\|\eta\|_{\mathrm{Th}}$ as the infimum of $\chi_-(S)$ over all surfaces S which are properly embedded in M and represent η. The function $\|\cdot\|_{\mathrm{Th}}$ can be extended to $H_2(M, \partial M; \mathbb{Q})$ by requiring linearity on each ray through the origin in $H_2(M, \partial M; \mathbb{R})$, and then to $H_2(M, \partial M; \mathbb{R})$ by requiring continuity [50, Section 1]. If every surface representing a nonzero class in $H_2(M, \partial M; \mathbb{Z})$ has negative Euler characteristic then $\|\cdot\|_{\mathrm{Th}}$ is a norm [50, Theorem 1], called the *Thurston norm*. If a properly embedded surface S does not have any homologically trivial components and satisfies $\chi_-(S) = -\|\eta\|_{\mathrm{Th}}$ then it is called a *taut representative* of η or a *Thurston norm minimizing representative* of η. The unit norm ball $\mathbb{B}_{\mathrm{Th}}$ of $\|\cdot\|_{\mathrm{Th}}$ is a polytope with rational vertices [50, Theorem 2]. Thus we can speak about *faces* of the Thurston norm ball.

A connection between the Thurston norm and pseudo-Anosov flows on closed hyperbolic 3-manifolds was established by Fried and Mosher in the 80s and 90s. The first result in that direction concerned only fibered faces:

Theorem 2.23 (Fried [13, Theorem 7]) *Let N be a closed hyperbolic 3-manifold. Let $\mathtt{F}$ be a fibered face of the Thurston norm ball in $H_2(N; \mathbb{R})$. There is a unique, up to isotopy and reparametrization, circular pseudo-Anosov flow Ψ such that a class $\eta \in H_2(N; \mathbb{Z})$ can be represented by a cross-section to Ψ if and only if η is in the interior of $\mathbb{R}_+ \cdot \mathtt{F}$.* □

The importance of this result lies in the fact that when $b_1(N) > 1$ there are infinitely many fibrations lying over $\mathtt{F}$, and thus one can construct infinitely many suspension flows on N: one for each fibration. Theorem 2.23 implies that all these flows are the same up to isotopy and reparametrization. We will say that the unique flow Ψ associated to a fibered face $\mathtt{F}$ *dynamically represents* $\mathtt{F}$.

Mosher extended Fried's result by showing that if a circular flow Ψ dynamically represents a fibered face $\mathtt{F}$ then $\eta \in H_2(N; \mathbb{Z})$ can be represented by a surface that is *almost transverse* to Ψ if and only if η is in $\mathbb{R}_+ \cdot \mathtt{F}$ [40, Theorem 1.4]. Almost transversality means that the surface is transverse to a slightly modified flow $\Psi^\#$ obtained by *dynamically blowing up* finitely many closed orbits of Ψ into a collection of annuli; see [41, Section 1.3] for details. If S is almost transverse to Ψ then the algebraic intersection of $[S]$

with the homology class $[\gamma]$ of every closed orbit γ of Ψ is nonnegative. Conversely, if $\eta \in H_2(M;\mathbb{Z})$ is such that $\langle \eta, [\gamma]\rangle \geq 0$ for every closed orbit γ of Ψ then η can be represented by a taut surface which is almost transverse to Ψ [41, Theorem 1.3.2]. Using these facts, Mosher extended the notion of dynamical representation of faces of the Thurston norm ball to nonfibered faces [41]. Given a pseudo-Anosov flow Ψ on a closed 3-manifold N let $\mathcal{C}(\Psi) \subset H_2(N;\mathbb{R})$ be the nonnegative span of the second homology classes whose algebraic intersection with the homology class of every closed orbit of Ψ is nonnegative. By the aforementioned result [41, Theorem 1.3.2], we can think of $\mathcal{C}(\Psi)$ as the cone of homology classes of surfaces that are almost transverse to Ψ. Associated to Ψ there is also a second cohomology class, the *Euler class* of the normal plane bundle to Ψ, denoted by χ_Ψ; see [41, Section 2.4] for a formula for the computation of χ_Ψ. Its main feature is that it correctly computes the Thurston norm of surfaces that are almost transverse to Ψ in the sense that if $\eta \in \mathcal{C}(\Psi)$ then $\|\eta\|_{\mathrm{Th}} = -\chi_\Psi(\eta)$ [41, page 262].

Definition 2.24 We say that a pseudo-Anosov flow Ψ on a closed hyperbolic 3-manifold N *dynamically represents* a face $\mathtt{F}$ of the Thurston norm ball $\mathbb{B}_{\mathrm{Th}}$ in $H_2(N;\mathbb{R})$ if $\mathcal{C}(\Psi) = \mathbb{R}_+ \cdot \mathtt{F}$ and $\mathtt{F}$ is the maximal face of $\mathbb{B}_{\mathrm{Th}}$ over which the Thurston norm agrees with $-\chi_\Psi$.

By our earlier discussion and the fact that fibered faces are always top dimensional, the circular flow Ψ associated to a fibered face $\mathtt{F}$ as in Theorem 2.23 dynamically represents $\mathtt{F}$. Furthermore, Mosher found sufficient conditions on a noncircular pseudo-Anosov flow to dynamically represent a nonfibered face of the Thurston norm ball in [41, Theorem 2.7]. In [41, Section 4] he presented an example of a noncircular pseudo-Anosov flow which dynamically represents a top-dimensional nonfibered face of the Thurston norm ball, as well as an example of a pseudo-Anosov flow which does not dynamically represent any face of the Thurston norm ball. In the latter case $\mathcal{C}(\Psi)$ is properly contained in the cone on some face of the Thurston norm ball [40, Theorem 2.8]. The results of Mosher raise the following two questions:

Question 1 Let N be a closed 3-manifold. Given a nonfibered face $\mathtt{F}$ of the Thurston norm ball in $H_2(N;\mathbb{R})$, is there a pseudo-Anosov flow Ψ on N which dynamically represents $\mathtt{F}$?

Question 2 Suppose that a nonfibered face $\mathtt{F}$ of the Thurston norm ball is dynamically represented by Ψ. Is the flow Ψ unique, up to isotopy and reparametrization?

Question 1 is still open. In Section 5 we will use veering triangulations to show that for some nonfibered faces the answer to Question 2 is negative. Once we know that a face of the Thurston norm ball can be represented by multiple distinct flows, we may ask another question:

Question 3 Suppose that a face $\mathtt{F}$ of the Thurston norm ball is dynamically represented by two topologically inequivalent flows Ψ and Ψ'. How are Ψ and Ψ' related?

This problem also can be approached by employing veering triangulations. We partially answer this question in the case of nonfibered faces of manifolds with nonempty boundary in Section 5.

Both Fried and Mosher worked in the setup of closed 3-manifolds. However, the Thurston norm can be defined for any compact oriented atoroidal 3-manifold M. Thus we can ask Questions 1–3 also in the context of faces of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ when $\partial M \neq \varnothing$. Since we will work with veering triangulations, considering 3-manifolds with $\partial M \neq \varnothing$ is in fact more natural, and is a necessary intermediate step when trying to answer the questions in the closed case. Instead of pseudo-Anosov flows we then consider blown-up pseudo-Anosov flows. Definition 2.24 can be generalized to these flows in a natural way. The fact that suspension flows of pseudo-Anosov homeomorphisms of surfaces with boundary dynamically represent fibered faces of the Thurston norm ball of their mapping tori (ie an analogue of Fried's Theorem 2.23) was proved by Landry in [30, Theorem 3.5].

We will rely on results of Landry, Minsky and Taylor, stated below, connecting the Thurston norm directly with veering triangulations. Recall from Section 2.1.3 that a veering triangulation $\mathcal{V}$ of M may carry surfaces properly embedded in M. Each such surface is a Thurston norm minimizing representative of its homology class [27, Theorem 3], and is transverse to the blown-up flow encoded by $\mathcal{V}$ [31, Theorem 5.1] (stated here as Theorem 2.22). In analogy to the cone $\mathcal{C}(\Psi)$ of homology classes of surfaces almost transverse to a flow Ψ, let $\mathcal{C}(\mathcal{V})$ be the cone of homology classes of surfaces carried by $\mathcal{V}$.

Definition 2.25 A veering triangulation $\mathcal{V}$ of M *combinatorially represents* a face $\mathtt{F}$ of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ if $\mathcal{C}(\mathcal{V}) = \mathbb{R}_+ \cdot \mathtt{F}$.

When defining dynamical representation (Definition 2.24) we require not only equality of appropriate cones, but also that $\mathtt{F}$ is the maximal face over which the Thurston norm agrees with minus the Euler class of the normal plane bundle to the flow. In the setup of veering triangulations, maximality is always satisfied: $\mathcal{C}(\mathcal{V})$ is equal to the cone where the Thurston norm agrees with minus the Euler class of $\mathcal{V}$ [32, Theorem 5.12].

Theorem 2.26 (Landry, Minsky and Taylor [32, Theorems 5.12 and 5.15]) *If $\mathcal{V}$ is a layered or measurable veering triangulation of M, then there is a (not necessarily top-dimensional) face $\mathtt{F}$ of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ with $\mathcal{C}(\mathcal{V}) = \mathbb{R}_+ \cdot (\mathtt{F})$. Furthermore, $\mathtt{F}$ is fibered if and only if $\mathcal{V}$ is layered.*

The above theorem says that layered and measurable veering triangulations always combinatorially represent some face of the Thurston norm ball. This is in contrast with pseudo-Anosov flows, for which it is possible that $\mathcal{C}(\Psi)$ is nonempty, but is a proper subset of the cone on some face of the Thurston norm ball [41, Section 4]. It is also known that if a fibered face is combinatorially represented by a veering triangulation, then this veering triangulation is unique [38, Proposition 2.7]. Note that not every fibered face is combinatorially represented by some veering triangulation, because the circular flow associated to the face might have singular orbits.

Remark 2.27 In the proof of [32, Theorem 5.12] the authors show that if $\mathbb{R}_+ \cdot \mathtt{F} = \mathcal{C}(\mathcal{V})$ then $\mathcal{V}$ not only carries *some* taut representative of every $\eta \in \mathbb{R}_+ \cdot \mathtt{F}$, but it carries *every* taut representative of η. Thus in particular if $\mathcal{C}(\mathcal{V}) = \mathcal{C}(\mathcal{V}')$ for some veering triangulations $\mathcal{V}$ and $\mathcal{V}'$ of a fixed manifold, then $\mathcal{V}$ carries S if and only if $\mathcal{V}'$ carries S.

Remark 2.28 For a taut triangulation $(\mathcal{T}, \alpha)$ let $\mathrm{Aut}^+(\mathcal{T} \mid \alpha)$ denote the group of orientation-preserving combinatorial automorphisms of $\mathcal{T}$ which preserve α. Recall from Section 2.2.3 that if $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ is veering and $\phi \in \mathrm{Aut}^+(\mathcal{T} \mid \alpha)$ then $\mathcal{V}$ and $\phi(\mathcal{V}) = (\phi(\mathcal{T}), \phi(\alpha), \phi(\mathcal{B}))$ have the same taut signature in the veering census. Thus the same entry in the veering census may encode multiple veering triangulations representing different faces of the Thurston norm ball which lie in the same orbit of the action of $\mathrm{Homeo}^+(M)$ on $H_2(M, \partial M; \mathbb{R})$. Furthermore, veering triangulations $\mathcal{V}$ and $-\mathcal{V}$ also have the same taut signature. They satisfy $\mathcal{C}(\mathcal{V}) = -\mathcal{C}(\mathcal{V})$ and represent a pair of opposite faces of the Thurston norm ball.

Landry, Minsky and Taylor defined the *flow graph* of a veering triangulation $\mathcal{V}$ whose oriented cycles correspond to orbits of the flow encoded by $\mathcal{V}$. We refer the reader to [32, Section 4] for the definition of the flow graph, and to [31, Section 6] for an explanation of the relationship between the flow graph of $\mathcal{V}$ and orbits of the flow encoded by $\mathcal{V}$. From their results it follows that the blown-up pseudo-Anosov flow on the 3-manifold underlying a layered or measurable veering triangulation dynamically encodes a face of the Thurston norm ball.

Theorem 2.29 (Landry, Minsky and Taylor [31, Theorem 6.1; 32, Theorem 5.1]) *Let $\mathcal{V}$ be a veering triangulation of M. Suppose that $\mathcal{V}$ encodes a pseudo-Anosov flow Ψ on some closed Dehn filling N of M. Let Ψ° be the associated blown-up flow on M. Then*

$$\mathcal{C}(\Psi^\circ) = \mathcal{C}(\mathcal{V}).$$

Under additional assumptions, we also have an analogous theorem concerning the pseudo-Anosov flow Ψ on N. There is a (potentially empty) subcone $\mathcal{C}(\mathcal{V}|N) \subset \mathcal{C}(\mathcal{V}) \subset H_2(M, \partial M; \mathbb{R})$ of homology classes of surfaces carried by $\mathcal{V}$ whose boundary components have slopes consistent with the Dehn filling slopes yielding N out of M. These surfaces cap off to embedded surfaces in N. We denote by $\mathcal{C}_N(\mathcal{V}) \subset H_2(N; \mathbb{R})$ the nonnegative span of homology classes of these capped-off surfaces.

Theorem 2.30 (Landry [29, Theorem A] and Landry, Minsky and Taylor [31, Theorem 6.1]) *Let $\mathcal{V}$ be a veering triangulation of M. Suppose that $\mathcal{V}$ encodes a pseudo-Anosov flow Ψ on some closed Dehn filling N of M such that the core curves of the filling solid tori are singular orbits of Ψ with at least three prongs. Then N is hyperbolic. Furthermore, if $\mathcal{C}_N(\mathcal{V}) \neq \varnothing$, there is a face $\mathtt{F}$ of the Thurston norm ball in $H_2(N; \mathbb{R})$ such that $\mathbb{R}_+ \cdot \mathtt{F} = \mathcal{C}_N(\mathcal{V}) = \mathcal{C}(\Psi)$.*

The face $\mathtt{F}$ is determined by the Euler class χ_Ψ of the normal plane bundle to Ψ in the sense that $\chi_\Psi(\eta) = -\|\eta\|_{\mathrm{Th}}$ for every $\eta \in \mathbb{R}_+ \cdot \mathtt{F}$. Mosher's example of a pseudo-Anosov flow Ψ on a closed hyperbolic 3-manifold N which does not dynamically represent a whole face of the Thurston norm ball in $H_2(N;\mathbb{R})$ is such that there is a class $\eta \in H_2(N;\mathbb{R})$ for which $\chi_\Psi(\eta) = -\|\eta\|_{\mathrm{Th}}$, but which pairs negatively with the homology class of some nonsingular closed orbit of Ψ [41, Section 4]. In this case $\mathcal{C}(\Psi)$ is a proper subset of the cone on some face of the Thurston norm ball in $H_2(N;\mathbb{R})$. This "pathology" does not happen under the assumptions of Theorem 2.30 because of Theorem 2.29 and the fact that if $\langle \eta, [\gamma]\rangle < 0$ for a singular orbit γ with at least three prongs then $\chi_\Psi(\eta) < \|\eta\|_{\mathrm{Th}}$; see [41, pages 259–261].

3 Mutations of veering triangulations

For the remainder of this paper by $X|Y$ we denote the metric completion of $X - Y$ with respect to the path metric on $X - Y$ induced from X.

Let S be an oriented surface properly embedded in an oriented compact 3-manifold M so that any two boundary components of S contained in the same boundary component of M have the same orientation. We say that $M|S$ is the *cut manifold* obtained from M by *decomposing* it along S. The orientation on S determines a transverse orientation on S via the right-hand rule. By S^+ we denote the boundary copy of S in $M|S$ which is cooriented out of $M|S$, and by S^- we denote the boundary copy of S in $M|S$ which is cooriented into $M|S$. Any homeomorphism $\varphi\colon S^+ \to S^-$ gives rise to a *mutant manifold* M^φ obtained from $M|S$ by gluing S^+ to S^- via φ. This manifold admits an embedded surface S^φ homeomorphic to S. We say that M^φ is obtained from M by *mutating* it *along* S via φ. We also say that S is a *mutating surface*.

In this section we approach the following problem: assuming that M admits a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$, when does the mutant manifold M^φ admit a veering triangulation? We restrict our considerations to the case when the mutating surface is carried by $\mathcal{V}$ with weights $w = (w_f)_{f\in F}$ and the homeomorphism φ comes from an orientation-preserving combinatorial automorphism φ of the ideal triangulation $\mathcal{Q}_{\mathcal{V},w}$ of S_w; we discuss this triangulation in detail in Section 3.2. In Sections 3.1 and 3.3 we recall the notions of combinatorial isomorphisms of triangulations and sutured manifolds, respectively. Section 3.4 is devoted to analyzing a certain *cut triangulation* $\mathcal{T}|F_w$ and its relation to the cut manifold $M|S_w^\epsilon$. In Section 3.5, given a pair (S_w, φ), we define a *mutant triangulation* $\mathcal{T}^\varphi$ of $\mathcal{T}$. In Section 3.6 we find a sufficient and necessary condition on φ that guarantees $\mathcal{T}^\varphi$ is an ideal triangulation of M^φ. In Section 3.7 we find a sufficient and necessary condition for the existence of a taut structure α^φ on $\mathcal{T}^\varphi$, and furthermore conditions which ensure that $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure $\mathcal{B}^\varphi$. In Section 3.8 we generalize this result to give sufficient and necessary conditions on $(\mathcal{T}^\varphi, \alpha^\varphi)$ to admit a veering structure. We also generalize a *veering mutation* $(\mathcal{T}, \alpha, \mathcal{B}) \rightsquigarrow (\mathcal{T}^\varphi, \alpha^\varphi, B^\varphi)$ to a *veering mutation with insertion*.

3.1 Triangulations: face identifications and combinatorial automorphisms

We will use a taut ideal triangulation $(\mathcal{T}, \alpha)$ of a 3-manifold to construct another ideal triangulation of a, typically different, 3-manifold using a combinatorial automorphism of a surface carried by $(\mathcal{T}, \alpha)$. In this subsection we explain the standard conventions used to encode triangulations and their combinatorial automorphisms.

Recall from Section 2 that by an ideal triangulation of a compact 3-manifold M we mean an expression of $M - \partial M$ as a collection of finitely many ideal tetrahedra with triangular faces identified in pairs by homeomorphisms which send vertices to vertices. An identification between a face f of tetrahedron t and a face f' of tetrahedron t' can be encoded by a bijection between their vertices. In our construction we will "forget" identifications between a subset of faces of the triangulation, and replace them with different ones. The whole procedure will be governed by a combinatorial automorphism of a carried surface.

Suppose that S_w is a surface carried by $(\mathcal{T}, \alpha)$ as in Section 2.1.3. Since S_w is built out of triangles and edges of $\mathcal{T}$, it inherits an ideal triangulation from $\mathcal{T}$. We discuss this triangulation in detail in Section 3.2. For now, we will denote this triangulation of S_w by $\mathcal{Q}$. Similarly as in the case of a 3-dimensional triangulation, $\mathcal{Q}$ is an expression of $S_w - \partial S_w$ as a collection of finitely many ideal triangles with edges identified in pairs by homeomorphisms which send vertices to vertices.

We recall the definition of a combinatorial isomorphism between a pair of 2-dimensional triangulations:

Definition 3.1 Let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be finite (ideal) 2-dimensional triangulations. For $i = 1, 2$ let F_i and E_i denote the sets of triangles and edges of $\mathcal{Q}_i$, respectively. A *combinatorial isomorphism* from $\mathcal{Q}_1$ to $\mathcal{Q}_2$ consists of

- a bijection $\varphi\colon F_1 \to F_2$,
- for each $f \in F_1$ a bijection φ_f between the edges of f and edges of $\varphi(f)$ such that if edges e of f and e' of f' are identified in $\mathcal{Q}_1$ then edges $\varphi_f(e)$ of $\varphi(f)$ and $\varphi_{f'}(e')$ of $\varphi(f')$ are identified in $\mathcal{Q}_2$.

If for every triangle of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ we fix a bijection between its edges and vertices, we can view φ_f as a bijection between vertices of f and vertices of f'. It is standard to fix this bijection so that a vertex v of f is associated to the edge of f opposite to v.

Different combinatorial isomorphisms from $\mathcal{Q}_1$ to $\mathcal{Q}_2$ may have the same bijection $\varphi\colon F_1 \to F_2$. Nonetheless, for simplicity we often abuse the notation and denote a combinatorial isomorphism by $\varphi\colon \mathcal{Q}_1 \to \mathcal{Q}_2$, understanding that it carries information about both a bijection $\varphi\colon F_1 \to F_2$ and bijections $\{\varphi_f \mid f \in F_1\}$.

If $\mathcal{Q}$ is an ideal triangulation then a combinatorial isomorphism $\varphi\colon \mathcal{Q} \to \mathcal{Q}$ is called a *combinatorial automorphism* of $\mathcal{Q}$. We denote the group of orientation-preserving combinatorial automorphism of $\mathcal{Q}$ by $\mathrm{Aut}^+(\mathcal{Q})$. It follows directly from Definition 3.1 that $\mathrm{Aut}^+(\mathcal{Q})$ is finite.

We sometimes refer to two 3-dimensional ideal triangulations as either combinatorially isomorphic or not. A combinatorial isomorphism between a pair of 3-dimensional ideal triangulations $\mathcal{T}_1$ and $\mathcal{T}_2$ is determined by a bijection φ from the set of tetrahedra of $\mathcal{T}_1$ to the set of tetrahedra of $\mathcal{T}_2$ and a collection of permutations φ_t, one for each tetrahedron t of $\mathcal{T}_1$, between the faces of tetrahedron t of $\mathcal{T}_1$ and faces of tetrahedron $\varphi(t)$ such that if faces f of t and f' of t' are identified in $\mathcal{T}_1$ then faces $\varphi_t(f)$ of $\varphi(t)$ and $\varphi_{t'}(f')$ of $\varphi(t')$ are identified in $\mathcal{T}_2$. In other words, it suffices to replace in Definition 3.1 triangles by

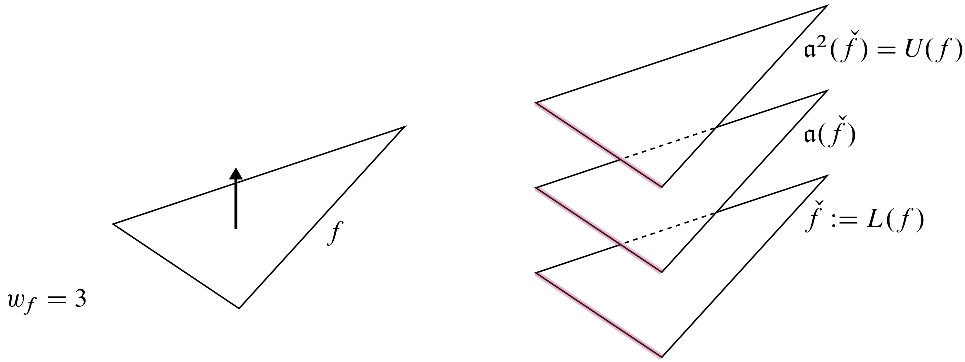


Figure 7: A face f with weight 3. Coorientation on f determines an order on the copies of f in $\mathcal{Q}_{\mathcal{V},w}$. Shaded edges are $\mathfrak{a}^k(\check{e})$, $\mathfrak{a}^{k+1}(\check{e})$ and $\mathfrak{a}^{k+2}(\check{e})$, from bottom to top, for an edge e of f and some $0 \le k \le w_e - 3$.

tetrahedra and edges by triangular faces. The pair $(\varphi, \{\varphi_t\})$ determines images of edges of $\mathcal{T}_1$ because, under the bijection which sends face f of t to a vertex of t opposite to f, permutations φ_t contain information about where the vertices are sent, and thus where the pairs of vertices are sent.

3.2 Triangulation of a surface carried by a veering triangulation

Let $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ be a finite veering triangulation of M with the set T of tetrahedra, the set F of 2-dimensional faces and the set E of edges. Let $w = (w_f)_{f \in F}$ be a weight system on $(\mathcal{T}, \alpha)$. We denote the triangulation of S_w inherited from $\mathcal{T}$ by $\mathcal{Q}_{\mathcal{V},w}$. A properly embedded surface S_w^ϵ obtained by slightly pulling apart overlapping regions of S_w also can be seen as triangulated by $\mathcal{Q}_{\mathcal{V},w}$. Recall that each triangle of $\mathcal{V}$ is equipped with a trivalent train track; see Figure 6. Therefore the surfaces S_w and S_w^ϵ come equipped with a train track dual to their triangulation $\mathcal{Q}_{\mathcal{V},w}$. We will denote this train track by $\tau_{\mathcal{V},w}$ and call it the *stable train track* of S_w or S_w^ϵ.

Let $F_w = \{f \in F \mid w_f > 0\}$ and $E_w = \{e \in E \mid w_e > 0\}$. Given $f \in F_w$ there are w_f copies of f in the triangulation $\mathcal{Q}_{\mathcal{V},w}$, and given $e \in E_w$ there are w_e copies of e in $\mathcal{Q}_{\mathcal{V},w}$. Coorientation on the faces of $\mathcal{V}$ and a fixed embedding of S_w^ϵ in M determine a linear order on the copies of a given simplex of $\mathcal{V}$ in $\mathcal{Q}_{\mathcal{V},w}$: from the lowermost to the uppermost; see Figure 7. Denote by $F_{\mathcal{V},w}$ and $E_{\mathcal{V},w}$ the sets of triangles and edges of $\mathcal{Q}_{\mathcal{V},w}$, respectively. We define maps

$$L\colon F_w \cup E_w \to F_{\mathcal{V},w} \cup E_{\mathcal{V},w} \quad \text{and} \quad U\colon F_w \cup E_w \to F_{\mathcal{V},w} \cup E_{\mathcal{V},w}$$

such that, given a simplex $x \in F_w \cup E_w$, the simplex $L(x)$ denotes the *lowermost copy* of x in $\mathcal{Q}_{\mathcal{V},w}$ and $U(x)$ denotes the *uppermost copy* of x in $\mathcal{Q}_{\mathcal{V},w}$. Observe that both L and U are injective. To simplify notation we will sometimes denote $L(x)$ by $\check{x}$. Given $f \in F_w$ we will denote by σ_f^L and σ_f^U the bijections between the edges of f and the corresponding edges of $L(f)$ and $U(f)$, respectively. Furthermore we define

$$\mathfrak{a}\colon (F_{\mathcal{V},w} \cup E_{\mathcal{V},w}) - U(F_w \cup E_w) \to (F_{\mathcal{V},w} \cup E_{\mathcal{V},w}) - L(F_w \cup E_w)$$

such that if $y \notin U(F_w \cup E_w)$ then $\mathfrak{a}(y)$ is the simplex of $\mathcal{Q}_{\mathcal{V},w}$ which is a copy of the same simplex of $\mathcal{V}$ as y and lies *immediately above* y in M; see Figure 7. Simplex $\mathfrak{a}(y)$ exists by the assumption that $y \notin U(F_w \cup E_w)$ (is not uppermost), and is never in $L(F_w \cup E_w)$, because lowermost copies do not have copies of the same simplex below them. If $w_x = k \geq 0$ then $\mathfrak{a}^0(\check{x})$, $\mathfrak{a}(\check{x}), \dots, \mathfrak{a}^{k-1}(\check{x})$ are defined, where by $\mathfrak{a}^0(\check{x})$ we mean $\check{x}$.

3.3 Sutured manifolds

Let S be an oriented surface properly embedded in an oriented 3-manifold M with empty or toroidal boundary. Assume additionally that any two boundary components of S contained in the same boundary component of M have the same orientation. Then the cut manifold $M|S$ is an example of a *sutured manifold*, defined below.

Definition 3.2 [19, Definition 3.1] A *sutured manifold* (N,γ) is a compact oriented 3-manifold N together with a set $\gamma \subset \partial N$ of pairwise-disjoint annuli $A(\gamma)$, called *sutured annuli*, and tori $T(\gamma)$, called *sutured tori*, such that

- the interior of each component of $A(\gamma)$ contains a homologically nontrivial oriented simple closed curve called a *suture*,
- every connected component of $R(\gamma) = \partial N - \mathrm{int}(\gamma)$ is oriented so that every connected component of $\partial R(\gamma)$ when equipped with the boundary orientation represents the same homology class in $H_1(\gamma)$ as some suture.

A fixed orientation of (N,γ) endows $R(\gamma)$ with coorientation. This determines a decomposition of $R(\gamma)$ into $R^+(\gamma)$, where the coorientation points out of N, and $R^-(\gamma)$ where the coorientation points into N. We call $R^+(\gamma)$ the *top boundary* of the sutured manifold N, and $R^-(\gamma)$ its *bottom boundary*. We also denote them by $\partial^+ N$ and $\partial^- N$, respectively. A boundary component of a sutured annulus A of (N,γ) that is contained in $\partial^+ N$ (respectively, $\partial^- N$) is called its *top* (respectively, *bottom*) *boundary* and denoted by $\partial^+ A$ (respectively, $\partial^- A$).

The pair $(M|S, \partial M|\partial S)$ is an example of a sutured manifold. Its sutured tori correspond to the boundary tori of M that are disjoint from S. A boundary torus of M containing k boundary components of S gives rise to k sutured annuli in $(M|S, \partial M|\partial S)$. For brevity, we often say that $M|S$ is a sutured manifold, without explicitly indicating its sutured tori and annuli.

3.4 Cutting veering triangulations along carried surfaces

Let $\mathcal{V}$ be a finite veering triangulation of a 3-manifold M with the set T of tetrahedra, the set F of triangular faces and the set E of edges. Let S_w be a surface properly carried by $\mathcal{V}$ with weights $(w_f)_{f \in F}$. As in Section 3.2, we denote by F_w the set of $f \in F$ for which $w_f > 0$.

Recall from Section 2.2 that the calligraphic letter $\mathcal{V}$ implicitly denotes three pieces of combinatorial data: an ideal triangulation $\mathcal{T}$, a taut structure α and a veering structure $\mathcal{B}$. We denote the result of decomposing $\mathcal{T}$ along F_w by $\mathcal{T}|F_w$. All faces of $\mathcal{T}|F_w$ inherit coorientations from $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$. We denote this choice of coorientations on the faces of $\mathcal{T}|F_w$ by $\alpha|F_w$. By F_w^+ we denote the boundary triangles of $\mathcal{T}|F_w$ which are cooriented out of $\mathcal{T}|F_w$ and by F_w^- the boundary triangles of $\mathcal{T}|F_w$ which are cooriented into $\mathcal{T}|F_w$. Finally, after cutting the stable branched surface $\mathcal{B}$ along $\mathcal{B} \cap F_w$, we obtain a branched surface $\mathcal{B}|F_w$. For simplicity, we denote the triple $(\mathcal{T}|F_w, \alpha|F_w, \mathcal{B}|F_w)$ by $\mathcal{V}|F_w$. In Sections 3.4–3.6 we do not make use of this (partial) veering structure, and consider only the pair $(\mathcal{T}|F_w, \alpha|F_w)$.

The set F_w can be seen as a branched surface embedded in M which fully carries S_w^ϵ. We denote the (sutured) manifold underlying $\mathcal{T}|F_w$ by $M|F_w$. Note that $\mathcal{T}|F_w$ is not an ideal triangulation of $M|F_w$ in the sense introduced at the beginning of Section 2. The manifold $M|F_w$ is expressed as a union of ideal tetrahedra of $\mathcal{T}|F_w$, but only some of their faces are identified in pairs by homeomorphisms sending vertices to vertices. The remaining faces make up the triangulations of the top and bottom boundaries of $M|F_w$. To avoid any confusion we call $\mathcal{T}|F_w$ a *cut triangulation*. In this section we will establish a relationship between $M|F_w$ and $M|S_w^\epsilon$.

Recall from Section 3.2 that S_w^ϵ is triangulated by $\mathcal{Q}_{\mathcal{V},w}$. We denote the corresponding triangulations of $S_w^{\epsilon+}$ and $S_w^{\epsilon-}$ in $M|S_w^\epsilon$ by $\mathcal{Q}_{\mathcal{V},w}^+$ and $\mathcal{Q}_{\mathcal{V},w}^-$, respectively. Given a simplex x of $\mathcal{Q}_{\mathcal{V},w}$ we denote by x^+ and x^- the corresponding simplices of $\mathcal{Q}_{\mathcal{V},w}^+$ and $\mathcal{Q}_{\mathcal{V},w}^-$, respectively. Notation that we use below was introduced in Section 3.2.

Let $e \in E_w = \{e \in E \mid w_e > 0\}$. Recall that $\check{e}$ is a shorthand for $L(e)$, the lowermost copy of e in $\mathcal{Q}_{\mathcal{V},w}$. Suppose that $w_e = k \geq 2$. Then for $i = 1, 2, \ldots, k-1$ there is a disk D_i^e properly embedded in $M|S_w^\epsilon$ whose boundary decomposes into four arcs: one arc corresponding to the edge $\mathfrak{a}^{i-1}(\check{e})^-$ of $\mathcal{Q}_{\mathcal{V},w}^-$, one arc corresponding to the edge $\mathfrak{a}^i(\check{e})^+$ of $\mathcal{Q}_{\mathcal{V},w}^+$ and two arcs each of which joins $\mathfrak{a}^{i-1}(\check{e})^-$ to $\mathfrak{a}^i(\check{e})^+$ and intersects a suture of $(M|S_w^\epsilon, \partial M|\partial S_w^\epsilon)$ exactly once; see Figure 8. We call $\mathfrak{a}^{i-1}(\check{e})^-$ the *bottom base* of the disk D_i^e and $\mathfrak{a}^i(\check{e})^+$ its *top base*. We denote them by $\partial^- D_i^e$ and $\partial^+ D_i^e$, respectively. The set $\partial_v D_i^e = \partial D_i^e - \mathrm{int}(\partial^+ D_i^e) - \mathrm{int}(\partial^- D_i^e)$ is called the *vertical boundary* of D_i^e.

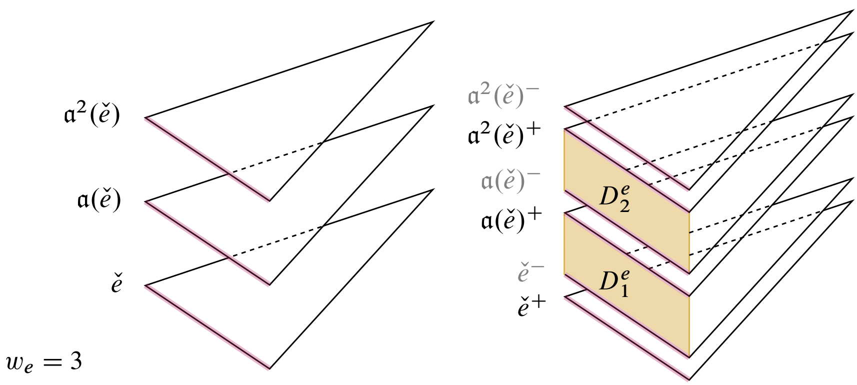


Figure 8: Two edge product disks D_1^e and D_2^e associated to an edge $e \in E$ of weight 3. For each copy of e in $\mathcal{Q}_{\mathcal{V},w}$ we draw only one triangle attached to it so that the disks D_1^e and D_2^e are clearly visible. To simplify notation we denote $L(e)$ by $\check{e}$.

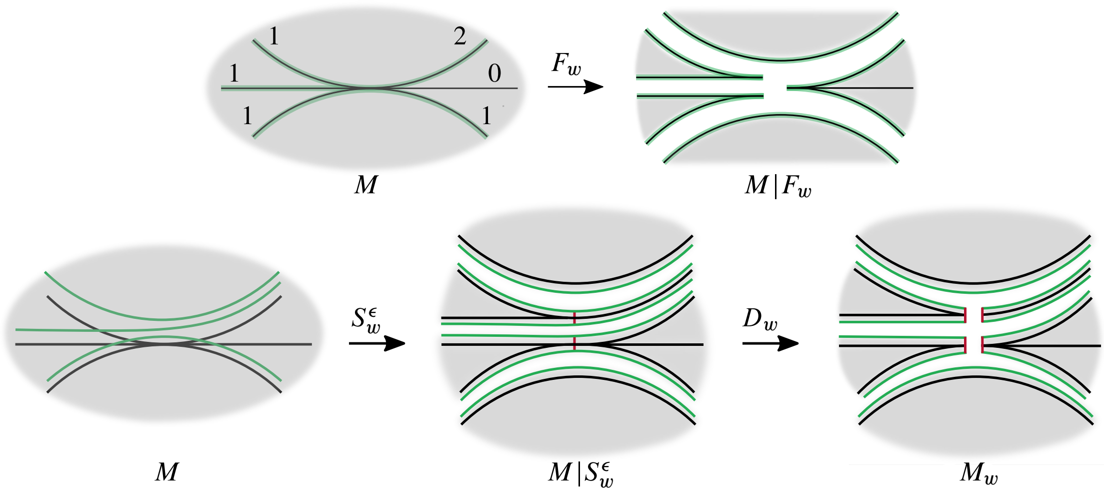


Figure 9: Top: Cutting along F_w. A weight on a face is indicated by the number immediately above the face. Bottom: First arrow: cutting along an embedded surface S_w^ϵ. Second arrow: cutting along edge product disks.

Disk D_i^e intersects the sutures of $(M|S_w^\epsilon, \partial M|\partial S_w^\epsilon)$ exactly twice. In the theory of sutured manifolds, properly embedded disks with this property are called *product disks*.

Let D_w denote the set of product disks in $M|S_w^\epsilon$ associated to the edges of $\mathcal{V}$ with $w_e > 1$. We say that an element of D_w is an *edge product disk*. Note that $M|S_w^\epsilon$ can admit more product disks, which are not elements of D_w. Let $M_w = (M|S_w^\epsilon)|D_w$. Since M_w arises as a result of decomposing a sutured manifold along finitely many product disks, it is also a sutured manifold; see [19, Definition 3.8]. Figure 9 illustrates the relationship between $M|F_w$ and M_w that we formalize in Lemma 3.4, after introducing necessary notation below.

Triangulations $\mathcal{Q}^+_{\mathcal{V},w}, \mathcal{Q}^-_{\mathcal{V},w} \subset M|S_w^\epsilon$ determine a pair of triangulations $\overline{\mathcal{Q}^+_{\mathcal{V},w}}$ and $\overline{\mathcal{Q}^-_{\mathcal{V},w}}$ in the top and the bottom boundary of M_w, respectively. For any triangle $g^\pm$ of $\mathcal{Q}^\pm_{\mathcal{V},w}$ there is an associated triangle $\overline{g^\pm}$ of $\overline{\mathcal{Q}^\pm_{\mathcal{V},w}}$. We will always assume that the indexing of edges/vertices of $\overline{g^\pm}$ is the same as in $g^\pm$. The only difference between $\mathcal{Q}^\pm_{\mathcal{V},w}$ and $\overline{\mathcal{Q}^\pm_{\mathcal{V},w}}$ is that there might be triangles $g_1^\pm$ and $g_2^\pm$ of $\mathcal{Q}^\pm_{\mathcal{V},w}$ which are identified along an edge e_1 of $g_1^\pm$ and an edge e_2 of $g_2^\pm$ such that the corresponding edges $\bar{e}_1$ of $\overline{g_1^\pm}$ and $\bar{e}_2$ of $\overline{g_2^\pm}$ are not identified in $\overline{\mathcal{Q}^\pm_{\mathcal{V},w}}$. This happens if and only if the common edge of $g_1^\pm$ and $g_2^\pm$ is the top or bottom base of an edge product disk from D_w.

For any $f \in F$ with $w_f > 1$, the sutured manifold M_w admits $(w_f - 1)$ connected components of the form $f \times [0,1]$. We call them *triangular prisms*, and denote the set of such triangular prisms in M_w by P_w. Each $P \in P_w$ is a sutured 3-ball, so we can speak about its top and bottom boundaries $\partial^+ P$ and $\partial^- P$, respectively. Observe that if $\partial^- P = \overline{g^-} \in \overline{F^-_{\mathcal{V},w}}$ then $\partial^+ P = \overline{\mathfrak{a}(g)^+} \in \overline{F^+_{\mathcal{V},w}}$. We call $\partial_v P = \partial P - \operatorname{int}(\partial^+ P) - \operatorname{int}(\partial^- P)$ the *vertical boundary* of P.

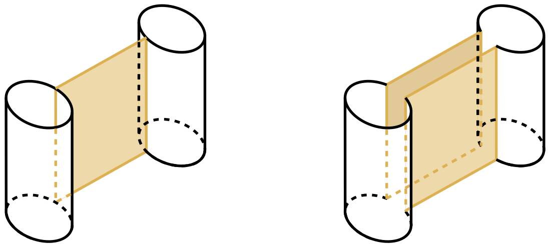


Figure 10: Left: the product D disk in $M|S_w^\epsilon$. Right: two disks contained in a sutured annulus of M_w arising from cutting $M|S_w^\epsilon$ along D.

Each edge product disk $D \in D_w$ gives rise to two disks D' and D'' contained in the sutured annuli of M_w; see Figure 10. We denote the set of such disks contained in the sutured annuli of M_w by $D(M_w)$. Let

(3.3) $$\mathrm{coll}\colon M_w \to \mathrm{coll}(M_w)$$

be the map which vertically collapses every $D \in D(M_w)$ to $\partial^- D$ and every $P \in P_w$ to $\partial^- P$.

Lemma 3.4 *The image of $M_w - P_w$ under* coll *is homeomorphic to $M|F_w$. Furthermore, for any $f \in F_w$,*

- *if $\overline{g^+}$ is a triangle of $\partial^+(M_w - P_w)$ with $\mathrm{coll}(\overline{g^+}) = f^+$ then $\overline{g^+} = \overline{L(f)^+}$,*
- *if $\overline{g^-}$ is a triangle of $\partial^-(M_w - P_w)$ with $\mathrm{coll}(\overline{g^-}) = f^-$ then $\overline{g^-} = \overline{U(f)^-}$.*

Proof The manifold M_w can be seen as a sutured manifold $M|(S_w^\epsilon \cup D_w)$. Observe that $S_w^\epsilon \cup D_w$ can be obtained from F_w in the following three steps:

(1) Replace every edge e in the branch locus of F_w by $e \times [0,1]$, keeping the triangles that were on the two sides of e attached to $e \times \{0\} \subset e \times [0,1]$.

(2) For every $f \in F_w$ with $w_f > 1$ add an additional $(w_f - 1)$ copies of f and for every edge e of f attach them along $e \times \{0\}$ immediately above f.

(3) For every $e \times [0,1]$ that occurs as a result of (1), spread the triangles that are on the two sides of $e \times \{0\}$ evenly along $e \times [0,1]$.

The last step is possible because edges in the branch locus of F_w correspond to the edges of $\mathcal{V}$ with weight greater than one, and thus there are at least two triangles attached to either side of $e \times [0,1]$.

It follows that collapsing every $D \in D(M_w)$ vertically to $\partial^- D$ collapses $S_w^\epsilon \cup D_w$ into a branched surface whose branch locus can be identified with that of F_w, but which has multiple parallel copies of $f \in F_w$ whenever $w_f > 1$. Each region between two parallel copies of f corresponds a triangular prism P with its vertical boundary collapsed and such that $\partial^- P = \overline{\mathfrak{a}^k(\check{f})^-}$ and $\partial^+ P = \overline{\mathfrak{a}^{k+1}(\check{f})^+}$ for some $0 \le k \le w_f - 2$ (recall that $\check{f} = L(f)$). Therefore collapsing each such P into $\partial^- P$ results in collapsing them all to the lowermost copy of f. Performing this for all $f \in F_w$ yields F_w. It follows that $\mathrm{coll}(M_w - P_w)$ can be identified with $M|F_w$.

The "furthermore" part follows from the observation that $\overline{L(f)^+}$ is the only copy of f in $\overline{\mathcal{Q}^+_{\mathcal{V},w}}$ which does not have a triangular prism below it, and $\overline{U(f)^-}$ is the only copy of f in $\overline{\mathcal{Q}^-_{\mathcal{V},w}}$ which does not have a triangular prism above it. $\square$

3.5 The mutant triangulation

In this subsection we explain how to construct the *mutant triangulation* $\mathcal{T}^\varphi$ out of the cut triangulation $(\mathcal{T}|F_w, \alpha|F_w)$, defined in Section 3.4, and a combinatorial automorphism $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$. This mutant triangulation is not guaranteed to be veering even when we do have a partial veering structure $\mathcal{B}|F_w$ on $(\mathcal{T}|F_w, \alpha|F_w)$. In Section 3.6 we give sufficient and necessary conditions on φ that guarantee $\mathcal{T}^\varphi$ is an ideal triangulation of M^φ. In Section 3.7 we put additional restrictions on φ which allow us to define taut and veering structures on $\mathcal{T}^\varphi$, thus resulting in a veering triangulation $\mathcal{V}^\varphi = (\mathcal{T}^\varphi, \alpha^\varphi, \mathcal{B}^\varphi)$.

Recall from Section 3.2 that $F_{\mathcal{V},w}$ denotes the set of triangles of $\mathcal{Q}_{\mathcal{V},w}$. A combinatorial automorphism $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ gives a bijection $\varphi\colon F_{\mathcal{V},w} \to F_{\mathcal{V},w}$ and a set of bijections $\{\varphi_g\}_{g \in F_{\mathcal{V},w}}$ between the edges of $g \in F_{\mathcal{V},w}$ and the edges of $\varphi(g) \in F_{\mathcal{V},w}$. Using the natural correspondence between the triangulations $\mathcal{Q}^+_{\mathcal{V},w}$ and $\mathcal{Q}^-_{\mathcal{V},w}$ in the top and bottom boundaries of $M|S^\epsilon_w$, respectively, we can view $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ as a combinatorial isomorphism $\varphi\colon \mathcal{Q}^+_{\mathcal{V},w} \to \mathcal{Q}^-_{\mathcal{V},w}$. Thus we can use φ to construct a mutant manifold M^φ out of $M|S^\epsilon_w$.

However, as explained in Section 3.4, when we work with the triangulation $(\mathcal{T}, \alpha)$ of M, we generally do not cut along S^ϵ_w, but along F_w. Thus to construct the mutant triangulation $\mathcal{T}^\varphi$ we need to specify a *regluing map* $r(\varphi) = (r^\varphi\colon F^+_w \to F^-_w, (r^\varphi_{f^+})_{f^+ \in F^+_w})$ determined by φ, consisting of a bijection $r^\varphi\colon F^+_w \to F^-_w$ and a family of bijections $(r^\varphi_{f^+})_{f^+ \in F^+_w}$ between edges of $f^+ \in F^+_w$ and edges of $r^\varphi(f^+) \in F^-_w$. The map $r(\varphi)$ has to be such that $\mathcal{T}^\varphi$ obtained from $(\mathcal{T}|F_w, \alpha|F_w)$ by identifying F^+_w with F^-_w via $r(\varphi)$ is, at least under certain conditions, a triangulation of M^φ. In this section we will define $r(\varphi)$. A sufficient and necessary condition on φ for the mutant triangulation to be a triangulation of M^φ appears in Theorem 3.10.

Below we use notation introduced in Section 3.2. In particular, recall that given $f \in F_w$, by $U(f)$ and $L(f)$ we denote the uppermost and the lowermost copies of f in $\mathcal{Q}_{\mathcal{V},w}$, respectively. Furthermore, let $\iota\colon M_w \to M|S^\epsilon_w$ be the surjective immersion, induced by cutting $M|S^\epsilon_w$ along D_w, which sends $\overline{g^+}$ to g^+ and $\overline{g^-}$ to g^- for every $g \in F_{\mathcal{V},w}$. Given $P \in P_w$ we will say that $\iota(P)$ is a triangular prism in $M|S^\epsilon_w$. To simplify notation, we set $P^\iota = \iota(P)$ and $\partial^\pm P^\iota = \iota(\partial^\pm P)$.

Before we formally define $r(\varphi)$ let us briefly explain the idea behind its definition. A triangle g^+ of $\mathcal{Q}^+_{\mathcal{V},w}$ does not have a triangular prism below it if and only if it is in $L(F_w)^+$; see Figure 7. Thus there is a natural identification between $L(F_w)^+$ and F^+_w. In particular, $r^\varphi(f^+)$ will depend on $\varphi(L(f)^+)$. Similarly, a triangle g^- of $\mathcal{Q}^-_{\mathcal{V},w}$ does not have a triangular prism above it if and only if it is in $U(F_w)^-$. Therefore we can identify $U(F_w)^-$ with F^-_w. If $\varphi(L(f)^+)$ is the bottom base of a triangular prism P^ι

then it is not immediately clear to which triangle of F_w^- the map r^φ should send f^+. In this case we flow upwards through the prism P^ι and look at $\varphi(\partial^+ P^\iota)$. If it is in $U(F_w)^-$ then the image of f^+ under r^φ will be the triangle f'^- with $U(f')^- = \varphi(\partial^+ P^\iota)$. Otherwise, we continue flowing upwards through triangular prisms. Below we describe this procedure more formally and prove that it always terminates.

Given $f \in F_w$ we define a sequence $g^\varphi(f) = (g_i)_{i \geq 1}$ of triangles of $\mathcal{Q}_{\mathcal{V},w}$ as follows. The first element g_1 is equal to $\varphi(L(f))$. For $i \geq 1$ if $g_i \in U(F_w)$ we are done. Otherwise, there is another triangle $\mathfrak{a}(g_i)$ of $\mathcal{Q}_{\mathcal{V},w}$ which is a copy of the same triangle of $(\mathcal{T}, \alpha)$ as g_i and lies immediately above g_i. Then we set $g_{i+1} = \varphi(\mathfrak{a}(g_i))$.

Lemma 3.5 *For every $f \in F_w$ the sequence $g^\varphi(f)$ is finite. Furthermore, if $f, f' \in F_w$ are distinct then the last elements of $g^\varphi(f)$ and $g^\varphi(f')$ are distinct.*

Proof Since the triangulation $\mathcal{Q}_{\mathcal{V},w}$ consists of finitely many triangles and g_i completely determines g_{i+1}, if the sequence $g^\varphi(f)$ is infinite then it is eventually periodic. That is, there are integers $m \geq 0$ and $N \geq 1$ such that $g_{m+j} = g_{m+kN+j}$ for any $j \in \{1, 2, \dots, N\}$ and $k \geq 0$. Pick minimal such m and N. We can write $g^\varphi(f)$ as

$$g^\varphi(f) = g_1, g_2, \dots, g_m, (h_1, \dots, h_N), (h_1, h_2, \dots, h_N), \dots.$$

First observe that we can assume that $m < N$. Otherwise, using the definition of $g^\varphi(f)$ and the fact that φ is a bijection on the set of triangles of $\mathcal{Q}_{\mathcal{V},w}$, we get that $g_{m-k} = h_{N-k}$ for any $k \in \{0, 1, \dots, N-1\}$. This means that the period $(h_1, \dots, h_N)$ starts immediately after g_{m-N} if $m > N$, or there is no preperiodic sequence at all if $m = N$. This is a contradiction with the minimality of m. On the other hand, if $m < N$ we obtain the equality $L(f) = \mathfrak{a}(h_{N-m})$. This is a contradiction, because the lowermost copy of f in $\mathcal{Q}_{\mathcal{V},w}$ does not have any copies of f below it, so in particular it cannot lie immediately above h_{N-m}. Thus $g^\varphi(f)$ is finite.

Now suppose that for some $1 \leq k \leq l$ we have

$$g^\varphi(f) = (g_1, g_2, \dots, g_k) \quad \text{and} \quad g^\varphi(f) = (g'_1, g'_2, \dots, g'_l).$$

If $g_k = g'_l$ then $g_1 = g'_{l-k+1}$. If $l < k$ we get $L(f) = \mathfrak{a}(g'_{l-k})$, which is a contradiction, because $L(f)$ is the lowermost copy of f in $\mathcal{Q}_{\mathcal{V},w}$ while $\mathfrak{a}(g'_{l-k})$ lies above g'_{l-k}. If $k = l$ we get $L(f) = L(f')$ and thus $f = f'$ by the injectivity of L. □

Let $f \in F_w$. Denote by $k \geq 1$ the length of the sequence $g^\varphi(f)$. By definition, $g_i \notin U(F_w)$ for all $i \in \{1, 2, \dots, k-1\}$ and $g_k \in U(F_w)$. Since U is injective, there is a unique $f' \in F_w$ such that $g_k = U(f')$. Let f^+ and f'^- be the triangles corresponding to f and f' in F_w^+ and F_w^-, respectively, and set

$$r^\varphi(f^+) = f'^- \in F_w^-.$$

Lemma 3.5 and injectivity of U imply that r^φ is a bijection. Thus it determines a pairing between faces of the top boundary F_w^+ of the cut triangulation $(\mathcal{T}|F_w, \alpha|F_w)$ and faces of the bottom boundary F_w^- of $(\mathcal{T}|F_w, \alpha|F_w)$. To define a mutant triangulation built out of $(\mathcal{T}|F_w, \alpha|F_w)$ it therefore remains to specify, for every $f \in F_w$, a bijection $r^{\varphi}_{f^+}$ between the edges of f^+ and that of f'^-. As in Section 3.2, to simplify notation we will denote $L(f)$ by $\check{f}$. Recall from Definition 3.1 that φ associates to $\check{f}$ a bijection $\varphi_{\check{f}}$ between the edges of $\check{f}$ and edges of $g_1 = \varphi(\check{f})$. Analogously, for $i = 1, 2, \dots, k-1$ there is a bijection φ_i between the edges of $\mathfrak{a}(g_i)$ and edges of $g_{i+1} = \varphi(\mathfrak{a}(g_i))$. Let δ_i be the bijection between the edges of g_i and edges of $\mathfrak{a}(g_i)$ such that $\delta_i(e) = \mathfrak{a}(e)$ for any edge e of g_i. Recall from Section 3.2 that σ^L_f and σ^U_f denote the bijections between the edges of f and edges of $L(f)$ and $U(f)$, respectively. Using this we set

$$r^{\varphi}_{f^+} = (\sigma^U_{f'})^{-1} \circ \varphi_{k-1} \circ \delta_{k-1} \circ \varphi_{k-2} \circ \delta_{k-2} \circ \cdots \circ \varphi_1 \circ \delta_1 \circ \varphi_{\check{f}} \circ \sigma^L_f.$$

We will also write $r^{\varphi}_{f^+} = (\sigma^U_{f'})^{-1} \circ (\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}} \circ \sigma^L_f$ for brevity. We define the *mutant triangulation* $\mathcal{T}^{\varphi}$ as the triangulation obtained from $(\mathcal{T}|F_w, \alpha|F_w)$ by identifying a triangle $f^+ \in F^+_w$ with the triangle $r^{\varphi}(f^+) \in F^-_w$ in such a way that an edge e of f^+ is identified with the edge $r^{\varphi}_{f^+}(e)$ of $r^{\varphi}(f^+)$.

Observe that $\mathcal{T}^{\varphi}$ is an ideal triangulation of $M^{r(\varphi)}$ which is not necessarily homeomorphic to M^{φ}. We explore this problem in the next subsection. For now, we state the relationship between $r(\varphi) = (r^{\varphi}\colon F^+_w \to F^-_w, (r^{\varphi}_{f^+})_{f^+ \in F^+_w})$ and $\varphi = (\varphi\colon \mathcal{Q}^+_{\mathcal{V},w} \to \mathcal{Q}^-_{\mathcal{V},w}, (\varphi_g)_{g^+ \in F^+_{\mathcal{V},w}})$.

Lemma 3.6 *Let $f \in F_w$. Suppose that*

$$r^{\varphi}(f^+) = f'^- \quad \text{and} \quad r^{\varphi}_{f^+} = (\sigma^U_{f'})^{-1} \circ (\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}} \circ \sigma^L_f.$$

Then:

(1) $k = 1$ *if and only if* $\varphi(L(f)^+) = U(f')^-$ *and their vertices are identified by* $\varphi_{\check{f}}$.

(2) $k \geq 2$ *if and only if there is a sequence* $(P^{\iota}_1, \dots, P^{\iota}_{k-1})$ *of triangular prisms in* $M|S^{\epsilon}_w$ *with the following properties*:

- $\varphi(L(f)^+) = \partial^- P^{\iota}_1$ *and their vertices are identified by* $\varphi_{\check{f}}$,
- *the vertex* v *of* $\partial^- P^{\iota}_i$ *is below the vertex* $\delta_i(v)$ *of* $\partial^+ P^{\iota}_i$ *for* $i = 1, 2, \dots, k-1$,
- $\varphi(\partial^+ P^{\iota}_i) = \partial^- P^{\iota}_{i+1}$ *and their vertices are identified by* φ_i *for* $i = 1, 2, \dots, k-2$,
- $\varphi(\partial^+ P^{\iota}_{k-1}) = U(f')^-$ *and their vertices are identified by* φ_{k-1}.

Denote by $P(f)$ *the quotient space of* $L(f)^+ \cup \big(\bigcup_{i=1}^{k-1} P^{\iota}_i\big) \cup U(f')^-$ *by the identifications listed above. Vertically collapsing* $P(f)$ *results in an identification of* $L(f)^+$ *with* $U(f')^-$ *with vertex correspondence given by* $(\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}}$.

Proof We view φ as a combinatorial automorphism $\varphi\colon \mathcal{Q}^+_{\mathcal{V},w} \to \mathcal{Q}^-_{\mathcal{V},w}$. The assumption on $r^{\varphi}_{f^+}$ implies that $g^{\varphi}(f)$ has length k. Let $g^{\varphi}(f) = (g_1, \dots, g_k)$. By the definition of $g^{\varphi}(f)$ we get that $\varphi(L(f)^+) \in U(F_w)^-$ if and only if $k = 1$. In this case indeed $\varphi(L(f)^+) = g^-_1 = U(f')^-$ and the vertex correspondence between these triangles is given by $\varphi_{\check{f}}$.

For the case $k \geq 2$ the existence of $P^{\iota}_1, P^{\iota}_2, \dots, P^{\iota}_{k-1}$ follows from the definition of $g^{\varphi}(f)$. Namely, since $g^-_i \notin U(F_w)^-$ and $g^-_{i+1} = \varphi(\mathfrak{a}(g_i)^+)$ for $1 \leq i \leq k-1$, there are prisms $P^{\iota}_1, \dots, P^{\iota}_{k-1}$ in $M|S^{\epsilon}_w$ satisfying $\partial^- P^{\iota}_i = g^-_i$ and $\partial^+ P^{\iota}_i = \mathfrak{a}(g_i)^+$, and such that identifications on $\partial^{\pm} P^{\iota}_i$ are as required. ◻

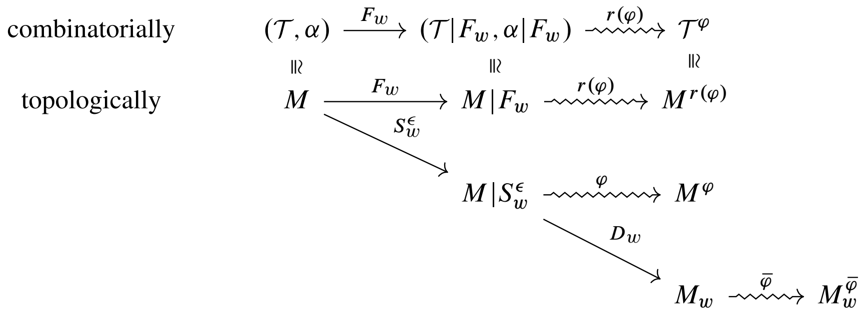


Figure 11: Each straight arrow corresponds to cutting the 3-manifold on the left of the arrow open along the set specified above the arrow. Each squiggly arrow corresponds to gluing the top boundary of the sutured manifold on the left of the arrow to its bottom boundary via the map specified above the arrow.

3.6 The manifold underlying the mutant triangulation

The mutant triangulation $\mathcal{T}^\varphi$ defined in Section 3.5 is an ideal triangulation of $M^{r(\varphi)}$. In this section we study the relationship between $M^{r(\varphi)}$ and M^φ, and derive sufficient and necessary condition for them to be homeomorphic.

Recall from Lemma 3.4 that $M|F_w$ is more closely related to $M_w=(M|S_w^\epsilon)|D_w$ than it is to $M|S_w^\epsilon$. As in Section 3.4, we will denote the triangulations in the top and bottom boundary of M_w by $\overline{\mathcal{Q}_{\mathcal{V},w}^+}$ and $\overline{\mathcal{Q}_{\mathcal{V},w}^-}$, respectively. A combinatorial isomorphism $\varphi\colon \mathcal{Q}_{\mathcal{V},w}^+\to\mathcal{Q}_{\mathcal{V},w}^-$ determines a map $\overline{\varphi}\colon \overline{\mathcal{Q}_{\mathcal{V},w}^+}\to\overline{\mathcal{Q}_{\mathcal{V},w}^-}$ via $\overline{\varphi}(\bar{f})=\overline{\varphi(f)}$ and $\overline{\varphi}_{\bar{f}}=\varphi_f$. Note that $\overline{\varphi}$ is not a combinatorial automorphism in the sense of Definition 3.1, as it can map a pair nonadjacent triangles to a pair of adjacent triangles, and vice versa. Nonetheless, we can use $\overline{\varphi}$ to construct a mutant manifold $M_w^{\overline{\varphi}}$ out of M_w. Figure 11 summarizes relationships between M^φ, $M^{r(\varphi)}$ and $M_w^{\overline{\varphi}}$.

The relationship between M^φ and $M_w^{\overline{\varphi}}$ is the easiest to state:

Lemma 3.7 *$M_w^{\overline{\varphi}}$ is obtained from M^φ by cutting it along finitely many (potentially zero) disks, annuli and Möbius bands.*

Proof Recall that $M_w=(M|S_w^\epsilon)|D_w$ is obtained from $M|S_w^\epsilon$ by cutting it along finitely many product disks. These cuts persists in the mutant manifold $M_w^{\overline{\varphi}}$. Since $\overline{\varphi}$ maps $\bar{f}$ to $\overline{\varphi(f)}$ with vertex correspondence φ_f, these cuts are the only difference between M^φ and $M_w^{\overline{\varphi}}$.

Observe that under φ the top boundary of an edge product disk $D\in D_w$ can be mapped to the bottom boundary of an edge product disk $D'\in D_w$. Thus edge product disks can match up into annuli or Möbius bands in M^φ which are cut in $M_w^{\overline{\varphi}}$. □

We will say that the disks, annuli and Möbius bands in M^φ coming from edge product disks are *vertical*. Below we define a property of φ which ensures the existence of vertical annuli or Möbius bands:

Definition 3.8 Let S_w be a surface properly carried by a veering triangulation $\mathcal{V}$. We say $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ *aligns edge product disks* if there is a sequence of edge product disks $(D_i)_{i \in I} \subset D_w$ in $M|S_w^\epsilon$ which glue up to an annulus or a Möbius band in M^φ. Otherwise we say that φ *misaligns edge product disks*.

In Theorem 3.10 we will prove that the mutant triangulation $\mathcal{T}^\varphi$ is an ideal triangulation of M^φ if and only if φ misaligns edge product disks. The forward direction will rely on an observation that when φ aligns edge product disks, M^φ and $M^{r(\varphi)}$ either have different numbers of connected components or have nonhomeomorphic boundaries. The boundary of M^φ is composed of sutured annuli and tori of $M|S_w^\epsilon$. Since S_w^ϵ is an oriented surface in an oriented 3-manifold M, it induces orientations on the boundaries of sutured annuli of $M|S_w^\epsilon$. Furthermore, since φ is orientation preserving, it sends a top boundary of a sutured annulus of $M|S_w^\epsilon$ to a bottom boundary of a sutured annulus of $M|S_w^\epsilon$ in an orientation-preserving way. It follows that all boundary components of M^φ are tori.

Lemma 3.4 implies that by identifying, for every $f \in F_w$, f^+ with $\overline{L(f)^+}$ and f^- with $\overline{U(f)^-}$, we can view $M^{r(\varphi)}$ as a quotient space of $\mathrm{coll}(M_w - P_w)$. Therefore boundary components of $M^{r(\varphi)}$ are composed of the images of sutured annuli and tori of $M_w - P_w$ under the collapsing map (3.3). Given a sutured annulus A of $M|S_w^\epsilon$ the image $\mathrm{coll}(\iota^{-1}(A))$ is either an annulus or a disjoint union of bigon disks and intervals. The latter option happens if and only if $A \cap D_w \neq \varnothing$. In this case D_w separates A into finitely many rectangles that we call *D_w-rectangles*.

Definition 3.9 Let A be a sutured annulus of $M|S_w^\epsilon$. We say that a subset $R \subset A$ is a *D_w-rectangle* if there are edge product disks $D, D' \in D_w$ such that the boundary of R decomposes into four arcs: one arc $\partial^+ R$ contained in $\partial^+ A$, one arc $\partial^- R$ contained in $\partial^- A$, one arc contained in $\partial_v D$ and one arc contained in $\partial_v D'$. We call the last two arcs in the boundary of R the *vertical sides* of R.

We say that a D_w-rectangle R is *prismatic* if there is a triangular prism $P \in P_w$ such that $P \cap \iota^{-1}(R) = \iota^{-1}(R)$. Otherwise we say that R is *nonprismatic*. If R is nonprismatic then $\mathrm{coll}(\iota^{-1}(R))$ is a bigon disk. In this case the boundary of $\mathrm{coll}(\iota^{-1}(R))$ decomposes into the *positive boundary* $\partial^+ \mathrm{coll}(\iota^{-1}(R)) = \mathrm{coll}(\iota^{-1}(\partial^+ R))$, and the *negative boundary* $\partial^- \mathrm{coll}(\iota^{-1}(R)) = \mathrm{coll}(\iota^{-1}(\partial^- R))$; see Figure 22. When R is prismatic, $\mathrm{coll}(\iota^{-1}(R))$ is an interval.

Theorem 3.10 *The mutant triangulation $\mathcal{T}^\varphi$ is an ideal triangulation of M^φ if and only if φ misaligns edge product disks.*

Proof By Lemma 3.4, we can view $M^{r(\varphi)}$ as a quotient space of $\mathrm{coll}(M_w - P_w)$. The map

$$\partial^+(\mathrm{coll}(M_w - P_w)) \to \partial^-(\mathrm{coll}(M_w - P_w))$$

by which we quotient $\mathrm{coll}(M_w - P_w)$ to get $M^{r(\varphi)}$ is obtained from

$$r(\varphi) = (r^\varphi \colon F_w^+ \to F_w^-, (r_{f^+}^\varphi)_{f^+ \in F_w^+})$$

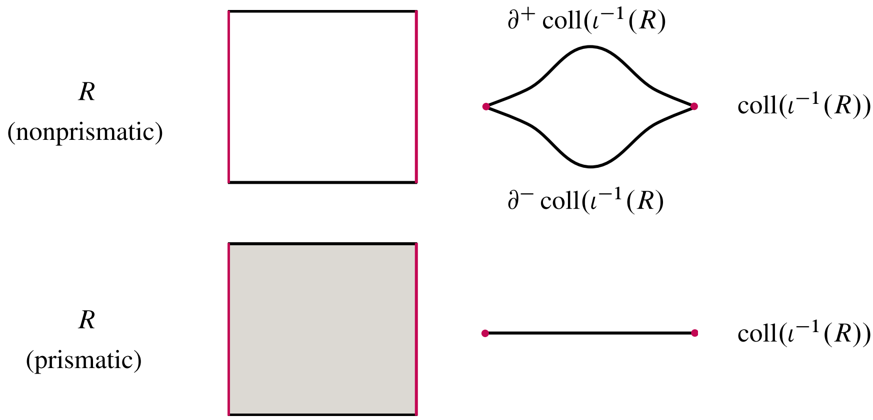


Figure 12: Top: a nonprismatic D_w-rectangle and its image under coll. Bottom: a prismatic D_w-rectangle and its image under coll. Red vertical intervals correspond to the intersection of R with D_w.

by modifying the bijections r^{φ}_{f+} to make up for the identifications $f^+ \sim \overline{L(f)^+}$ and $f^- \sim \overline{U(f)^-}$. For simplicity, we abuse the notation and denote this map by $r(\varphi)$. Let V^{φ}_w denote the set of vertical annuli and Möbius bands in M^{φ}. This set is nonempty if and only if φ aligns edge product disks.

First suppose that φ aligns edge product disks.

Case 1 There is a boundary torus T of M^{φ} such that $T \cap V^{\varphi}_w \neq \varnothing$ and T is not composed entirely of prismatic D_w-rectangles.

Observe that $T \cap V^{\varphi}_w$ consists of finitely many parallel simple closed curves in T. We denote the connected components of $T \cap V^{\varphi}_w$ by $d_1, \dots, d_r$ for $r \geq 1$, and we assume that they are circularly ordered so that d_i and d_{i+1} cobound an annulus $X_i \subset T$ whose interior is disjoint from V^{φ}_w (the subscript r is taken modulo r). Let $A_1, \dots, A_N$ be sutured annuli of $M|S^{\epsilon}_w$ such that $\varphi(\partial^+ A_j) = \partial^- A_{j+1}$ for every $j = 1, 2, \dots, N$ (the subscript j is taken modulo N) and T is the quotient of $A_1 \sqcup A_2 \sqcup \cdots \sqcup A_N$ by φ. Each $A_j \cap X_i$ consists of finitely many D_w-rectangles; see Figure 13, left. Let $\mathcal{R}^j_i$ be the collection of D_w-rectangles making up $A_j \cap X_i$ (in Figure 13, left, that would be all rectangles in one row) and let $\mathcal{R}_i$ be the union of all $\mathcal{R}^j_i$. By the assumption that T is not composed entirely of prismatic D_w-rectangles, there must be $i \in \{1, 2, \dots, r\}$ such that $\mathcal{R}_i$ contains a nonprismatic D_w-rectangle.

If every nonprismatic D_w-rectangle $R \in \mathcal{R}_i$ has both vertical sides contained in $d_i \cup d_{i+1}$ then each $\mathcal{R}^j_i$ either contains only prismatic D_w-rectangles or contains exactly one nonprismatic D_w-rectangle. Furthermore, since there is at least one nonprismatic D_w-rectangle in $\mathcal{R}_i$, there is at least one $\mathcal{R}^j_i$ of the latter type. Thus, by the definition of $r(\varphi)$, there are nonprismatic D_w-rectangles $R_1, \dots, R_n \in \mathcal{R}_i$ for $1 \leq n \leq N$ such that $r(\varphi)(\partial^+ \operatorname{coll}(\iota^{-1}(R_j))) = \partial^- \operatorname{coll}(\iota^{-1}(R_{j+1}))$, where the subscript j is taken modulo n. Since the image of a nonprismatic D_w-rectangle under coll is a bigon disk (see Figure 12), this gives us a sequence of bigon disks glued to each other top to bottom. The bigons $\operatorname{coll}(\iota^{-1}(R_j))$

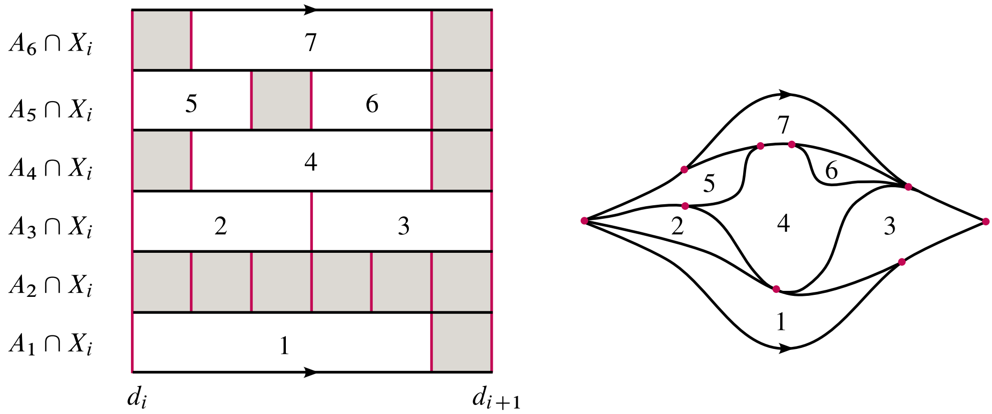


Figure 13: Left: An annular subset X_i of a boundary torus T of M^φ cobounded by a pair of vertical annuli or Möbius bands. Red intervals correspond to the intersection of X_i with edge product disks. Prismatic D_w-rectangles are shaded gray. Nonprismatic D_w-rectangles are numbered. Right: Spherical boundary component of $M^{r(\varphi)}$ corresponding to X_i. A bigon arising as a result of collapsing a nonprismatic D_w-rectangle labeled with i on the left is labeled with i.

inherit orientation on their boundary from R_j. Thus the assumption that φ is orientation preserving together with Lemma 3.6 imply that the quotient space of $\operatorname{coll}(\iota^{-1}(\mathcal{R}_i))$ by $r(\varphi)$ is a sphere. This means that $M^{r(\varphi)}$ admits a spherical boundary component. Since M^φ has only toroidal boundary components, these manifolds cannot be homeomorphic.

Now suppose that $R \in \mathcal{R}_i$ is a nonprismatic D_w-rectangle which has a vertical side d that is disjoint from $d_i \cup d_{i+1}$. An example of such a situation is presented in Figure 13. The assumption that $\operatorname{int}(X_i) \cap V_w^\varphi = \varnothing$ implies that there is a nonprismatic D_w-rectangle $R' \in \mathcal{R}_i$ such that $r(\varphi)$ identifies $\operatorname{coll}(\iota^{-1}(d))$ with an interior point of $\partial^- \operatorname{coll}(\iota^{-1}(R'))$. Therefore the quotient space of $\operatorname{coll}(\iota^{-1}(\mathcal{R}_i))$ by $r(\varphi)$ is again a sphere, and thus $M^{r(\varphi)}$ is not homeomorphic to M^φ.

Case 2 For every boundary torus T of M^φ either $T \cap V_w^\varphi = \varnothing$ or T is composed entirely of prismatic D_w-rectangles.

Since $V_w^\varphi \neq \varnothing$, the assumption of this case implies that there is a connected component of M^φ consisting entirely of the images of triangular prisms under ι. In particular, M^φ has strictly more connected components than $M^{r(\varphi)}$, so these manifolds are not homeomorphic. (Note, however, that $M^{r(\varphi)}$ may be homeomorphic to the union of other connected components of M^φ.)

Now suppose that φ misaligns edge product disks. Then there are no vertical annuli or Möbius bands in M^φ, and hence, by Lemma 3.7, $M_w^{\overline{\varphi}}$ can be obtained from M^φ by cutting it along finitely many vertical disks. Equivalently, M^φ is the quotient space of $\operatorname{coll}(M_w)$ by $\overline{\varphi}$.

Lemma 3.6 explains how the definition of the regluing map simulates the process of collapsing triangular prisms into their bottom triangles. Thus $r(\varphi)$ on F_w^+ respects not only the identification between $L(f)^+$

and $\varphi(L(f)^+)$, for all $f \in F_w$, but also the identification between g^+ and $\varphi(g^+)$ for all $g \in F_{\mathcal{V},w}$ such that either g or $\mathfrak{a}^{-1}(g)$ appears in the sequence $g^{\varphi}(f)$ for some $f \in F_w$. If there is $g \in F_{\mathcal{V},w} - L(F_w)$ such that for every $f \in F_w$ neither g nor $\mathfrak{a}^{-1}(g)$ appears in $g^{\varphi}(f)$, then there are triangular prisms of $M_w = (M|S_w^{\epsilon})|D_w$ through which we have not passed when defining $r(\varphi)$. These triangular prisms would arrange into solid tori components of $M_w^{\overline{\varphi}}$ consisting entirely of triangular prisms. However, the assumption that φ misaligns edge product disks implies that $M_w^{\overline{\varphi}}$ does not admit such solid tori components. Therefore when φ misaligns edge product disks Lemma 3.6 implies that the quotient space of $\operatorname{coll}(M_w)$ by $\overline{\varphi}$ is homeomorphic to the quotient space of $\operatorname{coll}(M_w - P_w)$ by $r(\varphi)$. The latter is $M^{r(\varphi)}$, while the former — as explained in the previous paragraph — is M^{φ}. Thus $M^{r(\varphi)}$ is homeomorphic to M^{φ}. □

Remark 3.11 In the proof of Theorem 3.10 we constructed a sphere out of bigon disks. This may look like a contradiction to the fact, mentioned in Section 2.1, that only surfaces with zero Euler characteristic can admit a bigon train track. However, the obtained decomposition of S^2 into bigons is not a bigon track in the usual sense. If τ is a train track on S then for every switch v of τ which is not contained in ∂S there must be two complementary regions of τ which meet v along a smooth point in their boundary. In the construction we get two points in the sphere which meet only cusps of bigons.

3.7 Veeringness of the mutant triangulation

In Section 3.6 we found a sufficient and necessary condition on φ for the mutant triangulation $\mathcal{T}^{\varphi}$ to be a triangulation of M^{φ}. In this subsection we are interested in endowing $\mathcal{T}^{\varphi}$ with a veering structure.

By tautness, edges of the dual spine $\mathcal{D}$ of $(\mathcal{T}, \alpha)$ admit orientations such that every vertex v of $\mathcal{D}$ has exactly two incoming edges and two outgoing edges; this is Definition 2.1(1). When we construct $\mathcal{T}^{\varphi}$ out of $(\mathcal{T}|F_w, \alpha|F_w)$ we always identify a face $f^+ \in F_w^+$ with a face $r^{\varphi}(f^+) \in F_w^-$. Therefore there is a natural orientation on the edges of the dual spine $\mathcal{D}^{\varphi}$ of $\mathcal{T}^{\varphi}$ that is induced from the orientation on the edges of the dual spine of $\mathcal{T}$. With this orientation $\mathcal{D}^{\varphi}$ satisfies Definition 2.1(1). To obtain a taut structure on $\mathcal{T}^{\varphi}$ it suffices to find a sufficient condition on $r(\varphi)$ that guarantees every 2-cell of $\mathcal{D}^{\varphi}$ has exactly one top vertex and exactly one bottom vertex. To derive such a condition it is helpful to analyze the structure of $\mathcal{D}|F_w$ and its relationship to $(\mathcal{T}|F_w, \alpha|F_w)$. First, observe that edges of $(\mathcal{T}|F_w, \alpha|F_w)$ can be classified into four types. We say that an edge e of $(\mathcal{T}|F_w, \alpha|F_w)$, or $\mathcal{V}|F_w$, is

- *internal* if e is neither an edge of a triangle from F_w^+ nor an edge of a triangle from F_w^-,
- *positive* if e is an edge of a triangle from F_w^+ and is not edge of a triangle from F_w^-,
- *negative* if e is an edge of a triangle from F_w^- and is not an edge of a triangle from F_w^+,
- *mixed* if e is an edge of a triangle from F_w^+ and also an edge of a triangle from F_w^-.

Assume that $\mathcal{T}$ and $\mathcal{D}$ are embedded in M so that they are dual to one another. For every 2-cell p of $\mathcal{D}|F_w$ there is a 2-cell s of $\mathcal{D}$ such that p is a connected component of $s|F_w$. If $s \cap F_w = \varnothing$ then we say that p is an *internal cell* of $\mathcal{D}|F_w$. Now suppose that $s \cap F_w \neq \varnothing$. If p contains the top vertex of s we say

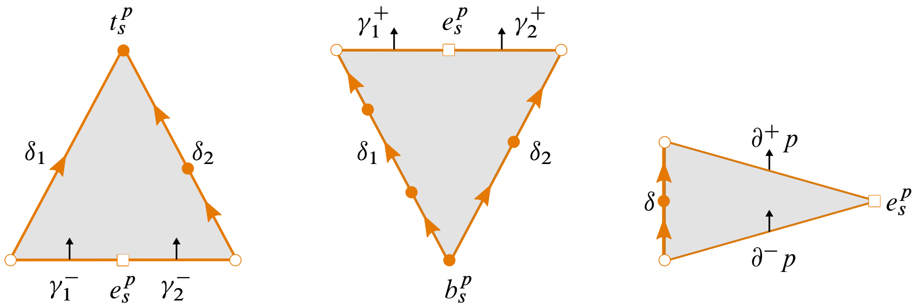


Figure 14: Left: negative cell of $\mathcal{D}|F_w$. Center: positive cell of $\mathcal{D}|F_w$. Right: mixed cell of $\mathcal{D}|F_w$. The number of vertices in the interiors of δ_1, δ_2 and δ may vary.

that p is a *negative cell* of $\mathcal{D}|F_w$; see Figure 14, left. If p contains the bottom vertex of s we say that p is a *positive cell* of $\mathcal{D}|F_w$; see Figure 14, center. If p contains neither the top nor the bottom vertex of s we say that p is a *mixed cell* of $\mathcal{D}|F_w$; see Figure 14, right. Naturally, an internal/positive/negative/mixed cell of $\mathcal{D}|F_w$ is dual to an internal/positive/negative/mixed edge of $(\mathcal{T}|F_w, \alpha|F_w)$.

For every cell s of $\mathcal{D}$ such that $s \cap F_w \neq \varnothing$, we denote by e_s the intersection of s with its dual edge in $(\mathcal{T}, \alpha)$. Then every connected component p of $s|F_w$ has a point in its boundary corresponding to e_s that we will denote by e_s^p. If p is negative then there is a point t_s^p in the boundary of p corresponding to the top vertex t_s of s. If p is positive then there is a point b_s^p in the boundary of p corresponding to the bottom vertex b_s of s.

If p is a negative cell of $\mathcal{D}|F_w$ then its boundary decomposes into

- two arcs γ_1^- and γ_2^- meeting at e_s^p, both cooriented into p,
- two arcs δ_1 and δ_2 meeting at t_s^p, both oriented so that they point into t_s^p.

See Figure 14, left. We say that γ_1^- and γ_2^- are *maximal negative arcs* in the boundary of p.

If p is a positive cell of $\mathcal{D}|F_w$ then its boundary decomposes into

- two arcs γ_1^+ and γ_2^+ meeting at e_s^p, both cooriented out of p,
- two arcs δ_1 and δ_2 meeting at b_s^p, both oriented so that they point out of b_s^p.

See Figure 14, center. We say that γ_1^+ and γ_2^+ are *maximal positive arcs* in the boundary of p.

If p is a mixed cell of $\mathcal{D}|F_w$ then its boundary decomposes into

- two arcs $\partial^+ p$ and $\partial^- p$ meeting at e_s^p, such that $\partial^+ p$ is cooriented out of p and $\partial^- p$ is cooriented into p,
- one arc δ oriented from $\partial^- p$ to $\partial^+ p$.

See Figure 14, right. We say that $\partial^+ p$ is the *maximal positive arc* in the boundary of p, and that $\partial^- p$ is the *maximal negative arc* in the boundary of p.

If $\gamma^{\pm}$ is a maximal positive/negative arc in the boundary of a cell p of $\mathcal{D}|F_w$ we say that e_s^p is the *internal endpoint* of $\gamma^{\pm}$.

Via duality between internal/positive/negative/mixed cells of $\mathcal{D}|F_w$ and internal/positive/negative/mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$, respectively, we get a combinatorial sufficient condition on $r(\varphi)$ that guarantees $\mathcal{T}^{\varphi}$ admits a taut structure in Lemma 3.12. In Lemma 3.15 we explain how this condition relates to the matching of edge product disks of $M|S_w^{\epsilon}$ under φ.

Lemma 3.12 *If for every mixed edge e of $(\mathcal{T}|F_w, \alpha|F_w)$ there is a positive edge e^+ of $(\mathcal{T}|F_w, \alpha|F_w)$ and a negative edge e^- of $(\mathcal{T}|F_w, \alpha|F_w)$ such that e, e^+ and e^- are identified in $\mathcal{T}^{\varphi}$ then the triangulation $\mathcal{T}^{\varphi}$ admits a taut structure.*

Proof We assume that the 1-skeleton of D^{φ} is equipped with the orientation induced by $\alpha|F_w$. It suffices to show that under the assumption of the lemma, every 2-cell of D^{φ} has exactly one top vertex and exactly one bottom vertex; see Definition 2.1.

Denote by Γ^+ (respectively, Γ^-) the set of maximal positive (respectively, negative) arcs in the boundaries of cells of $\mathcal{D}|F_w$. Recall that $\mathcal{T}^{\varphi}$ is the quotient space of $(\mathcal{T}|F_w, \alpha|F_w)$ under the regluing map $r(\varphi)$. Dually we get a regluing map $\Gamma(\varphi) = \{\Gamma^{\varphi}\colon \Gamma^+ \to \Gamma^-, (\Gamma^{\varphi}_{\gamma^+})_{\gamma^+\in\Gamma^+}\}$, where Γ^{φ} is a bijection between Γ^+ and Γ^- and $\Gamma^{\varphi}_{\gamma^+}$ is a bijection between the endpoints of γ^+ and the endpoints of $\Gamma^{\varphi}(\gamma^+)$, such that the quotient of $\mathcal{D}|F_w$ by $\Gamma(\varphi)$ gives $\mathcal{D}^{\varphi}$. Since the 1-skeleton of $\mathcal{D}^{\varphi}$ arises from recombining the 1-skeleton of $\mathcal{D}$, we get that $\Gamma^{\varphi}_{\gamma^+}$ must send the internal endpoint of γ^+ to the internal endpoint of $\Gamma^{\varphi}(\gamma^+)$.

Let p be a positive cell of $\mathcal{D}|F_w$. Denote by γ_1^+ and γ_2^+ the two distinct maximal positive arcs in the boundary of p. For $i = 1, 2$ the bijection Γ^{φ} can send γ_i^+ only to a maximal negative arc in the boundary of a mixed cell or to a maximal negative arc in the boundary of a negative cell. Let $q_1, \dots, q_m$ be the maximal collection of mixed cells of $\mathcal{D}|F_w$ such that $\Gamma^{\varphi}(\gamma_1^+) = \partial^- q_1$ and $\Gamma^{\varphi}(\partial^+ q_i) = \partial^- q_{i+1}$ for $i = 1, 2, \dots, m-1$. Let $q_1', \dots, q_n'$ be the maximal collection of mixed cells of $\mathcal{D}|F_w$ such that $\Gamma^{\varphi}(\gamma_2^+) = \partial^- q_1'$ and $\Gamma^{\varphi}(\partial^+ q_j') = \partial^- q_{j+1}'$ for $j = 1, 2, \dots, n-1$. First assume that these collections of mixed cells are nonempty, that is $m, n \geq 1$. By maximality and the fact that positive cells do not have arcs in Γ^-, there are negative cells p' and p'' of $\mathcal{D}|F_w$ such that $\Gamma^{\varphi}(\partial^+ q_m)$ is a maximal negative arc in the boundary of p' and $\Gamma^{\varphi}(\partial^+ q_n')$ is a maximal negative arc in the boundary of p''. The cells $q_1, \dots, q_m, q_1', \dots, q_n'$ must all be distinct, and hence $\Gamma^{\varphi}(\partial^+ q_m)$ and $\Gamma^{\varphi}(\partial^+ q_n')$ are distinct. Since $\Gamma(\varphi)$ sends internal endpoints of arcs to internal endpoints of arcs, we must have $p' = p''$. Thus we obtain a cell of $\mathcal{D}^{\varphi}$ composed of $p, q_1, \dots, q_m, q_1', \dots, q_n', p' = p''$. Such a cell has exactly one top vertex (coming from $p' = p''$) and exactly one bottom vertex (coming from p). If $m = 0$ it suffices to replace $\Gamma^{\varphi}(\partial^+ q_m)$ with $\Gamma^{\varphi}(\gamma_1^+)$, and if $n = 0$ it suffices to replace $\Gamma^{\varphi}(\partial^+ q_n')$ with $\Gamma^{\varphi}(\gamma_2^+)$, to still get a cell of $\mathcal{D}^{\varphi}$ with precisely one top vertex and precisely one bottom vertex.

It follows that every 2-cell of $\mathcal{D}^{\varphi}$ is either

- composed of one internal cell of $\mathcal{D}|F_w$, or

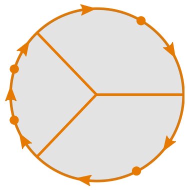


Figure 15: Gluing mixed cells of $\mathcal{D}|F_w$ cyclically yields a cell of $\mathcal{D}^\varphi$ whose edges are cyclically oriented.

- composed of one positive cell of $\mathcal{D}|F_w$, one negative cell of $\mathcal{D}|F_w$ and finitely many (potentially zero) mixed cells of $\mathcal{D}|F_w$, or
- composed of finitely many mixed cells of $\mathcal{D}|F_w$.

The last type of cells of $\mathcal{D}^\varphi$ have cyclically oriented edges in their boundary; see Figure 15. These are the only cells of $\mathcal{D}^\varphi$ that do not satisfy Definition 2.1. Using the duality between positive/negative/mixed cells of $\mathcal{D}|F_w$ and positive/negative/mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$ it is easy to see that if for every mixed edge e of $(T|F_w, \alpha|F_w)$ there is a positive edge e^+ of $(T|F_w, \alpha|F_w)$ and a negative edge e^- of $(T|F_w, \alpha|F_w)$ such that e, e^+ and e^- are identified in $\mathcal{T}^\varphi$ then $\mathcal{D}^\varphi$ does not admit such cells. Thus under this assumption, $\alpha|F_w$ induces a taut structure on $\mathcal{T}^\varphi$. □

We will devote the rest of this subsection to restate the condition of Lemma 3.12 in terms of φ and then prove that it is not only sufficient but also necessary for the existence of a taut structure on $\mathcal{T}^\varphi$. Recall that by $D(M_w)$ we denote the set of disks contained in the sutured annuli of M_w arising from cutting $M|S_w^\epsilon$ along D_w. Let $D(M_w - P_w)$ be the subset of $D(M_w)$ consisting only of those disks which are not contained in the sutured annuli of triangular prisms of M_w. Lemma 3.4 implies the following relationship between mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$ and disks $D(M_w - P_w)$ contained in the sutured annuli of $M_w - P_w$:

Corollary 3.13 *For every one sided-edge e of $(\mathcal{T}|F_w, \alpha|F_w)$ there is precisely one $D' \in D(M_w - P_w)$ such that $e = \mathrm{coll}(D')$.* □

However, we are mainly interested in the relationship between mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$ and edge product disks in $M|S_w^\epsilon$.

Definition 3.14 Let $D', D'' \in D(M_w)$ be such that $\iota(D') = \iota(D'') = D \in D_w$. We say that the edge product disk D

- *has prisms on both sides* if there are triangular prisms $P', P'' \in P_w$ such that D' is contained in the sutured annulus of P' and D'' is contained in the sutured annulus of P'',
- *has a prism on one side* if exactly one disk out of D' and D'' is contained in the sutured annulus of some triangular prism of M_w,
- *does not have a prism on either side* if $D, D' \in D(M_w - P_w)$.

Using this definition, we can restate Corollary 3.13 and say that an edge product disk $D \in D_w$ corresponds to two, one or zero mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$ if and only if D does not have a prism on either side, has a prism on one side or has prisms on both sides, respectively.

Let V be a vertical annulus or a vertical Möbius band in M^φ. Let $D_1, \dots, D_N$ be edge product disks of $M|S_w^\epsilon$ such that V is the quotient space of $D_1 \sqcup \dots \sqcup D_N$ by φ. We say that V *lies in a prismatic region of* M^φ if D_i has prisms on both sides for every $1 \le i \le N$. Using this terminology we can now restate the assumption of Lemma 3.12 in terms of topological properties of M^φ:

Lemma 3.15 *The following statements are equivalent*:

1. *For every mixed edge e of $(\mathcal{T}|F_w, \alpha|F_w)$ there is a positive edge e^+ of $(\mathcal{T}|F_w, \alpha|F_w)$ and a negative edge e^- of $(\mathcal{T}|F_w, \alpha|F_w)$ such that e, e^+ and e^- are identified in $\mathcal{T}^\varphi$.*
2. *Every vertical annulus or Möbius band in M^φ lies in a prismatic region of M^φ.*

Proof We show both directions by contraposition. First we show that (2) implies (1). Observe that if e is an edge of triangle from F_w^+ then the edge $r(\varphi)(e)$ might not be well defined, because a positive edge can be mapped to two different mixed edges. However, if e is a mixed edge then $r(\varphi)(e)$ is well defined, because e is an edge of only one $f^+ \in F_w^+$.

In what follows the subscript i is taken modulo n. Suppose that there is a collection $e_1, \dots, e_n$ of mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$ such that $r(\varphi)(e_i) = e_{i+1}$ for every $1 \le i \le n$. By Corollary 3.13, there is a collection $D_1', \dots, D_n' \in D(M_w - P_w)$ of disks contained in the sutured annuli of $M_w - P_w$ such that $\mathrm{coll}(D_i') = e_i$. Let $D_i = \iota(D_i') \in D_w$. By the definition of $r(\varphi)$, for every $1 \le i \le n$ either $\varphi(\partial^+ D_i) = \partial^- D_{i+1}$ or there is $k_i \ge 1$ and a collection of edge product disks $D_i^{(1)}, \dots, D_i^{(k_i)} \in D_w$ (which all have prisms on at least one side) such that $\varphi(\partial^+ D_i) = \partial^- D_i^{(1)}$ and $\varphi(\partial^+ D_i^{(j)}) = \partial^- D_i^{(j+1)}$ for every $1 \le j \le k_i - 1$ and $\varphi(\partial^+ D_i^{(k_i)}) = \partial^- D_{i+1}$. Therefore the quotient of $\bigsqcup_{i=1}^n D_i \sqcup \bigsqcup_{j=1}^{k_i} D_i^{(j)}$ by φ is a vertical annulus or Möbius band V in M^φ. Since $D_i' \in D(M_w - P_w)$, D_i does not have prisms on both sides. It follows that V does not lie in a prismatic region of M^φ.

Now let $D_1, \dots, D_n \in D_w$ be a collection of edge product disks in $M|S_w^\epsilon$ such that $\varphi(\partial^+ D_i) = \partial^- D_{i+1}$ for every $1 \le i \le n$. Denote by V the quotient space of $D_1 \sqcup \dots \sqcup D_n$ by φ. If V does not lie in a prismatic region of M^φ there is $1 \le k \le n$ and a sequence $1 \le i_1 < i_2 < \dots < i_k \le n$ such that D_{i_j} does not have prisms on both sides for every $1 \le j \le k$, and for every $l \notin \{i_1, i_2, \dots, i_k\}$ the edge product disk D_l has prisms on both sides. For $1 \le j \le k$ if D_{i_j} does not have a prism on either side there are disks $D_{i_j}', D_{i_j}'' \in D(M_w - P_w)$ such that $\iota(D_{i_j}') = \iota(D_{i_j}'') = D_{i_j}$. If D_{i_j} has a prism on one side, there is $D_{i_j}' \in D(M_w - P_w)$ such that $\iota(D_{i_j}') = D_{i_j}$. By Corollary 3.13 there are mixed edges $e_{i_j} = \mathrm{coll}(D_{i_j}')$ and $e_{i_j}' = \mathrm{coll}(D_{i_j}'')$ of $(\mathcal{T}|F_w, \alpha|F_w)$. By the definition of $r(\varphi)$, after possibly switching e_{ij} with e_{i_j}' for some $1 \le j \le k$, there is $m \ge 1$ and a sequence $1 \le l_1 < l_2 < \dots < l_m \le k$ such that $r(\varphi)(e_{i_{l_j}}) = e_{i_{l_{j+1}}}$ for every $1 \le j \le m$. This gives an edge of $\mathcal{T}^\varphi$ which is composed entirely of mixed edges of $(\mathcal{T}|F_w, \alpha|F_w)$. □

Proposition 3.16 *The mutant triangulation $\mathcal{T}^\varphi$ admits a taut structure if and only if every vertical annulus or Möbius band in M^φ lies in a prismatic region of M^φ.*

Proof The fact that when every vertical annulus or Möbius band in M^φ lies in a prismatic region of M^φ then $\mathcal{T}^\varphi$ admits a taut structure follows from Lemmas 3.12 and 3.15.

We prove the other direction by contraposition. Suppose that there is a vertical annulus or a Möbius band V in M^φ which does not lie in a prismatic region of M^φ. Let T be a boundary torus of M such that $T \cap V \neq \varnothing$. Since V does not consist entirely of edge product disks which have prisms on both sides, T does not consist entirely of prismatic D_w-rectangles. It follows from the proof of Theorem 3.10 (Case 1) that $M^{r(\varphi)}$, the manifold underlying $\mathcal{T}^\varphi$, admits a spherical boundary component. Therefore, by Lemma 2.3, $M^{r(\varphi)}$ does not have a taut triangulation. In particular, $\mathcal{T}^\varphi$ does not admit a taut structure. □

It follows that $\mathcal{T}^\varphi$ admits a taut structure if and only if orientations on the edges of the dual spine $\mathcal{D}^\varphi$ inherited from $(\mathcal{T}|F_w, \alpha|F_w)$ determine a taut structure on $\mathcal{T}^\varphi$. In this case we denote the taut structure on $\mathcal{T}^\varphi$ by α^φ. We also say that $(\mathcal{T}, \alpha)$ and $(\mathcal{T}^\varphi, \alpha^\varphi)$ are *taut mutants*.

The assumption that φ misaligns edge product disks is stronger than the assumption that every vertical annulus or Möbius band in M^φ lies in a prismatic region of M^φ. However, since it is a necessary condition for $\mathcal{T}^\varphi$ to be an ideal triangulation of M^φ (Theorem 3.10), it is reasonable to assume this stronger condition for the rest of the paper. Below we also prove that when φ aligns edge product disks then M^φ does not admit a veering triangulation. This further justifies restricting our considerations only to automorphisms which misalign edge product disks.

Proposition 3.17 *If $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ aligns edge product disks then M^φ does not admit a veering triangulation.*

Proof Since only hyperbolic manifolds can admit veering triangulations [23, Theorem 1.5] (stated here as Theorem 2.11), it suffices to show that M^φ is not hyperbolic. Recall that when φ aligns edge product disks M^φ must admit a vertical annulus or a vertical Möbius band. By passing to a finite cover, we can assume that there is a vertical annulus A in M^φ. We will show that A is essential, that is, incompressible, boundary-incompressible and not boundary parallel. First we will prove that M^φ is irreducible and boundary-irreducible.

Claim 3.17.1 *M^φ is irreducible and boundary-irreducible.*

Proof of Claim 3.17.1 Lackenby proved that every 3-manifold with a taut triangulation is irreducible and boundary-irreducible [27, Proposition 10]. Therefore it suffices to show that M^φ has a taut triangulation.

If every vertical annulus and Möbius band lies in a prismatic region of M^φ then, by Proposition 3.16, $\mathcal{T}^\varphi$ admits a taut structure, and we are done. To construct a taut triangulation on M^φ when there are vertical annuli or Möbius bands outside of the prismatic region of M^φ we first modify the triangulation $(\mathcal{T}, \alpha)$

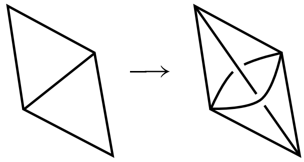


Figure 16: The 0-2 Pachner move. If f and g are on opposite sides of e we can cut $(\mathcal{T}, \alpha)$ along $f \cup g$ and glue in a "taut pillow", producing a new taut triangulation of M.

of M using 0-2 *Pachner moves*. Such a move can be performed on any pair of adjacent triangles f and g which are on opposite sides of an edge e of $(\mathcal{T}, \alpha)$. First, we cut $(\mathcal{T}, \alpha)$ along $f \cup g$. The resulting cut triangulation has four boundary faces: f^+, g^+, f^- and g^-. Using two additional taut tetrahedra t and t' we construct a "taut pillow" by identifying the square consisting of the top faces of t with the square consisting of the bottom faces of t' so that the top diagonal of t is identified with the bottom diagonal of t'; see Figure 16, right. Now there is a natural way to identify $f^+ \cup g^+$ with the bottom faces of t and the top faces of t' with $f^- \cup g^-$ to get a new taut triangulation of M. The key feature of the 0-2 moves is that by applying them finitely many times above pairs of triangles adjacent to edges with weight greater than one, we can construct a taut triangulation $(\mathcal{T}_*, \alpha_*)$ of M with the following properties:

(a) $(\mathcal{T}_*, \alpha_*)$ carries a surface S_{w_*} whose induced triangulation $\mathcal{Q}_*$ is combinatorially isomorphic to $\mathcal{Q}_{\mathcal{V},w}$.

(b) The weight system w_* on $(\mathcal{T}_*, \alpha_*)$ is such that no edge of $(\mathcal{T}_*, \alpha_*)$ has weight greater than one.

(c) Each edge product disk $D \in D_w$ gives a pair of edges of $(\mathcal{T}_*, \alpha_*)$ which are homotopic to each other (while keeping their endpoints on the boundary tori); one of these edges corresponds to $\partial^- D$, and the other to $\partial^+ D$; see Figure 16.

From (a) it follows that there is $\varphi_* \in \mathrm{Aut}^+(\mathcal{Q})$ such that $M^\varphi = M^{\varphi_*}$. Part (b) implies that $M|S^\epsilon_{w_*}$ does not admit any edge product disks and thus φ_* misaligns edge product disks. Therefore, by Proposition 3.16 and Theorem 3.10, we can use φ_* to construct a taut triangulation $(\mathcal{T}_*^{\varphi_*}, \alpha_*^{\varphi_*})$ of $M^\varphi = M^{\varphi_*}$. □

By construction, the core curve c of the vertical annulus A of M^φ is homotopic to an essential simple closed curve on some boundary torus T of M^φ. Claim 3.17.1 implies that T is incompressible, and thus A is incompressible.

Since A is composed entirely of edge product disks in $M|S^\epsilon_w$, an essential arc γ in A is homotopic to the image of the bottom boundary $\partial^- D$ of some edge product disk $D \in D_w$ under the quotient map $M|S^\epsilon_w \to M^\varphi$. By property (c) of the triangulation $(\mathcal{T}_*, \alpha_*)$ constructed in the proof of Claim 3.17.1, there is an edge of $(\mathcal{T}_*^{\varphi_*}, \alpha_*^{\varphi_*})$ which is homotopic to γ. This implies that A cannot be boundary parallel, because in a taut triangulation an edge is never homotopic into a vertex neighborhood while keeping its ends in the vertex neighborhood [24, Theorem 6.1]. If there was a boundary-compression disk for A, A would compress into a properly embedded disk D. By boundary-irreducibility of M^φ (Claim 3.17.1), D must be boundary parallel. Hence A must be boundary parallel, which contradicts our previous observation that it is not. Thus A is boundary-incompressible.

We showed that, up to passing to a finite cover, M^φ admits an essential annulus. So it either contains an essential torus or is a small Seifert fibered manifold [33, Lemma 11.2.10]. In either case, M^φ is not hyperbolic, and thus cannot admit a veering triangulation. □

From now on we assume that φ misaligns edge product disks. The taut triangulation $(\mathcal{T}^\varphi, \alpha^\varphi)$ is veering if and only if its dual spine $\mathcal{D}^\varphi$ has a smoothing into a branched surface which locally looks like in Figure 4; see Definition 2.5. One way of ensuring that is to construct the required branched surface structure on $\mathcal{D}^\varphi$ using the branched surface structure $\mathcal{B}|F_w$ on $\mathcal{D}|F_w$. Extending the (partial) veering structure $\mathcal{B}|F_w$ to a veering structure on $\mathcal{D}^\varphi$ is possible only if for every $f^+ \in F_w^+$ the regluing map $r(\varphi)$ maps the large edge of f^+ to the large edge of $r^\varphi(f^+)$. For this reason we define the group $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ of orientation-preserving combinatorial automorphisms of $\mathcal{Q}_{\mathcal{V},w}$ which preserve $\tau_{\mathcal{V},w}$.

Lemma 3.18 *Let $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$. If e is the large edge of $f^+ \in F_w^+$ then $r^\varphi_{f^+}(e)$ is the large edge of $r^\varphi(f^+) \in F_w^-$.*

Proof Suppose that $g^\varphi(f) = (g_1, \dots, g_k)$. Then $g_1 = \varphi(L(f))$, $g_{i+1} = \varphi(\mathfrak{a}(g_i))$ for $i = 1, \dots, k-1$, and $r^\varphi(f^+)$ is the triangle $f'^- \in F_w^-$ such that $U(f') = g_k$. Furthermore, $r(\varphi)$ maps e to

$$((\sigma^U_{f'})^{-1} \circ (\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}} \circ \sigma^L_f)(e).$$

We refer the reader to Sections 3.2 and 3.5 to recall the notation.

The large edge of $L(f)$ is given by $\sigma^L_f(e)$. Since $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$, we get that $(\varphi_{\check{f}} \circ \sigma^L_f)(e)$ is the large edge of g_1. By the definition of δ_i, $(\delta_1 \circ \varphi_{\check{f}} \circ \sigma^L_f)(e)$ is the large edge of $\mathfrak{a}(g_1)$. Again, the assumption that $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ implies that $(\varphi_1 \circ \delta_1 \circ \varphi_{\check{f}} \circ \sigma^L_f)(e)$ is the large edge of g_2. Continuing this way, $((\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}} \circ \sigma^L_f)(e)$ is the large edge of g_k, and thus $((\sigma^U_{f'})^{-1} \circ (\varphi_i \circ \delta_i)_{i=1}^{k-1} \circ \varphi_{\check{f}} \circ \sigma^L_f)(e)$ is the large edge of $r^\varphi(f^+) = f'^-$. □

The above lemma gives a sufficient condition for when the dual spine of $\mathcal{T}^\varphi$ admits a smoothing into a branched surface. Combining it with Proposition 3.16 gives sufficient conditions for the existence of a veering structure on $\mathcal{T}^\varphi$.

Theorem 3.19 *Let S_w be a surface properly carried by a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ of M. Suppose that $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ misaligns edge product disks. If additionally $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ then $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure.*

Proof By Proposition 3.16, the assumption that φ misaligns edge product disks implies the existence of a taut structure α^φ on $\mathcal{T}^\varphi$. If $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ then, by Lemma 3.18, for every $f^+ \in F_w^+$ the regluing map $r(\varphi)$ maps the large edge of f^+ to the large edge of $r^\varphi(f^+)$. Thus the branched surface $\mathcal{B}|F_w$ can be extended to a branched surface $\mathcal{B}^\varphi$ which combinatorially is just the dual spine $\mathcal{D}^\varphi$ of $(\mathcal{T}^\varphi, \alpha^\varphi)$. For every tetrahedron t of $\mathcal{T}^\varphi$ the branched surface $\mathcal{B}^\varphi_t = \mathcal{B}^\varphi \cap t$ looks as in Figure 4, because there is a tetrahedron t' of $\mathcal{V}$ with $\mathcal{B}_{t'} = \mathcal{B} \cap t' = \mathcal{B}^\varphi_t$ and $(\mathcal{T}, \alpha, \mathcal{B})$ is veering. Thus $\mathcal{B}^\varphi$ satisfies Definition 2.5 and $\mathcal{V}^\varphi = (\mathcal{T}^\varphi, \alpha^\varphi, \mathcal{B}^\varphi)$ is veering. □

We say that $\mathcal{V}^\varphi = (\mathcal{T}^\varphi, \alpha^\varphi, \mathcal{B}^\varphi)$ is obtained from $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ by a *veering mutation*, or that $\mathcal{V}^\varphi$ and $\mathcal{V}$ are *veering mutants*. For instance, the first two veering triangulations in the veering census, the veering triangulation `cPcbbbdxm_10` of the figure eight knot sister (manifold m003 in the SnapPy census [8]) and the veering triangulation `cPcbbbiht_12` of the figure eight knot (m004), are veering mutants.

Remark 3.20 By Lemma 3.18, a combinatorial automorphism $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ is uniquely determined by the associated bijection $\varphi\colon F_{\mathcal{V},w} \to F_{\mathcal{V},w}$. For this reason, when discussing examples of veering mutants in Section 4 we will not label the vertices of tetrahedra nor talk about bijections between vertices of identified triangles.

Theorem 3.19 gives a sufficient condition for veeringness of a taut mutant, but this condition is not necessary. It is possible that $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure even though $\varphi \notin \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$. We discuss this possibility briefly, and give an example of this phenomenon, in the next subsection.

3.8 Generalizations

We say that $(\mathcal{T}|F_w, \alpha|F_w)$ admits a veering structure if it is possible to smooth its dual spine $\mathcal{D}|F_w$ into a branched surface which locally around every vertex looks like one of the options in Figure 4. Suppose that $(\mathcal{T}|F_w, \alpha|F_w)$ admits a veering structure $\mathcal{B}^*|F_w$. Let t be a tetrahedron of $(\mathcal{T}|F_w, \alpha|F_w)$. Let $\mathcal{B}^*_t = \mathcal{B}^*|F_w \cap t$. By $-\mathcal{B}^*_t$ we denote the other possible veering structure on t; see Figure 4 to see the two options. Lemma 2.12 implies that if t has a top face f which is a bottom face of some tetrahedron of $(\mathcal{T}|F_w, \alpha|F_w)$ then we cannot change the veering structure on t from $\mathcal{B}^*_t$ to $-\mathcal{B}^*_t$ without destroying veeringness. On the other hand, if both top faces of t are in F_w^+ then we can freely change $\mathcal{B}^*_t$ to $-\mathcal{B}^*_t$ and the resulting branched surface still defines a veering structure on $(\mathcal{T}|F_w, \alpha|F_w)$.

Let $\tau^{*\,+}_{\mathcal{V},w}$ and $\tau^{*\,-}_{\mathcal{V},w}$ be the train tracks in $\mathcal{Q}^+_{\mathcal{V},w}$ and $\mathcal{Q}^-_{\mathcal{V},w}$, respectively, induced by $\mathcal{B}^*|F_w$. Using the same arguments as in the proof of Theorem 3.19 we can show that if $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ misaligns edge product disks and sends $\tau^{*\,+}_{\mathcal{V},w}$ to $\tau^{*\,-}_{\mathcal{V},w}$ then the veering structure $\mathcal{B}^*|F_w$ on $(\mathcal{T}|F_w, \alpha|F_w)$ glues up into a veering structure on $(\mathcal{T}^\varphi, \alpha^\varphi)$. The advantage of considering this more general setup is that now we can derive both sufficient and necessary conditions for veeringness of a taut mutant.

Theorem 3.21 *Let S_w be a surface properly carried by a veering triangulation $\mathcal{V} = (\mathcal{T}, \alpha, \mathcal{B})$ of M. Suppose that $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w})$ misaligns edge product disks. The taut triangulation $(\mathcal{T}^\varphi, \alpha^\varphi)$ admits a veering structure if and only if there is a veering structure $\mathcal{B}^*|F_w$ on $(\mathcal{T}|F_w, \alpha|F_w)$ such that the isomorphism $\varphi\colon \mathcal{Q}^+_{\mathcal{V},w} \to \mathcal{Q}^-_{\mathcal{V},w}$ sends $\tau^{*\,+}_{\mathcal{V},w}$ to $\tau^{*\,-}_{\mathcal{V},w}$.*

Proof The backward direction can be proved exactly as Theorem 3.19. If $(\mathcal{T}^\varphi, \alpha^\varphi)$ has a veering structure $\mathcal{B}^*$ then $(\mathcal{T}|F_w, \alpha|F_w)$ must have a veering structure $\mathcal{B}^*|F_w$ such that $r(\varphi)$ sends the train track on F_w^+ induced by $\mathcal{B}^*|F_w$ to the train track induced by $\mathcal{B}^*|F_w$ on F_w^-. Since φ misaligns edge product disks, for every $g \in \mathcal{Q}_{\mathcal{V},w}$ we have a trichotomy: $g \in L(F_w)$, g appears in $g^\varphi(f)$ for some $f \in F_w$, or $\mathfrak{a}(g)$ appears in $g^\varphi(f)$ for some $f \in F_w$. Equivalently, when constructing $r(\varphi)$ from φ we have passed through every triangular prism of M_w. Therefore Lemma 3.6 implies that φ sends $\tau^{*\,+}_{\mathcal{V},w}$ to $\tau^{*\,-}_{\mathcal{V},w}$. □

The more general setup of Theorem 3.21 is not just theoretical. There are veering triangulations $(\mathcal{T}, \alpha, \mathcal{B})$ and $(\mathcal{T}^{\varphi}, \alpha^{\varphi}, \mathcal{B}')$ which are taut mutants but not veering mutants. One such pair is given by the veering triangulation `gLMzQbcdefffhhhqxdu_122100` of the manifold s463 and the veering triangulation `gLMzQbcdefffhhhqxti_122100` of the manifold s639.

Another generalization we might consider is a *veering mutation with insertion*. Let $(\mathcal{T}|F_w, \alpha|F_w, \mathcal{B}|F_w)$ be a veering cut triangulation. If there are two triangles $f_1^+, f_2^+ \in F_w^+$ which are adjacent along an edge which is large in both f_1^+ and f_2^+, we might stack another veering tetrahedron on top of $f_1^+ \cup f_2^+$. We then obtain another cut triangulation with a veering structure which has more tetrahedra than $\mathcal{T}|F_w$. We can also add a new veering tetrahedron on top of two faces $f_1^+, f_2^+ \in F_w^+$ whose large edges are mixed.

Suppose that $(\mathcal{T}^*|F_{w*}, \alpha^*|F_{w*}, \mathcal{B}^*|F_{w*})$ is obtained from $(\mathcal{T}|F_w, \alpha|F_w, \mathcal{B}|F_w)$ by adding finitely many veering tetrahedra on top of F_w^+. If there is a map $r\colon F_{w*}^+ \to F_{w*}^-$ such that identifying F_{w*}^+ with F_{w*}^- yields a veering triangulation $\mathcal{V}^r = (\mathcal{T}^r, \alpha^r, \mathcal{B}^r)$ then we say that $\mathcal{V}^r$ is obtained from $\mathcal{V}$ by a *veering mutation with insertion*. For instance, the veering triangulation `dLQbccchhfo_122` of the manifold m009 is obtained from the veering triangulation `cPcbbbiht_12` of the manifold m004 (the figure eight knot complement) by a veering mutation with insertion.

4 Homeomorphic veering mutants

A manifold M and its mutant M^{φ} can be homeomorphic. This can happen for instance for many graph manifolds mutated along one of their decomposing tori. If veering mutants $\mathcal{V}$ and $\mathcal{V}^{\varphi}$ live on the same manifold, they might be combinatorially isomorphic or combinatorially distinct. For instance, the veering triangulation `eLMkbcddddedde_2100` of the 6_2^2 link complement carries a four-times punctured sphere such that mutating the triangulation along it via an involution yields `eLMkbcddddedde_2100` back. We discuss a few examples of combinatorially distinct veering mutants of the same manifold in Sections 4.1, 4.2 and 4.4.

Recall from Theorem 2.26 that veering triangulations combinatorially represent faces of the Thurston norm ball. A pair of veering mutants on a 3-manifold M may represent either the same face or different faces of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$. The main obstacle to finding examples of measurable veering mutants which represent the same face of the Thurston norm ball is that when $b_1(M) > 1$ there are infinitely many distinct bases for $H_2(M, \partial M; \mathbb{Z})$. While it is relatively easy to find the cones $\mathcal{C}(\mathcal{V})$ and $\mathcal{C}(\mathcal{V}^{\varphi})$ of homology classes of surfaces carried by $\mathcal{V}$ and $\mathcal{V}^{\varphi}$, respectively (this is explained in [44, Section 11.2]), it is not always straightforward to figure out whether they are the same up to a change of basis.

The above problem does not appear in the $b_1(M) = 1$ case. Then, up to $\eta \mapsto -\eta$, there is only one (0-dimensional) face of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$. If M admits a pair of veering mutants $\mathcal{V}$ and $\mathcal{V}^{\varphi}$ then neither $\mathcal{C}(\mathcal{V})$ nor $\mathcal{C}(\mathcal{V}^{\varphi})$ is empty. Hence $\mathcal{V}$ and $\mathcal{V}^{\varphi}$ must combinatorially represent the same face of the Thurston norm ball; see Remark 4.1. When $b_1(M) > 1$ it is sometimes possible to verify if $\mathcal{C}(\mathcal{V}) = \mathcal{C}(\mathcal{V}^{\varphi})$ using the combinatorics of the Thurston norm ball. We do this in Sections 4.2 (where the faces are the same) and 4.4 (where the faces are different).

Remark 4.1 Recall that veering triangulations come in pairs $\mathcal{V}$ and $-\mathcal{V}$ having the same taut signature and representing opposite faces of the Thurston norm ball; see Remarks 2.10 and 2.28. In what follows we will refer to veering triangulations using their taut signatures, without specifying coorientations on the faces. Thus when we say that two veering triangulations $\mathcal{V}$ and $\mathcal{V}'$ represent the same face of the Thurston norm ball, or write $\mathcal{C}(\mathcal{V}) = \mathcal{C}(\mathcal{V}')$, we really mean $\mathcal{C}(\mathcal{V}) \cup \mathcal{C}(-\mathcal{V}) = \mathcal{C}(\mathcal{V}') \cup \mathcal{C}(-\mathcal{V}')$.

In this section we will establish the following facts connecting veering mutations and faces of the Thurston norm ball:

Fact 4.2 (Veering mutations and faces of the Thurston norm ball) (1) *There are nonfibered faces of the Thurston norm ball that can be represented by two combinatorially nonisomorphic veering mutants.*

(2) *A veering mutation along a surface representing a class lying at the boundary of the cone on a fibered face may yield a veering triangulation representing a nonfibered face of the Thurston norm ball of the mutant manifold.*

Proof In Section 4.1 we discuss four veering mutants $\mathcal{V}$, $\mathcal{V}^{\varrho}$, $\mathcal{V}^{\sigma}$ and $\mathcal{V}^{\varrho\sigma}$ such that $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ represent the same nonfibered face of the Thurston norm ball in a certain manifold M with $b_1(M) = 1$, and $\mathcal{V}^{\varrho}$ and $\mathcal{V}^{\sigma}$ represent adjacent fibered faces of $M^{\varrho} \cong M^{\sigma}$ with $b_1(M^{\varrho}) = 2$. Triangulations $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ prove (1) in the $b_1(M) = 1$ case, while triangulations $\mathcal{V}^{\varrho}$ and $\mathcal{V}^{\sigma\varrho}$ prove (2); see also Proposition 4.4. Veering mutants proving (1) in the case $b_1(M) > 1$ are discussed in Section 4.2. □

4.1 Two veering mutants representing the same face of the Thurston norm ball when $b_1(M) = 1$

Let M be the manifold t12488 from the SnapPy census. This manifold is not fibered and $H_1(M;\mathbb{Z}) = \mathbb{Z} \oplus \mathbb{Z}/8$. It also admits a pair of distinct measurable veering triangulations which, since $b_1(M) = 1$, must represent the same face of the Thurston norm ball; see Remark 4.1. We will show that they differ by a veering mutation.

Let $\mathcal{V}$ be a veering triangulation of M with the taut signature

`iLLLPQccdgefhhghqrqqssvof_02221000.`

We present the tetrahedra of $\mathcal{V}$ in Figure 17.

By solving the system of branch equations associated to $\mathcal{V}$ one can verify that $\mathcal{V}$ carries four surfaces that can be expressed as the following (relative) 2-cycles (which we identify with the induced triangulations):

$$\mathcal{Q}_0 = f_2 + f_5 + f_7 + f_{11}, \quad \mathcal{Q}_1 = f_1 + f_5 + f_8 + f_{11},$$
$$\mathcal{Q}_2 = f_2 + f_7 + f_{10} + f_{12}, \quad \mathcal{Q}_3 = f_1 + f_8 + f_{10} + f_{12}.$$

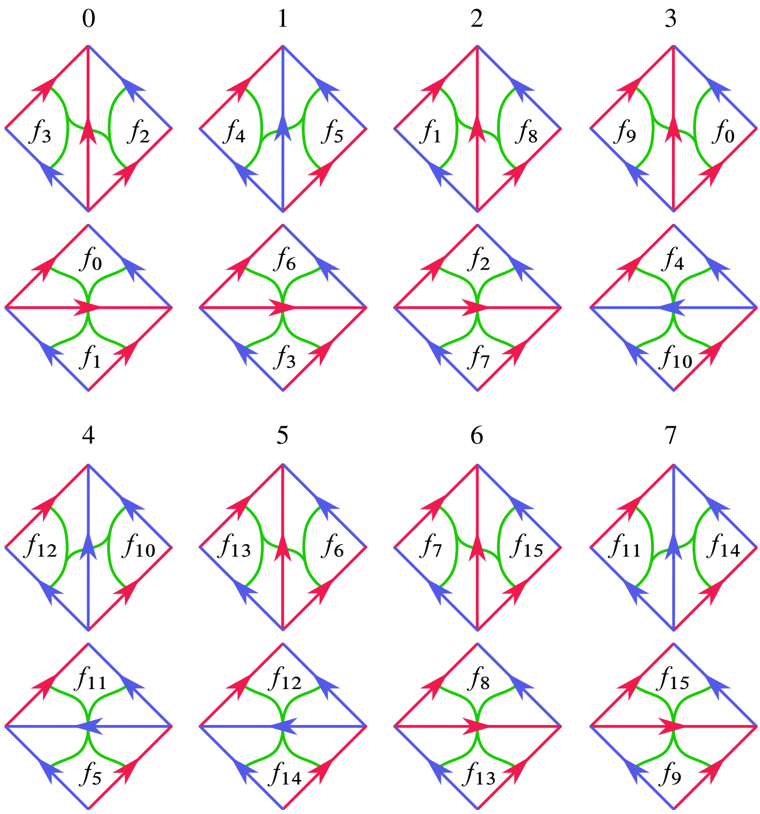


Figure 17: Veering triangulation `iLLLPQccdgefhhghqrqqssvof_02221000` of the manifold t12488.

All these surfaces are twice punctured tori. Triangulations $\mathcal{Q}_0$ and $\mathcal{Q}_3$ are presented in Figure 18. Since the first Betti number of M is equal to one, these punctured tori are homologous. In fact, it is easy to see that they are all homotopic. $\mathcal{Q}_0$ consists of two bottom faces of tetrahedron 2 and two bottom faces of tetrahedron 4. By performing the diagonal exchange corresponding to tetrahedron 2, one obtains triangulation $\mathcal{Q}_1$. By performing the diagonal exchange corresponding to tetrahedron 4, one obtains triangulation $\mathcal{Q}_2$. Triangulation $\mathcal{Q}_3$ can be obtained from $\mathcal{Q}_0$ by performing diagonal exchanges through both tetrahedra 2 and 4.

Let τ_i be the train track dual to $\mathcal{Q}_i$ induced by the stable train track of $\mathcal{V}$. That is, $\tau_i = \tau_{\mathcal{V},w_i}$ where w_i is the weight system on $\mathcal{V}$ determining $\mathcal{Q}_i$. As visible in Figure 18, the complementary regions of τ_i are punctured bigons, and thus, by Corollary 2.14, $\mathcal{Q}_i$ is properly carried. Let $\mathrm{Aut}^+(\mathcal{Q}_i \mid \tau_i)$ be the group of orientation-preserving combinatorial automorphisms of $\mathcal{Q}_i$ which preserve τ_i. Then

$$\mathrm{Aut}^+(\mathcal{Q}_0 \mid \tau_0) = \mathbb{Z}/2 \oplus \mathbb{Z}/2, \quad \mathrm{Aut}^+(\mathcal{Q}_1 \mid \tau_1) = \mathbb{Z}/2,$$
$$\mathrm{Aut}^+(\mathcal{Q}_2 \mid \tau_2) = \mathbb{Z}/2, \qquad \mathrm{Aut}^+(\mathcal{Q}_3 \mid \tau_3) = \mathbb{Z}/2.$$

The group $\mathrm{Aut}^+(\mathcal{Q}_0 \mid \tau_0)$ is generated by a rotation by π about the center of the red edge between faces f_7 and f_{11} and a "shift by one square" map; see Figure 18, left. We denote these combinatorial

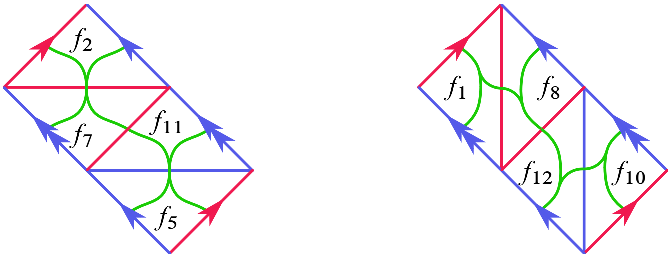


Figure 18: Left: the triangulation $\mathcal{Q}_0 = f_2 + f_5 + f_7 + f_{11}$ and its dual stable track τ_0. The group $\mathrm{Aut}^+(\mathcal{Q}_0 \mid \tau_0)$ is generated by a rotation ϱ by π about the center of the red edge between faces f_7 and f_{11} and a "shift by one square" map σ. Right: the triangulation $\mathcal{Q}_3 = f_1 + f_8 + f_{10} + f_{12}$ and its dual stable track τ_3. The group $\mathrm{Aut}^+(\mathcal{Q}_3 \mid \tau_3)$ is generated by $\varrho\sigma$ only.

isomorphisms of $\mathcal{Q}_0$ by ϱ and σ, respectively. For $i = 1, 2, 3$ the group $\mathrm{Aut}^+(\mathcal{Q}_i \mid \tau_i)$ is generated by $\varrho\sigma$; see Figure 18, right, for the $i = 3$ case. The fact that $\mathrm{Aut}^+(\mathcal{Q}_0 \mid \tau_0)$ is the largest is not surprising; the surface underlying $\mathcal{Q}_0$ is the lowermost carried representative of the generator of $H_2(M, \partial M; \mathbb{Z})$, and thus the stable train track τ_0 has the most large branches (is minimally split).

Since $\mathcal{Q}_0$ does not traverse any edge of $\mathcal{V}$ more than once, we automatically get that ϱ, σ and $\varrho\sigma$ do not align edge product disks. Thus by Theorem 3.19 we get three veering mutants $\mathcal{V}^{\varrho}$, $\mathcal{V}^{\sigma}$ and $\mathcal{V}^{\varrho\sigma}$ of $\mathcal{V}$. Information about the regluing maps $r(\varrho)$, $r(\sigma)$ and $r(\varrho\sigma)$ is presented in Table 1. Recall from Remark 3.20 that since we are looking only at elements of $\mathrm{Aut}^+(\mathcal{Q}_0 \mid \tau_0)$, the regluing maps are uniquely determined by their associated bijections $\{f_2^+, f_5^+, f_7^+, f_{11}^+\} \to \{f_2^-, f_5^-, f_7^-, f_{11}^-\}$. In Figure 19 we present taut signatures of the four mutants, as well some additional information about their underlying manifolds.

Proposition 4.3 *A nonfibered face* $\mathtt{F}$ *of the Thurston norm ball of the* **t12488** *manifold can be combinatorially represented by two combinatorially distinct veering mutants.*

Proof Triangulations $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ are two combinatorially distinct measurable veering mutants. Since there is a sequence of Pachner moves from $\mathcal{V}$ to $\mathcal{V}^{\varrho\sigma}$, their underlying manifolds are homeomorphic. (One can use Regina [7] to verify that the shortest such path has length four and consists of two 2-3 moves and two 3-2 moves.) Both these triangulations carry a twice punctured torus, which in particular means that the cones $\mathcal{C}(\mathcal{V})$ and $\mathcal{C}(\mathcal{V}^{\varrho\sigma})$ are nonempty. Since the first Betti number of M is equal to 1, up to $\eta \mapsto -\eta$ there is only one face $\mathtt{F}$ of the Thurston norm ball in $H^1(M; \mathbb{R})$. After possibly switching coorientations on faces of one of the triangulations we get that $\mathcal{C}(\mathcal{V}) = \mathcal{C}(\mathcal{V}^{\varrho\sigma}) = \mathbb{R}_+ \cdot (\mathtt{F})$; see Remark 4.1. □

f^+	f_2^+	f_5^+	f_7^+	f_{11}^+
$r^{\varrho}(f^+)$	f_5^-	f_2^-	f_{11}^-	f_7^-
$r^{\sigma}(f^+)$	f_{11}^-	f_7^-	f_5^-	f_2^-
$r^{\varrho\sigma}(f^+)$	f_7^-	f_{11}^-	f_2^-	f_5^-

Table 1: The regluing maps determined by ϱ, σ and $\varrho\sigma$.

iLLLPQccdgefhhghqrqqssvof 02221000 $t12488$ edge-orientable measurable $\mathbb{Z} \oplus \mathbb{Z}/8$	$\xrightarrow{\varrho}$	ivLLQQccfhfeghghwadiwadrv 20110220 $t12487$ not edge-orientable layered $\mathbb{Z} \oplus \mathbb{Z}$
$\sigma\downarrow$		$\sigma\downarrow$
ivLLQQccdhghgfhggrqlipigb 12020011 $t12487$ edge-orientable layered $\mathbb{Z} \oplus \mathbb{Z}$	$\xrightarrow{\varrho}$	iLLLPQccdgefhhghhrhajsvss 02221000 $t12488$ not edge-orientable measurable $\mathbb{Z} \oplus \mathbb{Z}/8$

Figure 19: Four veering mutants. Each dataset consists of the isomorphism signature of the triangulation (first row), the taut angle structure (second row), the name of the underlying manifold in the SnapPy's census (third row), information about edge-orientability (fourth row), the type of the triangulation (fifth row) and the first homology group with integer coefficients of the underlying manifold (sixth row).

Another conclusion that we can draw from Figure 19 is that a mutant of a measurable veering triangulation does not have to be measurable. Observe however that if a layered veering triangulation $\mathcal{V}$ admits a measurable veering mutant $\mathcal{V}^\varphi$ then the homology class of the mutating surface must lie in the boundary of the fibered cone represented by $\mathcal{V}$.

Proposition 4.4 *Let $\mathcal{V}$ be a finite layered veering triangulation of a 3-manifold M. Suppose that $\mathcal{V} \to \mathcal{V}^\varphi$ is a veering mutation such that $\mathcal{V}^\varphi$ is measurable. Then the homology class of the mutating surface lies in the boundary of the cone on the fibered face represented by $\mathcal{V}$.*

Proof Since $\mathcal{V}$ is layered, the face $\mathtt{F}$ of the Thurston norm ball represented by $\mathcal{V}$ is fibered; see [32, Theorem 5.15], stated here as Theorem 2.26. Denote by S_w the surface carried by $\mathcal{V}$ that can be used to mutate $\mathcal{V}$ into $\mathcal{V}^\varphi$, and by S_w^ϵ the embedded surface obtained from S_w by slightly pulling apart overlapping regions of S_w. The surface S_w^ϵ is a Thurston norm minimizing representative of its homology class [27, Theorem 3]. If that homology class lies in the interior of $\mathbb{R}_+ \cdot (\mathtt{F})$ then $M | S_w^\epsilon$ is a product sutured manifold [50, Theorem 3]. Therefore the mutant manifold M^φ is fibered over the circle with the mutating surface being the fiber. The assumption that $\mathcal{V} \to \mathcal{V}^\varphi$ is a veering mutation implies that φ misaligns edge product disks, and therefore, by Theorems 3.10 and 3.19, $\mathcal{V}^\varphi$ is a veering triangulation of M^φ. Therefore $\mathcal{V}^\varphi$ carries a fiber of a fibration of M^φ over the circle. But a veering triangulation that carries fibers of fibrations over the circle is layered [32, Theorem 5.15]. This is a contradiction with the assumption that $\mathcal{V}^\varphi$ is measurable. □

The other two mutants, $\mathcal{V}^\varrho$ and $\mathcal{V}^\sigma$, both live on the same 3-manifold t12487, which is the L11n222 link complement. Since we cannot have two combinatorially distinct veering triangulations representing the same fibered face of the Thurston norm ball [38, Proposition 2.7], we deduce that $\mathcal{V}^{\varrho}$ and $\mathcal{V}^{\sigma}$ represent different faces of the Thurston norm ball. In particular, it is possible that two different fibered faces of the Thurston norm ball of the same manifold are related by a veering mutation.

Let $\mathtt{F}^{\varrho}$ and $\mathtt{F}^{\sigma}$ be the fibered faces represented by $\mathcal{V}^{\varrho}$ and $\mathcal{V}^{\sigma}$, respectively. The Thurston norm ball of the L11n222 link complement is a quadrilateral with two pairs of fibered faces. Therefore the mutating twice punctured torus S represents the primitive integral class lying either on the ray $\mathbb{R}_+ \cdot \mathtt{F}^{\varrho} \cap \mathbb{R}_+ \cdot \mathtt{F}^{\sigma}$ or on the ray $\mathbb{R}_+ \cdot \mathtt{F}^{\varrho} \cap (-\mathbb{R}_+ \cdot \mathtt{F}^{\sigma})$. Since the stable train tracks τ^{ϱ} and τ^{σ} on S induced from $\mathcal{V}^{\varrho}$ and $\mathcal{V}^{\sigma}$, respectively, are equal, under the same mutation (eg by ϱ) two different veering triangulations on t12487 mutate into two different veering triangulations on t12488. It is the fact that the first Betti number of t12488 is equal to one that makes these two distinct veering triangulations represent the same top-dimensional face of the Thurston norm ball. In other words, in this example the phenomenon of a top-dimensional nonfibered face of the Thurston norm ball represented by multiple distinct veering triangulations arises from mutating a fibered 3-manifold with a higher first Betti number along a surface representing a class lying at the intersection of multiple fibered faces.

Remark 4.5 In the veering census there are 110 manifolds with the first Betti number equal to one which admit two measurable veering triangulations. Among those, (at least) 87 differ by a veering mutation along a connected surface. The mutating surface is either a four times punctures sphere (8 cases), a twice punctured torus (75 cases) or a four times punctured torus (4 cases).

4.2 Two veering mutants representing the same face of the Thurston norm ball when $b_1(M) = 2$

As explained at the beginning of this section, finding examples of two different measurable veering triangulations representing the same face of the Thurston norm ball is harder when $b_1(M) > 1$ because then $H_2(M, \partial M; \mathbb{R})$ admits infinitely many distinct bases. A possible approach to overcome this problem is to focus on manifolds for which any two nonfibered nonopposite faces of the Thurston norm ball have different combinatorics, or which have only one pair of opposite nonfibered faces. For instance, we searched for a cusped hyperbolic 3-manifold M such that

- $b_1(M) = 2$,
- the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ is a quadrilateral,
- M is fibered,
- M admits at least two measurable veering triangulations $\mathcal{V}$ and $\mathcal{V}'$ with different taut signatures and such that the cones $\mathcal{C}(\mathcal{V})$ and $\mathcal{C}(\mathcal{V}')$ are 2-dimensional.

When the first three conditions are satisfied, the Thurston norm ball of M admits only one pair of top-dimensional nonfibered faces. Therefore if additionally M admits at least two measurable veering triangulations whose cones of homology classes of carried surfaces are 2-dimensional, they either represent the same nonfibered face or opposite nonfibered faces; see Remark 4.1. If they represent opposite faces

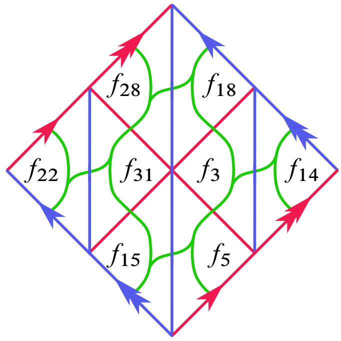


Figure 20: Ideal triangulation $\mathcal{Q}_{\mathcal{V}_1,w}$ and the stable train track $\tau_{\mathcal{V}_1,w}$ of a four times punctured torus carried by $\mathcal{V}_1$.

then switching coorientations on faces of one of the triangulations makes them represent the same face. In the veering census [20] there is a 3-manifold M which satisfies all these conditions. It admits (at least) three veering triangulations $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$ with the taut signatures

`qLLLzvQMQLMkbeeekljjlmljonppphhhhaaahhahhaahha_0111022221111001`,

`qLLLzvQMQLMkbeeekljjlmljonppphhhhaaahhahhaahha_1200111112020112`,

`qLLLzvQMQLMkbeeekljjlmljonppphhhhaaahhahhaahha_2111200001111221`,

respectively. Observe that $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$ are combinatorially isomorphic as triangulations, but they have different taut structures. Triangulations $\mathcal{V}_1$ and $\mathcal{V}_3$ are measurable, and $\mathcal{V}_2$ is layered. Using `tnorm` [54] we can verify that M indeed has only two pairs of faces of the Thurston norm ball. One pair has to be fibered because M admits a layered veering triangulation. Thus, after possibly replacing $\mathcal{V}_1$ by $-\mathcal{V}_1$, we must have that $\mathcal{C}(\mathcal{V}_1) = \mathcal{C}(\mathcal{V}_3)$.

Triangulations $\mathcal{V}_1$ and $\mathcal{V}_3$ not only represent the same nonfibered face of the Thurston norm ball, but they are also each other's mutants. Triangulation $\mathcal{V}_1$ carries a four times punctured torus which, using the same labels as in the veering census, can be represented by the 2-cycle

$$S_w = f_3 + f_5 + f_{14} + f_{15} + f_{18} + f_{22} + f_{28} + f_{31}.$$

To save some space, we do not include the picture of the tetrahedra of $\mathcal{V}_1$ (this is a triangulation with 16 tetrahedra). We do, however, present the induced triangulation $\mathcal{Q}_{\mathcal{V}_1,w}$ and the induced train track $\tau_{\mathcal{V}_1,w}$ in Figure 20. Since all complementary regions of $\tau_{\mathcal{V}_1,w}$ are punctured bigons, S_w is properly carried by $\mathcal{V}_1$; see Corollary 2.14. We have

$$\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V}_1,w} \mid \tau_{\mathcal{V}_1,w}) = \mathbb{Z}/2 \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

The group $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V}_1,w} \mid \tau_{\mathcal{V}_1,w})$ is generated by the rotation by π around the center of Figure 20, which we denote by ϱ, the shift by one "layer" in the northeast direction σ_+, and the shift by one "layer" in the northwest direction σ_-. The surface S_w does not traverse any edge of $\mathcal{V}_1$ more than once, and thus there are no edge product disks in $M|S_w^\epsilon$. Consequently, we can construct eight veering mutants of $\mathcal{V}_1$. They

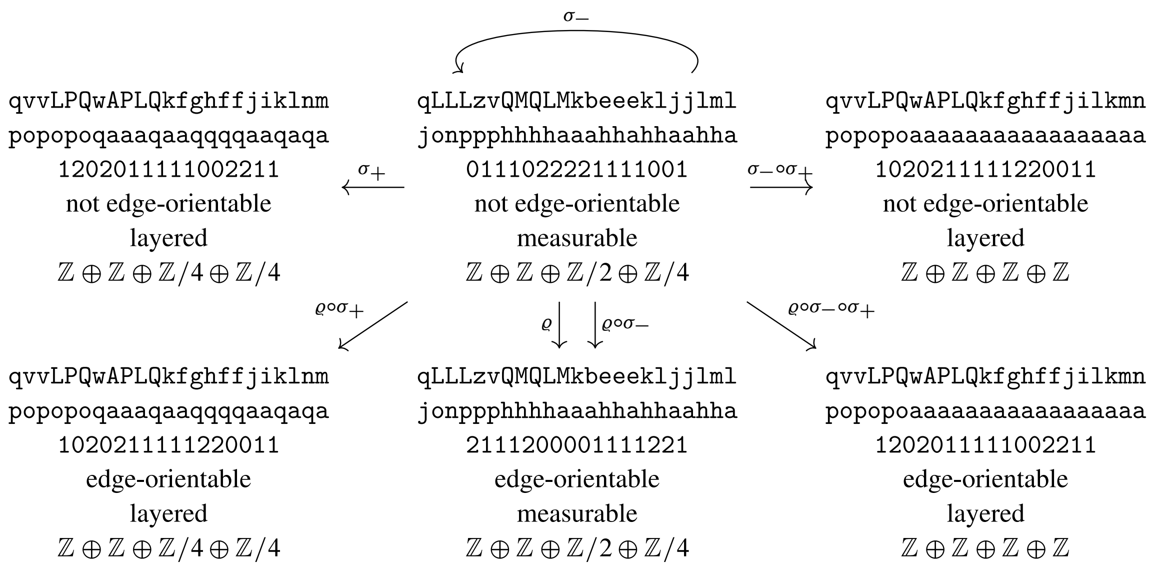


Figure 21: Six veering mutants of $\mathcal{V}_1$. Each dataset consists of the isomorphism signature of the triangulation (split into the first and second row), the taut angle structure (third row), information about edge-orientability (fourth row), the type of the triangulation (fifth row) and the first homology group with integer coefficients of the underlying manifold (sixth row).

do not have pairwise-distinct taut signatures. In particular, $\mathcal{V}_3$ has the same taut signature as both $\mathcal{V}_1^{\varrho}$ and $\mathcal{V}_1^{\varrho\sigma_-}$. Data on the remaining mutants of $\mathcal{V}_1$ is available in Figure 21. Observe that in each column we have veering triangulations with the same isomorphism signature but different taut structure. Thus in each column we have two veering triangulations of the same manifold. The manifold in the right column is the complement of the L14n62847 link.

It is worth mentioning that in this example the mutating surface represents a homology class that lies in the interior of the cone $\mathcal{C}(\mathcal{V}_1) = \mathcal{C}(\mathcal{V}_3)$, and thus in the interior of the cone over a face of the Thurston norm ball, but over a vertex of the Alexander norm ball. See [35] for the relationship between the Thurston and Alexander norms on $H_2(M, \partial M; \mathbb{R})$. The Alexander polynomial of M is equal to

$$\Delta_M = a^2b^2 + 2a^2b + 4ab + 4a + 2b + 6 + 2b^{-1} + 4a^{-1} + 4a^{-1}b^{-1} + 2a^{-2}b^{-1} + a^{-2}b^{-2}.$$

Using this we present the Alexander norm ball of M in Figure 22. We also marked the cones $\mathcal{C}(\mathcal{V}_1)$, $\mathcal{C}(\mathcal{V}_2)$ and $\mathcal{C}(\mathcal{V}_3)$ of homology classes carried by $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$, respectively. We can see that $\mathcal{C}(\mathcal{V}_1) = \mathcal{C}(\mathcal{V}_3)$, and that this cone is a cone on two adjacent faces of the Alexander norm ball, but one face of the Thurston norm ball. The homology class of the mutating surface lies over a vertex of the Alexander norm ball.

4.3 Nonmutants representing the same nonfibered face

Even though the focus of this paper is on veering mutants, it is important to note that there are faces of the Thurston norm ball which are combinatorially represented by two combinatorially nonisomorphic veering triangulations which do not differ by a mutation.

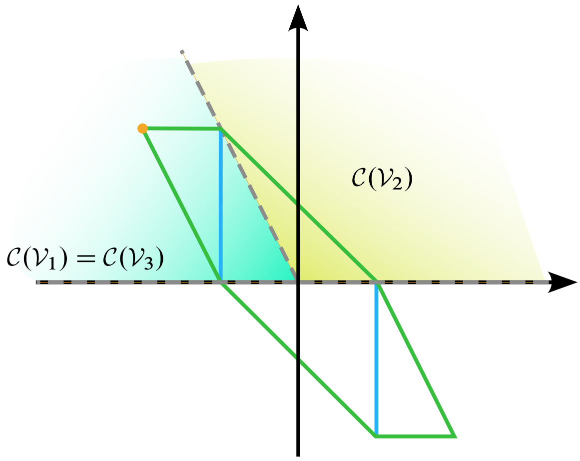


Figure 22: The unit norm ball of the Alexander norm on $H_2(M, \partial M; \mathbb{R})$ has vertices at $\left(\pm\frac{1}{2}, \mp\frac{1}{2}\right)$, $\left(\pm\frac{1}{4}, \mp\frac{1}{2}\right)$ and $\left(\pm\frac{1}{4}, 0\right)$. Its boundary is marked green. The unit norm ball of the Thurston norm has vertices at $\left(\pm\frac{1}{4}, \mp\frac{1}{2}\right)$ and $\left(\pm\frac{1}{4}, 0\right)$. The part of its boundary which does not overlap with the boundary of the Alexander norm ball is marked blue. The cones of homology classes carried by $\mathcal{V}_1$ and $\mathcal{V}_3$ are equal to the cone on two adjacent faces of the Alexander norm ball. The surface mutating $\mathcal{V}_1$ to $\mathcal{V}_3$ represents the class $(-1, 1)$ lying over the vertex $\left(-\frac{1}{2}, \frac{1}{2}\right)$ of the Alexander norm ball (marked orange).

Fact 4.6 *There are nonfibered faces of the Thurston norm ball that can be combinatorially represented by two distinct veering triangulations which do not differ by a mutation or a mutation with insertion.*

Proof Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be veering triangulations with the taut signatures

`lLLvLMQQccdjgkihhijkkqrwsdcfkfjdq_02221000012`,

`pvLLALLAPQQcdhehlkjmonmoonnwrawwaewaamgwwvn_122221111122002`,

respectively. These are two measurable veering triangulations on a 3-manifold M with $H_1(M; \mathbb{Z}) = \mathbb{Z} \oplus \mathbb{Z}/4$. Thus they must represent the same face of the Thurston norm ball. Since $\mathcal{V}_1$ and $\mathcal{V}_2$ have different numbers of tetrahedra (the strings describing their taut angle structures have different length), they cannot be mutants.

To show that $\mathcal{V}_2$ is not obtained from $\mathcal{V}_1$ by a single mutation with insertion, consider the set $F_{\max}$ of faces of $\mathcal{V}_1$ which have a nonzero weight for some weight system on $\mathcal{V}_1$. If $\mathcal{V}_2$ was obtained from $\mathcal{V}_1$ by a single mutation with insertion then the dual graph Γ_2 of $\mathcal{V}_2$ would have a subgraph isomorphic to the graph $\Gamma_1^{\max} = \Gamma_1 - F_{\max}$ obtained from the dual graph Γ_1 of $\mathcal{V}_1$ by deleting all its edges dual to the faces from $F_{\max}$. The graph $\Gamma_1^{\max}$ has three simple cycles: one of length three, and two of length four. The unique simple cycle of length three shares an edge with one of the cycles of length four. The graph Γ_2 has four simple cycles of length three and six simple cycles of length four, but none of the cycles of length three shares an edge with any cycle of length four. Thus $\mathcal{V}_2$ cannot be obtained from $\mathcal{V}_1$ by a single mutation with insertion. □

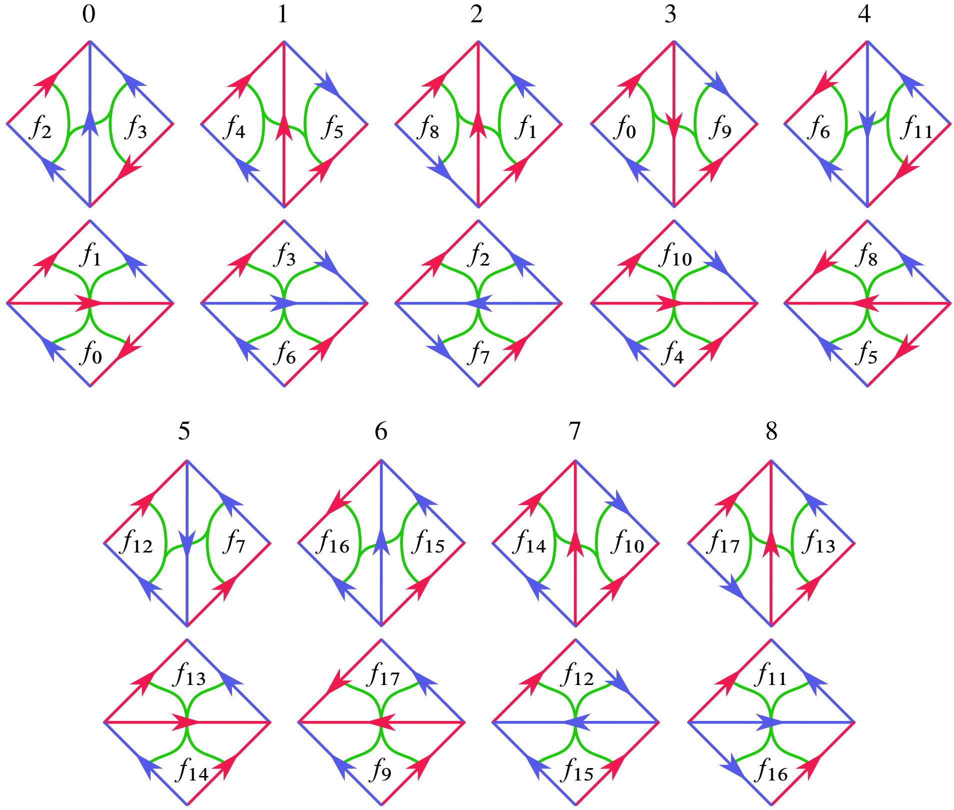


Figure 23: Veering triangulation `jLLAvQQcedehihiihiinasmkutn_011220000`.

4.4 Mutating along a higher-genus surface

In Sections 4.1 and 4.2 we discussed pairs of homeomorphic veering mutants for which the mutating surface was of genus one. In Remark 4.5 we also mentioned homeomorphic mutants with mutating surface of genus zero. In this subsection we discuss a pair of veering triangulations of the same manifold which differ by a mutation along a surface of genus two. This example differs from the previous ones not only by the genus of the mutating surface, but also by the fact that this surface is a fiber of a fibration over the circle. Moreover, the sutured manifold $M|S_w^\epsilon$ has edge product disks.

Let $\mathcal{V}$ and $\mathcal{V}'$ be veering triangulations with taut signatures

`jLLAvQQcedehihiihiinasmkutn_011220000`, `jvLLAQQdfghhfgiiijttmtltrcr_201102102`,

respectively. These are two veering triangulations of the 10^3_{12} link complement. We present tetrahedra of $\mathcal{V}$ in Figure 23.

Let $S_w = 2f_0 + f_2 + f_6 + 2f_7 + 2f_9 + 2f_{11} + f_{12} + f_{16}$. This is a genus-two surface with four punctures; see Figure 24. The stable train track $\tau_{\mathcal{V},w}$ has two complementary regions with five cusps and

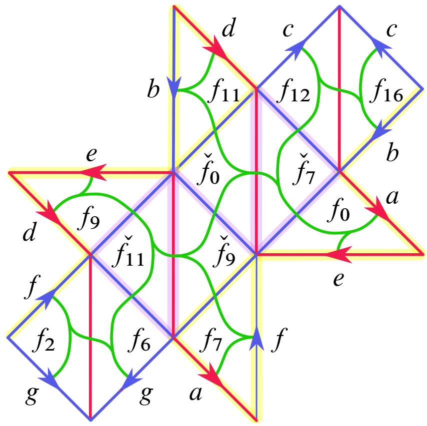


Figure 24: Ideal triangulation $\mathcal{Q}_{\mathcal{V},w}$ and the stable train track $\tau_{\mathcal{V},w}$ of a four times punctured genus-two surface S_w properly carried by $\mathcal{V}$. An edge e of $\mathcal{Q}_{\mathcal{V},w}$ is shaded yellow (respectively, purple) if the edge e^+ of $\mathcal{Q}^+_{\mathcal{V},w}$ (respectively, edge e^- of $\mathcal{Q}^-_{\mathcal{V},w}$) is the top (respectively, the bottom) base of an edge product disk in $M|S^\epsilon_w$. To distinguish between the two copies of $f_i \in F_w$ in $\mathcal{Q}_{\mathcal{V},w}$ when $w_{f_i} > 1$, we denote the lowermost copy of f_i by $\check{f}_i$. The letters a, b, c, d, e, f and g indicate side identifications. The only nontrivial element of $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ is the rotation by π around the center of the edge between $\check{f}_9$ and $\check{f}_0$. It misaligns edge product disks, because no edge shaded yellow is mapped to an edge shaded purple.

two complementary regions with one cusp. Thus, by Corollary 2.14, S_w is properly carried. Let ϱ be the generator of $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w}) = \mathbb{Z}/2$. Information on the bijection $r^\varrho\colon F^+_w \to F^-_w$ determined by ϱ is included in Table 2. In this example $M|S^\epsilon_w$ admits edge product disks — in Figure 24 their top bases are shaded yellow and their bottom bases are shaded purple. We can directly check that no edge which is the top base of some edge product disk in $M|S^\epsilon_w$ is mapped by ϱ to an edge which is the bottom base of some edge product disk in $M|S^\epsilon_w$. This means that ϱ misaligns edge product disks. Therefore, by Theorems 3.10 and 3.19, $\mathcal{V}^\varrho$ is a veering triangulation of M^ϱ. Using Regina [7] we can verify that $\mathcal{V}^\varrho$ is combinatorially isomorphic to $\mathcal{V}'$, and thus M^ϱ is homeomorphic to M.

With some choice of basis for $H_2(M, \partial M; \mathbb{Z})$ the cone $\mathcal{C}(\mathcal{V})$ is spanned by $(0,0,1)$, $(0,1,-1)$ and $(1,0,0)$, and the homology class of S_w is then given by $(1,2,-1)$. In particular, S^ϵ_w is a fiber of a fibration of M over the circle. The taut polynomial of $\mathcal{V}$ and its specialization at $[S^\epsilon_w]$ are equal to

$$\Theta(a,b,c) = a^2b^3c^2 - a^2b^2c^2 - ab^3c^2 - a^2b^2c + ab^2c + abc - bc - a - b + 1,$$
$$\Theta^{(1,2,-1)}(z) = \Theta(z^1, z^2, z^{-1}) = z^6 - 2z^5 - 2z + 1,$$

f^+	f_0^+	f_2^+	f_6^+	f_7^+	f_9^+	f_{11}^+	f_{12}^+	f_{16}^+
$r^\sigma(f^+)$	f_0^-	f_{16}^-	f_{12}^-	f_7^-	f_9^-	f_{11}^-	f_6^-	f_2^-

Table 2: The regluing map determined by σ.

respectively. It follows from [32, Theorem 7.1; 34, Theorem 4.2] that the stretch factor λ of the monodromy f of the fibration with fiber S_w^{ϵ} is equal to the largest real root of $\Theta^{(1,2,-1)}(z)$, that is

$$\lambda = \tfrac{1}{4}(1+\sqrt{17}+\sqrt{2(1+\sqrt{17})}) \approx 2.081.$$

The mutating surface in $\mathcal{V}^{\varrho}$ is a fiber of a fibration of M over the circle with monodromy ϱf and thus the same stretch factor. The fibered faces $\mathtt{F}$ and $\mathtt{F}^{\varrho}$ represented by $\mathcal{V}$ and $\mathcal{V}^{\varrho}$, respectively, must be different because there is at most one veering triangulation associated to a fibered face $\mathtt{F}$ of the Thurston norm ball (zero if the associated circular flow has singular orbits); see [38, Proposition 2.7]. In this case we can actually deduce a stronger statement, that there is no automorphism Φ of $H_2(M, \partial M; \mathbb{R})$ that sends $\mathtt{F}$ to $\mathtt{F}^{\varrho}$. This follows from the fact that $\mathtt{F}$ is a triangle, while $\mathtt{F}^{\varrho}$ is a pentagon. Thus ϱf and f are not conjugate in the mapping class group of a genus-two surface with four punctures.

Fact 4.7 *Let M be the complement of the 10^3_{12} link.*

- *M fibers in two different ways with fiber being a genus-two surface with four punctures and such that the monodromy of one fibration is obtained from the monodromy of the other fibration by postcomposing it with an involution ϱ.*
- *The two fibrations lie over different faces of the Thurston norm ball, and no automorphism of $H_2(M, \partial M; \mathbb{R})$ sends one face to the other. Thus the monodromies are not conjugate in the mapping class group of a genus-two surface with four punctures.*

5 Flows representing the same face of the Thurston norm ball

We will show that the flows built from some of the combinatorially nonisomorphic veering triangulations discussed in Section 4 are topologically inequivalent using the following lemma:

Lemma 5.1 *Let Ψ_1 and Ψ_2 be two pseudo-Anosov flows on a closed 3-manifold N. If the stable lamination of Ψ_1 is transversely orientable and the stable lamination of Ψ_2 is not, then Ψ_1 and Ψ_2 are not topologically equivalent. An analogous statement holds for blown-up flows Ψ_1° and Ψ_2°.*

Proof If Ψ_1 and Ψ_2 are topologically equivalent there is a homeomorphism $h\colon M \to M$ taking oriented orbits of Ψ_1 to oriented orbits of Ψ_2 (see Definition 2.18). This homeomorphism must take leaves of the stable lamination of Ψ_1 to the leaves of the stable lamination of Ψ_2. But the stable lamination of Ψ_2 admits Möbius band leaves, while the stable lamination of Ψ_1 has only planar and annular leaves. Therefore Ψ_1 and Ψ_2 cannot be topologically equivalent. □

Thus if a face $\mathtt{F}$ of the Thurston norm ball is combinatorially represented by two veering triangulations one of which is edge-orientable and the other is not, Lemma 5.1 together with Corollary 2.21 imply that the flows built out of these veering triangulations have to be topologically inequivalent.

Theorem 5.2 *There are nonfibered faces of the Thurston norm ball that can be dynamically represented by two topologically inequivalent flows.*

Proof For the cusped case apply Lemma 5.1 and Corollary 2.21 to the blown-up flows built from the veering mutants $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ discussed in Section 4.1 (for the $b_1 = 1$ case) or from the veering mutants $\mathcal{V}_1, \mathcal{V}_3 = \mathcal{V}_1^{\varrho}$ discussed in Section 4.2 (for the $b_1 = 2$ case) using Theorem 2.20. The maximality condition from Definition 2.24 is satisfied in both cases because the faces are top dimensional.

Both these pairs of veering mutants differ by a mutation along a punctured torus. Thus Dehn filling their underlying 3-manifolds along the slopes determined by the boundaries of these tori yields toroidal 3-manifolds. Consequently, these veering triangulations cannot be used to construct two distinct pseudo-Anosov flows on a closed hyperbolic 3-manifold which represent the same face of the Thurston norm ball. However, to do so we can use veering triangulations $\mathcal{V}_1$ and $\mathcal{V}_2$ that we discussed in the proof of Fact 4.6:

(a) $\mathcal{V}_1$ and $\mathcal{V}_2$ are two measurable veering triangulations of the same manifold M with $H_1(M;\mathbb{Z}) = \mathbb{Z} \oplus \mathbb{Z}/4$.

(b) $\mathcal{V}_1$ is edge-orientable, while $\mathcal{V}_2$ is not.

(c) For $i = 1, 2$ every connected surface S carried by $\mathcal{V}_i$ has genus three and four punctures, and all complementary regions of its stable train track have four cusps.

It follows from (a) that $\mathcal{V}_1$ and $\mathcal{V}_2$ represent the same face of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$. Part (c) and Remark 2.27 imply that both $\mathcal{V}_1$ and $\mathcal{V}_2$ carry a surface S of genus three with four punctures such that each boundary component of S intersects the ladderpole curves of $\mathcal{V}_i$ four times. Thus, by Theorem 2.30, the 3-manifold N obtained from M by Dehn filling it along the slope determined by ∂S is hyperbolic.

For $i = 1, 2$ let Ψ_i be the pseudo-Anosov flow on N built from $\mathcal{V}_i$ via the Agol–Tsang construction. Since triangulations $\mathcal{V}_1$ and $\mathcal{V}_2$ represent the same face of the Thurston norm ball in $H_2(M, \partial M; \mathbb{R})$ and both carry S, Theorem 2.30 implies that $\mathcal{C}(\Psi_1) = \mathcal{C}(\Psi_2) \neq \varnothing$. By (a) and the fact that N contains an incompressible surface, $b_1(N) = 1$. Thus the maximality condition from Definition 2.24 must be satisfied. Therefore Ψ_1 and Ψ_2 dynamically represent the same face of the Thurston norm ball in $H_2(N; \mathbb{R})$. The fact that they are not topologically equivalent follows from (b), Lemma 5.1 and Corollary 2.21. □

Remark 5.3 In Section 4.1 we constructed a pair of veering mutants $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ on the manifold t12488. Let N denote the manifold obtained from t12488 by Dehn filling it along the boundary of the mutating surface. Using Regina [7] it is possible to verify that N is a graph manifold obtained from the orientable circle bundle N_0 over a 2-holed $\mathbb{R}P^2$ by identifying its two toroidal boundary components. Thus N is a so-called *BL-manifold*, as defined by Barbot in [2].

Langevin and Bonatti constructed an Anosov flow on one BL-manifold in [5]; a description of this flow written in English can be found in [25]. Barbot generalized the construction to most other BL-manifolds

[2, Theorem A] and called the resulting Anosov flows *BL-flows*. If a BL-manifold is not a circle bundle then the constructed flow is not $\mathbb{R}$-covered, because it is not circular but is transverse to a torus.

In [2, Theorem B(2)] Barbot claims that all non-$\mathbb{R}$-covered Anosov flows on a fixed BL-manifold which is not a circle bundle are topologically equivalent. This is in contradiction with our results. It follows from Theorem 2.20 and Lemma 5.1 that N admits a pair of topologically inequivalent BL-flows — one constructed from $\mathcal{V}$ and the other constructed from $\mathcal{V}^{\varrho\sigma}$. We denote them by Ψ and $\Psi^{\varrho\sigma}$, respectively. The flows Ψ and $\Psi^{\varrho\sigma}$ are constructed from the same semiflow Φ_0 on N_0, but — unsurprisingly, given that $\mathcal{V}$ and $\mathcal{V}^{\varrho\sigma}$ are mutants — by gluing the two boundary tori in different ways. More specifically, if for $i = 1, 2$ we choose a basis (o_i, f_i) on the boundary torus T_i of N_0 so that f_i is a fiber of a Seifert fibration while o_i corresponds to the boundary of the 2-holed $\mathbb{R}P^2$ contained in T_i, then one gluing $T_1 \to T_2$ can be represented by a matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},$$

and the other by $-A$. These two gluings result in the same manifold N because N_0 admits an involution which fixes (o_1, f_1) and sends (o_2, f_2) to $(-o_2, -f_2)$. This involution can be obtained as the composition of the reflection across the stable leaf through the periodic orbit of Φ_0 missing T_i and the reflection which fixes every fiber of the Seifert fibration, but reverses their orientation.

In Remark 4.5 we mentioned 79 pairs of veering mutants on manifolds with first Betti number equal to one for which the mutating surface is a punctured torus. In most cases Regina recognizes their appropriate Dehn fillings as BL-manifolds.

6 Polynomial invariants of veering triangulations

In [34] McMullen introduced a polynomial invariant of fibered faces of the Thurston norm ball called the *Teichmüller polynomial*. Recall that associated to a fibered face $\mathtt{F}$ there is a unique circular flow Ψ; see [13, Theorem 7], stated here as Theorem 2.23. The Teichmüller polynomial of $\mathtt{F}$ is an invariant of the module of transversals to the preimage of the stable lamination of Ψ in the maximal free abelian cover of the manifold [34, Section 3]. Its main feature is that it can be used to compute the stretch factors of monodromies of all fibrations lying over $\mathtt{F}$ [34, Theorem 4.2].

McMullen asked whether it is possible to define a similar invariant for nonfibered faces. If a nonfibered face is dynamically represented by a pseudo-Anosov flow Ψ, one could try to replicate the definition of the Teichmüller polynomial using the stable lamination of Ψ. Landry, Minsky and Taylor used veering triangulations to devise such a polynomial invariant [32]. In fact, they defined two polynomial invariants of veering triangulations: the *taut polynomial* and the *veering polynomial*. Furthermore, they showed that if a face $\mathtt{F}$ of the Thurston norm ball represented by a veering triangulation $\mathcal{V}$ is fibered, then the taut polynomial of $\mathcal{V}$ is equal to the Teichmüller polynomial of $\mathtt{F}$ [32, Theorem 7.1]. Therefore the taut

	iLLLPQccdgefhhghqrqqssvof_02221000
Θ	$4(a+1)$
$\mathbb{V}$	$(a-1)^3\Theta$
	iLLLPQccdgefhhghhrhajsvss_02221000
Θ	$4(a-1)$
$\mathbb{V}$	$(a-1)^2(a+1)\Theta$
	qLLLzvQMQLMkbeeekljjlmljonppphhhhaaahhahhaahha_0111022221111001
Θ	$a^2b^2-2a^2b+4ab-4a-2b+6-2b^{-1}-4a^{-1}+4a^{-1}b^{-1}-2a^{-2}b^{-1}+a^{-2}b^{-2}$
$\mathbb{V}$	0
	qLLLzvQMQLMkbeeekljjlmljonppphhhhaaahhahhaahha_2111200001111221
Θ	$a^2b^2+2a^2b+4ab+4a+2b+6+2b^{-1}+4a^{-1}+4a^{-1}b^{-1}+2a^{-2}b^{-1}+a^{-2}b^{-2}$
$\mathbb{V}$	0
	lLLvLMQQccdjgkihhijkkqrwsdcfkfjdq_02221000012
Θ	$(a+1)(a^2+1)(a^2-a+1)^2$
$\mathbb{V}$	$(a+1)(a-1)^2(a^2+1)(a^4+a^3+a^2+a+1)(a^6+a^5+a^4+a^3+a^2+a+1)\Theta$
	pvLLALLAPQQcdhehlkjmonmoonnwrawwaewaamgwwvn_122221111122002
Θ	$(a-1)(a^2+1)(a^2+a+1)^2$
$\mathbb{V}$	0

Table 3: The taut and veering polynomials of pairs of veering triangulations representing the same face of the Thurston norm ball discussed in Section 4. Θ denotes the taut polynomial and $\mathbb{V}$ denotes the veering polynomial. The variables correspond to the basis elements of $H = H_1(M;\mathbb{Z})/\text{torsion}$. The invariants are well defined up to a change of basis of H and multiplication by $\pm h$ for $h \in H$.

polynomial (and its specializations under Dehn fillings) can be seen as a generalization of the Teichmüller polynomial to (some) nonfibered faces.

However, in Section 4 we showed that a veering triangulation representing a nonfibered face of the Thurston norm ball is not necessarily unique. This means that the taut and veering polynomials of a veering triangulation might actually not be invariants of the face represented by the triangulation.

An algorithm to compute the taut and veering polynomials of a veering triangulation is explained in [46]. A much faster algorithm for the computation of the taut polynomial follows from the fact that it is equal to the Alexander polynomial of the underlying manifold twisted by a certain representation $\omega\colon \pi_1(M) \to \mathbb{Z}/2$ [45, Proposition 5.7] and can therefore be computed using *Fox calculus*. Both algorithms have been implemented by the author, Saul Schleimer and Henry Segerman; see Veering on GitHub [47]. Using this software, we computed the taut and veering polynomials of the pairs of veering triangulations representing the same face of the Thurston norm ball discussed in Section 4. We include this data in Table 3.

Fact 6.1 *A nonfibered face of the Thurston norm ball can be combinatorially represented by two distinct veering triangulations with different taut polynomials, and different veering polynomials.*

Proof See Table 3. □

The only pair of veering triangulations from Table 3 which have different both taut and veering polynomials are not veering mutants; see Fact 4.6. Hence the question still remains whether two homeomorphic veering mutants representing the same face of the Thurston norm ball can have different both taut and veering polynomials. The answer to this question is positive. One such pair consists of veering triangulations is

`mvLLMvQQQegffhijkllkklreuegggvvrggr_120200111111`,

`mvLLMvQQQegffhjikllkklreuegrrvvrwwr_120200111111`.

They differ by a veering mutation along a four times punctured torus. Their taut polynomials are $8(a+1)$ and $8(a-1)$, respectively, and their veering polynomials are $8(a-1)(a+1)^3$ and $8(a-1)^3(a+1)$, respectively.

7 Further questions

7.1 Operations on flows underlying veering mutations

Throughout the paper we worked combinatorially with veering triangulations and used existing literature [1; 31; 32] to deduce statements about pseudo-Anosov flows on closed manifolds or their blow-ups on manifolds with toroidal boundary. We have intentionally avoided discussing how the flows underlying veering mutants are related. A naive expectation would be that the flows differ by a *mutation of* (*blown-up*) *pseudo-Anosov flows*. In the closed case this would be a mutation along a surface transverse to a pseudo-Anosov flow whose intersections with the stable and unstable foliations of the flow are invariant under some nontrivial symmetry. Mutating these foliations via this symmetry gives a pair of 2-dimensional singular foliations intersecting along "recombined flow lines". It remains to find sufficient conditions for a flow along recombined flow lines (a *mutant flow*) to admit a parametrization which makes it pseudo-Anosov. When $\partial M \neq \varnothing$ one could expect a similar operation performed on the stable and unstable laminations of a blown-up pseudo-Anosov flow.

However, examples presented in Section 4.1 suggest that the problem may be more complicated. Namely, let $\mathcal{V}$ be a veering triangulation with taut signature `iLLLPQccdgefhhghqrqqssvof_02221000`. We showed that $\mathcal{V}$ admits four veering mutants: $\mathcal{V}$, $\mathcal{V}^{\varrho}$, $\mathcal{V}^{\sigma}$ and $\mathcal{V}^{\varrho\sigma}$. Denote by M the manifold underlying $\mathcal{V}$. By [1, Theorem 5.1], there is a transitive Anosov flow Ψ on the manifold obtained from M by Dehn filling it along the boundary of the mutating surface, and the blown-up flow Ψ° on M. It can be deduced from Figure 18, right, that the intersection $\mathcal{L}_{\Psi^\circ,S}$ of the mutating twice punctured torus S with the stable lamination of Ψ° has two closed leaves and the remaining leaves spiral into them; we approximate this lamination in Figure 25. It is clear from Figure 25 that $\mathcal{L}_{\Psi^\circ,S}$ is invariant only under the identity and $\varrho\sigma$. This means that even though we can mutate the veering branched surface carrying the stable lamination of Ψ° in four different ways, the lamination itself can only be mutated in two different ways.

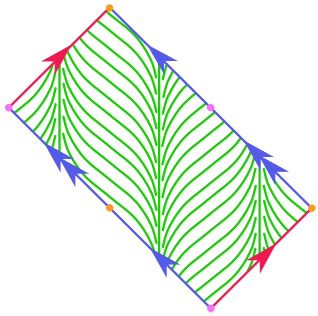


Figure 25: The intersection of the stable lamination of the blown-up Anosov flow determined by the veering triangulation `iLLLPQccdgefhhghqrqqssvof_02221000` and a twice punctured torus carried by this triangulation.

Working with veering triangulations as opposed to working directly with flows has both advantages and disadvantages. On one hand, it allowed us to find explicit examples of topologically inequivalent flows on the same manifold which differ by a veering mutation and represent the same face of the Thurston norm ball (Theorem 5.2). On the other hand, the fact that veering triangulations exist only on hyperbolic 3-manifolds means that if a mutation along a surface transverse to some (blown-up) pseudo-Anosov flow yields a (blown-up) pseudo-Anosov flow on a nonhyperbolic manifold, there will not be a corresponding mutation on the level of triangulations. This can happen when we mutate along $\varphi \in \mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w})$ which aligns edge product disks; see Proposition 3.17. Another obstruction for a veering mutation that would not be an obstruction for a mutation of flows is the "no perfect fits" condition. It is possible that a flow which is without perfect fits relative to a finite collection Λ of closed orbits mutates into a flow which does have perfect fits relative to the recombined collection of orbits Λ^φ. In this case again we do not have a corresponding veering mutation. For these reasons, it is still of interest to properly define and study mutations of pseudo-Anosov flows (and possibly other operations underlying veering mutations) without referring to veering triangulations. This would fit into a more general framework of constructing new flows out of old, similarly to the Goodman–Fried surgery [15; 21], and Handel–Thurston shearing along tori [22].

7.2 The orbit spaces of mutant flows and recognizing mutative flows

Associated to a pseudo-Anosov flow Ψ there is a bifoliated plane called the *orbit space* of Ψ; see [11, Proposition 4.1]. Suppose that flows Ψ and Ψ^φ differ by a mutation along a transverse surface S in the sense introduced in Section 7.1. If S is a fiber of a fibration over the circle there is a homeomorphism from the orbit space of Ψ to the orbit space of Ψ^φ which sends foliations of one to the foliations of the other; this follows from the fact these orbit spaces are the universal covers of S and S^φ equipped with the invariant foliations lifted from S and S^φ, respectively. It is not immediately clear how the orbit spaces differ when S is not a virtual fiber. More generally, it would be advantageous to have an invariant which is equal for flows which are *mutative*, that is, differ by a finite number of mutations, and distinguishes flows which are not mutative.

7.3 A general result on the relationship between two flows representing the same face of the Thurston norm ball

We showed that two blown-up Anosov flows representing the same face of the Thurston norm ball can differ by a veering mutation (Sections 4.1 and 4.2). However, we also noted that there are examples of veering triangulations that represent the same face of the Thurston norm ball and do not differ by a veering mutation or even a veering mutation with insertion (Fact 4.6). We have not explained how these veering triangulations, or their underlying flows, are related. Ideally, we would like to have a theorem that describes all possible ways in which two distinct flows can represent the same face of the Thurston norm ball.

7.4 Homology classes versus free homotopy classes of closed orbits of flows

In recent work Barthelmé, Frankel and Mann found an invariant which distinguishes distinct transitive pseudo-Anosov flows, provided that their orbit spaces satisfy a technical condition called *no tree of scalloped regions*; see [3, Definition 3.21]. More precisely, they showed that two such flows Ψ_1 and Ψ_2 on N are isotopically equivalent if and only if the sets $\mathcal{P}(\Psi_1)$ and $\mathcal{P}(\Psi_2)$ of unoriented free homotopy classes of their closed orbits are equal, and topologically equivalent if these sets differ by an automorphism of $\pi_1(N)$; see [3, Theorem 1.1].

In the proof of Theorem 5.2 we discussed veering triangulations which, after appropriate Dehn filling, yield topologically inequivalent transitive pseudo-Anosov flows Ψ_1 and Ψ_2 on a closed hyperbolic 3-manifold N representing the same face of the Thurston norm ball in $H_2(N, \mathbb{R})$. By [3, Proposition 1.2], the orbit spaces of Ψ_1 and Ψ_2 do not have trees of scalloped regions. Thus it follows from [3, Theorem 1.1] that $\mathcal{P}(\Psi_1) \neq \Phi\mathcal{P}(\Psi_2)$ for any $\Phi \in \mathrm{Aut}(\pi_1(N))$. On the other hand, we know that the homology classes of closed orbits of Ψ_1 and Ψ_2 span the same rational cone in $H_1(N;\mathbb{R})$. This motivates the question of how exactly the sets $\mathcal{P}(\Psi_1)$ and $\mathcal{P}(\Psi_2)$ differ, not just for these particular flows from the proof of Theorem 5.2, but for any two flows which represent the same face of the Thurston norm ball.

7.5 Many distinct flows representing the same face of the Thurston norm ball in the $b_1(M) = 1$ case

Suppose that S is a Thurston norm minimizing surface representing a primitive integral class lying at the intersection of two fibered cones $\mathbb{R}_+ \cdot \mathtt{F}_1$ and $\mathbb{R}_+ \cdot \mathtt{F}_2$. Let Ψ_1 and Ψ_2 be the circular flows associated to $\mathtt{F}_1$ and $\mathtt{F}_2$ as in Theorem 2.23. If the intersections of S with the stable and unstable foliations of Ψ_1 and Ψ_2 are isotopic, we may be able to perform a mutation along S which yields two distinct noncircular flows Ψ_1^{φ} and Ψ_2^{φ} representing the same top-dimensional nonfibered face in the mutant manifold. A combinatorial version of this phenomenon occurs for manifolds t12487 and t12488; see Section 4.1. This leads to a question: given $k > 2$ is there a 3-manifold M with $b_1(M) > 2$ such that

- M admits k fibered faces, intersecting at a point α, dynamically represented by topologically inequivalent circular flows $\Psi_1, \Psi_2, \dots, \Psi_k$,

taut signature	volume
gLLPQccdfeffhggaagb_201022	5.33348956689812
gLLPQccdfeffhwraarw_201022	5.33348956689812
gLLPQbefefefhhxhqhh_211120	5.07470803204827
gLLPQbefefefhhhhhha_011102	5.07470803204827

Table 4: Veering mutants of gLLPQccdfeffhggaagb_201022.

- the primitive integral class on $\mathbb{R}_+ \cdot \alpha$ can be represented by a Thurston norm minimizing surface S such that mutating M along S gives a nonfibered 3-manifold M^φ with $b_1(M^\varphi) = 1$ and $[S^\varphi] \in \mathcal{C}(\Psi_i^\varphi)$ for $i = 1, 2, \ldots, k$?

More generally, can a face of the Thurston norm ball be dynamically represented by more than two topologically inequivalent flows? Can it be represented by infinitely many flows?

7.6 Veering mutants and hyperbolic geometry

Recall that if $\mathcal{V}$ and $\mathcal{V}^\varphi$ are veering mutants then they are both hyperbolic [23, Theorem 1.5]. However, they do not always have the same hyperbolic volume. In Table 4 we list the taut signatures and volumes of certain veering mutants of the veering triangulation gLLPQccdfeffhggaagb_201022 of the 6_3^2 link complement. The mutating surface is a twice punctured torus with the induced triangulation $\mathcal{Q}_{\mathcal{V},w}$ and the stable train track $\tau_{\mathcal{V},w}$ satisfying $\mathrm{Aut}^+(\mathcal{Q}_{\mathcal{V},w} \mid \tau_{\mathcal{V},w}) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$.

Ruberman studied mutations of hyperbolic 3-manifolds and found sufficient conditions for a mutant of a hyperbolic 3-manifold to be a hyperbolic 3-manifold of the same volume. One of his results concerns only mutating via very special types of involutions of certain surfaces [48, Theorem 1.3]; another concerns only mutating along surfaces which are not virtual fibers [48, Theorem 4.4]. Conditions of neither of these theorems are satisfied when mutating the first triangulation from the table to either the third or the fourth one; the mutating involutions are not of the type required by [48, Theorem 1.3], and the mutating surface is a fiber of a fibration over the circle. Nonetheless, there are plenty of veering mutants with the same hyperbolic volume. In particular, all nonhomeomorphic veering mutants discussed in Section 4 have the same volume. It would be interesting to know if it is possible to figure out purely combinatorially when a veering mutation along a carried surface results in a hyperbolic 3-manifold of the same volume. The relationship between a veering structure and a hyperbolic structure is still not well understood, and perhaps analyzing it in this fairly narrow setup of mutations would give some new insight on the matter.

References

[1] **I Agol**, **C C Tsang**, *Dynamics of veering triangulations: infinitesimal components of their flow graphs and applications*, Algebr. Geom. Topol. 24 (2024) 3401–3453 MR Zbl

[2] **T Barbot**, *Generalizations of the Bonatti–Langevin example of Anosov flow and their classification up to topological equivalence*, Comm. Anal. Geom. 6 (1998) 749–798 MR Zbl

[3] **T Barthelmé**, **S Frankel**, **K Mann**, *Orbit equivalences of pseudo-Anosov flows*, Invent. Math. (online publication March 2025)

[4] **C Bonatti**, **I Iakovoglou**, *Anosov flows on 3-manifolds: the surgeries and the foliations*, Ergodic Theory Dynam. Systems 43 (2023) 1129–1188 MR Zbl

[5] **C Bonatti**, **R Langevin**, *Un exemple de flot d'Anosov transitif transverse à un tore et non conjugué à une suspension*, Ergodic Theory Dynam. Systems 14 (1994) 633–643 MR Zbl

[6] **B A Burton**, *The Pachner graph and the simplification of 3-sphere triangulations*, from "Computational geometry", ACM, New York (2011) 153–162 MR Zbl

[7] **B A Burton**, **R Budney**, **W Pettersson**, et al., *Regina*: *software for low-dimensional topology* (1999) Available at `https://regina-normal.github.io`

[8] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy*, *a computer program for studying the geometry and topology of* 3*-manifolds* Available at `http://snappy.computop.org`

[9] **N M Dunfield**, *Alexander and Thurston norms of fibered 3-manifolds*, Pacific J. Math. 200 (2001) 43–58 MR Zbl

[10] **N M Dunfield**, **S Garoufalidis**, **A Shumakovitch**, **M Thistlethwaite**, *Behavior of knot invariants under genus 2 mutation*, New York J. Math. 16 (2010) 99–123 MR Zbl

[11] **S Fenley**, **L Mosher**, *Quasigeodesic flows in hyperbolic 3-manifolds*, Topology 40 (2001) 503–537 MR Zbl

[12] **S Frankel**, **S Schleimer**, **H Segerman**, *From veering triangulations to link spaces and back again*, preprint (2019) arXiv 1911.00006

[13] **D Fried**, *Fibrations over S^1 with pseudo-Anosov monodromy*, from "Travaux de Thurston sur les surfaces" (A Fathi, F Laudenbach, V Poénaru, editors), Astérisque 66-67, Soc. Math. France, Paris (1979) 251–266 Zbl

[14] **D Fried**, *The geometry of cross sections to flows*, Topology 21 (1982) 353–371 MR Zbl

[15] **D Fried**, *Transitive Anosov flows and pseudo-Anosov maps*, Topology 22 (1983) 299–303 MR Zbl

[16] **S Friedl**, **W Lück**, *The L^2-torsion function and the Thurston norm of 3-manifolds*, Comment. Math. Helv. 94 (2019) 21–52 MR Zbl

[17] **D Futer**, **F Guéritaud**, *Explicit angle structures for veering triangulations*, Algebr. Geom. Topol. 13 (2013) 205–235 MR Zbl

[18] **D Gabai**, *Foliations and the topology of 3-manifolds*, J. Differential Geom. 18 (1983) 445–503 MR Zbl

[19] **D Gabai**, *Foliations and genera of links*, Topology 23 (1984) 381–394 MR Zbl

[20] **A Giannopolous**, **S Schleimer**, **H Segerman**, *A census of veering triangulations*, electronic reference (2019) Available at `https://math.okstate.edu/people/segerman/veering.html`

[21] **S Goodman**, *Dehn surgery on Anosov flows*, from "Geometric dynamics", Lecture Notes in Math. 1007, Springer (1983) 300–307 MR Zbl

[22] **M Handel**, **W P Thurston**, *Anosov flows on new three manifolds*, Invent. Math. 59 (1980) 95–103 MR Zbl

[23] **C D Hodgson**, **J H Rubinstein**, **H Segerman**, **S Tillmann**, *Veering triangulations admit strict angle structures*, Geom. Topol. 15 (2011) 2073–2089 MR Zbl

[24] **C D Hodgson**, **J H Rubinstein**, **H Segerman**, **S Tillmann**, *Triangulations of 3-manifolds with essential edges*, Ann. Fac. Sci. Toulouse Math. 24 (2015) 1103–1145 MR Zbl

[25] **S Kamatani**, **H Kodama**, **T Noda**, *A Birkhoff section for the Bonatti–Langevin example of Anosov flow*, from "Foliations 2005", World Sci., Hackensack, NJ (2006) 229–243 MR Zbl

[26] **P Kirk**, **C Livingston**, *Concordance and mutation*, Geom. Topol. 5 (2001) 831–883 MR Zbl

[27] **M Lackenby**, *Taut ideal triangulations of 3-manifolds*, Geom. Topol. 4 (2000) 369–395 MR Zbl

[28] **M Landry**, *Taut branched surfaces from veering triangulations*, Algebr. Geom. Topol. 18 (2018) 1089–1114 MR Zbl

[29] **M P Landry**, *Veering triangulations and the Thurston norm: homology to isotopy*, Adv. Math. 396 (2022) art. id. 108102 MR Zbl

[30] **M Landry**, *Stable loops and almost transverse surfaces*, Groups Geom. Dyn. 17 (2023) 35–75 MR Zbl

[31] **M P Landry**, **Y N Minsky**, **S J Taylor**, *Flows, growth rates, and the veering polynomial*, Ergodic Theory Dynam. Systems 43 (2023) 3026–3107 MR Zbl

[32] **M P Landry**, **Y N Minsky**, **S J Taylor**, *A polynomial invariant for veering triangulations*, J. Eur. Math. Soc. 26 (2024) 731–788 MR Zbl

[33] **B Martelli**, *An introduction to geometric topology*, self published, Pisa, Italy (2016)

[34] **C T McMullen**, *Polynomial invariants for fibered 3-manifolds and Teichmüller geodesics for foliations*, Ann. Sci. École Norm. Sup. 33 (2000) 519–560 MR Zbl

[35] **C T McMullen**, *The Alexander polynomial of a 3-manifold and the Thurston norm on cohomology*, Ann. Sci. École Norm. Sup. 35 (2002) 153–171 MR Zbl

[36] **C T McMullen**, **C H Taubes**, *4-manifolds with inequivalent symplectic forms and 3-manifolds with inequivalent fibrations*, Math. Res. Lett. 6 (1999) 681–696 MR Zbl

[37] **C Millichap**, *Mutations and short geodesics in hyperbolic 3-manifolds*, Comm. Anal. Geom. 25 (2017) 625–683 MR Zbl

[38] **Y N Minsky**, **S J Taylor**, *Fibered faces, veering triangulations, and the arc complex*, Geom. Funct. Anal. 27 (2017) 1450–1496 MR Zbl

[39] **H R Morton**, **P Traczyk**, *The Jones polynomial of satellite links around mutants*, from "Braids", Contemp. Math. 78, Amer. Math. Soc., Providence, RI (1988) 587–592 MR Zbl

[40] **L Mosher**, *Surfaces and branched surfaces transverse to pseudo-Anosov flows on 3-manifolds*, J. Differential Geom. 34 (1991) 1–36 MR Zbl

[41] **L Mosher**, *Dynamical systems and the homology norm of a 3-manifold, II*, Invent. Math. 107 (1992) 243–281 MR Zbl

[42] **L Mosher**, *Laminations and flows transverse to finite depth foliations*, preprint (1996) Zbl

[43] **P Ozsváth**, **Z Szabó**, *Link Floer homology and the Thurston norm*, J. Amer. Math. Soc. 21 (2008) 671–709 MR Zbl

[44] **A Parlak**, *Veering triangulations and polynomial invariants of three-manifolds*, PhD thesis, University of Warwick (2021) Available at `http://wrap.warwick.ac.uk/162096`

[45] **A Parlak**, *The taut polynomial and the Alexander polynomial*, J. Topol. 16 (2023) 720–756 MR Zbl

[46] **A Parlak**, *Computation of the taut, the veering and the Teichmüller polynomials*, Exp. Math. 33 (2024) 1–26 MR Zbl

[47] **A Parlak**, **S Schleimer**, **H Segerman**, *Veering code for studying taut and veering ideal triangulations* Available at `https://github.com/henryseg/Veering`

[48] **D Ruberman**, *Mutation and volumes of knots in* S^3, Invent. Math. 90 (1987) 189–215 MR Zbl

[49] **M Shannon**, *Hyperbolic models for transitive topological Anosov flows in dimension three*, preprint (2021) arXiv 2108.12000

[50] **W P Thurston**, *A norm for the homology of* 3*-manifolds*, Mem. Amer. Math. Soc. 339, Amer. Math. Soc., Providence, RI (1986) 99–130 MR Zbl

[51] **C C Tsang**, *Veering branched surfaces, surgeries, and geodesic flows*, New York J. Math. 29 (2023) 1425–1495 MR Zbl

[52] **C C Tsang**, *Veering triangulations and pseudo-Anosov flows*, PhD thesis, University of California, Berkeley (2023) Available at `https://escholarship.org/uc/item/26h9h6dp`

[53] **C C Tsang**, *Constructing Birkhoff sections for pseudo-Anosov flows with controlled complexity*, Ergodic Theory Dynam. Systems 44 (2024) 2308–2360 MR Zbl

[54] **W Worden**, *Tnorm*, Sage package, version 1.0.2 (2021) Available at `https://pypi.org/project/tnorm`

Department of Mathematics, University of California, Davis
Davis, CA, United States

anna.parlak@gmail.com

Proposed: Ian Agol
Seconded: Leonid Polterovich, David Fisher

Received: 4 May 2023
Revised: 21 February 2024

Geometry & Topology **29**:4 (2025) 2175–2216
DOI: 10.2140/gt.2025.29.2175
Published: 27 June 2025

Unit inclusion in a (nonsemisimple) braided tensor category and (noncompact) relative TQFTs

BENJAMIN HAÏOUN

The inclusion of the unit in a braided tensor category $\mathcal{V}$ induces a 1-morphism in the Morita 4-category of braided tensor categories **BrTens**. We give criteria for the dualizability of this morphism.

When $\mathcal{V}$ is a semisimple (resp. nonsemisimple) modular category, we show that the unit inclusion induces, under the cobordism hypothesis, a (resp. noncompact) relative 3-dimensional topological quantum field theory. Following Jordan, Reutter and Safronov, we conjecture that these relative field theories together with their bulk theories recover Witten–Reshetikhin–Turaev (resp. De Renzi–Gainutdinov–Geer–Patureau-Mirand–Runkel) theories, in a fully extended setting. In particular, we argue that these theories can be obtained by the cobordism hypothesis.

18M20, 57R56

1 Introduction

This paper is motivated by the quest to bridge topological and higher-categorical constructions of Topological Quantum Field Theories (TQFTs). In the first approach one explicitly defines an n-manifold invariant and works their way to a TQFT, adding structure or extra conditions as necessary. This is the approach behind Reshetikhin and Turaev's construction [1994] of the 3-TQFTs predicted by Witten [1989], and their nonsemisimple variants [De Renzi et al. 2022]. The second approach classifies "vanilla" TQFTs, ie fully extended and without the extra structures/conditions of the above examples, using the cobordism hypothesis [Baez and Dolan 1995; Lurie 2009]. This classification is in terms of fully dualizable objects in a higher category. To bridge the two approaches, we must answer the questions:

Can the cobordism hypothesis recover the interesting, hand-built examples that we know?

If so, do we know what the relevant dualizable objects are?

There is evidence that the answer is yes if one allows for relative and noncompact versions of the cobordism hypothesis. The relevant dualizable objects were predicted to be those induced by the unit inclusion mentioned in this paper's title. This follows ideas of Walker, Freed and Teleman in the semisimple case, and was predicted to extend to the nonsemisimple case by Jordan, Reutter and Safronov. The whole story has mainly been communicated in talks; the only written references we are aware of are Walker's notes [2006] and Freed's slides [2011]. An obstacle to obtaining the Witten–Reshetikhin–Turaev theories from the cobordism hypothesis is that these theories are defined on a category of *cobordisms equipped with some extra structure*, which morally comes from the data of a bounding higher manifold. It was noticed by Walker and Freed–Teleman that this extra data is actually obtained from the Crane–Yetter 4-TQFT on the bounding manifold. Therefore the WRT theory should be thought of as a boundary theory for the Crane–Yetter theory. We do not know of a formal proof of this statement. An adequate description of relative field theory was given by Freed and Teleman [2014] and formalized by Johnson-Freyd and Scheimbauer [2017].

Another obstacle is that the nonsemisimple variants are defined on a *restricted class of cobordisms*, namely every 3-cobordism must have nonempty incoming boundary in every connected component. We need to use a noncompact version of the cobordism hypothesis to work with this restricted category of cobordisms. This noncompact version appears as an intermediate step in the sketch of proof of the cobordism hypothesis proposed by Hopkins and Lurie; see [Lurie 2009]. Note that there is independent work in progress of Reutter–Walker and Schommer-Pries in this direction.

A final obstacle is that WRT theories are *not fully extended*. It is known that they extend to the circle, but work of Douglas, Schommer-Pries and Snyder [Douglas et al. 2020] (see also [Freed and Teleman 2021]) shows that they extend to the point if and only if they are of Turaev–Viro type. This can be explained by the fact that they come from a relative setting, namely are defined on a category of cobordisms equipped with a bounding higher manifold, which we call filled cobordisms, and the point cannot be equipped with a bounding 1-manifold.

Summing up, one should be able to recover the WRT theories (resp. their nonsemisimple variants) from a 4-TQFT and a boundary (resp. noncompact boundary) theory for this 4-TQFT, both of which are fully extended and obtained from the cobordism hypothesis. It was proposed by Freed and Teleman in the semisimple case that the 4-TQFT is induced by the modular tensor category $\mathscr{V}$, and the boundary theory by the inclusion if the unit in $\mathscr{V}$; see the last slide of [Freed 2011]. It was proven in [Brochier et al. 2021a] that a possibly nonsemisimple modular tensor category $\mathcal{V} = \mathrm{Ind}(\mathscr{V})$ is indeed 4-dualizable in the even higher Morita 4-category of braided tensor categories **BrTens**, and therefore induces a 4-TQFT under the cobordism hypothesis. It was conjectured by Jordan, Reutter, Safronov and Walker in 2019 that this 4-TQFT together with the relative theory induced by the unit inclusion will also recover the nonsemisimple variants of WRT.

This paper gives the first step towards executing the above program. We use the framework of [Johnson-Freyd and Scheimbauer 2017] to prove that the unit inclusion is 3-dualizable (resp. noncompact-3-dualizable), and therefore induces a (resp. noncompact) relative 3-TQFT under the cobordism hypothesis.

In the last section we explain how one can obtain a theory defined on filled cobordisms from a relative theory together with its bulk theory. We state the conjectures that these recover the WRT theories and their nonsemisimple variants. Proving these conjectures would answer both questions above in the affirmative.

Context

The cobordism hypothesis formulated in [Baez and Dolan 1995], see [Lurie 2009] for a sketch of proof, provides a new angle to study and construct topological quantum field theories. One simply has to find a fully dualizable object in a higher category, and it induces a framed fully extended TQFT. We will study here one particular example of target category, the 4-category $\mathbf{BrTens}$ of braided tensor categories and bimodules between them, or more precisely the even higher Morita category $\mathrm{Alg}_2(\mathbf{Pr})$ of $\mathbb{E}_2$-algebras in some 2-category of cocomplete categories.

Even higher Morita categories are defined in [Johnson-Freyd and Scheimbauer 2017]; see also [Scheimbauer 2014] and [Haugseng 2017]. They form an $(n+k)$-fold Segal space, which we will abbreviate $(n+k)$-category, $\mathrm{Alg}_n(\mathscr{S})$ for $\mathscr{S}$ a k-category. It is shown in [Gwilliam and Scheimbauer 2018] that every object in $\mathrm{Alg}_n(\mathscr{S})$ is n-dualizable. We study the case $\mathscr{S} = \mathbf{Pr}$, the 2-category locally presentable $\Bbbk$-linear categories, cocontinuous functors and natural transformations, over a field $\Bbbk$ of characteristic zero. It was shown in [Brochier et al. 2021b, Theorem 5.21] that fusion categories provide a family of fully dualizable objects in $\mathbf{BrTens}$,[1] and later in [Brochier et al. 2021a, Theorem 1.1] that possibly nonsemisimple modular tensor categories are invertible, and hence also fully dualizable. Provided that we can endow these objects with the extra orientation structure needed for the oriented cobordism hypothesis, the TQFTs induced by fusion categories are expected to coincide with the Crane–Yetter theories. The TQFTs induced by nonsemisimple modular tensor categories are expected to coincide with the ones constructed in [Costantino et al. 2023].

There is another version of the cobordism hypothesis for relative field theories. Actually, there are multiple versions as there are multiple notions of relative TQFT. Lurie [2009, Example 4.3.23] proposed a definition of a domain wall based on a category of bipartite cobordisms, and proves the relative cobordism hypothesis under the assumption that the ambient category has duals. Stolz and Teichner [2011] define a notion of twisted quantum field theories in the context of topological algebras. Freed and Teleman [2014] describe relative n-TQFTs as morphisms between truncations of fully extended, ideally invertible, $(n+1)$-TQFTs. Johnson-Freyd and Scheimbauer [2017] give an explicit definition of what a morphism between TQFTs is and exhibit three different notions of strong-, lax- and oplax-twisted quantum field theories, which also makes sense when the bulk theories are only $(n+\varepsilon)$-TQFTs. We will be mostly interested in their notion of oplax-twisted quantum field theory.

We consider the category $\mathbf{BrTens}^{\rightarrow}$ of arrows in our chosen target category, where morphisms are "oplax" squares filled by a 2-morphism. There is a well-defined notion of source and target for objects and morphisms in this arrow category. Given $\mathfrak{Z}$ a fully extended $(n+\varepsilon)$-TQFT, a relative theory to $\mathfrak{Z}$ is a

[1] This is where the characteristic-zero assumption is needed.

symmetric monoidal functor $\mathcal{R}\colon \mathbf{Bord}_n \to \mathbf{BrTens}^{\to}$ whose source is trivial and whose target coincides with $\mathcal{Z}$. The cobordism hypothesis applies directly in this context, namely $\mathcal{R}$ can be reconstructed from its value on the point $\mathcal{R}(\mathrm{pt})\colon \mathbb{1} \to \mathcal{Z}(\mathrm{pt})$ which has to be fully dualizable in $\mathbf{BrTens}^{\to}$. We say that the 1-morphism $\mathcal{R}(\mathrm{pt})$ has to be fully oplax-dualizable.

In this paper we study the dualizability of the 1-morphism induced by the inclusion of the unit $\eta\colon \mathrm{Vect}_{\mathbb{k}} \to \mathcal{V}$. The braided monoidal functor η induces a $\mathrm{Vect}_{\mathbb{k}}$-$\mathcal{V}$-central algebra $\mathcal{A}_\eta$ which is $\mathcal{V}$ as a tensor category with bimodule structure induced by η. When it is oplax-dualizable, it induces a relative TQFT to the one of $\mathcal{V}$. There is a stronger notion of dualizability for a 1-morphism needed to induced a domain wall in the sense of Lurie. It is already known that the 1-morphism $\mathcal{A}_\eta$ is always 1-dualizable in the strong sense, by [Gwilliam and Scheimbauer 2018], 2-dualizable as soon as $\mathcal{V}$ is cp-rigid, and 3-dualizable as soon as $\mathcal{V}$ is fusion, by [Brochier et al. 2021b].

We show that fusion is a criterion for 3-dualizability, but not for 3-oplax-dualizability, emphasizing the difference between these notions. We study oplax-dualizability in detail, including the nonsemisimple cases. It is expected that the induced oplax-twisted theory corresponds to the Witten–Reshetikhin–Turaev TQFT seen as relative to the Crane–Yetter 4-TQFT in Walker's [2006] and Freed–Teleman's picture as in [Freed 2011].

In the modular nonsemisimple case, we will show that $\mathcal{A}_\eta$ is *not* 3-oplax-dualizable. We can only hope for a partial dualizability, and a partially defined TQFT. It turns out that Lurie's sketch of proof of the cobordism hypothesis is building a TQFT inductively from the dualizability data of the value on the point, and if this object is not fully dualizable, this process will simply stop partway through. Note again related ongoing work of Reutter–Walker and Schommer-Pries. In our case this process will stop before the very last step, and we obtain a theory defined on cobordisms with outgoing boundary in every connected component, which we call a noncompact TQFT; see Section 2.3.2.

Results

The unit inclusion in a braided tensor category $\mathcal{V} \in \mathbf{BrTens}$ gives a braided monoidal functor $\eta\colon \mathrm{Vect}_{\mathbb{k}} \to \mathcal{V}$. We work over a field $\mathbb{k}$ of characteristic zero. Using Definition 3.8, it induces a $\mathrm{Vect}_{\mathbb{k}}$-$\mathcal{V}$-central algebra $\mathcal{A}_\eta$, ie a 1-morphism in $\mathbf{BrTens}$. When we see this morphism as an object of the oplax arrow category $\mathbf{BrTens}^{\to}$, we denote it by $\mathcal{A}_\eta^\flat$. We are interested in the adjunctibility of $\mathcal{A}_\eta$ and in its oplax-dualizability, ie in the dualizability of $\mathcal{A}_\eta^\flat$.

We recall Lurie's sketch of proof of the noncompact cobordism hypothesis, and introduce the corresponding notion of noncompact-n-dualizable object in Section 2.3.2. We characterize oplax-dualizability of the unit inclusion:

Theorem 1.1 (Theorems 3.18, 3.21 and 3.20) *Let $\mathcal{V} \in \mathbf{BrTens}$ be a braided tensor category, and $\mathcal{A}_\eta^\flat$ the object of $\mathbf{BrTens}^{\to}$ induced by the inclusion of the unit. Then*:

(1) *$\mathcal{A}_\eta^\flat$ is always 2-dualizable.*

If $\mathcal{V}$ has enough compact-projectives, then:

(2) *$\mathscr{A}^{\flat}_{\eta}$ is noncompact-3-dualizable if and only if $\mathcal{V}$ is cp-rigid.*

(3) *$\mathscr{A}^{\flat}_{\eta}$ is 3-dualizable if and only if $\mathcal{V}$ is the free cocompletion of a small rigid braided monoidal category if and only if $\mathcal{V}$ is cp-rigid with compact-projective unit.*

In particular, for the examples of interest for Section 4:

Corollary 1.2 *Let $\mathscr{V}$ be a modular tensor category in the sense of [De Renzi et al. 2022], let $\mathcal{V} := \mathrm{Ind}(\mathscr{V})$ be its Ind-completion, and $\mathscr{A}^{\flat}_{\eta}$ be induced by the unit inclusion in $\mathcal{V}$. Then:*

(1) *If $\mathscr{V}$ is semisimple, $\mathscr{A}^{\flat}_{\eta}$ is 3-dualizable and induces a 3-TQFT $\mathscr{R}_{\mathcal{V}}$ with values in $\mathbf{BrTens}^{\rightarrow}$.*

(2) *If $\mathscr{V}$ is nonsemisimple, $\mathscr{A}^{\flat}_{\eta}$ is not 3-dualizable, but is noncompact-3-dualizable and induces a noncompact-3-TQFT $\mathscr{R}_{\mathcal{V}}$ with values in $\mathbf{BrTens}^{\rightarrow}$.*

We describe the dualizability data of $\mathscr{A}^{\flat}_{\eta}$ explicitly, which gives the values of $\mathscr{R}_{\mathcal{V}}$ on elementary handles. In dimension 2, the handle of index 2 is mapped to some mate of the unit inclusion η. The handle of index 1 is mapped to some mate of the tensor product T. And the handle of index 0 is mapped to some mate of the "balanced tensor product"

$$T_{\mathrm{bal}} : \mathcal{V} \underset{\mathcal{V}\boxtimes\mathcal{V}}{\boxtimes} \mathcal{V} \to \mathcal{V}$$

which is induced by T on the relative tensor product.

To determine this dualizability data, we use the fact that the dualizability of $\mathscr{A}^{\flat}_{\eta}$ is equivalent to the dualizability of $\mathcal{V}$ and the right-adjunctibility of $\mathscr{A}_{\eta}$; see [Johnson-Freyd and Scheimbauer 2017, Theorem 7.6] and Section 2.2.3. We remark that the adjunctibility of $\mathscr{A}_{\eta}$ often implies the relevant dualizability of $\mathcal{V}$, see Remark 3.25, which is a priori a phenomenon specific to the unit inclusion.

Freed and Teleman [2021, Theorem B] study the dualizability of the unit inclusion in the 3-category $\mathrm{Alg}_1(\mathrm{Rex}_{\mathbb{C}})$. Here $\mathrm{Rex}_{\mathbb{C}}$ is the 2-category of finitely cocomplete categories and finitely cocontinuous functors. They show that $\mathscr{V} \in \mathrm{Alg}_1(\mathrm{Rex}_{\mathbb{C}})$ is finite rigid semisimple if and only if $\mathscr{M}_{\eta}$ is 2-dualizable, ie lies in a subcategory with duals. The forward implication is [Douglas et al. 2020]. We can give an analogous statement one categorical number higher:

Theorem 1.3 (Theorem 3.22) *Suppose $\mathcal{V} \in \mathbf{BrTens}$ has enough compact-projectives. Then $\mathscr{A}_{\eta}$ is 3-dualizable if and only if $\mathcal{V}$ is finite rigid semisimple.*

Note that the full dualizability of $\mathscr{A}_{\eta}$ is indeed stronger that its oplax-dualizability. They are however expected to be equivalent when $\mathcal{V}$ is itself fully dualizable; see Remark 2.22. This actually implies some nondualizability results on $\mathcal{V}$; see Remark 3.24

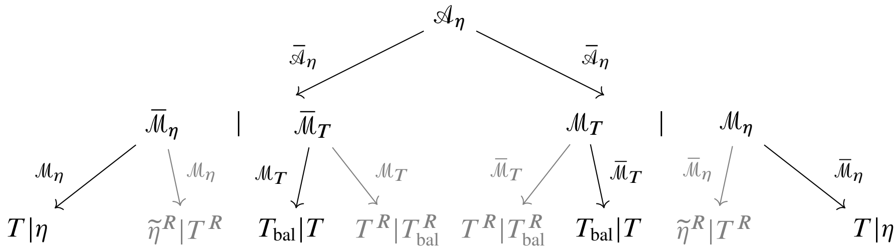
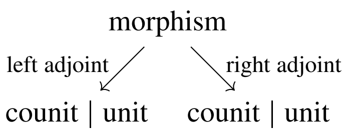


Figure 1: Adjunctibility data of the unit inclusion. The whole description (including gray) holds for $\mathcal{V}$ cp-rigid (Proposition 3.26), and the black subset holds when $\mathcal{V}$ has enough compact-projectives (Theorem 3.18). The functor $\widetilde{\eta}^R$ is the essentially unique cocontinuous functor that agrees with η^R on the compact-projectives.

We study in Section 3.2 the dualizability of the 1- (resp. 2-) morphism induced by a braided monoidal (resp. bimodule monoidal) functor F, which we denote by $\mathcal{A}_F$ (resp. $\mathcal{M}_F$). This allows us to give explicitly the adjunctibility data of $\mathcal{A}_\eta$; see Figure 1. We will use the notation



to depict adjunctibility data. We will give a further example explaining this notation in Figure 2.

We studied the unit inclusion, but a version of our arguments still work for any bimodule induced by a braided monoidal functor. Instead of a necessary and sufficient condition, we only have a sufficient condition:

Theorem 1.4 (Theorem 3.28) *Let $F\colon \mathcal{V} \to \mathcal{W}$ be a braided monoidal functor between two objects of* **BrTens**. *Then the object $\mathcal{A}_F^\flat \in$ **BrTens**$^\to$ induced by F is 2-dualizable. It is noncompact-3-dualizable as soon as $\mathcal{V}$ and $\mathcal{W}$ are cp-rigid. In this case, it is 3-dualizable if and only if F preserves compact-projectives.*

Applications

Nonsemisimple variants of Witten–Reshetikhin–Turaev TQFTs were introduced in [Blanchet et al. 2016] and [De Renzi et al. 2022]. They are defined on a restricted class of decorated cobordisms, including in particular cobordisms with incoming boundary in every connected component. This matches the notion of noncompact TQFT from Lurie, up to orientation reversal, and this is the part of the theory that we expect to obtain. We will focus on the TQFTs from [De Renzi et al. 2022] here, and actually their extension to the circle by [De Renzi 2021]. They are defined for possibly nonsemisimple modular categories, whose Ind-completions have been found to be 4-dualizable, and actually invertible, in **BrTens** by [Brochier et al. 2021a]. Using the cobordism hypothesis, for $\mathscr{V}$ a modular tensor category and $\mathcal{V} :=$

$\operatorname{Ind}(\mathscr{V})$ its Ind-completion, there is an essentially unique framed 4-TQFT $\mathfrak{X}_{\mathcal{V}}\colon \mathbf{Bord}_4^{\mathrm{fr}} \to \mathbf{BrTens}$ with $\mathfrak{X}_{\mathcal{V}}(\mathrm{pt}) = \mathcal{V}$.

We know from Corollary 1.2 that if $\mathcal{V}$ is semisimple (resp. nonsemisimple) the unit inclusion is 3-oplax-dualizable (resp. noncompact-3-oplax-dualizable). Using the cobordism hypothesis, it induces a framed (resp. noncompact) 3-TQFT $\mathfrak{R}_{\mathcal{V}}\colon \mathbf{Bord}_3^{\mathrm{fr}} \to \mathbf{BrTens}^{\rightarrow}$ (resp. $\mathfrak{R}_{\mathcal{V}}\colon \mathbf{Bord}_3^{\mathrm{fr,nc}} \to \mathbf{BrTens}^{\rightarrow}$) relative to $\mathfrak{X}_{\mathcal{V}}$. We give conjectures that these theories can be oriented.

Conjecture 1.5 (Conjectures 4.11 and 4.13) *Let $\mathscr{V}$ be a modular tensor category and $\mathcal{V} = \operatorname{Ind}(\mathscr{V})$, which is invertible and in particular 5-dualizable by [Brochier et al. 2021a]. Then*:

1. *The ribbon structure of $\mathscr{V}$ induces an $\mathrm{SO}(3)$-homotopy-fixed-point structure on $\mathcal{V}$.*
2. *The ribbon structure of η induces an $\mathrm{SO}(3)$-homotopy-fixed-point structure on $\mathscr{A}_\eta^\flat$.*
3. *A choice of modified trace t on $\mathscr{V}$ induces an $\mathrm{SO}(4)$-homotopy-fixed-point structure on $\mathcal{V}$.*

Given a modified trace t, let $d(\mathscr{V})_{\mathsf{t}}$ denote the global dimension of $\mathscr{V}$ computed using t, defined as the value on S^4 of the $(3+1)$-TQFT of [Costantino et al. 2023] with the same input.

4. *Exactly two modified traces induce $\mathrm{SO}(5)$-homotopy-fixed-point structures on $\mathcal{V}$, namely $\pm\mathscr{D}_{\mathsf{t}}^{-1}\mathsf{t}$ for $\mathscr{D}_{\mathsf{t}}$ a square root of the global dimension $d(\mathscr{V})_{\mathsf{t}}$.*

Let us include here the conjecture that the TQFTs of [Costantino et al. 2023] compute the (3+1)-part of the fully extended TQFT associated with $\mathscr{V}$, which we can state now that we have conjectured orientation structures.

Conjecture 1.6 *Let $\mathscr{V}$ be a modular tensor category. Choose t a modified trace on $\mathscr{V}$ and let $\mathfrak{X}_{\mathcal{V}}$ be the associated oriented 4-TQFT. Then one has a natural isomorphism*

$$\mathscr{S}_{\mathscr{V}} \simeq h_1\Omega^3\mathfrak{X}_{\mathcal{V}}$$

between the $(3+1)$-TQFT defined in [Costantino et al. 2023] and the $(3+1)$-part of $\mathfrak{X}_{\mathcal{V}}$.

We construct the "anomalous" theory $\mathscr{A}_{\mathcal{V}}\colon \mathbf{Bord}_3^{\mathrm{filled}} \to \mathbf{Tens}$ associated with $\mathfrak{R}_{\mathcal{V}}$ and $\mathfrak{X}_{\mathcal{V}}$. It is defined on a 3-category of cobordisms equipped with a filling, ie a bounding higher manifold, which degenerates to the more usual $\widetilde{\mathbf{Cob}}$ on which WRT-type theories are defined. The anomalous theory is noncompact when $\mathfrak{R}_{\mathcal{V}}$ is.

We can now state the main conjectures. We claim that one can recover WRT and DGGPR theories from the cobordism hypothesis using the construction we described:

Conjecture 1.7 (Conjecture 4.16) *Let $\mathscr{V}$ be a semisimple modular tensor category with a chosen square root of its global dimension. The anomalous theory $\mathscr{A}_{\mathcal{V}}$ induced by the associated oriented 4-TQFT $\mathfrak{X}_{\mathcal{V}}$ and oriented oplax-$\mathfrak{X}_{\mathcal{V}}$-twisted 3-TQFT $\mathfrak{R}_{\mathcal{V}}$ recovers the once-extended Witten–Reshetikhin–Turaev theory as its 321-part.*

Conjecture 1.8 (Conjecture 4.17) *Let $\mathscr{V}$ be a nonsemisimple modular tensor category with a chosen modified trace* $\mathtt{t}$ *and square root of its global dimension. The anomalous theory $\mathscr{A}_{\mathscr{V}}$ induced by the associated oriented 4-TQFT $\mathscr{Z}_{\mathscr{V}}$ and oriented noncompact oplax-$\mathscr{Z}_{\mathscr{V}}$-twisted 3-TQFT $\mathscr{R}_{\mathscr{V}}$ recovers the once-extended DGGPR theory for cobordisms with trivial decoration as its 321-part.*

We show that the values on the circle coincide by computing $\mathscr{R}_{\mathscr{V}}$ explicitly and using the factorization homology description of $\mathscr{Z}_{\mathscr{V}}$. For higher dimensions, one needs to identify the values of $\mathscr{Z}_{\mathscr{V}}$ on 3-manifolds with skein modules, which is at this day still conjectural [Johnson-Freyd 2021, Conjecture 9.10].

An interesting consequence of these conjectures is that since the WRT and DGGPR theories only correspond to the 321-part of $\mathscr{A}_{\mathscr{V}}$, they actually extend down. They do not descend to the point, which is not null-bordant, but to the pair of points S^0.

Note that we use both the oplax-twisted 3-TQFT and the 4-TQFT in this construction. Therefore, not every case of 3-oplax-dualizability in Theorem 1.1 induces such a theory. We also need $\mathscr{V}$ to be 4-dualizable. The assumption that $\mathscr{Z}_{\mathscr{V}}$ is invertible however can be dropped. The anomalous theory would then strongly depend on the filling, and give interesting invariants of 4-manifolds with boundary.

The construction of the anomalous theory $\mathscr{A}_{\mathscr{V}}$ using a bounding manifold is needed to recover the usual constructions of WRT and DGGPR theories. It is also necessary for some applications, eg to obtain a scalar invariant of 3-manifold. However, one could argue that the more fundamental object is the fully extended twisted 3-TQFT $\mathscr{R}_{\mathscr{V}}$. It does not assign a scalar to a 3-manifold, but an element in a one-dimensional vector space: the state space of the invertible Crane–Yetter TQFT. If one is happy to allow this feature, then one can argue that WRT is a fully extended theory, with values in the oplax arrow category $\mathbf{BrTens}^{\rightarrow}$.

Acknowledgements I would like to thank my advisors Francesco Costantino and David Jordan for their invaluable help and encouragements. I am grateful to Pavel Safronov for many helpful discussions and to the anonymous referee for many useful remarks on a first version of this article. This research is part of my PhD thesis and took place both at the Université Toulouse 3 Paul Sabatier and at the University of Edinburgh.

2 Relative and noncompact TQFTs

In this paper we will study the dualizability of a 1-morphism. What exact kind of dualizability we are interested in is dictated by the relative cobordism hypothesis: we want a 1-morphism that will induce a relative TQFT. It turns out that there are multiple notions of relative TQFTs, and therefore multiple interesting notions of dualizability for a 1-morphism.

Throughout, we will use the expression n-category to mean (∞, n)-category, and more precisely complete n-fold Segal space. For $j \geq k$, we write $\circ_k$ for the composition of j-morphisms in the direction of k-morphisms. We write Id_f^k for taking k-times the identity of f.

2.1 Review of relative TQFTs

We recall the notions of relative TQFTs that will be our motivation. Let $\mathscr{C}$ be a symmetric monoidal n-category. We distinguish two flavors.

The first is purely topological. Lurie [2009, Example 4.3.23] defines a category $\mathbf{Bord}_n^{dw}$ of bipartite cobordisms with two different colors for the bulk and interfaces between them. There are in particular manifolds with only one color and without interfaces. This induces two inclusions $\mathbf{Bord}_n \to \mathbf{Bord}_n^{dw}$.

Definition 2.1 (Lurie) A domain wall between two theories $\mathcal{Z}_1, \mathcal{Z}_2 \colon \mathbf{Bord}_n \to \mathscr{C}$ is a symmetric monoidal functor $\mathbf{Bord}_n^{dw} \to \mathscr{C}$ that restricts to $\mathcal{Z}_1$ and $\mathcal{Z}_2$ on manifolds with one color.

In particular, the interval with an interface point in the middle induces a morphism $\mathcal{Z}_1(\mathrm{pt}) \to \mathcal{Z}_2(\mathrm{pt})$. Freed and Teleman [2014] describe a notion of relative TQFT by means of such morphisms for every values of $\mathcal{Z}_1$ and $\mathcal{Z}_2$ on manifolds of dimension strictly less than n. They mention that their notion should be equivalent to Lurie's notion of domain wall. A more detailed comparison will appear in William Stewart's PhD thesis.

The second notion focuses on the algebraic flavor of Freed–Teleman's description. One can drop the assumption that $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are well defined on n-manifolds because these don't appear. Johnson-Freyd and Scheimbauer define three different notions of an n-category of arrows in an n-category. We will focus on the oplax one $\mathscr{C}^{\to}$.

Definition 2.2 (sketch, see Definition 5.14 in [Johnson-Freyd and Scheimbauer 2017]) Let $\mathscr{C}$ be a symmetric monoidal n-category. The *symmetric monoidal n-category $\mathscr{C}^{\to}$ of oplax arrows in $\mathscr{C}$* has:

- **Objects** Triples $f = (s_f, t_f, f^{\#})$, where s_f and t_f are objects of $\mathscr{C}$ and $f^{\#} \colon s_f \to t_f$ is a 1-morphism.
- **1-morphisms, $f \to g$** Triples $h = (s_h, t_h, h^{\#})$, where $s_h : s_f \to s_g$ and $t_h : t_f \to t_g$ are 1-morphisms, and $h^{\#} \colon g^{\#} \circ s_h \Rightarrow t_h \circ f^{\#}$ is a 2-morphism.

$\vdots$

- **k-morphisms, $a \to b$** Triples $f = (s_f, t_f, f^{\#})$, where $s_f : s_a \to s_b$ and $t_f : t_a \to t_b$ are k-morphisms in $\mathscr{C}$, and $f^{\#}$ is a $k+1$-morphism in $\mathscr{C}$ from the composition of some whiskerings of $b^{\#}$ and s_f to the composition of some whiskerings of t_f and $a^{\#}$.

It has two symmetric monoidal functors $s, t : \mathscr{C}^{\to} \to \mathscr{C}$.

To avoid confusion, when we see a 1-morphism f of $\mathscr{C}$ as an object of $\mathscr{C}^{\to}$ we will denote it by $f^{\flat}$. The notation comes from $(f^{\flat})^{\#} = f$.

Definition 2.3 [Johnson-Freyd and Scheimbauer 2017, Definition 5.16] Let $\mathscr{C}$ be a symmetric monoidal n-category and $\mathfrak{X}_1, \mathfrak{X}_2 \colon \mathbf{BORD}_{n-1} \to \mathscr{C}$ two categorified $(n-1)$-TQFTs. An *oplax-$\mathfrak{X}_1$-$\mathfrak{X}_2$-twisted* $(n-1)$-*TQFT* is a symmetric monoidal functor

$$\mathfrak{R} \colon \mathbf{BORD}_{n-1} \to \mathscr{C}^{\to}$$

such that $s(\mathfrak{R}) = \mathfrak{X}_1$ and $t(\mathfrak{R}) = \mathfrak{X}_2$.

The name and strategy come from [Stolz and Teichner 2011].

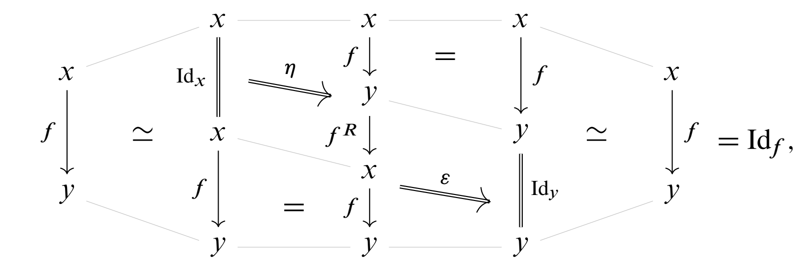
We will use the formalism of Johnson-Freyd and Scheimbauer in this paper. For applications, see Section 4, we are interested in the case where $\mathfrak{X}_1$ is the trivial theory and $\mathfrak{X}_2$ is well-defined on n-manifolds. If $\mathfrak{X} \colon \mathbf{BORD}_n \to \mathscr{C}$ is defined on n-manifolds, we will say oplax-$\mathfrak{X}$-twisted theory for oplax-Triv-$\mathfrak{X}|_{\mathbf{BORD}_{n-1}}$-twisted theory. Under this extra hypothesis, which was made in [Freed and Teleman 2014], the notion of oplax-twisted field theory should agree with Lurie's notion of domain walls; see Remark 2.22.

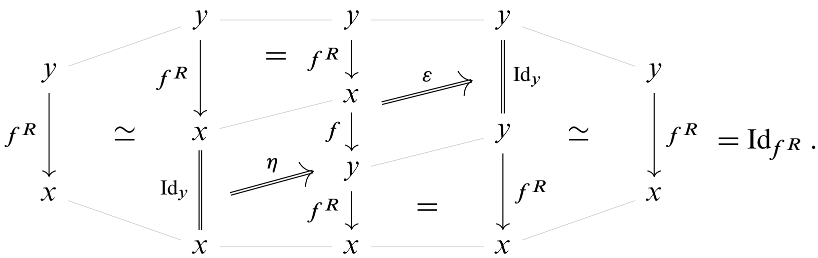
2.2 Dualizability data

Let us first recall the multiple notions of dualizability and adjunctibility for morphisms in a symmetric monoidal n-category $\mathscr{C}$.

2.2.1 Definitions

Definition 2.4 Let $\mathscr{C}$ be a bicategory, and $f \colon x \to y$ a 1-morphism in $\mathscr{C}$. A right adjoint for f is a morphism $f^R \colon y \to x$ together with two 2-morphisms $\varepsilon \colon f \circ f^R \Rightarrow \mathrm{Id}_y$ called the counit and $\eta \colon \mathrm{Id}_x \Rightarrow f^R \circ f$ called the unit, satisfying the so-called snake identities:





We say that f has a right adjoint f^R and that f^R has a left adjoint f.

This definition extends to higher categories. Let $\mathscr{C}$ be an n-category, $2 \leq k \leq n$ and $f: x \to y$ a k-morphism between two $k-1$-morphisms $x, y: a \to b$ in $\mathscr{C}$. A right adjoint for f is a right adjoint for f seen as a 1-morphism in the bicategory $h_2(\mathrm{Hom}_{\mathscr{C}}(a,b))$. If $k = 1$ then we demand a right adjoint of f in $h_2(\mathscr{C})$. If $k = 0$ and $\mathscr{C}$ is a monoidal category then we demand a right adjoint of f in $h_2(B\mathscr{C})$, where $B\mathscr{C}$ is the one object $n+1$-category with endomorphisms of the object being $\mathscr{C}$, and composition the monoidal structure of $\mathscr{C}$, ie $X \circ_0 Y := X \otimes Y$.

Definition 2.5 Let $\mathscr{C}$ be a symmetric monoidal n-category. It is said to have duals up to level m if every k-morphism of $\mathscr{C}$, for $0 \leq k < m$, has both a left and a right adjoint. It is said to have duals if it has duals up to level n.

An object $X \in \mathscr{C}$ is said m-dualizable if it lies in a sub-n-category with duals up to level m. It is called fully dualizable if it is n-dualizable.

There are multiple notions of dualizability, or adjunctibility, for morphisms and higher morphisms in $\mathscr{C}$. Following [Lurie 2009] one defines:

Definition 2.6 A k-morphism f of $\mathscr{C}$ is said *m-dualizable* if it lies in a sub-n-category with duals up to level $m+k$. It is called *fully dualizable* if it is $(n-k)$-dualizable.

Following [Johnson-Freyd and Scheimbauer 2017] one gets a few more notions. For simplicity we focus on 1-morphisms.

Definition 2.7 A 1-morphism $f: X \to Y$ of $\mathscr{C}$ is said *m-oplax-dualizable* if it is m-dualizable as an object $f^\flat$ of $\mathscr{C}^\to$. It is said *m-lax-dualizable* if it is m-dualizable as an object of $\mathscr{C}^\downarrow$, where $\mathscr{C}^\downarrow$ is the category of lax arrows defined in [Johnson-Freyd and Scheimbauer 2017, Definition 5.14].

Definition 2.8 A k-morphism f is said to be *left* (resp. *right*) *adjunctible* if it has a left (resp. right) adjoint, and *adjunctible* if it has arbitrary left and right adjoints ($(f^L)^L$, $(f^R)^R$ and so on). It is said to be *m-times* (resp. *left*, *right*) *adjunctible* if it is $m-1$-times (resp. left, right) adjunctible and every unit/counit witnessing this are themselves (resp. left, right) adjunctible. We sometimes abbreviate m-times adjunctible as m-adjunctible.

Note that being (left, right) adjunctible is only a condition on the morphism while being (lax, oplax) dualizable is also a condition on its source and target.

Theorem 2.9 [Johnson-Freyd and Scheimbauer 2017, Theorem 7.6] *A 1-morphism $f: X \to Y$ of $\mathscr{C}$ is m-oplax-dualizable if and only if X and Y are both m-dualizable and f is m-times right adjunctible.*

Similarly, it is m-lax-dualizable if and only if X and Y are both m-dualizable and f is m-times left adjunctible.

Similarly, a 1-morphism $f: X \to Y$ is m-dualizable if and only if it is m-times adjunctible and its source and targets are $(m+1)$-dualizable.

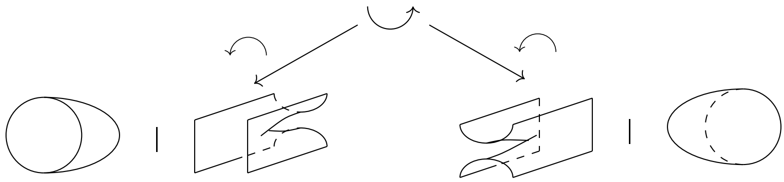


Figure 2: Example of the notation for the adjunctibility data of the cup $\circlearrowleft : \varnothing \to \bullet^- \sqcup \bullet^+$ in the bicategory $\mathbf{Cob}_{0,1,2}$ of 0, 1 and 2-dimensional oriented cobordisms.

2.2.2 Redundancy in the dualizability data The dualizability data of a morphism grows very fast: there are four units/counits for the left and right adjunctions, and this does not consider taking the right adjoint of the right adjoint and so on. In particular, checking n-adjunctibility of a morphism seems tedious. It turns out that there is a lot of redundancy in this data, especially if we are only interested in dualizability properties.

Let us begin with some notation. Let f be a k-morphism in an n-category. We say that $\mathrm{Radj}(f)$ (resp. $\mathrm{Ladj}(f)$) exists if f has a right (resp. left) adjoint, in which case we denote this adjoint by $\mathrm{Radj}(f)$ (resp. $\mathrm{Ladj}(f)$), and the unit and counit of the adjunction by $\mathrm{Ru}(f)$ and $\mathrm{Rco}(f)$ (resp. $\mathrm{Lu}(f)$ and $\mathrm{Lco}(f)$). When these adjoints exist we will display the right and left dualizability data as

$$\begin{array}{c} f \\ \searrow^{\mathrm{Radj}(f)} \\ \mathrm{Rco}(f)\,|\,\mathrm{Ru}(f) \end{array} \quad \text{and} \quad \begin{array}{c} f \\ {}^{\mathrm{Ladj}(f)}\swarrow \\ \mathrm{Lco}(f)\,|\,\mathrm{Lu}(f) \end{array}.$$

See Figure 2 for an example. Note that adjoints, units and counits are only defined up to isomorphism, and we may write $\mathrm{Radj}(f)$, $\mathrm{Ru}(f)$, $\mathrm{Rco}(f)$ for any choice of adjoint, unit, counit. The fact that these choices do not matter will be shown in point (1) of Proposition 2.13.

Definition 2.10 We say that two k-morphisms f and g have same dualizability properties, which we denote by $f \doteq g$, if for every finite sequence $(a_i)_{i\in\{1,\dots,m\}}$, with $a_i \in \{\mathrm{Radj}, \mathrm{Ladj}, \mathrm{Ru}, \mathrm{Rco}, \mathrm{Lu}, \mathrm{Lco}\}$,

$$a_m(\cdots a_2(a_1(f))\cdots) \text{ exists if and only if } a_m(\cdots a_2(a_1(g))\cdots) \text{ exists,}$$

and this for any choice of adjoints, units and counits.

We will show that dualizability properties are preserved by isomorphisms and "higher mating" defined in Definition 2.12. Let us describe formally this second notion.

Proposition 2.11 *Let $f : x \to y$ be a k-morphism in an n-category $\mathscr{C}$ with adjoint (f^R, ε, η). Then for any other k-morphisms $g : z \to x$ and $h : z \to y$, one has an equivalence between the $(n-k-1)$-categories of $(k+1)$-morphisms*:

$$\Phi^f_{g,h} : \left\{ \begin{array}{rcl} \mathrm{Hom}_{\mathscr{C}}(f \circ_k g, h) & \xrightarrow{\sim} & \mathrm{Hom}_{\mathscr{C}}(g, f^R \circ_k h), \\ N : f \circ_k g \to h & \mapsto & (\mathrm{Id}_{f^R} \circ_k N) \circ_{k+1} (\eta \circ_k \mathrm{Id}_g), \\ (k+j)\text{-morphism } \alpha & \mapsto & (\mathrm{Id}^j_{f^R} \circ_k \alpha) \circ_{k+1} (\mathrm{Id}^{j-1}_{\eta} \circ_k \mathrm{Id}^j_g). \end{array} \right\}$$

Similarly, for any $g\colon x\to z$ and $h\colon y\to z$, one gets an equivalence

$$\Psi^f_{g,h}=(-\circ_k \mathrm{Id}_f)\circ_{k+1}(\mathrm{Id}_g\circ_k\eta)\colon \mathrm{Hom}_{\mathscr{C}}(g\circ_k f^R,h)\xrightarrow{\sim}\mathrm{Hom}_{\mathscr{C}}(g,h\circ_k f).$$

Proof Its inverse is given by

$$(\Phi^f_{g,h})^{-1}:\left\{\begin{array}{rcl}\mathrm{Hom}_{\mathscr{C}}(g,f^R\circ_k h) & \xrightarrow{\sim} & \mathrm{Hom}_{\mathscr{C}}(f\circ_k g,h),\\ M:g\to f^R\circ_k h & \mapsto & (\varepsilon\circ_k\mathrm{Id}_h)\circ_{k+1}(\mathrm{Id}_f\circ_k M),\\ (k+j)\text{-morphism }\beta & \mapsto & (\mathrm{Id}^{j-1}_\varepsilon\circ_k\mathrm{Id}^j_h)\circ_{k+1}(\mathrm{Id}^j_f\circ_k\beta).\end{array}\right\}$$

The composition $\Phi^f_{g,h}\circ(\Phi^f_{g,h})^{-1}$ (resp. $(\Phi^f_{g,h})^{-1}\circ\Phi^f_{g,h}$) is postcomposition (resp. precomposition) by a snake identity. Similarly, we have $(\Psi^f_{g,h})^{-1}=(\mathrm{Id}_g\circ_k\varepsilon)\circ_{k+1}(-\circ_k\mathrm{Id}_{f^R})$. □

Definition 2.12 For a $(k+1)$-morphism $N\colon f\circ_k g\to h$, we say that N and $\Phi^f_{g,h}(N)$ are *mates*. For a higher morphism α in $\mathrm{Hom}_{\mathscr{C}}(f\circ_k g,h)$, we say that α and $\Phi^f_{g,h}(\alpha)$ are *higher mates*. Similarly, for N and α in $\mathrm{Hom}_{\mathscr{C}}(g\circ_k f^R,h)$ we call N and $\Psi^f_{g,h}(N)$ mates, and α and $\Psi^f_{g,h}(\alpha)$ higher mates. More generally we say that N and M are mates (resp. α and β are higher mates) if they can be linked by a chain of matings (resp. higher matings) and isomorphisms.

For a k-morphism f, we say that g is *obtained from f by whiskering* if it can be written as a composition of f with identities of lower morphisms. Note that if α and β are higher mates, each is obtained from the other by whiskering.

Proposition 2.13 *Let f and g be k-morphisms in an n-category. Then*:

1. $f\doteq f$.
2. *If $f\overset{\varphi}{\simeq}g$ are isomorphic, then $f\doteq g$.*
3. *If $f=g\circ_k h$ for an isomorphism h, then $f\doteq g$.*
4. *If f and g are higher mates, then $f\doteq g$.*

Proof What we have to prove for point (1) is that existence of higher adjoints in the adjunctibility data does not depend on the choices made in the adjunctions. This kind of result is known as a coherence statement in the literature and is usually stated as contractibility of a space of dualizability data; see [Lurie 2017, Lemma 4.6.1.10] or [Riehl and Verity 2016], though our statement here is more elementary.

Let us discuss only right adjoints and right units below, every other notion being related by taking appropriate opposite categories.

The adjunctibility data of f is unique up to isomorphism. This means that if g_1 and g_2 are both right adjoints to f, with units u_1 and u_2, then there is an isomorphism $\varphi\colon g_1\xrightarrow{\sim}g_2$ such that u_2 is isomorphic to $(\varphi\circ_k\mathrm{Id}_f)\circ_{k+1}u_1$.

We observe that any choice of right adjoint or right unit of f are related by a sequence of either point (2) (isomorphism) or point (3) (composing with an isomorphism). The strategy of the proof is to show more generally that if f and g are related by a sequence of either point (2) or point (3), then f is right adjunctible if and only if g is, and their adjoints, units and counits are again related by a sequence of either point (2) or point (3). The result then follows by induction on m the number of letters in $a_m(\cdots a_2(a_1(f))\cdots)$. For point (4) we show similarly that if f is related to g by a sequence of points (2), (3) and (4), then so are their adjoints, units and counits.

(2) If $f \overset{\varphi}{\simeq} g$ are isomorphic, then f has a right (resp. left) adjoint if and only if g does, in which case one can choose

$$\mathrm{Radj}(g)=\mathrm{Radj}(f), \quad \mathrm{Ru}(g)=(\mathrm{Id}_{\mathrm{Radj}(f)}\circ_k\varphi)\circ_{k+1}\mathrm{Ru}(f), \quad \mathrm{Rco}(g)=(\varphi^{-1}\circ_k\mathrm{Id}_{\mathrm{Radj}(f)})\circ_{k+1}\mathrm{Rco}(f).$$

(3) If $f=g\circ_k h$ is obtained as a composition, then f has a right (resp. left) adjoint as soon as g and h do, in which case one can choose $\mathrm{Radj}(f)=\mathrm{Radj}(h)\circ_k\mathrm{Radj}(g)$, $\mathrm{Ru}(f)=(\mathrm{Id}_{\mathrm{Radj}(h)}\circ_k \mathrm{Ru}(g)\circ_k\mathrm{Id}_h)\circ_{k+1}\mathrm{Ru}(h)$ and $\mathrm{Rco}(f)=(\mathrm{Id}_g\circ_k\mathrm{Rco}(h)\circ_k\mathrm{Id}_{\mathrm{Radj}(g)})\circ_{k+1}\mathrm{Rco}(h)$. In particular, if h is an isomorphism, then $g\simeq f\circ_k h^{-1}$, and f has a right (resp. left) adjoint if and only if g does.

(4) If $f=g\circ_j h$ is obtained as a composition in the direction of j-morphisms for $j<k$, then f has a right (resp. left) adjoint as soon as g and h do, in which case one can choose $\mathrm{Radj}(f)=\mathrm{Radj}(g)\circ_j\mathrm{Radj}(h)$, $\mathrm{Ru}(f)=\mathrm{Ru}(g)\circ_j\mathrm{Ru}(h)$ and $\mathrm{Rco}(f)=\mathrm{Rco}(g)\circ_j\mathrm{Rco}(h)$. In particular, if h is an identity of a lower morphism, then f has a right (resp. left) adjoint as soon as g does. So, if f and g are higher mates, they both can be obtained as composition of the other with identities of lower morphisms, and f has a right (resp. left) adjoint if and only if g does. □

We can now describe the redundancy in the dualizability data:

Proposition 2.14 *Let f be a k-morphism in an n-category $\mathscr{C}$, and suppose that* $\mathrm{Radj}(f)$, $\mathrm{Radj}(\mathrm{Rco}(f))$ *and* $\mathrm{Radj}(\mathrm{Ru}(f))$ *exist. Then*

(1) *f is 1-adjunctible, and one can choose* $\mathrm{Ladj}(f)=\mathrm{Radj}(f)$, $\mathrm{Lu}(f)=\mathrm{Radj}(\mathrm{Rco}(f))$ *and* $\mathrm{Lco}(f)=\mathrm{Radj}(\mathrm{Ru}(f))$, *and*

(2) $\mathrm{Rco}(\mathrm{Ru}(f))\doteq\mathrm{Ru}(\mathrm{Rco}(f))$.

Suppose moreover that $\mathrm{Radj}(\mathrm{Lco}(f))$ *and* $\mathrm{Radj}(\mathrm{Lu}(f))$ *exist. Then*

(3) *f is 2-adjunctible, and* $\mathrm{Rco}(f)\doteq\mathrm{Radj}(\mathrm{Radj}(\mathrm{Rco}(f)))$ *and* $\mathrm{Ru}(f)\doteq\mathrm{Radj}(\mathrm{Radj}(\mathrm{Ru}(f)))$.

In particular, if $f=X$ is an object in a symmetric monoidal n-category, then

(4) *X is 1-adjunctible if and only if it has a dual. It is 2-adjunctible if and only if* $\mathrm{ev}_X:=\mathrm{Rco}(X)$ *and* $\mathrm{coev}_X:=\mathrm{Ru}(X)$ *have right adjoints. More generally, it is m-adjunctible if and only if* $\mathrm{Radj}(\mathrm{Rco}^k(\mathrm{Ru}^{m-1-k}(X)))$ *exist for all* $0\le k\le m-1$.

Proof Point (1) is [Lurie 2009, Remark 3.4.22], or [Schommer-Pries 2014, Lemma 20.1]. One directly checks that the right adjoints of the right counit and unit satisfy the snake relations, because taking right adjoints behaves well with composition, and exhibit $\mathrm{Radj}(f)$ as the left adjoint of f.

Point (2) is [Lurie 2009, Proposition 3.4.21]. It is shown that $\mathrm{Rco}(\mathrm{Ru}(f))$ and $\mathrm{Ru}(\mathrm{Rco}(f))$ are higher mates, so in particular $\mathrm{Rco}(\mathrm{Ru}(f)) \doteq \mathrm{Ru}(\mathrm{Rco}(f))$.

Point (3) is [Johnson-Freyd and Scheimbauer 2017, Lemma 7.11]. Let us recall the argument to illustrate what we mean by redundancy in the dualizability data. First, by point (1) we can take $\mathrm{Lco}(f)$ and $\mathrm{Lu}(f)$ to be the right adjoints of $\mathrm{Ru}(f)$ and $\mathrm{Rco}(f)$. Now by a left-handed version of point (1) we can also obtain a unit and counit for the right adjunction of f as the right adjoints of $\mathrm{Lco}(f)$ and $\mathrm{Lu}(f)$. Therefore both $\mathrm{Rco}(f)$ and its double right adjoint are counits for the right adjunction of f: there is a redundancy in the dualizability data. By the proposition above, they have the same dualizability properties, and similarly for the right unit.

For point (4), by [Johnson-Freyd and Scheimbauer 2017, Corollary 7.12] we have to check that X is dualizable and that its evaluation and coevaluation maps are $(m-1)$-times right adjunctible, ie that $\mathrm{Radj}(a_1(\cdots a_{m-1}(X)\cdots)$ exists for any $(a_i)_i \in \{\mathrm{Rco}, \mathrm{Ru}\}^{m-1}$. Using point (2), we know that Ru and Rco commute as far as existence of adjoints is concerned, so there are only m different $(m-1)$-morphisms whose adjunctibility should be checked, $\mathrm{Rco}^k(\mathrm{Ru}^{m-1-k}(X))$ for $0 \le k \le m-1$. □

2.2.3 Oplax dualizability data We investigate the proof of Theorem 2.9 and explain how to get from adjunctibility data in $\mathscr{C}$ to dualizability data in $\mathscr{C}^{\to}$.

Theorem 2.15 (Johnson-Freyd–Scheimbauer) *Let $f = (s_f, t_f, f^\#) : a = (s_a, t_a, a^\#) \to b = (s_b, t_b, b^\#)$ be a k-morphism in $\mathscr{C}^{\to}$ so $s_f : s_a \to s_b$ and $t_f : t_a \to t_b$ are k-morphism in $\mathscr{C}$, and $f^\#$ is a $(k+1)$-morphism in $\mathscr{C}$ from the composition of some whiskerings of $b^\#$ and s_f to the composition of some whiskerings of t_f and $a^\#$. Then*

$$f \text{ has a right adjoint in } \mathscr{C}^{\to} \text{ if and only if } s_f,\ t_f \text{ and } f^\# \text{ have right adjoints in } \mathscr{C}.$$

In this case,

- $\mathrm{Radj}(f) = (\mathrm{Radj}(s_f), \mathrm{Radj}(t_f), g)$ *where g is a mate of* $\mathrm{Radj}(f^\#)$,
- $\mathrm{Ru}(f) = (\mathrm{Ru}(s_f), \mathrm{Ru}(t_f), u)$ *where u is a higher mate of* $\mathrm{Rco}(f^\#)$, *and*
- $\mathrm{Rco}(f) = (\mathrm{Rco}(s_f), \mathrm{Rco}(t_f), v)$ *where v is a higher mate of* $\mathrm{Ru}(f^\#)$.

In particular, if we only look at the right dualizability data, and only take right adjoints once, then

for all $i, j \in \mathbb{N}$, $\mathrm{Radj}(\mathrm{Rco}^i(\mathrm{Ru}^j(f)))$ exists if and only if $\mathrm{Radj}(\mathrm{Rco}^i(\mathrm{Ru}^j(s_f)))$, $\mathrm{Radj}(\mathrm{Rco}^i(\mathrm{Ru}^j(t_f)))$ and $\mathrm{Radj}(\mathrm{Ru}^i(\mathrm{Rco}^j(f^\#)))$ exist.

Proof The description of the right adjunctibility of a morphism in $\mathscr{C}^{\rightarrow}$ is [Johnson-Freyd and Scheimbauer 2017, Proposition 7.13], in the oplax case.

For the last statement, remember that $u \doteq \mathrm{Rco}(f^{\#})$ and $v \doteq \mathrm{Ru}(f^{\#})$ by Proposition 2.13.4. The first statement implies by induction that $\mathrm{Rco}^i(\mathrm{Ru}^j(f))$ is of the form (s,t,w), where $s \doteq \mathrm{Rco}^i(\mathrm{Ru}^j(s_f))$, $t \doteq \mathrm{Rco}^i(\mathrm{Ru}^j(t_f))$ and $w \doteq \mathrm{Ru}^i(\mathrm{Rco}^j(f))$.

Indeed increasing j for $i=0$ we have $\mathrm{Ru}^{j+1}(f) = (\mathrm{Ru}(s), \mathrm{Ru}(t), U)$, which have the same dualizability properties as, respectively, $(\mathrm{Ru}^{j+1}(s_f), \mathrm{Ru}^{j+1}(t_f), \mathrm{Rco}^{j+1}(f^{\#}))$. Increasing i we have $\mathrm{Rco}^{i+1}\mathrm{Ru}^j(f) = (\mathrm{Rco}(s), \mathrm{Rco}(t), V)$ which have the same dualizability properties as, respectively, $(\mathrm{Rco}^{i+1}\mathrm{Ru}^j(s_f), \mathrm{Rco}^{i+1}\mathrm{Ru}^j(t_f), \mathrm{Ru}^{i+1}\mathrm{Rco}^j(f^{\#}))$. □

Example 2.16 ($k=0$) An object $f = (X, Y, A\colon X \to Y)$ of $\mathscr{C}^{\rightarrow}$ is dualizable if and only if X and Y are dualizable, and A has a right adjoint $\mathrm{Radj}(A)$. Then

- $f^* = (X^*, Y^*, \mathrm{Radj}(A)^* := (\mathrm{Id}_{Y^*} \otimes \mathrm{ev}_X) \circ (\mathrm{Id}_Y^* \otimes \mathrm{Radj}(A) \otimes \mathrm{Id}_{X^*}) \circ (\mathrm{coev}_Y \otimes \mathrm{Id}_{X^*}))$,
- $\mathrm{coev}_f = (\mathrm{coev}_X, \mathrm{coev}_Y, (\mathrm{Rco}(A) \otimes \mathrm{Id}_{\mathrm{Id}_{Y^*}}) \circ_1 \mathrm{Id}_{\mathrm{coev}_Y})$, and
- $\mathrm{ev}_f = (\mathrm{ev}_X, \mathrm{ev}_Y, \mathrm{Id}_{\mathrm{ev}_X} \circ_1 (\mathrm{Ru}(A) \otimes \mathrm{Id}_{\mathrm{Id}_{X^*}}))$.

A surprising consequence of this result is that if f is 2-dualizable, the right counit and unit of A are biadjoints up to isomorphisms and mating. A drawing for this is given in Figure 4.

2.3 Cobordism hypotheses

The cobordism hypothesis describes fully extended topological quantum field theories with values in a higher category $\mathscr{C}$ in terms of fully dualizable objects of $\mathscr{C}$. We also recall relative versions that describes relative TQFTs, and a noncompact version that describes partially defined TQFTs. The cobordism hypothesis was formulated in [Baez and Dolan 1995]. A sketch of proof was given in [Lurie 2009], a more formal version is work in progress of Schommer-Pries. An independent proof of a more general result appears in the preprint [Grady and Pavlov 2021]. Another independent proof using factorization homology is work in progress, see [Ayala and Francis 2017].

Conjecture 2.17 (the cobordism hypothesis, Theorems 2.4.6 and 2.4.26 in [Lurie 2009]) *Let $\mathscr{C}$ be a symmetric monoidal n-category. Evaluation at the point induces equivalences of ∞-groupoids*

$$\mathrm{Fun}^{\otimes}(\mathbf{Bord}_n^{\mathrm{fr}}, \mathscr{C}) \simeq (\mathscr{C}^{\mathrm{fd}})^{\sim}$$

between framed fully extended n-TQFTs with values in $\mathscr{C}$ and the underlying ∞-groupoid of the subcategory of fully dualizable objects of $\mathscr{C}$, and

$$\mathrm{Fun}^{\otimes}(\mathbf{Bord}_n, \mathscr{C}) \simeq ((\mathscr{C}^{\mathrm{fd}})^{\sim})^{\mathrm{SO}(n)}$$

between oriented fully extended n-TQFTs with values in $\mathscr{C}$ and $\mathrm{SO}(n)$-homotopy-fixed-points in $(\mathscr{C}^{\mathrm{fd}})^{\sim}$, where $\mathrm{SO}(n)$ acts on the n-category $\mathbf{Bord}_n^{\mathrm{fr}}$ by changing the framing and therefore on $(\mathscr{C}^{\mathrm{fd}})^{\sim}$ by the first equivalence.

For $X \in \mathcal{C}$ a fully dualizable object, we denote by $\mathcal{Z}_X$ (a choice of representative of) the associated fully extended framed n-TQFT.

2.3.1 The relative cobordism hypothesis Lurie proposes a result classifying his notion of domain wall.

Conjecture 2.18 [Lurie 2009, Theorem 4.3.11 and Example 4.3.23] *Let $\mathcal{C}$ be a symmetric monoidal n-category and $X, Y \in \mathcal{C}^{\mathrm{fd}}$. There is a bijection between isomorphism classes of framed domain walls between $\mathcal{Z}_X$ and $\mathcal{Z}_Y$ and isomorphism classes of fully dualizable 1-morphisms $f : X \to Y$, given by evaluation at the interval with an interface point in the middle.*

There is an oriented version asking that f preserve orientation structures.

On the other hand, the notions in [Johnson-Freyd and Scheimbauer 2017] of a twisted quantum field theory are already classified by the usual cobordism hypothesis. Note however that [loc. cit., Definition 5.16] is surprisingly strict because it demands that the source and target of the functor $\mathcal{R} : \mathbf{Bord}_{n-1} \to \mathcal{C}$ agree strictly with $\mathcal{Z}_1$ and $\mathcal{Z}_2$. Equivalently, we could have asked that $\mathcal{R}$ come equipped with isomorphisms $s(\mathcal{R}) \simeq \mathcal{Z}_1$ and $t(\mathcal{R}) \simeq \mathcal{Z}_2$. In both cases, it is clear that the cobordism hypothesis does not apply on the nose. The fix is easy.

Definition 2.19 Let $\mathcal{C}$ be a symmetric monoidal n-category and $X, Y \in \mathcal{C}$. Denote by $(\mathcal{C}^{\to})^{\sim}_{X,Y}$ the homotopy pullback

$$\begin{array}{ccc} (\mathcal{C}^{\to})^{\sim}_{X,Y} & \longrightarrow & (\mathcal{C}^{\to})^{\sim} \\ \downarrow & \lrcorner h & \downarrow{\scriptstyle s,t} \\ * & \xrightarrow{X,Y} & (\mathcal{C}^{\sim})^{\times 2} \end{array}$$

Similarly, for $\mathcal{Z}_1, \mathcal{Z}_2 : \mathbf{Bord}^{\mathrm{fr}}_{n-1} \to \mathcal{C}$ denote by $\mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathcal{C}^{\to})_{\mathcal{Z}_1,\mathcal{Z}_2}$ the homotopy pullback

$$\begin{array}{ccc} \mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathcal{C}^{\to})_{\mathcal{Z}_1,\mathcal{Z}_2} & \longrightarrow & \mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathcal{C}^{\to}) \\ \downarrow & \lrcorner h & \downarrow{\scriptstyle s,t} \\ * & \xrightarrow{\mathcal{Z}_1,\mathcal{Z}_2} & (\mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathcal{C}))^{\times 2} \end{array}$$

called the space of framed oplax-$\mathcal{Z}_1$-$\mathcal{Z}_2$-twisted-$(n-1)$-TQFTs.

Note that both are also strict pullbacks, as taking source and target induces a fibration of spaces.

Corollary 2.20 (of the cobordism hypothesis) *Let $\mathcal{C}$ be a symmetric monoidal n-category and $X, Y \in \mathcal{C}$. Choose two TQFTs $\mathcal{Z}_X, \mathcal{Z}_Y : \mathbf{Bord}^{\mathrm{fr}}_{n-1} \to \mathcal{C}$ associated with X and Y by the cobordism hypothesis. Evaluation at the point induces an equivalence*

$$\mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathcal{C}^{\to})_{\mathcal{Z}_X,\mathcal{Z}_Y} \simeq (\mathcal{C}^{\to})^{\sim}_{X,Y}.$$

Proof The cobordism hypothesis on $\mathscr{C}$ and $\mathscr{C}^{\to}$ gives a commutative diagram of horizontal equivalences

$$\begin{array}{ccc}
\mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathscr{C}^{\to}) & \xrightarrow{\mathrm{ev}_{\mathrm{pt}}} & (\mathscr{C}^{\to})^{\sim} \\
\downarrow{\scriptstyle s,t} & & \downarrow{\scriptstyle s,t} \\
(\mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathscr{C}))^{\times 2} & \xrightarrow{\mathrm{ev}_{\mathrm{pt}} \times \mathrm{ev}_{\mathrm{pt}}} & (\mathscr{C}^{\sim})^{\times 2} \\
{\scriptstyle \mathscr{Z}_X, \mathscr{Z}_Y}\uparrow & & {\scriptstyle X,Y}\uparrow \\
* & \longrightarrow & *
\end{array}$$

inducing an equivalence between homotopy pullbacks. □

Remark 2.21 There is an oriented version well. The maps

$$s, t\colon \mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathscr{C}^{\to}) \to \mathrm{Fun}^{\otimes}(\mathbf{Bord}^{\mathrm{fr}}_{n-1}, \mathscr{C})$$

are $\mathrm{SO}(n-1)$-equivariant because $\mathrm{SO}(n-1)$ acts on the source $\mathbf{Bord}^{\mathrm{fr}}_{n-1}$. The maps $s, t\colon (\mathscr{C}^{\to})^{\sim} \to \mathscr{C}^{\sim}$ are therefore also equivariant, and descend to maps between the $\mathrm{SO}(n-1)$-homotopy-fixed points $s, t\colon (\mathscr{C}^{\to})^{\sim, \mathrm{SO}(n-1)} \to \mathscr{C}^{\sim, \mathrm{SO}(n-1)}$. Given two objects $X, Y \in \mathscr{C}$ equipped with $\mathrm{SO}(n-1)$-homotopy-fixed point structure, one can reproduce exactly the whole paragraph above and define $(\mathscr{C}^{\to})^{\sim, \mathrm{SO}(n-1)}_{X,Y}$ as a pullback. We get

$$\mathrm{Fun}^{\otimes}(\mathbf{Bord}_{n-1}, \mathscr{C}^{\to})_{\mathscr{Z}_X, \mathscr{Z}_Y} \simeq (\mathscr{C}^{\to})^{\sim, \mathrm{SO}(n-1)}_{X,Y}$$

by the same proof, using the oriented cobordism hypothesis.

Remark 2.22 Results to appear in the PhD thesis of William Stewart show that if we assume that the source and target objects X and Y are fully dualizable then a morphism $f\colon X \to Y$ is $(n-1)$-oplax dualizable if and only if it is $(n-1)$-dualizable. In particular, if we restrict the notion of oplax twisted TQFTs to the case where the "twisting" theories $\mathscr{Z}_1$ and $\mathscr{Z}_2$ extend to $\mathbf{Bord}_n$, which is the setting in [Freed and Teleman 2014], then this notion, using the cobordism hypothesis twice, is equivalent to Lurie's notion of domain walls.

2.3.2 Noncompact TQFTs To study nonsemisimple variants of Witten–Reshetikhin–Turaev TQFTs, we will be interested in theories defined on a restricted class of cobordisms, namely where top-dimensional cobordisms have nonempty outgoing boundary in every connected component.

Lurie's sketch of proof of the cobordism hypothesis is done by induction on the handle indices allowed. One starts with only opening balls, then allows more and more complex cobordisms. Eventually one allows every cobordism but closing balls, namely cobordisms with outgoing boundary in every connected component. Finally one allows every cobordism, obtaining a TQFT. We call it a noncompact TQFT when we stop at this ante-last step. Lurie's proof then gives an algebraic criterion classifying these.

We follow [Lurie 2009, Section 3.4] and state the results there in a form fitted for our use. It should be noted that the proofs of the statements below are not very formal.

Definition 2.23 Let $\mathbf{Bord}_n^{\mathrm{fr,nc}} \subseteq \mathbf{Bord}_n^{\mathrm{fr}}$ denote the subcategory where n-dimensional bordisms have nonempty outgoing boundary in every connected component.

A *framed fully extended noncompact n-TQFT* with values in a symmetric monoidal n-category $\mathscr{C}$ is a symmetric monoidal functor $\mathscr{Z}\colon \mathbf{Bord}_n^{\mathrm{fr,nc}} \to \mathscr{C}$.

Lurie [2009, Definition 3.4.9] defines an n-category $\mathscr{F}_k$ of $\leq n$-dimensional bordisms where all n-manifolds are equipped with a decomposition into handles of index $\leq k$. Here bordisms are actually equipped with a framed function without certain kinds of critical points.

We denote by $\alpha_k^m = D^k \times D^{m-k}\colon S^{k-1} \times D^{m-k} \to D^k \times S^{m-k-1}$ the m-dimensional index k handle attachment, seen as an m-morphism in $\mathbf{Bord}_m^{\mathrm{fr}}$, or in $\mathscr{F}_k$ if $m = n$. Let $x = S^{k-2} \times D^{n-k}$ and $y = D^{k-1} \times S^{n-k-1}$ be seen as $(n-2)$-morphisms $\varnothing \to S^{k-2} \times S^{n-k-1}$ in $\mathbf{Bord}_{n-1}^{\mathrm{fr}}$. Note that $\alpha_{k-1}^{n-1}\colon x \to y$ and $\alpha_{n-k}^{n-1}\colon y \to x$ for $1 \leq k \leq n$. Then, α_{k-1}^n can be seen (up to higher mating) as a morphism $\mathrm{Id}_x \to \alpha_{n-k}^{n-1} \circ \alpha_{k-1}^{n-1}$ and α_k^n as a morphism $\alpha_{k-1}^{n-1} \circ \alpha_{n-k}^{n-1} \to \mathrm{Id}_y$, and they form a unit/counit pair in $\mathscr{F}_k$; see [Lurie 2009, Claim 3.4.17]. Namely, $\mathrm{Radj}(\alpha_{k-1}^{n-1}) = \alpha_{n-k}^{n-1}$, $\mathrm{Ru}(\alpha_{k-1}^{n-1}) = \alpha_{k-1}^n$ and $\mathrm{Rco}(\alpha_{k-1}^{n-1}) = \alpha_k^n$, or, in our diagrammatic notation,

$$\begin{array}{ccc} \alpha_{k-1}^{n-1} & & \\ & \searrow & \alpha_{n-k}^{n-1} \\ & & \alpha_k^n \mid \alpha_{k-1}^n \end{array}$$

By induction, $\mathrm{Rco}^k(\mathrm{Ru}^{m-k}(\mathrm{pt})) = \alpha_k^m$.

Conjecture 2.24 (index-k cobordism hypothesis, Lurie) *A symmetric monoidal functor*

$$\mathscr{Z}_0\colon \mathbf{Bord}_{n-1}^{\mathrm{fr}} \to \mathscr{C}$$

extends to $\mathscr{Z}\colon \mathscr{F}_k \to \mathscr{C}$ for $1 \leq k \leq n$ if and only if the images of every $(n-1)$-dimensional handle of index $\leq k-1$ is right adjunctible.

This extension is essentially unique: there is an isomorphism $\mathscr{Z} \Rightarrow \mathscr{Z}'$ between any two such extensions. This isomorphism may not be the identity on $\mathbf{Bord}_{n-1}^{\mathrm{fr}}$.

Sketch For $k = 0$, one can extend $\mathscr{Z}_0\colon \mathbf{Bord}_{n-1}^{\mathrm{fr}} \to \mathscr{C}$ with any n-morphism $\mathscr{Z}(\alpha_0^n)\colon 1 \to \mathscr{Z}_0(S^{n-1})$; see [Lurie 2009, Claim 3.4.13]. Note that Lurie works in the unoriented case there, and demands an $O(n)$-equivariant morphism, and we look at the framed case.

Now, for $1 \leq k \leq n$, a symmetric monoidal functor $\mathscr{Z}_0\colon \mathscr{F}_{k-1} \to \mathscr{C}$ extends to $\mathscr{Z}\colon \mathscr{F}_k \to \mathscr{C}$ if and only if α_{k-1}^n is mapped to a unit of an adjunction between α_{k-1}^{n-1} and α_{n-k}^{n-1}; see [loc. cit., Proposition 3.4.19]. In this case, the extension is essentially unique, and α_k^n is mapped to the counit of the adjunction.

For $k = 1$, this gives little choice for the n-morphism $\mathscr{Z}(\alpha_0^n)$: it has to be the unit of an adjunction and is therefore determined up to isomorphism. Then, α_1^n will be sent to the counit.

For $k \geq 2$, we want $\mathfrak{X}(\alpha^n_{k-1})$, which is so far defined as the counit of the adjunction between $\mathfrak{X}(\alpha^{n-1}_{k-2})$ and $\mathfrak{X}(\alpha^{n-1}_{n-k+1})$, to be also the unit of the adjunction between $\mathfrak{X}(\alpha^{n-1}_{k-1})$ and $\mathfrak{X}(\alpha^{n-1}_{n-k})$. This in particular implies that the $(n-1)$-dimensional handle of index $k-1$ is right adjunctible, as stated in the conjecture. For the converse, we use [Lurie 2009, Proposition 3.4.20] (which we recalled in Proposition 2.14(2)), which states that provided the adjunction exists, α^n_{k-1} must map to (a higher mate of) the unit. □

Definition 2.25 Let $\mathscr{C}$ be a symmetric monoidal n-category. An object X in $\mathscr{C}$ is (n,k)*-dualizable* if it is $(n-1)$-dualizable and the $(n-1)$-morphisms $\mathrm{Ru}^{n-1}(X), \mathrm{Rco}(\mathrm{Ru}^{n-2}(X)), \dots, \mathrm{Rco}^{k-1}(\mathrm{Ru}^{n-k}(X))$ have right adjoints. We say X is *noncompact-n-dualizable* if it is $(n, n-1)$-dualizable.

For example, for $n = 3$ and $k = 2$, we want X to have a dual $(X^*, \mathrm{ev}_X, \mathrm{coev}_X)$, both its evaluation and coevaluation maps to have right adjoints (ev^R_X, a, b) and $(\mathrm{coev}^R_X, c, d)$, and the unit and counit of the right adjunction of the coevaluation to have right adjoints c^R and d^R.

We can now state the noncompact version of the cobordism hypothesis, which we will assume in Section 4. A formal proof is work-in-progress of Schommer-Pries.

Conjecture 2.26 (noncompact cobordism hypothesis) *Let $\mathscr{C}$ be a symmetric monoidal n-category, with $n \geq 2$. There is a bijection between isomorphism classes of framed fully extended noncompact n-TQFTs with values in $\mathscr{C}$ and isomorphism classes of noncompact-n-dualizable objects of $\mathscr{C}$, given by evaluation at the point.*

There is an oriented version as well, stating that oriented noncompact theories are classified by $\mathrm{SO}(n)$-homotopy-fixed points in the space of noncompact-n-dualizable objects.

3 Dualizability of the unit inclusion

Let $\mathscr{V} \in \mathbf{BrTens}$ be a braided tensor category. We consider the inclusion of the unit $\eta\colon \mathrm{Vect}_{\Bbbk} \to \mathscr{V}$. It is a braided monoidal functor and we define an associated $\mathrm{Vect}_{\Bbbk}$-$\mathscr{V}$-central algebra $\mathscr{A}_\eta$, which is simply the category $\mathscr{V}$ seen as the regular right $\mathscr{V}$-module; see Definition 3.8. We study the dualizability of this 1-morphism in $\mathbf{BrTens}$. First, we recall some context and develop some properties of bimodules induced by functors. Then we describe all the dualizability data explicitly and give criteria for dualizability.

3.1 Cocomplete braided tensor categories

We will work in the even higher Morita category $\mathrm{Alg}_2(\mathbf{Pr})$. This category have been formally defined in [Johnson-Freyd and Scheimbauer 2017], and described more explicitly in [Brochier et al. 2021b] under the name $\mathbf{BrTens}$.

3.1.1 Cocomplete categories We begin by recalling some properties of the 2-category $\mathbf{Pr}$. Let $\Bbbk$ be a field of characteristic zero.

Let $\mathrm{Cat}_{\Bbbk}$ denote the 2-category of small $\Bbbk$-linear categories, and $\mathbf{Pr}$ denote the 2-category of cocomplete locally presentable $\Bbbk$-linear categories [Adámek and Rosický 1994, Defintion 1.17], $\Bbbk$-linear cocontinuous functors and $\Bbbk$-linear natural transformations, equipped with the Kelly tensor product $\boxtimes$. We denote by $\mathrm{Free} = \mathrm{Hom}_{\mathrm{Cat}_{\Bbbk}}((-)^{\mathrm{op}}, \mathrm{Vect}_{\Bbbk}) \colon \mathrm{Cat}_{\Bbbk} \to \mathbf{Pr}$ the symmetric monoidal free cocompletion functor. Its essential image is denoted by $\mathrm{Bimod}_{\Bbbk}$.[2]

An object $C \in \mathscr{C}$ is called compact-projective (which we abbreviate cp) if the functor $\mathrm{Hom}_{\mathscr{C}}(C, -)$ is cocontinuous. The category $\mathscr{C}$ is said to have enough compact-projectives if its full subcategory $\mathscr{C}^{\mathrm{cp}}$ of cp objects generates $\mathscr{C}$ under colimits, or equivalently if the canonical functor $\mathrm{Free}(\mathscr{C}^{\mathrm{cp}}) \to \mathscr{C}$ is an equivalence, or if it lies in $\mathrm{Bimod}_{\Bbbk}$. A monoidal category $\mathscr{C}$ is called cp-rigid if it has enough cp and all its cp objects are left and right dualizable.

Proposition 3.1 *A 1-morphism $F \colon \mathscr{C} \to \mathscr{D}$ in $\mathbf{Pr}$ between two categories with enough cp has a cocontinuous right adjoint if and only if it preserves cp.*

Proof If F^R is cocontinuous, then for $C \in \mathscr{C}^{\mathrm{cp}}$ and $D = \mathrm{colim}_i\, D_i$ obtained as a colimit,

$$\begin{aligned}\mathrm{Hom}_{\mathscr{D}}(F(C), D) &\simeq \mathrm{Hom}_{\mathscr{C}}(C, F^R(D)) \\ &\simeq \mathrm{Hom}_{\mathscr{C}}(C, \mathrm{colim}_i\, F^R(D_i)) \quad \text{since } F^R \text{ is cocontinuous} \\ &\simeq \mathrm{colim}_i\, \mathrm{Hom}_{\mathscr{C}}(C, F^R(D_i)) \quad \text{since } C \text{ is cp} \\ &\simeq \mathrm{colim}_i\, \mathrm{Hom}_{\mathscr{D}}(F(C), D_i),\end{aligned}$$

and $F(C)$ is compact-projective.

The other direction is a classical construction; see [Bartlett et al. 2015, Lemma 2.10]. □

Remark 3.2 The condition of F preserving cp, namely of F being the cocontinuous extension of a functor f on the subcategories of cp objects, is very similar to that of a bimodule being induced by a functor in Section 3.2. When F preserves cp then F^R is associated with the "mirrored" bimodule induced by f, with unit induced by f, and counit induced by composition in $\mathscr{D}$. This is again very similar to what happens in Section 3.2.

3.1.2 BrTens and Alg₂(Pr) The higher Morita n-category $\mathrm{Alg}_n(\mathscr{S})$ associated with an ∞-category $\mathscr{S}$ was introduced in [Haugseng 2017] using a combinatorial/operadic description. A pointed version was introduced in [Scheimbauer 2014] using very geometric means, namely factorization algebras. This geometric description allows for a good description of dualizability in $\mathrm{Alg}_n(\mathscr{S})$, but the pointing prevents any higher dualizability; see [Gwilliam and Scheimbauer 2018]. Even higher Morita categories are defined in [Johnson-Freyd and Scheimbauer 2017], for pointed and unpointed versions. They form an

[2] The name comes from the Eilenberg–Watts theorem, which describes cocontinuous functors between categories of modules over two algebras as bimodules.

$(n+k)$-category $\mathrm{ALG}_n(\mathcal{S})$ for $\mathcal{S}$ a k-category. We consider the unpointed even higher Morita 4-category $\mathrm{ALG}_2(\mathbf{PR})$, which we denote by $\mathbf{BRTENS}$ for reasons that will be made explicit below. Even though we are not formally in this context, we will sometimes use factorization algebra drawings to illustrate our point.

One represents an $\mathbb{E}_2$-algebra $\mathcal{V}$ as  and the $\mathrm{Vect}_{\Bbbk}$–$\mathcal{V}$-algebra $\mathcal{A}_\eta$ as . Let us recall the description of $\mathbf{BRTENS}$ from [Brochier et al. 2021b].

Definition 3.3 [Brochier et al. 2021b, Section 2.4] An object $\mathcal{V}$ of $\mathbf{BRTENS}$ is a locally presentable cocomplete $\Bbbk$-linear braided monoidal category. We call these *braided tensor categories* here, even though this name has many uses. Equivalently, it is an $\mathbb{E}_2$-algebra in $\mathbf{PR}$.

In the factorization algebra picture , one can read that a 1-morphism between two braided tensor categories $\mathcal{V}$ (in red, on top) and $\mathcal{W}$ (in blue, below) is a monoidal category $\mathcal{A} \in \mathbf{PR}$ (the horizontal line) with a top $\mathcal{V}$-action and a bottom $\mathcal{W}$-action that commute with respect to each other and that commute with the monoidal structure of $\mathcal{A}$ in a coherent way. Note that as $\mathcal{A}$ is monoidal, such a $\mathcal{V}$-action $\triangleright$ is determined by a monoidal functor $\mathcal{V} \to \mathcal{A}$ that maps V to $V \triangleright \mathbf{1}_{\mathcal{A}}$. See [Brochier et al. 2021b, Figure 2].

Definition 3.4 [Brochier et al. 2021b, Definition–Proposition 3.2] A 1-morphism between $\mathcal{V}$ and $\mathcal{W}$ in $\mathbf{BRTENS}$ is a *$\mathcal{V}$-$\mathcal{W}$-central algebra* $\mathcal{A}$. Namely, it is an monoidal category $\mathcal{A} \in \mathbf{PR}$ equipped with a braided monoidal functor

$$(F_{\mathcal{A}}, \sigma^{\mathcal{A}}) \colon \mathcal{V} \boxtimes \mathcal{W}^{\sigma\,\mathrm{op}} \to Z(\mathcal{A})$$

to the Drinfeld center of $\mathcal{A}$.

Recall that the Drinfeld center of $\mathcal{A}$ has objects pairs (y, β), where y is an object of $\mathcal{A}$ and $\beta \colon - \otimes y \xRightarrow{\sim} y \otimes -$ is a natural isomorphism. Here $F_{\mathcal{A}}$ gives the object and $\sigma^{\mathcal{A}}$ gives the half braiding. We write $V \triangleright A := F_{\mathcal{A}}(V) \otimes A$ and $A \triangleleft V := A \otimes F_{\mathcal{A}}(V)$.

Composition of 1-morphism is relative tensor product over the corresponding braided tensor category, see [Brochier et al. 2021b, Section 3.4].

Again in the factorization algebra picture , a 2-morphism $\mathcal{M}$ between two $\mathcal{V}$-$\mathcal{W}$-central algebras $\mathcal{A}$ and $\mathcal{B}$ is a $\mathcal{A}$-$\mathcal{B}$-bimodule category where $\mathcal{V}$ (resp. $\mathcal{W}$) acts similarly when acting through $\mathcal{A}$ or through $\mathcal{B}$.

Definition 3.5 [Brochier et al. 2021b, Definition 3.9] A 2-morphism between $\mathcal{A}$ and $\mathcal{B}$ in $\mathbf{BRTENS}$ is a *$\mathcal{V}$-$\mathcal{W}$-centered $\mathcal{A}$-$\mathcal{B}$-bimodule category*. Namely, it is an $\mathcal{A}$-$\mathcal{B}$-bimodule category $\mathcal{M}$ equipped with natural isomorphisms

$$\eta_{v,m} \colon F_{\mathcal{A}}(v) \triangleright m \xrightarrow{\sim} m \triangleleft F_{\mathcal{B}}(v) \quad \text{for } v \in \mathcal{V} \text{ and } m \in \mathcal{M},$$

satisfying coherences with the tensor product in $\mathcal{V}$ and with the half braidings in $\mathcal{A}$ and $\mathcal{B}$.

Horizontal and vertical composition are again relative tensor product over the corresponding monoidal category.

Definition 3.6 [Brochier et al. 2021b, Section 3.6] A 3-morphism $F\colon \mathcal{M} \to \mathcal{N}$ is a functor of $\mathcal{A}$-$\mathcal{B}$-bimodules categories that preserves the $\mathcal{V}$-$\mathcal{W}$-centered structure.

A 4-morphism $\eta\colon F \Rightarrow G$ is a natural transformation of bimodule functors.

3.1.3 Previous dualizability results We will define in Definition 3.8 the 1-morphism $\mathcal{A}_\eta\colon \mathrm{Vect}_{\Bbbk} \to \mathcal{V}$ induced by the unit inclusion in a braided tensor category $\mathcal{V}$. It is $\mathcal{V}$ as a monoidal category with obvious actions. Let us recall previously known results about its dualizability. The following is [Gwilliam and Scheimbauer 2018, Theorem 5.1], [Brochier et al. 2021b, Theorem 5.16] and [Brochier et al. 2021b, Theorem 5.21], respectively.

Theorem 3.7 *The* 1*-morphism* $\mathcal{A}_\eta$ *is always* 1*-dualizable. It is* 2*-dualizable as soon as* $\mathcal{V}$ *is cp-rigid, and* 3*-dualizable as soon as* $\mathcal{V}$ *is fusion.*

Note that the fusion requirement can easily be relaxed to rigid finite semisimple, without the assumption that the unit is simple; see the proof of Theorem 3.22.

3.2 Bimodules induced by functors

We give basic definitions and facts about bimodules induced by (braided) monoidal functors, and show how to compute their adjoints.

3.2.1 Definition and coherence We show that the notion of bimodules induced by functors behaves as expected in **BrTens**. Namely, the Morita category, whose morphisms are bimodules, extends the category whose morphisms are functors.

Definition 3.8 Let $\mathcal{A}$ and $\mathcal{B}$ be two objects of **BrTens**. A braided monoidal functor $F\colon \mathcal{A} \to \mathcal{B}$ induces an $\mathcal{A}$-$\mathcal{B}$-central algebra $\mathcal{A}_F$, which is given by $\mathcal{B}$ as a monoidal category on which $\mathcal{A}$ acts on the top using $F(-)\otimes -$ and $\mathcal{B}$ acts on the bottom using $-\otimes -$. More formally its structure of $\mathcal{A}$-$\mathcal{B}$-central algebra is given by

$$\mathcal{A} \boxtimes \mathcal{B}^{\sigma\,\mathrm{op}} \to Z(\mathcal{B}), \quad (A, B) \mapsto (F(A) \otimes B, (\mathrm{Id}_{F(A)} \otimes \sigma^{-1}_{B,-}) \circ (\sigma_{-,F(A)} \otimes \mathrm{Id}_B)),$$

where σ is the braiding in $\mathcal{B}$. It is braided monoidal because F is braided monoidal.

It also induces a $\mathcal{B}$-$\mathcal{A}$-central algebra $\overline{\mathcal{A}}_F$ which is also given by $\mathcal{B}$ as a monoidal category on which $\mathcal{A}$ acts on the bottom using $-\otimes F(-)$ and $\mathcal{B}$ acts on the top using $-\otimes -$.

When the functor F is understood, we may write ${}_{\mathcal{A}}\mathcal{B}_{\mathcal{B}}$ for $\mathcal{A}_F$ and ${}_{\mathcal{B}}\mathcal{B}_{\mathcal{A}}$ for $\overline{\mathcal{A}}_F$.

Proposition 3.9 *The above induced-central-algebra construction preserves composition. Given two braided monoidal functors* $F\colon \mathcal{A} \to \mathcal{B}$ *and* $G\colon \mathcal{B} \to \mathcal{C}$*, one has* $\mathcal{A}_G \circ \mathcal{A}_F \simeq \mathcal{A}_{G\circ F}$ *and* $\overline{\mathcal{A}}_F \circ \overline{\mathcal{A}}_G \simeq \overline{\mathcal{A}}_{G\circ F}$*.*

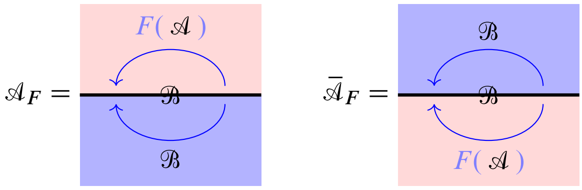


Figure 3: The 1-morphisms $\mathcal{A}_F$ and $\bar{\mathcal{A}}_F$.

Proof We want to prove that ${}_{\mathcal{A}}\mathcal{B}_{\mathcal{B}} \underset{\mathcal{B}}{\boxtimes} {}_{\mathcal{B}}\mathcal{C}_{\mathcal{C}} \simeq {}_{\mathcal{A}}\mathcal{C}_{\mathcal{C}}$. This is true on the underlying categories as $\mathcal{B} \underset{\mathcal{B}}{\boxtimes} \mathcal{C} \overset{\Phi}{\simeq} \mathcal{C}$ with equivalence given on pure tensors by $\Phi(B \boxtimes C) = G(B) \otimes C$. This assignment is balanced as G is monoidal:

$$\Phi((B \otimes B') \boxtimes C) = G(B \otimes B') \otimes C \simeq G(B) \otimes G(B') \otimes C = \Phi(B \boxtimes (G(B') \otimes C)).$$

It is monoidal (the monoidal structure on the relative tensor product is described in [Brochier et al. 2021b, Definition–Proposition 3.6]) by

$$\Phi(B\boxtimes C)\otimes\Phi(B'\boxtimes C') = G(B)\otimes C\otimes G(B')\otimes C' \xrightarrow[\sim]{\sigma_{C,B'}} G(B)\otimes G(B')\otimes C\otimes C' \simeq \Phi((B\boxtimes C)\otimes(B'\boxtimes C')).$$

The bottom action of $\mathcal{C}$ is unchanged, and the top action of $\mathcal{A}$ is preserved by Φ:

$$A \rhd (B \boxtimes C) := (A \rhd \mathbb{1}) \otimes (B \boxtimes C) = (F(A) \otimes B) \boxtimes C \overset{\Phi}{\mapsto} G(F(A)) \otimes G(B) \otimes C = A \rhd \Phi(B \boxtimes C).$$

Finally, let us show that Φ preserves the central structure. The central structure in the composed bimodule $\mathcal{A}_F \underset{\mathcal{B}}{\boxtimes} \mathcal{A}_G$ is given by

$$(B\boxtimes C) \lhd A := (B\boxtimes C)\otimes(F(A)\boxtimes 1_{\mathcal{C}}) = (B\otimes F(A))\boxtimes C \xrightarrow[\sim]{\sigma^{\mathcal{B}}_{B,F(A)}\boxtimes \mathrm{Id}_C} (F(A)\otimes B)\boxtimes C = A \rhd (B\boxtimes C).$$

This maps under Φ, using that G is braided monoidal, to $\sigma^{\mathcal{C}}_{G(B),G(F(A))} \otimes \mathrm{Id}_C$. And indeed, the following diagram, where the horizontal arrows are the central structures and the vertical arrow monoidality of Φ, commutes:

$$\begin{array}{ccc}
\Phi(B \boxtimes C) \otimes \Phi(\mathbb{1} \lhd A) & \xrightarrow{\sigma^{\mathcal{C}}_{(G(B)\otimes C),G(F(A))}} & \Phi(A \rhd \mathbb{1}) \otimes \Phi(B \boxtimes C) \\
{\scriptstyle \mathrm{Id}_{G(B)} \otimes \sigma^{\mathcal{C}}_{C,G(F(A))}} \downarrow & & \downarrow {\scriptstyle \mathrm{Id}} \\
\Phi((B \boxtimes C) \otimes (\mathbb{1} \lhd A)) & \xrightarrow{\sigma^{\mathcal{C}}_{G(B),G(F(A))} \otimes \mathrm{Id}_C} & \Phi((A \rhd \mathbb{1}) \otimes (B \boxtimes C))
\end{array}$$

The $\bar{\mathcal{A}}$ case is similar. □

Definition 3.10 Let $\mathcal{C}$ and $\mathcal{D}$ be $\mathcal{A}$-$\mathcal{B}$-central algebras, ie 1-morphisms of **BrTens**. A bimodule monoidal functor $F\colon \mathcal{C} \to \mathcal{D}$ preserving the $\mathcal{A}$-$\mathcal{B}$-central structures induces an $\mathcal{A}$-$\mathcal{B}$-centered $\mathcal{C}$-$\mathcal{D}$-bimodule $\mathcal{M}_F$, which is given by $\mathcal{D}$ as a category on which $\mathcal{C}$ acts on the left using $F(-) \otimes -$ and $\mathcal{D}$ act on the right using $-\otimes-$. The $\mathcal{A}$-$\mathcal{B}$-centered structure on $\mathcal{M}_F$ is induced by the $\mathcal{A}$-$\mathcal{B}$-central structure of $\mathcal{D}$, and the fact that F is a bimodule functor:

$$F(A \triangleright \mathbb{1}_{\mathcal{C}} \triangleleft B) \otimes M \simeq (A \triangleright 1_{\mathcal{D}} \triangleleft B) \otimes M \overset{\sigma^{\mathcal{D}}}{\simeq} M \otimes (A \triangleright 1_{\mathcal{D}} \triangleleft B).$$

It also induces an $\mathcal{A}$-$\mathcal{B}$-centered $\mathcal{D}$-$\mathcal{C}$-bimodule $\overline{\mathcal{M}}_F$ which is again given by $\mathcal{D}$ as a monoidal category on which $\mathcal{C}$ acts on the right using $-\otimes F(-)$ and $\mathcal{D}$ act on the left using $-\otimes-$.

When the functor F is understood, we may write ${}_{\mathcal{C}}\mathcal{D}_{\mathcal{D}}$ for $\mathcal{M}_F$ and ${}_{\mathcal{D}}\mathcal{D}_{\mathcal{C}}$ for $\overline{\mathcal{M}}_F$

Proposition 3.11 *The above induced-bimodule construction preserves*:

(1) ***Horizontal composition*** *Given two $\mathcal{A}$-$\mathcal{B}$-bimodule monoidal functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{E}$ preserving central structures, one has $\mathcal{M}_G \circ \mathcal{M}_F \simeq \mathcal{M}_{G\circ F}$ and $\overline{\mathcal{M}}_F \circ \overline{\mathcal{M}}_G \simeq \overline{\mathcal{M}}_{G\circ F}$.*

(2) ***Vertical composition*** *Given $\mathcal{C}$ and $\mathcal{D}$ two $\mathcal{A}_1$-$\mathcal{A}_2$-central algebras, $\mathcal{C}'$ and $\mathcal{D}'$ two $\mathcal{A}_2$-$\mathcal{A}_3$ central algebras, $F\colon \mathcal{C} \to \mathcal{D}$ an $\mathcal{A}_1$-$\mathcal{A}_2$-bimodule monoidal functor and $F'\colon \mathcal{C}' \to \mathcal{D}'$ an $\mathcal{A}_2$-$\mathcal{A}_3$-bimodule monoidal functor preserving central structures, one has*

$$\mathcal{M}_F \underset{\mathcal{A}_2}{\boxtimes} \mathcal{M}_{F'} \simeq \mathcal{M}_{F \underset{\mathcal{A}_2}{\boxtimes} F'} \quad \text{and} \quad \overline{\mathcal{M}}_F \underset{\mathcal{A}_2}{\boxtimes} \overline{\mathcal{M}}_{F'} \simeq \overline{\mathcal{M}}_{F \underset{\mathcal{A}_2}{\boxtimes} F'}.$$

Proof The first point is similar to the last proposition. We proved that ${}_{\mathcal{C}}\mathcal{D}_{\mathcal{D}} \underset{\mathcal{D}}{\boxtimes} {}_{\mathcal{D}}\mathcal{E}_{\mathcal{E}} \overset{\Phi}{\simeq} {}_{\mathcal{C}}\mathcal{E}_{\mathcal{E}}$, as bimodules. Recall from [Brochier et al. 2021b, Definition–Proposition 3.13] that the centered structure on the composition of bimodules $\mathcal{D} \underset{\mathcal{D}}{\boxtimes} \mathcal{E}$ is given by the composition of the centered structure and a balancing. In our case on some A, D and E, this is

$$D \boxtimes (E \otimes A) \xrightarrow[\sim]{\mathrm{Id}_D \boxtimes \sigma^{\mathcal{E}}_{E,A}} D \boxtimes (A \otimes E) \simeq (D \otimes A) \boxtimes E \xrightarrow[\sim]{\sigma^{\mathcal{D}}_{D,A} \boxtimes \mathrm{Id}_E} (A \otimes D) \boxtimes E,$$

which maps by Φ to $(G(\sigma^{\mathcal{D}}_{D,A}) \otimes \mathrm{Id}_E) \circ (\mathrm{Id}_{G(D)} \otimes \sigma^{\mathcal{E}}_{E,A})$. The centered structure of ${}_{\mathcal{C}}\mathcal{E}_{\mathcal{E}}$ is given by $\sigma^{\mathcal{E}}_{G(D)\otimes E,A}$. They coincide as G preserves central structures.

The second point is not surprising either. We want

$${}_{\mathcal{C}}\mathcal{D}_{\mathcal{D}} \underset{\mathcal{A}_2}{\boxtimes} {}_{\mathcal{C}'}\mathcal{D}'_{\mathcal{D}'} \simeq {}_{\mathcal{C} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{C}'}\mathcal{D} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{D}'_{\mathcal{D} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{D}'},$$

which is true on the underlying categories. Because F and F' are bimodule functors, the functor

$$F \boxtimes F'\colon \mathcal{C} \boxtimes \mathcal{C}' \to \mathcal{D} \boxtimes \mathcal{D}' \twoheadrightarrow \mathcal{D} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{D}'$$

is $\mathcal{B}$-balanced and descends to the relative tensor product $\mathcal{C} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{C}'$. We then see that the left $\mathcal{C} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{C}'$-action is the one induced by $F \boxtimes F'$ on the relative tensor product, namely action by $F \underset{\mathcal{A}_2}{\boxtimes} F'$. The centered structures are both given by the central structure of $\mathcal{D} \underset{\mathcal{A}_2}{\boxtimes} \mathcal{D}'$ and coincide. ◻

3.2.2 Dualizability Given a braided monoidal functor $F\colon \mathscr{A} \to \mathscr{B}$, we will prove that both adjoints of $\mathscr{A}_F$ are given by $\bar{\mathscr{A}}_F$. For the right adjunction, the counit should go:

$$\mathscr{A}_F \circ \bar{\mathscr{A}}_F = {}_{\mathscr{B}}\mathscr{B}_{\mathscr{A}} \underset{\mathscr{A}}{\boxtimes} {}_{\mathscr{A}}\mathscr{B}_{\mathscr{B}} \to \mathrm{Id}_{\mathscr{B}} = {}_{\mathscr{B}}\mathscr{B}_{\mathscr{B}}.$$

We actually have a functor going this way, the tensor product T in $\mathscr{B}$, which is $\mathscr{A}$-balanced and descends to the relative tensor product. We denote it by $T_{\mathrm{bal}}\colon \mathscr{B} \underset{\mathscr{A}}{\boxtimes} \mathscr{B} \to \mathscr{B}$, and it is indeed a $\mathscr{B}$-$\mathscr{B}$-bimodule monoidal functor. The central structures on both sides are given by braiding in $\mathscr{B}$, which is preserved by T. Hence we can construct a $\mathscr{B}$-$\mathscr{B}$-centered ${}_{\mathscr{B}}\mathscr{B}_{\mathscr{A}} \underset{\mathscr{A}}{\boxtimes} {}_{\mathscr{A}}\mathscr{B}_{\mathscr{B}}$-${}_{\mathscr{B}}\mathscr{B}_{\mathscr{B}}$-bimodule $\mathcal{M}_{T_{\mathrm{bal}}}$ using Definition 3.10.

The unit should go:

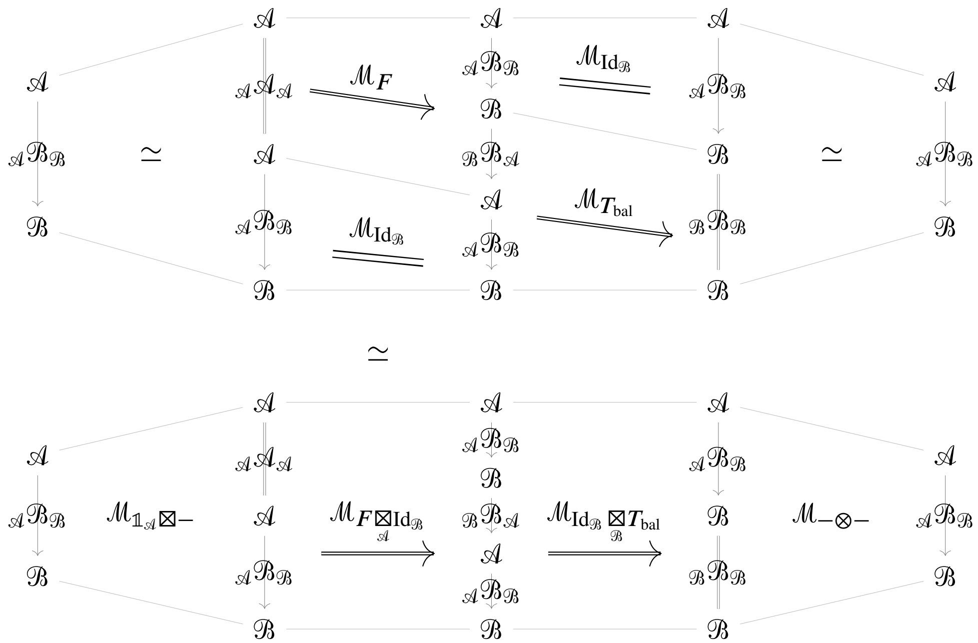
$$\mathrm{Id}_{\mathscr{A}} = {}_{\mathscr{A}}\mathscr{A}_{\mathscr{A}} \to \bar{\mathscr{A}}_F \circ \mathscr{A}_F = {}_{\mathscr{A}}\mathscr{B}_{\mathscr{B}} \underset{\mathscr{B}}{\boxtimes} {}_{\mathscr{B}}\mathscr{B}_{\mathscr{A}} \simeq {}_{\mathscr{A}}\mathscr{B}_{\mathscr{A}}.$$

Again we have a functor $F\colon \mathscr{A} \to \mathscr{B}$ which is an $\mathscr{A}$-$\mathscr{A}$-module monoidal functor. The central structure on the left is given by braiding in $\mathscr{A}$, and on the right by braiding in $\mathscr{B}$. The first is sent on the latter because F is braided monoidal, and the central structures are preserved. Therefore we also have an $\mathscr{A}$-$\mathscr{A}$-centered ${}_{\mathscr{A}}\mathscr{A}_{\mathscr{A}}$-${}_{\mathscr{A}}\mathscr{B}_{\mathscr{A}}$-bimodule $\mathcal{M}_F$.

Note also that the identity of $\mathscr{A}_F$ is the bimodule induced by $\mathrm{Id}_{\mathscr{B}}$ seen as an $\mathscr{A}$-$\mathscr{B}$-bimodule monoidal functor.

Proposition 3.12 *The* 1*-morphism* $\mathscr{A}_F$ *has right adjoint given by* $\bar{\mathscr{A}}_F$*, with counit* $\mathcal{M}_{T_{\mathrm{bal}}}$ *and unit* $\mathcal{M}_F$*. Its left adjoint is also given by* $\bar{\mathscr{A}}_F$*, with counit* $\bar{\mathcal{M}}_F$ *and unit* $\bar{\mathcal{M}}_{T_{\mathrm{bal}}}$*.*

Proof We directly check the snake. We repeatedly use Proposition 3.11:



which is the bimodule induced by the composition

$$\mathcal{B} \to \mathcal{A} \underset{\mathcal{A}}{\boxtimes} \mathcal{B} \to \mathcal{B} \underset{\mathcal{B}}{\boxtimes} \mathcal{B} \underset{\mathcal{A}}{\boxtimes} \mathcal{B} \to \mathcal{B} \underset{\mathcal{B}}{\boxtimes} \mathcal{B} \to \mathcal{B},$$
$$X \mapsto (\mathbb{1}_{\mathcal{A}}, X) \mapsto (\mathbb{1}_{\mathcal{B}}, \mathbb{1}_{\mathcal{B}}, X) \mapsto (\mathbb{1}_{\mathcal{B}}, X) \mapsto X,$$

which is indeed the identity.

Every other snake identity is similar, with functors going in the other direction for the left adjunction. □

Proposition 3.13 *Let $F\colon \mathcal{C} \to \mathcal{D}$ be an $\mathcal{A}$-$\mathcal{B}$-bimodule monoidal functor. The bimodule $\mathcal{M}_F$ has right adjoint given by $\overline{\mathcal{M}}_F$, with counit $T_{\mathrm{bal}}\colon \mathcal{D} \underset{\mathcal{C}}{\boxtimes} \mathcal{D} \to \mathcal{D}$ seen as a $\mathcal{D}$-$\mathcal{D}$-bimodule functor and unit F seen as a $\mathcal{C}$-$\mathcal{C}$-bimodule functor.*

Proof The proof is the same as above, except that the horizontal morphisms are now the functors instead of the bimodules induced by the functors. The snake identities read

$$(1)\qquad (\mathrm{Id}_{\mathcal{D}} \underset{\mathcal{D}}{\boxtimes} T_{\mathrm{bal}}) \circ (F \underset{\mathcal{C}}{\boxtimes} \mathrm{Id}_{\mathcal{D}}) \simeq \mathrm{Id}_{\mathcal{M}_F} \quad \text{and} \quad (T_{\mathrm{bal}} \underset{\mathcal{D}}{\boxtimes} \mathrm{Id}_{\mathcal{D}}) \circ (\mathrm{Id}_{\mathcal{D}} \underset{\mathcal{C}}{\boxtimes} F) \simeq \mathrm{Id}_{\overline{\mathcal{M}}_F},$$

as has been used above. Here $\mathrm{Id}_{\mathcal{D}}$ is seen alternatively as a $\mathcal{C}$-$\mathcal{D}$-bimodule functor and as a $\mathcal{D}$-$\mathcal{C}$-bimodule functor. □

We would like to apply Proposition 2.14.1, to have the left adjoint of $\mathcal{M}_F$. We need F and T_{bal} to have right adjoints in **BrTens**. There is a well-known sufficient condition for this.

Proposition 3.14 [Brochier et al. 2021b, Proposition 4.2 and Corollary 4.3] *Let $F\colon \mathcal{M} \to \mathcal{N}$ be an $\mathcal{A}$-$\mathcal{B}$-centered $\mathcal{C}$-$\mathcal{D}$-bimodule functor, so a 3-morphism in* **BrTens**. *Suppose that $\mathcal{M}$ and $\mathcal{N}$ have enough cp, that $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ are cp-rigid, and that F preserves cp. Then $F^R\colon \mathcal{N} \to \mathcal{M}$ is an $\mathcal{A}$-$\mathcal{B}$-centered $\mathcal{C}$-$\mathcal{D}$-bimodule functor, and is the right adjoint of F in* **BrTens**.

All we need to check is that both F and T_{bal} preserve cp.

Lemma 3.15 *Let $\mathcal{M}$ and $\mathcal{N}$ be right and left modules over $\mathcal{C}$ and $F\colon \mathcal{M} \boxtimes \mathcal{N} \to \mathcal{P}$ be a cocontinuous $\mathcal{C}$-balanced functor. Suppose $\mathcal{M}$ and $\mathcal{N}$ have enough cp, $\mathcal{C}$ is cp-rigid and F preserves cp. Then the induced functor $F_{\mathrm{bal}}\colon \mathcal{M} \underset{\mathcal{C}}{\boxtimes} \mathcal{N} \to \mathcal{P}$ preserves cp.*

In particular, if $\mathcal{A}$ and $\mathcal{B}$ are cp-rigid, then $T_{\mathrm{bal}}\colon \mathcal{B} \underset{\mathcal{A}}{\boxtimes} \mathcal{B} \to \mathcal{B}$ preserves cp.

Proof Following the proof of closure under composition of 1-morphisms [Brochier et al. 2021b, Section 4.2], the cp objects of $\mathcal{M} \underset{\mathcal{C}}{\boxtimes} \mathcal{N}$ are generated by pure tensors of cp objects. These are sent to cp objects in $\mathcal{P}$.

For the second point, T_{bal} is induced by T which preserves cp as $\mathcal{B}$ is cp-rigid. □

We can summarize the result as follows:

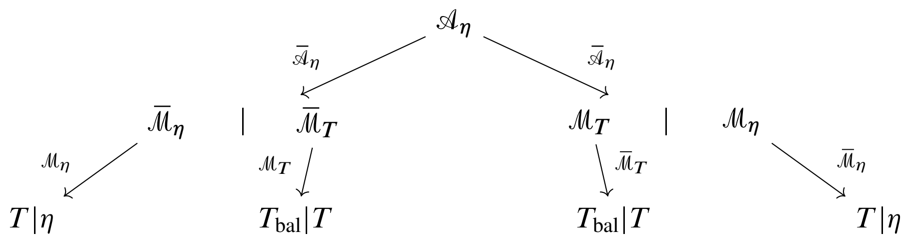
Proposition 3.16 *Let $F\colon \mathscr{C} \to \mathscr{D}$ be an $\mathscr{A}$-$\mathscr{B}$-bimodule monoidal functor which preserves cp, where $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ and $\mathscr{D}$ are cp-rigid. The bimodule $\mathscr{M}_F$ has left adjoint given by $\overline{\mathscr{M}}_F$, with counit F^R seen as a $\mathscr{C}$-$\mathscr{C}$-bimodule functor and unit T^R_{bal} seen as a $\mathscr{D}$-$\mathscr{D}$-bimodule functor.*

3.3 Unit inclusion

We give explicitly the dualizability data of the 1-morphism induced by the unit inclusion in a braided tensor category $\mathscr{V}$, and criteria for dualizability when $\mathscr{V}$ has enough cp.

Definition 3.17 Let $\mathscr{V} \in \mathbf{BrTens}$ be an $\mathbb{E}_2$-algebra in $\mathbf{Pr}$. We denote by $T\colon \mathscr{V} \boxtimes \mathscr{V} \to \mathscr{V}$ its monoidal structure, and $\eta\colon \mathrm{Vect}_{\Bbbk} \to \mathscr{V}$ the inclusion of the unit. The functor η is braided monoidal and induces a $\mathrm{Vect}_{\Bbbk}$-$\mathscr{V}$-central algebra $\mathscr{A}_\eta$, namely a 1-morphism in $\mathbf{BrTens}$. Recall that we denote by $\mathscr{A}^\flat_\eta \in \mathbf{BrTens}^{\to}$ the associated object in the oplax arrow category.

Theorem 3.18 *The 1-morphism $\mathscr{A}_\eta$ is both twice left and twice right adjunctible, with adjunctibility data as displayed:*



where $T_{\mathrm{bal}}\colon \mathscr{V} \underset{\mathscr{V}\boxtimes\mathscr{V}}{\boxtimes} \mathscr{V} \to \mathscr{V}$ is induced by T on the relative tensor product.

Proof We use the results of Section 3.2. By Proposition 3.12, the 1-morphism $\mathscr{A}_\eta$ has left and right adjoints given by $\overline{\mathscr{A}}_\eta$, with units and counits as displayed in the second line above, with $\eta\colon \mathrm{Vect}_{\Bbbk} \to \mathscr{V}$ now seen as a $\mathrm{Vect}_{\Bbbk}$-$\mathrm{Vect}_{\Bbbk}$-bimodule monoidal functor, and $T\colon \mathscr{V} \underset{\mathrm{Vect}_{\Bbbk}}{\boxtimes} \mathscr{V} \to \mathscr{V}$ the tensor product balanced over $\mathrm{Vect}_{\Bbbk}$ so not balanced.

Then by Proposition 3.13 each of these bimodules has either a left or a right adjoint, with units and counits as displayed, with $T_{\mathrm{bal}}\colon \overline{\mathscr{M}}_T \underset{\mathscr{V}\boxtimes\mathscr{V}}{\boxtimes} \mathscr{M}_T = \mathscr{V} \underset{\mathscr{V}\boxtimes\mathscr{V}}{\boxtimes} \mathscr{V} \to \mathscr{V}$ induced by T. □

Corollary 3.19 *The object $\mathscr{A}^\flat_\eta$ is 2-dualizable in $\mathbf{BrTens}^{\to}$, and:*

- $\mathrm{Ru}(\mathrm{Ru}(\mathscr{A}^\flat_\eta))$ *has a right adjoint if and only if both T_{bal} and* $\mathrm{Ru}(\mathrm{Ru}(\mathscr{V}))$ *do.*
- $\mathrm{Rco}(\mathrm{Ru}(\mathscr{A}^\flat_\eta))$ *has a right adjoint if and only if both T and* $\mathrm{Rco}(\mathrm{Ru}(\mathscr{V}))$ *do.*
- $\mathrm{Rco}(\mathrm{Rco}(\mathscr{A}^\flat_\eta))$ *has a right adjoint if and only if both η and* $\mathrm{Rco}(\mathrm{Rco}(\mathscr{V}))$ *do.*

Proof For 2-dualizability, we use the criterion of [Johnson-Freyd and Scheimbauer 2017, Theorem 7.6], we know that $\mathscr{V}$ is 2-dualizable by [Gwilliam and Scheimbauer 2018, Theorem 5.1] and $\mathscr{A}_\eta$ is twice right adjunctible by the theorem above. The rest is Theorem 2.15 on the right dualizability data of $\mathscr{A}_\eta$. □

Theorem 3.20 *Suppose that $\mathscr{V}$ has enough cp, then $\mathscr{A}_\eta^\flat$ is 3-dualizable if and only if $\mathscr{V}$ the free cocompletion of a small rigid braided monoidal category.*

Proof The heart of the proof is to notice that T appears in the dualizability data, and by [Brochier et al. 2021b, Proposition 4.1] when $\mathscr{V}$ has enough cp, it is cp-rigid if and only if T has a bimodule cocontinuous right adjoint.

If $\mathscr{A}_\eta^\flat$ is 3-dualizable then $\mathrm{Ru}(\mathrm{Ru}(\mathscr{A}_\eta^\flat))$, $\mathrm{Rco}(\mathrm{Ru}(\mathscr{A}_\eta^\flat))$ and $\mathrm{Rco}(\mathrm{Rco}(\mathscr{A}_\eta^\flat))$ have right adjoints, so T_{bal}, T and η have bimodule cocontinuous right adjoints. The functors T and η preserving cp mean that they are well-defined on $\mathcal{V} := \mathscr{V}^{\mathrm{cp}}$ and endow it with a monoidal structure, and $\mathcal{V}$ is rigid as $\mathscr{V}$ is cp-rigid. Therefore $\mathscr{V}$ is the free cocompletion of a small rigid braided monoidal category.

On the other hand if $\mathscr{V}$ is the free cocompletion of a small rigid braided monoidal category then it is cp-rigid and hence 3-dualizable [Brochier et al. 2021b, Theorem 5.16]. The functors T and η, and also T_{bal} by Lemma 3.15, preserve cp, and have bimodule cocontinuous right adjoints by Proposition 3.14. We get that $\mathscr{A}_\eta$ is 3-times right adjunctible and its source and targets are 3-dualizable, so $\mathscr{A}_\eta^\flat$ 3-dualizable by [Johnson-Freyd and Scheimbauer 2017, Theorem 7.6]. □

Theorem 3.21 *Suppose that $\mathscr{V}$ has enough cp. Then $\mathscr{A}_\eta^\flat$ is noncompact-3-dualizable if and only if $\mathscr{V}$ is cp-rigid.*

Proof If $\mathscr{V}$ is cp-rigid, then $\mathscr{V}$ is 3-dualizable and T and T_{bal} have right adjoints in **BrTens**. By Corollary 3.19, $\mathscr{A}_\eta^\flat$ is noncompact-3-dualizable.

Suppose now that $\mathscr{A}_\eta^\flat$ is noncompact-3-dualizable. Then T has a bimodule cocontinuous right adjoint, and $\mathscr{V}$ is cp-rigid. □

Theorem 3.22 *Let $\mathscr{V}$ be a braided tensor category with enough cp. Then the following are equivalent:*

(1) *$\mathscr{A}_\eta$ is 3-dualizable.*

(2) *$\mathscr{A}_\eta$ is 3-adjunctible.*

(3) *$\mathscr{V}$ is rigid finite semisimple.*

Proof The implication (1) $\Longrightarrow$ (2) is immediate: for a 1-morphism 3-dualizable demands 3-adjunctibility and 4-dualizablility of the source and target.

The implication (3) $\Longrightarrow$ (1) is essentially [Brochier et al. 2021b, Theorem 5.21]. If $\mathscr{V}$ is fusion, then $\mathscr{V}$ and $\mathscr{A}_\eta$ lie in BrFus, which has duals. Now fusion demands simplicity of the unit, which may not be the case here. This is easily solved by noticing that coproduct agrees with product in **Pr** and ought to be called direct sum [Brandenburg et al. 2015, Remark 2.5], and that braided rigid finite semisimple categories are direct sums of fusion categories [Etingof et al. 2015, Section 4.3].

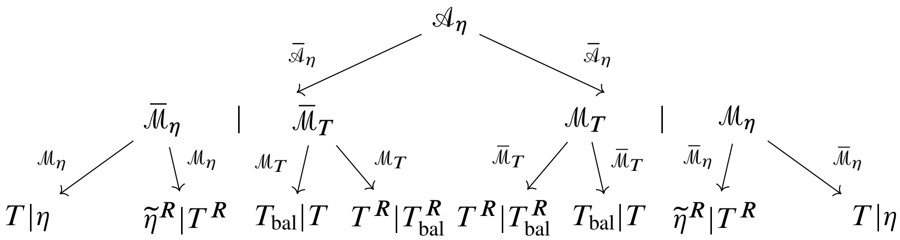
Let us prove (2) $\Longrightarrow$ (3). If $\mathcal{A}_\eta$ is 3-adjunctible, then $\mathcal{M}_\eta$ and $\overline{\mathcal{M}}_\eta$, which are respectively $\mathrm{Ru}(\mathcal{A}_\eta)$ and $\mathrm{Lco}(\mathcal{A}_\eta)$ by Theorem 3.20, must be 2-adjunctible. Hence their composite $\mathcal{M}_\eta \boxtimes_{\mathcal{V}} \overline{\mathcal{M}}_\eta$ has to be 2-adjunctible in the symmetric monoidal 2-category $\Omega\Omega\,\mathbf{BrTens} \simeq \mathbf{Pr}$. This composition is just $\mathcal{V} \boxtimes_{\mathcal{V}} \mathcal{V} \simeq \mathcal{V}$ as a category, and by our assumption that it has enough cp, it actually lies in the full subcategory $\mathrm{Bimod}_{\Bbbk} \subseteq \mathbf{Pr}$. By [Bartlett et al. 2015, Theorem A.22], the 2-dualizable objects of $\mathrm{Bimod}_{\Bbbk}$ are finite semisimple categories. We already saw that $\mathcal{V}$ has to be cp-rigid, so $\mathcal{V}^{\mathrm{cp}}$ is rigid finite semisimple, and so is $\mathcal{V} \simeq \mathrm{Free}(\mathcal{V}^{\mathrm{cp}})$. □

Remark 3.23 A very similar result one categorical dimension down, in $\mathrm{ALG}_1(\mathrm{Rex}_{\mathbb{C}})$, is proven in [Freed and Teleman 2021, Theorem B]. The proof is similar too, but we couldn't directly use their result on $\mathcal{M}_\eta$, as we work in $\mathrm{Bimod}_{\Bbbk}$ and not in $\mathrm{Rex}_{\mathbb{C}}$.

Remark 3.24 Both results need the full adjunctibility of $\mathcal{A}_\eta$: oplax dualizability does not imply semisimplicity, one can take the free cocompletion of a nonsemisimple ribbon category in Theorem 3.20. Semisimplicity is not needed for 4-dualizability either, as proven in [Brochier et al. 2021a]. However, if we assume that $\mathcal{V}$ is 4-dualizable and $\mathcal{A}_\eta$ is 3-oplax-dualizable, which is the case of interest for Section 4, then work-to-appear of William Stewart shows that $\mathcal{A}_\eta$ is 3-adjunctible. This has an interesting consequence: the free cocompletion of a ribbon category which is not semisimple cannot be 4-dualizable. Indeed if it were, Stewart's result would apply and $\mathcal{V}$ would have to be semisimple. This justifies that, given a nonsemisimple ribbon tensor category as in [Costantino et al. 2023], we want to work with its Ind-completions, and not its free cocompletion.

Remark 3.25 Being dualizable for a morphism is both a condition on its adjunctibility and on the dualizability of its source and target. However, we saw in the proof of Theorem 3.20 that $\mathcal{A}_\eta$ is 3-right-adjunctible if and only if $\mathcal{A}_\eta$ is 3-oplax-dualizable, and in the theorem above that $\mathcal{A}_\eta$ is 3-adjunctible if and only if $\mathcal{A}_\eta$ is 3-dualizable. This phenomenon seems to be specific to the unit inclusion.

Proposition 3.26 *Suppose that $\mathcal{V}$ is cp-rigid. Then $\mathcal{A}_\eta$ is 2-adjunctible with the following adjunctibility data in* **BrTens**:



where $\tilde{\eta}^R$ is the essentially unique cocontinuous functor that agrees with η^R on cp objects.

Proof The snake for T^R and $\tilde{\eta}^R$ comes from the following. Write $\mathscr{V} = \mathcal{V}^{\mathrm{cp}}$. Then T^R is computed as the coend

$$T^R(\mathbb{1}_{\mathcal{V}}) = \int^{(V,W)\in\mathscr{V}^{\otimes 2}} (V \boxtimes W) \otimes \mathrm{Hom}_{\mathcal{V}}(V \otimes W, \mathbb{1}_{\mathcal{V}}) \simeq \int^{V\in\mathscr{V}} V \boxtimes V^*,$$

and more generally

$$T^R(X) \simeq \int^{V\in\mathscr{V}} (X \otimes V) \boxtimes V^* \simeq \int^{V\in\mathscr{V}} V \boxtimes (V^* \otimes X).$$

For X cp, the snake goes

$$(\tilde{\eta}^R \underset{\mathrm{Vect}_{\Bbbk}}{\boxtimes} \mathrm{Id}_{\mathcal{V}}) \circ (\mathrm{Id}_{\mathcal{V}} \underset{\mathcal{V}}{\boxtimes} T^R)(X) \simeq \int^{V\in\mathscr{V}} \tilde{\eta}^R(X \otimes V) \boxtimes V^* = \int^{V\in\mathscr{V}} \mathrm{Hom}(\mathbb{1}_{\mathcal{V}}, X \otimes V) \otimes V^*$$
$$\simeq \int^{V\in\mathscr{V}} \mathrm{Hom}(V^*, X) \otimes V^* \simeq X.$$

The part with T^R and T^R_{bal} is given by Proposition 3.16. Indeed T, and hence T_{bal}, preserves cp as $\mathcal{V}$ is cp-rigid.

That this is sufficient for 2-adjunctibility is [Johnson-Freyd and Scheimbauer 2017, Lemma 7.11]. □

Remark 3.27 We studied the oplax-dualizability of $\mathcal{A}_\eta$ above, but Johnson-Freyd and Scheimbauer [2017] also define a notion of lax-dualizability. We are interested in the oplax-dualizability for our applications, but let us include the lax version of our results. By Theorem 3.18, $\mathcal{A}_\eta$ is always 2-lax-dualizable, and it is 3-times left adjunctible if and only if η, T and T_{bal} have left adjoints in **BrTens**. Using the proposition above, we can also get another characterization of adjunctibility: every morphism appearing there must have a right adjoint. If $\mathcal{V}$ has enough cp, then $\mathcal{A}_\eta$ is 3-adjunctible if and only if $\mathcal{V}$ is cp-rigid and η, η^R, T^R and T^R_{bal} preserve cp.

We studied the unit inclusion, but similar arguments work for any bimodule induced by a functor. Instead of a necessary and sufficient condition, we only have a sufficient condition because T no longer appears in the dualizability data, only some balanced version does.

Theorem 3.28 *Let $F : \mathcal{V} \to \mathcal{W}$ be a braided monoidal functor between two objects of* **BrTens**. *Then the object $\mathcal{A}_F^\flat \in \mathbf{BrTens}^{\to}$ induced by the 1-morphism $\mathcal{A}_F$ is 2-dualizable. It is noncompact-3-dualizable as soon as $\mathcal{V}$ and $\mathcal{W}$ are cp-rigid. In this case, it is 3-dualizable if and only if F preserves cp.*

Proof We know that $\mathrm{Radj}(\mathcal{A}_F) = \bar{\mathcal{A}}_F$ with $\mathrm{Ru}(\mathcal{A}_F) = \mathcal{M}_F$ and $\mathrm{Rco}(\mathcal{A}_F) = \mathcal{M}_{T_{\mathcal{V}\text{-bal}}}$ by Proposition 3.12, where $T_{\mathcal{V}\text{-bal}} : \mathcal{W} \underset{\mathcal{V}}{\boxtimes} \mathcal{W} \to \mathcal{W}$ is induced by the monoidal structure on $\mathcal{W}$.

Then, $\mathrm{Radj}(\mathcal{M}_{T_{\mathcal{V}\text{-bal}}}) = \bar{\mathcal{M}}_{T_{\mathcal{V}\text{-bal}}}$ with $\mathrm{Ru}(\mathcal{M}_{T_{\mathcal{V}\text{-bal}}}) = T_{\mathcal{V}\text{-bal}}$ and $\mathrm{Rco}(\mathcal{M}_{T_{\mathcal{V}\text{-bal}}}) = T_{2\,\mathrm{bal}}$ by Proposition 3.13, where

$$T_{2\,\mathrm{bal}} : \mathcal{W} \underset{\mathcal{W} \underset{\mathcal{V}}{\boxtimes} \mathcal{W}}{\boxtimes} \mathcal{W} \to \mathcal{W}$$

is induced by the monoidal structure on $\mathcal{W}$.

Similarly, $\mathrm{Radj}(\mathcal{M}_F) = \bar{\mathcal{M}}_F$ with $\mathrm{Ru}(\mathcal{M}_F) = F$ and $\mathrm{Rco}(\mathcal{M}_F) = T_{\mathcal{V}\text{-bal}}$.

We know by Theorem 2.15 that the existence and right adjunctibility of $\mathrm{Ru}(\mathrm{Ru}(\mathscr{A}_F^\flat))$, $\mathrm{Rco}(\mathrm{Ru}(\mathscr{A}_F^\flat))$ and $\mathrm{Rco}(\mathrm{Rco}(\mathscr{A}_F^\flat))$ is equivalent to that of, respectively, $T_{2\,\mathrm{bal}}$, $T_{\mathscr{V}\text{-bal}}$ and F, and of the same units/counits of the source and target. So $\mathscr{A}_F^\flat$ is noncompact-3-dualizable if and only if $T_{\mathscr{V}\text{-bal}}$ and $T_{2\,\mathrm{bal}}$ have right adjoints in **BrTens**, and both $\mathscr{V}$ and $\mathscr{W}$ are noncompact-3-dualizable. This is true as soon as $\mathscr{V}$ and $\mathscr{W}$ are cp-rigid by Lemma 3.15 and [Brochier et al. 2021b, Theorem 5.6].

It is 3-dualizable if and only if F, $T_{\mathscr{V}\text{-bal}}$ and $T_{2\,\mathrm{bal}}$ have right adjoints and $\mathscr{V}$ and $\mathscr{W}$ are 3-dualizable. If $\mathscr{V}$ and $\mathscr{W}$ are cp-rigid, this is true if and only if F preserves cp. □

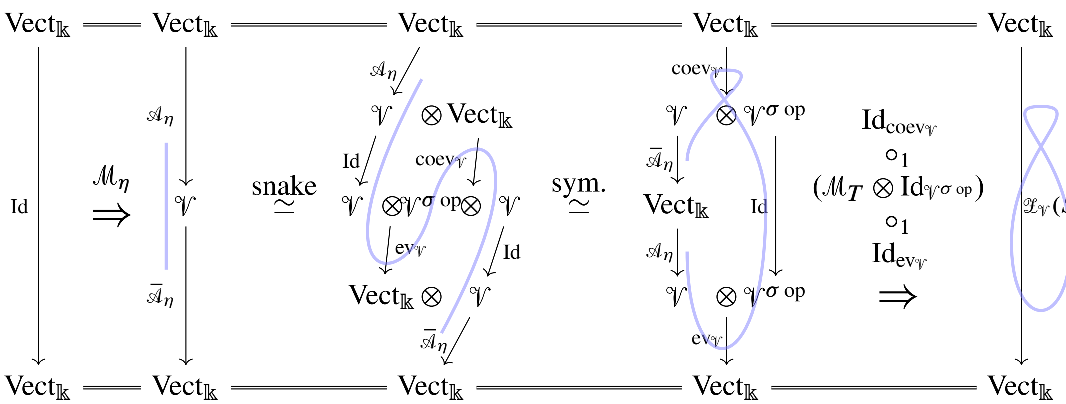
3.4 The relative theory on the circle

We compute the value on the circle of the relative TQFT $\mathscr{R}_{\mathscr{V}}$ induced by $\mathscr{A}_\eta^\flat$ under the cobordism hypothesis, for any $\mathscr{V}$. Namely, we write $S^1_{\mathrm{nb}} = \mathrm{ev}_{\mathrm{pt}} \circ \mathrm{coev}_{\mathrm{pt}}$, compute the images of $\mathrm{ev}_{\mathrm{pt}}$ and $\mathrm{coev}_{\mathrm{pt}}$ under $\mathscr{R}_{\mathscr{V}}$, which are $\mathrm{ev}_{\mathscr{A}_\eta^\flat}$ and $\mathrm{coev}_{\mathscr{A}_\eta^\flat}$, and compose them. Note that it is S^1 with nonbounding framing that we are computing.[3] We need the symmetric monoidal structure of $\mathscr{C}$ to compose $\mathrm{ev}_X : \mathbb{1} \to X \otimes X^*$ and $\mathrm{coev}_X : X \otimes X^* \simeq X^* \otimes X \to \mathbb{1}$. We know that the evaluation and coevaluation for $\mathscr{A}_\eta^\flat$ are mates of the unit and counit for the right adjunction of $\mathscr{A}_\eta$, namely $\mathscr{M}_\eta$ and $\mathscr{M}_T$. It might sound surprising that one can compose them, but indeed up to whiskering and mating they are composable; see Figure 4.

We know from [Brochier et al. 2021a, Theorem 2.19] that the evaluation and coevaluation for $\mathscr{V}$ are respectively ${}_{\mathscr{V}\boxtimes\mathscr{V}^{\sigma\,\mathrm{op}}}\mathscr{V}_{\mathrm{Vect}_\Bbbk}$ and ${}_{\mathrm{Vect}_\Bbbk}\mathscr{V}_{\mathscr{V}^{\sigma\,\mathrm{op}}\boxtimes\mathscr{V}}$. Then, Example 2.16 gives:

$$\mathscr{R}_{\mathscr{V}}(\mathrm{ev}_{\mathrm{pt}}) = \begin{array}{ccc} \mathrm{Vect}_\Bbbk & \xrightarrow{\mathscr{A}_\eta \otimes (\bar{\mathscr{A}}_\eta)^*} & \mathscr{V} \otimes \mathscr{V}^{\sigma\,\mathrm{op}} \\ {\scriptstyle \mathrm{Id}}\big\downarrow & \overset{\mathscr{M}_\eta}{\nearrow} & \big\downarrow{\scriptstyle \mathrm{ev}_{\mathscr{V}}} \\ \mathrm{Vect}_\Bbbk & \xrightarrow{\mathrm{Vect}_\Bbbk} & \mathrm{Vect}_\Bbbk \end{array} \quad \text{and} \quad \mathscr{R}_{\mathscr{V}}(\mathrm{coev}_{\mathrm{pt}}) = \begin{array}{ccc} \mathrm{Vect}_\Bbbk & \xrightarrow{\mathrm{Vect}_\Bbbk} & \mathrm{Vect}_\Bbbk \\ {\scriptstyle \mathrm{Id}}\big\downarrow & \overset{(\mathscr{M}_T \otimes \mathrm{Id}_{\mathscr{V}^{\sigma\,\mathrm{op}}}) \circ_1 \mathrm{Id}_{\mathrm{coev}_{\mathscr{V}}}}{\nearrow} & \big\downarrow{\scriptstyle \mathrm{coev}_{\mathscr{V}}} \\ \mathrm{Vect}_\Bbbk & \xrightarrow{\mathscr{A}_\eta \otimes (\bar{\mathscr{A}}_\eta)^*} & \mathscr{V} \otimes \mathscr{V}^{\sigma\,\mathrm{op}} \end{array}$$

Their composition is vertical stacking, and gives that $\mathscr{R}_{\mathscr{V}}(S^1_{\mathrm{nb}})$ is the following composition. The blue lines give the connection with Figure 4, with correct framing:



[3] In dimension 3, there are two framings on the circle, only one of which bounds a framed disk.

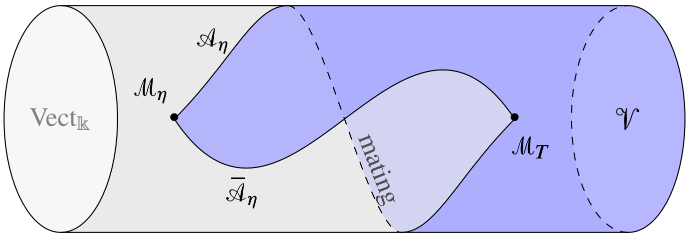


Figure 4: The unit and the counit compose up to mating. Beware that the framing is not faithfully represented in this picture.

Note that every bimodule above is induced by a functor as displayed here:

$$\begin{array}{ccccccccc} \mathrm{Vect}_{\Bbbk} & \xrightarrow{\eta} & \mathcal{V} \underset{\mathcal{V}}{\boxtimes} \mathcal{V} & \xrightarrow{\sim} & (\mathcal{V}\otimes\mathcal{V}) \underset{\mathcal{V}\otimes\mathcal{V}^{\sigma\,\mathrm{op}}\otimes\mathcal{V}}{\boxtimes} (\mathcal{V}\otimes\mathcal{V}) & \xrightarrow{\sim} & \mathcal{V} \underset{\mathrm{Vect}_{\Bbbk}\otimes\mathcal{V}^{\sigma\,\mathrm{op}}}{\boxtimes} \mathcal{V} & \xrightarrow{\mathrm{Id}_{\mathcal{V}}\boxtimes(T\otimes\mathrm{Id})\boxtimes\mathrm{Id}_{\mathcal{V}}} & \mathcal{V} \underset{\mathcal{V}\otimes\mathcal{V}^{\sigma\,\mathrm{op}}}{\boxtimes} \mathcal{V}, \\ \Bbbk & \longmapsto & \mathbb{1}\boxtimes\mathbb{1} & \longmapsto & (\mathbb{1}\otimes\mathbb{1})\boxtimes(\mathbb{1}\otimes\mathbb{1}) & \longmapsto & \mathbb{1}\boxtimes\mathbb{1} & \longmapsto & \mathbb{1}\boxtimes\mathbb{1}. \end{array}$$

So $\mathcal{R}_{\mathcal{V}}(S^1_{\mathrm{nb}})$ is induced by the monoidal functor given by inclusion of the unit in $\mathcal{Z}_{\mathcal{V}}(S^1_{\mathrm{nb}})$.

4 Nonsemisimple WRT relative to CY

We can now state the conjectures which are the main motivation for the study above. The main idea is that the Witten–Reshetikhin–Turaev theories and their nonsemisimple variants can be obtained in a fully extended setting from a 3D theory relative to an invertible 4D anomaly. In particular, they are defined in a setting where the cobordism hypothesis applies, and can be rebuilt out of their value at the point. These would be a not necessarily semisimple modular tensor category for the invertible 4-TQFT and the 1-morphism induced by the inclusion of the unit for the relative 3-TQFT. As shown above, in the nonsemisimple case the unit inclusion is only partially dualizable, and induces a noncompact TQFT.

These conjectures follow ideas of Walker [2006], Freed and Teleman [Freed 2011] in the semisimple case, and of Jordan, Reutter and Safronov in the nonsemisimple case. We do not know of a formal statement in the existing literature and propose one here.

4.1 Bulk + Relative = Anomalous

Remember that the WRT theories, and their nonsemisimple variants, are defined on a category of cobordisms equipped with some extra structure. They morally come from the data of a bounding higher manifold. Three-manifolds come equipped with an integer, which corresponds to the signature of the bounding 4-manifold, and surfaces come equipped with a Lagrangian in their first cohomology group, which corresponds to the data of the contractible curves in a bounding handlebody. In this setting, this extra structure is used to resolve an anomaly, and is due to Walker. We describe below how this kind of extra structure arises in the setting of relative field theories.

Definition 4.1 The $(n-1)$-*category of filled bordisms*

$$\mathbf{Bord}_{n-1}^{\mathrm{filled}} \subseteq \mathbf{Bord}_n^{\rightarrow}$$

is the $(n-1)$-subcategory of bordisms that map to the empty manifold under the target functor $\mathbf{Bord}_n^{\rightarrow} \to \mathbf{Bord}_n$ and to $\mathbf{Bord}_{n-1}$ under the source functor. These are k-bordisms, where $k \le n-1$, equipped with a bounding $(k+1)$-bordism which we call the filling. We denote by

$$\mathrm{Hollow}\colon \mathbf{Bord}_{n-1}^{\mathrm{filled}} \to \mathbf{Bord}_{n-1}$$

the functor that forgets the filling, namely the source functor.

The $(n-1)$-*category of noncompact filled bordisms*

$$\mathbf{Bord}_{n-1}^{\mathrm{nc,filled}} \subseteq \mathbf{Bord}_n^{\rightarrow}$$

is the $(n-1)$-subcategory of bordisms that map to the empty under the target functor and to $\mathbf{Bord}_{n-1}^{\mathrm{nc}}$ under the source functor.

Definition 4.2 An *n-relative pair* $(\mathcal{Z}, \mathcal{R})$ is the data of

- an n-TQFT $\mathcal{Z}\colon \mathbf{Bord}_n \to \mathcal{C}$, and
- an oplax-$\mathcal{Z}$-twisted-$(n-1)$-TQFT $\mathcal{R}\colon \mathbf{Bord}_{n-1} \to \mathcal{C}^{\rightarrow}$, namely an oplax transformation $\mathrm{Triv} \Rightarrow \mathcal{Z}|_{\mathbf{Bord}_{n-1}}$.

Such a pair is called a *noncompact n-relative pair* if $\mathcal{R}$ is a noncompact theory.

Given an n-relative pair $(\mathcal{Z}, \mathcal{R})$ one has two symmetric monoidal functors $\mathbf{Bord}_{n-1}^{\mathrm{filled}} \to \mathcal{C}^{\rightarrow}$. One is given by applying functoriality of $(-)^{\rightarrow}$ on $\mathcal{Z}$, namely applying $\mathcal{Z}$ to any diagram in $\mathbf{Bord}_n$ to get a diagram of the same shape in $\mathcal{C}$. It has trivial target and gives an oplax transformation

$$\mathcal{Z}^{\rightarrow \mathbb{1}}\colon \mathcal{Z}|_{\mathbf{Bord}_{n-1}} \circ \mathrm{Hollow} \Rightarrow \mathrm{Triv}$$

between functors $\mathbf{Bord}_{n-1}^{\mathrm{filled}} \to \mathcal{C}$.

The other one is given by applying the relative field theory on the hollowed out bordism, it is an oplax transformation

$$\mathcal{R} \circ \mathrm{Hollow}\colon \mathrm{Triv} \Rightarrow \mathcal{Z}|_{\mathbf{Bord}_{n-1}} \circ \mathrm{Hollow}\,.$$

Definition 4.3 The *anomalous* $(n-1)$-*theory* $\mathcal{A}$ induced by the n-relative pair $(\mathcal{Z}, \mathcal{R})$ is the composition $\mathcal{Z}^{\rightarrow \mathbb{1}} \circ (\mathcal{R} \circ \mathrm{Hollow})$ of these two oplax transformations. It gives an oplax transformation $\mathrm{Triv} \Rightarrow \mathrm{Triv}$, which by [Johnson-Freyd and Scheimbauer 2017, Theorem 7.4 and Remark 7.5] is equivalent to a symmetric monoidal functor

$$\mathcal{A}\colon \mathbf{Bord}_{n-1}^{\mathrm{filled}} \to (\Omega\mathcal{C})^{\mathrm{odd\,opp}},$$

where odd opp means we take opposite of k-morphisms for k odd, and $\Omega\mathcal{C} := \mathrm{End}_{\mathcal{C}}(\mathbb{1})$ is the delooping $(n-1)$-category.

If $(\mathcal{Z}, \mathcal{R})$ is a noncompact n-relative pair, the same construction on the appropriate subcategories gives an anomalous theory $\mathcal{A}\colon \mathbf{Bord}_{n-1}^{\mathrm{nc,filled}} \to (\Omega\mathcal{C})^{\mathrm{odd\,opp}}$.

For comparison with WRT theories, we will need to restrict to a once extended theory, namely look at endomorphisms of the trivial in $\mathbf{BORD}_{n-1}^{\text{filled}}$, and to check that the anomalous theory descends to the quotient where one only remembers signatures and Lagrangians out of the fillings. We will also move this odd opposite to the source category.

Definition 4.4 The *bicategory of simply filled* 3-2-1-*cobordisms*, which we denote by $\mathbf{Cob}_{321}^{\text{filled}}$, is the subcategory of $h_2(\Omega\,\mathbf{BORD}_3^{\text{filled,odd opp}})$ where circles can only be filled by disks, and surfaces by handlebodies. Taking the opposite orientation for 1- and 2-manifolds (which will have the effect of switching the source and target of a 3-bordism), one can identify this bicategory as:

$$\mathbf{Cob}_{321}^{\text{filled}} \simeq \begin{cases} \text{objects} & (\sqcup^n S^1, \sqcup^n \mathbb{D}^2 \colon \sqcup^n S^1 \to \varnothing) \text{ for } n \in \mathbb{N}, \\ \text{1-morphisms} & (\Sigma \colon \sqcup^{n_1} S^1 \to \sqcup^{n_2} S^1, H \colon \varnothing \to (\sqcup^{n_1} \overline{\mathbb{D}}^2) \cup \Sigma \cup (\sqcup^{n_2} \mathbb{D}^2)), \\ \text{2-morphisms} & (M \colon \Sigma_1 \to \Sigma_2, W \colon H_1 \cup M \cup \overline{H}_2 \to \varnothing). \end{cases}$$

The analogous definition in the noncompact case $\mathbf{Cob}_{321}^{\text{nc,filled}} \subseteq h_2(\Omega\mathbf{BORD}_3^{\text{nc,filled,odd opp}})$ will require a 3-bordism to have nonempty *incoming* boundary in every connected component, as source and targets of 3-manifolds are switched. To facilitate comparison with the existing literature, we also require that all surfaces have nonempty incoming boundary, although in our setting this is purely artificial.

This bicategory is to be compared with:

Definition 4.5 The bicategory $\widetilde{\mathbf{Cob}}_{321}$ (resp. $\widetilde{\mathbf{Cob}}_{321}^{\text{nc}}$) is the bicategory of circles, surface bordisms (resp. surface bordisms with nonempty incoming boundary) equipped with a Lagrangian subspace in their first homology group, and 3-bordisms (resp. 3-bordisms with nonempty incoming boundary) equipped with an integer. Composition is given by usual composition on the underlying bordisms, plus:

- taking the sum of the Lagrangian subspaces for composition of surfaces,
- adding the integers plus some Maslov index for composition of 3-bordisms,
- just adding the integers for composition of 3-bordisms in the direction of 1-morphisms.

See [De Renzi 2021, Section 3] for a precise definition. The bordisms there are decorated by objects of a ribbon category, and we are looking at the subcategory where every decoration is empty. The category $\widetilde{\mathbf{Cob}}_{321}^{\text{nc}}$ corresponds to admissible bordisms there.

Proposition 4.6 *The assignment*

$$\pi_{321} \colon \begin{cases} \mathbf{Cob}_{321}^{\text{filled}} \to \widetilde{\mathbf{Cob}}_{321}, \\ (\sqcup^n S^1, \sqcup^n \mathbb{D}^2) \mapsto \sqcup^n S^1, \\ (\Sigma, H) \mapsto (\Sigma, \ker(i_* \colon H_1(\Sigma) \to H_1(H))), \\ (M, W) \mapsto (M, \sigma(W)) \end{cases}$$

is a symmetric monoidal functor.

Proof For composition of 1-morphisms we want to show that the kernel of a gluing is the sum of the kernels. One inclusion is immediate and the other one follows by dimensions since both are Lagrangians; see [De Renzi 2017, Propositions B.6.5 and B.6.6].

For composition of 2-morphisms we use Wall's theorem; see [De Renzi 2017, Theorem B.6.1] for a statement in our context.

For composition of 2-morphisms in the direction of 1-morphisms we use that filled surfaces only glue on disks, and hence filled 3-manifolds on 3-balls, so the signature of the filling simply adds. □

Similarly, one can restrict to noncompact cobordisms and get

$$\pi_{321}^{\mathrm{nc}}: \mathbf{Cob}_{321}^{\mathrm{nc,filled}} \to \widetilde{\mathbf{Cob}}_{321}^{\mathrm{nc}}.$$

If we restrict $\widetilde{\mathbf{Cob}}_{321}$ to surfaces equipped with Lagrangians that are induced by some handlebody, these functors are essentially surjective, hence the name.

4.2 Conjectures

We want to relate the Witten–Reshetikhin–Turaev theories and their nonsemisimple variants to the ones induced by the cobordism hypothesis. We want to say that the anomalous theory induced the relative pair $(\mathscr{Z}_{\mathcal{V}}, \mathscr{R}_{\mathcal{V}})$ factors through $\widetilde{\mathbf{Cob}}_{321}$ and recovers WRT and DGGPR theories.

It has long been a folklore result that WRT theories extend to the circle [Walker 1991; Gelca 1997]; see also [Kirillov and Balsam 2010] for Turaev–Viro theories. Once-extended 3-TQFTs are classified in the preprint [Bartlett et al. 2015, Theorem 3], and the following result can be obtained from [Bartlett et al. 2015, Proposition 6.1] (in our case the unit is simple). We give the statement of [De Renzi 2017, Theorem 1.1.1] restricted to trivially decorated bordisms.

Theorem 4.7 *For a semisimple modular tensor category $\mathcal{V}$ with a chosen square root of its global dimension, the Witten–Reshetikhin–Turaev TQFT extends to the circle as a symmetric monoidal functor*

$$\mathrm{WRT}_{\mathcal{V}}: \widetilde{\mathbf{Cob}}_{321} \to \widehat{\mathrm{Cat}}_{\Bbbk},$$

where $\widehat{\mathrm{Cat}}_{\Bbbk}$ is the category of Cauchy-complete categories.

Similarly, restricting the statement of [De Renzi 2021] to trivially decorated bordisms:

Theorem 4.8 [De Renzi 2021, Theorem 1.1] *For a nonsemisimple modular tensor category $\mathcal{V}$ with a chosen square root of its global dimension, the nonsemisimple TQFT from [De Renzi et al. 2022] extends to the circle as a symmetric monoidal functor*

$$\mathrm{DGGPR}_{\mathcal{V}}: \widetilde{\mathbf{Cob}}_{321}^{\mathrm{nc}} \to \widehat{\mathrm{Cat}}_{\Bbbk}$$

On the other hand, using the cobordism hypothesis:

Theorem 4.9 (Brochier–Jordan–Safronov–Snyder) *For a semisimple or nonsemisimple modular tensor category $\mathscr{V}$, its Ind-cocompletion $\mathcal{V} \in \mathbf{BrTens}$ is 4-dualizable and induces under the cobordism hypothesis a 4-TQFT $\mathscr{Z}_{\mathcal{V}}\colon \mathbf{Bord}_4^{\mathrm{fr}} \to \mathbf{BrTens}$.*

The main result of this paper can be stated in this context.

Theorem 4.10 *For a semisimple modular tensor category $\mathscr{V}$, the arrow $\mathscr{A}_\eta^\flat \in \mathbf{BrTens}^{\to}$ induced by the unit inclusion $\eta\colon \mathrm{Vect}_{\Bbbk} \to \mathcal{V} := \mathrm{Ind}(\mathscr{V})$ is 3-dualizable and induces under the cobordism hypothesis a framed oplax-$\mathscr{Z}_{\mathcal{V}}$-twisted 3-TQFT*

$$\mathscr{R}_{\mathcal{V}}\colon \mathbf{Bord}_3^{\mathrm{fr}} \to \mathbf{BrTens}^{\to}.$$

For $\mathscr{V}$ a nonsemisimple modular tensor category, $\mathscr{A}_\eta^\flat$ is not 3-dualizable but is noncompact-3-dualizable and induces under the noncompact cobordism hypothesis a framed noncompact oplax-$\mathscr{Z}_{\mathcal{V}}$-twisted 3-TQFT

$$\mathscr{R}_{\mathcal{V}}\colon \mathbf{Bord}_3^{\mathrm{fr,nc}} \to \mathbf{BrTens}^{\to}.$$

Proof If $\mathscr{V}$ is semisimple, $\mathcal{V} = \mathrm{Ind}(\mathscr{V}) = \mathrm{Free}(\mathscr{V})$ and Theorem 3.20 applies. If $\mathscr{V}$ is not semisimple, the unit is not projective in $\mathscr{V}$, nor in $\mathcal{V} = \mathrm{Ind}(\mathscr{V})$, so $\mathscr{A}_\eta^\flat$ is not 3-dualizable. But $\mathcal{V}$ is cp-rigid and Theorem 3.21 applies. □

To compare the two sides, we need all theories to be oriented. We assume the following:

Conjecture 4.11 *Let $\mathscr{V}$ be a ribbon tensor category and $\mathcal{V} := \mathrm{Ind}(\mathscr{V})$. Then:*

- *The ribbon structure of $\mathscr{V}$ induces an $\mathrm{SO}(3)$-homotopy-fixed-point structure on $\mathcal{V}$.*
- *The ribbon structure of η induces an $\mathrm{SO}(3)$-homotopy-fixed-point structure on $\mathscr{A}_\eta^\flat$.*

The first statement is expected by experts. The second one follows [Lurie 2009, Example 4.3.23]. Note that in the second statement we really mean an SO(3)-homotopy-fixed-point structure compatible with the one on $\mathcal{V}$, as in Remark 2.21.

Remark 4.12 The fact that the anomalous theory $\mathscr{A}_{\mathcal{V}}$ would factor through $\widetilde{\mathbf{Cob}}_{321}$ is not too surprising. As was pointed to me by Pavel Safronov, we know from [Brochier et al. 2021a] that $\mathcal{V}$ is not only 4-dualizable, but invertible, and hence 5-dualizable. But $\mathbf{BrTens}$ has no nontrivial 5-morphisms, and hence the 5-theory induced by $\mathcal{V}$ is trivial on 5-bordisms. This means that $\mathscr{Z}_{\mathcal{V}}$ should give the same value on cobordant 4-manifolds. If this story can be made oriented, it means it depends only on the signature of 4-manifolds.

It was observed by Walker [2006, Chapter 9] in the semisimple case that there is a scalar choice of ways to extend $\mathscr{Z}_{\mathcal{V}}$ from $\mathbf{Bord}_3^{or}$ to $\mathbf{Bord}_4^{or}$, namely $\mathscr{Z}_{\mathcal{V}}(B^4)$, and that exactly two of these scalars yield theories which are cobordant-invariant on 4-manifolds. He observes that these scalars are exactly the two square roots of the global dimension among which one has to choose when defining WRT theories.

This motivates the following conjecture. In the nonsemisimple case, it is supported by the fact that the constructions of the (3+1)-TQFTs of [Costantino et al. 2023] need exactly the choice of a modified trace.

Conjecture 4.13 *Let $\mathscr{V}$ be a modular tensor category and $\mathcal{V} := \mathrm{Ind}(\mathscr{V})$. Then:*

- *A choice of modified trace on $\mathscr{V}$ induces an $\mathrm{SO}(4)$-homotopy-fixed-point structure on $\mathcal{V}$.*
- *A modified trace induces an $\mathrm{SO}(5)$-homotopy-fixed-point structure on $\mathcal{V}$ if and only if the global dimension $\mathscr{S}_{\mathscr{V},\mathsf{t}}(S^4) = 1$ with this choice of modified trace in the construction of [Costantino et al. 2023].*

In particular, we conjecture that every modular tensor category has an SO(5)-homotopy-fixed-point structure. Indeed let $\mathscr{V}$ be a modular tensor category and choose t a nondegenerate modified trace on $\mathscr{V}$, which exists and is unique up to scalar by [Geer et al. 2022, Corollary 5.6]. Choose a square root $\mathscr{D}_\mathsf{t}$ of its global dimension $d(\mathscr{V})_\mathsf{t} := \mathscr{S}_{\mathscr{V},\mathsf{t}}(S^4) = \Delta_+\Delta_-$ as defined in [Costantino et al. 2023] (and denoted by ζ there). Then the modified traces $\pm\mathscr{D}_\mathsf{t}^{-1}\mathsf{t}$ are the only two modified traces satisfying $\mathscr{S}_{\mathscr{V},\pm\mathscr{D}_\mathsf{t}^{-1}\mathsf{t}}(S^4) = 1$ by [Costantino et al. 2023, Proposition 5.7].

Remark 4.14 Let us try to give a conceptual reason for why SO(5)-structures correspond to square roots of the global dimension. As $\mathcal{V}$ is an invertible object, the oriented theory $\mathfrak{X}_\mathcal{V}$ is an invertible 4-TQFT and these are known to give invariants which only depend on the signature and Euler characteristic of 4-manifolds. Two closed 4-manifold are cobordant if and only if they have the same signature, so to get a cobordant-invariant theory we need to kill the dependence on the Euler characteristic. Changing the choice of modified traces by a scalar κ alters this dependence by a factor $\kappa^{\sigma(W)}$ [Costantino et al. 2023, Proposition 5.7],[4] and a well-chosen scalar κ will kill it. We need to ask that for any closed 4-manifold W with signature zero, $\mathfrak{X}_\mathcal{V}(W) = 1$. However there is no closed 4-manifold with signature 0 and Euler characteristic 1, they always have same parity. It is sufficient to ask that $\mathfrak{X}_\mathcal{V}(S^4) = 1$. The 4-sphere has Euler characteristic 2, hence there are exactly two solutions for κ, the two square roots of the global dimension.

Corollary 4.15 (of conjectures) *Both $\mathfrak{X}_\mathcal{V}$ and $\mathfrak{R}_\mathcal{V}$ give oriented TQFTs by the oriented cobordism hypothesis.*

We now assume that this corollary is true, that the choice of square root of the global dimension has been made, and that $\mathfrak{X}_\mathcal{V}$ and $\mathfrak{R}_\mathcal{V}$ are oriented.

In the semisimple case, the relative pair

$$(\mathfrak{X}_\mathcal{V} \colon \mathbf{Bord}_4 \to \mathbf{BrTens},\ \mathfrak{R}_\mathcal{V} \colon \mathbf{Bord}_3 \to \mathbf{BrTens}^{\to})$$

induces an anomalous theory

$$\mathcal{A}_\mathcal{V} \colon \mathbf{Bord}_3^{\text{filled,odd opp}} \to \mathbf{Tens} := \Omega\mathbf{BrTens}.$$

[4] We assume Conjecture 1.6 here.

Its restriction on $\mathbf{Cob}_{321}^{\text{filled}}$ gives a 2-functor

$$\mathcal{A}_{\mathcal{V}}^{321}\colon \mathbf{Cob}_{321}^{\text{filled}} \to \Omega\mathbf{TENS} \simeq \mathbf{PR}.$$

Conjecture 4.16 *For a semisimple modular tensor category $\mathcal{V}$, the anomalous theory induced by $(\mathcal{Z}_{\mathcal{V}}, \mathcal{R}_{\mathcal{V}})$ recovers the Witten–Reshetikhin–Turaev theory. Namely,*

$$\begin{array}{ccc} \mathbf{Cob}_{321}^{\text{filled}} & \xrightarrow{\mathcal{A}_{\mathcal{V}}^{321}} & \mathbf{PR} \\ \downarrow{\scriptstyle \pi_{321}} & & \uparrow{\scriptstyle \text{Free}} \\ \widetilde{\mathbf{Cob}}_{321} & \xrightarrow{\text{WRT}_{\mathcal{V}}} & \widehat{\text{Cat}}_{\Bbbk} \end{array}$$

commutes up to isomorphism.

In the nonsemisimple case, the relative pair

$$(\mathcal{Z}_{\mathcal{V}}\colon \mathbf{BORD}_4 \to \mathbf{BRTENS},\ \mathcal{R}_{\mathcal{V}}\colon \mathbf{BORD}_3^{\text{nc}} \to \mathbf{BRTENS}^{\to})$$

induces an anomalous theory

$$\mathcal{A}_{\mathcal{V}}\colon \mathbf{BORD}_3^{\text{nc,filled,odd opp}} \to \mathbf{TENS} := \Omega\mathbf{BRTENS}.$$

Its restriction on $\mathbf{Cob}_{321}^{\text{nc,filled}}$ gives a 2-functor

$$\mathcal{A}_{\mathcal{V}}^{321}\colon \mathbf{Cob}_{321}^{\text{nc,filled}} \to \Omega\mathbf{TENS} \simeq \mathbf{PR}.$$

Conjecture 4.17 *For a nonsemisimple modular tensor category $\mathcal{V}$, the noncompact anomalous theory induced by $(\mathcal{Z}_{\mathcal{V}}, \mathcal{R}_{\mathcal{V}})$ recovers the De Renzi–Gainutdinov–Geer–Patureau-Mirand–Runkel theory. Namely,*

$$\begin{array}{ccc} \mathbf{Cob}_{321}^{\text{nc,filled}} & \xrightarrow{\mathcal{A}_{\mathcal{V}}^{321}} & \mathbf{PR} \\ \downarrow{\scriptstyle \pi_{321}^{\text{nc}}} & & \uparrow{\scriptstyle \text{Free}} \\ \widetilde{\mathbf{Cob}}_{321}^{\text{nc}} & \xrightarrow{\text{DGGPR}_{\mathcal{V}}} & \widehat{\text{Cat}}_{\Bbbk} \end{array}$$

commutes up to isomorphism.

We know how to check these conjectures on the circle. We have $\text{WRT}_{\mathcal{V}}(S^1) = \mathcal{V}$ whose free cocompletion is equivalent to $\mathcal{V}$ because $\mathcal{V}$ is semisimple. Similarly, $\text{DGGPR}_{\mathcal{V}}(S^1) = \text{Proj}(\mathcal{V})$ whose free cocompletion is equivalent to $\mathcal{V}$. On the other side, we know that in dimension two $\mathcal{Z}_{\mathcal{V}}$ coincides with factorization homology, and we computed $\mathcal{R}_{\mathcal{V}}(S^1)$ in Section 3.4. So

$$\mathcal{A}_{\mathcal{V}}^{321}(S^1, \mathbb{D}^2) = \mathcal{R}_{\mathcal{V}}(S^1) \underset{\mathcal{Z}_{\mathcal{V}}(S^1)}{\boxtimes} \mathcal{Z}_{\mathcal{V}}(\mathbb{D}^2) \simeq {}_{\text{Vect}_{\Bbbk}}\mathcal{Z}_{\mathcal{V}}(S^1) \underset{\mathcal{Z}_{\mathcal{V}}(S^1)}{\boxtimes} \mathcal{V}_{\text{Vect}_{\Bbbk}} \simeq {}_{\text{Vect}_{\Bbbk}}\mathcal{V}_{\text{Vect}_{\Bbbk}}.$$

Computing the values of the theories induced by the cobordism hypothesis on higher-dimensional bordisms comes down to computing some adjoints in **BRTENS** and composing them in various ways. This will be carried out in future work.

Corollary 4.18 (of conjectures) *Both* $\mathrm{WRT}_{\mathscr{V}}$ *and* $\mathrm{DGGPR}_{\mathscr{V}}$ *extend to* S^0.

Proof Indeed, the anomalous theory $\mathscr{A}_{\mathscr{V}}$ is really defined as a functor between the 3-categories $\mathbf{Bord}_3^{\text{filled}} \to \mathbf{Tens}$ (resp. $\mathbf{Bord}_3^{\text{nc,filled}} \to \mathbf{Tens}$ in the nonsemisimple case). The two points S^0 are bordant, by a cap, and therefore give an object $(S^0, \cap) \in \mathbf{Bord}_3^{\text{filled}}$ (resp. $\mathbf{Bord}_3^{\text{nc,filled}}$).

It is easy to compute the value of the anomalous theory on this object, namely

$$\mathscr{A}_{\mathscr{V}}(S^0, \cap) = \mathscr{R}_{\mathscr{V}}(S^0) \circ \mathscr{Z}_{\mathscr{V}}(\cap) = (\mathscr{A}_\eta \boxtimes (\bar{\mathscr{A}}_\eta)^*) \underset{\mathscr{V} \boxtimes \mathscr{V}^{\sigma\,\mathrm{op}}}{\boxtimes} \mathscr{V} \simeq \mathscr{V}$$

seen as a $\mathrm{Vect}_{\Bbbk}$-$\mathrm{Vect}_{\Bbbk}$-central algebra. □

Remark 4.19 This corollary is to be compared with results of [Douglas et al. 2020], which shows that $\mathrm{WRT}_{\mathscr{V}}$ extends to the point if and only if $\mathscr{V} \simeq Z(\mathscr{C})$ is a Drinfeld center, in which case the point is mapped to $\mathscr{C}$. In the modular case, the Drinfeld center $Z(\mathscr{C})$ is isomorphic to $\mathscr{C} \otimes \mathscr{C}^{\sigma\,\mathrm{op}}$, and the two descriptions agree on S^0. Therefore it appears that $\mathrm{WRT}_{\mathscr{V}}$ always extends to S^0, and extends to the point if and only if one can find a "square root" for its value on S^0. This is also related to ongoing work of Freed, Teleman and Scheimbauer.

Note however that the statement above is a bit informal, because it is really $\mathrm{Free} \circ \mathrm{WRT}_{\mathscr{V}} \circ \pi_{321}$ that extends to S^0, so WRT indeed but with different source and target. In particular, the results of [Douglas et al. 2020] do not apply directly in this context.

References

[Adámek and Rosický 1994] **J Adámek**, **J Rosický**, *Locally presentable and accessible categories*, Lond. Math. Soc. Lect. Note Ser. 189, Cambridge Univ. Press (1994) MR Zbl

[Ayala and Francis 2017] **D Ayala**, **J Francis**, *The cobordism hypothesis*, preprint (2017) arXiv 1705.02240

[Baez and Dolan 1995] **J C Baez**, **J Dolan**, *Higher-dimensional algebra and topological quantum field theory*, J. Math. Phys. 36 (1995) 6073–6105 MR Zbl

[Bartlett et al. 2015] **B Bartlett**, **C L Douglas**, **C J Schommer-Pries**, **J Vicary**, *Modular categories as representations of the* 3*-dimensional bordism* 2*-category*, preprint (2015) arXiv 1509.06811

[Blanchet et al. 2016] **C Blanchet**, **F Costantino**, **N Geer**, **B Patureau-Mirand**, *Nonsemisimple TQFTs, Reidemeister torsion and Kashaev's invariants*, Adv. Math. 301 (2016) 1–78 MR Zbl

[Brandenburg et al. 2015] **M Brandenburg**, **A Chirvasitu**, **T Johnson-Freyd**, *Reflexivity and dualizability in categorified linear algebra*, Theory Appl. Categ. 30 (2015) art. id. 23 MR Zbl

[Brochier et al. 2021a] **A Brochier**, **D Jordan**, **P Safronov**, **N Snyder**, *Invertible braided tensor categories*, Algebr. Geom. Topol. 21 (2021) 2107–2140 MR Zbl

[Brochier et al. 2021b] **A Brochier**, **D Jordan**, **N Snyder**, *On dualizability of braided tensor categories*, Compos. Math. 157 (2021) 435–483 MR Zbl

[Costantino et al. 2023] **F Costantino**, **N Geer**, **B Haïoun**, **B Patureau-Mirand**, *Skein* (3+1)*-TQFTs from non-semisimple ribbon categories*, preprint (2023) arXiv 2306.03225

[De Renzi 2017] **M De Renzi**, *Construction of extended topological quantum field theories*, PhD thesis, Université Sorbonne Paris Cité (2017) Available at https://theses.hal.science/tel-02003025

[De Renzi 2021] **M De Renzi**, *Extended TQFTs from non-semisimple modular categories*, Indiana Univ. Math. J. 70 (2021) 1769–1811 MR Zbl

[De Renzi et al. 2022] **M De Renzi**, **A M Gainutdinov**, **N Geer**, **B Patureau-Mirand**, **I Runkel**, *3-dimensional TQFTs from non-semisimple modular categories*, Selecta Math. 28 (2022) art. id. 42 MR Zbl

[Douglas et al. 2020] **C L Douglas**, **C Schommer-Pries**, **N Snyder**, *Dualizable tensor categories*, Mem. Amer. Math. Soc. 1308, Amer. Math. Soc., Providence, RI (2020) MR Zbl

[Etingof et al. 2015] **P Etingof**, **S Gelaki**, **D Nikshych**, **V Ostrik**, *Tensor categories*, Math. Surv. Monogr. 205, Amer. Math. Soc., Providence, RI (2015) MR Zbl

[Freed 2011] **D S Freed**, *Remarks on fully extended* 3*-dimensional topological field theories*, lecture slides, String-Math (2011) Available at https://web.ma.utexas.edu/users/dafr/stringsmath_np.pdf

[Freed and Teleman 2014] **D S Freed**, **C Teleman**, *Relative quantum field theory*, Comm. Math. Phys. 326 (2014) 459–476 MR Zbl

[Freed and Teleman 2021] **D S Freed**, **C Teleman**, *Gapped boundary theories in three dimensions*, Comm. Math. Phys. 388 (2021) 845–892 MR Zbl

[Geer et al. 2022] **N Geer**, **J Kujawa**, **B Patureau-Mirand**, *M-traces in (non-unimodular) pivotal categories*, Algebr. Represent. Theory 25 (2022) 759–776 MR Zbl

[Gelca 1997] **R Gelca**, *Topological quantum field theory with corners based on the Kauffman bracket*, Comment. Math. Helv. 72 (1997) 216–243 MR Zbl

[Grady and Pavlov 2021] **D Grady**, **D Pavlov**, *The geometric cobordism hypothesis*, preprint (2021) arXiv 2111.01095

[Gwilliam and Scheimbauer 2018] **O Gwilliam**, **C Scheimbauer**, *Duals and adjoints in higher Morita categories*, preprint (2018) arXiv 1804.10924

[Haugseng 2017] **R Haugseng**, *The higher Morita category of* $\mathbb{E}_n$*-algebras*, Geom. Topol. 21 (2017) 1631–1730 MR Zbl

[Johnson-Freyd 2021] **T Johnson-Freyd**, *Heisenberg-picture quantum field theory*, from "Representation theory, mathematical physics, and integrable systems", Progr. Math. 340, Birkhäuser, Cham (2021) 371–409 MR Zbl

[Johnson-Freyd and Scheimbauer 2017] **T Johnson-Freyd**, **C Scheimbauer**, *(Op)lax natural transformations, twisted quantum field theories, and "even higher" Morita categories*, Adv. Math. 307 (2017) 147–223 MR Zbl

[Kirillov and Balsam 2010] **A Kirillov, Jr**, **B Balsam**, *Turaev–Viro invariants as an extended TQFT*, preprint (2010) arXiv 1004.1533

[Lurie 2009] **J Lurie**, *On the classification of topological field theories*, from "Current developments in mathematics, 2008", International, Somerville, MA (2009) 129–280 MR Zbl

[Lurie 2017] **J Lurie**, *Higher algebra*, book project (2017) Available at https://url.msp.org/Lurie-HA

[Riehl and Verity 2016] **E Riehl**, **D Verity**, *Homotopy coherent adjunctions and the formal theory of monads*, Adv. Math. 286 (2016) 802–888 MR Zbl

[Scheimbauer 2014] **C Scheimbauer**, *Factorization homology as a fully extended topological field theory*, PhD thesis, ETH Zürich (2014) Available at https://doi.org/10.3929/ethz-a-010399715

[Schommer-Pries 2014] **C J Schommer-Pries**, *Dualizability in low-dimensional higher category theory*, from "Topology and field theories", Contemp. Math. 613, Amer. Math. Soc., Providence, RI (2014) 111–176 MR Zbl

[Stolz and Teichner 2011] **S Stolz**, **P Teichner**, *Supersymmetric field theories and generalized cohomology*, from "Mathematical foundations of quantum field theory and perturbative string theory", Proc. Sympos. Pure Math. 83, Amer. Math. Soc., Providence, RI (2011) 279–340 MR Zbl

[Turaev 1994] **V G Turaev**, *Quantum invariants of knots and 3–manifolds*, de Gruyter Stud. Math. 18, de Gruyter, Berlin (1994) MR Zbl

[Walker 1991] **K Walker**, *On Witten's 3-manifold invariants*, unpublished notes (1991) Available at http://canyon23.net/math/1991TQFTNotes.pdf

[Walker 2006] **K Walker**, *TQFTs*, incomplete notes (2006) Available at http://canyon23.net/math/tc.pdf

[Witten 1989] **E Witten**, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. 121 (1989) 351–399 MR Zbl

Institut de Mathématiques de Toulouse, Université Toulouse 3 Paul Sabatier
Toulouse, France

benjamin.haioun@ens-lyon.fr

Proposed: Nathalie Wahl
Seconded: Ciprian Manolescu, Ulrike Tillmann

Received: 24 May 2023
Revised: 5 July 2024

Geometry & Topology **29**:4 (2025) 2217–2250
DOI: 10.2140/gt.2025.29.2217
Published: 27 June 2025

Closed geodesics in dilation surfaces

Adrien Boulanger
Selim Ghazouani
Guillaume Tahar

We prove that directions of closed geodesics in every dilation surface form a dense subset of the circle. The proof draws on a study of the degenerations of the Delaunay triangulation of dilation surfaces under the action of Teichmüller flow in the moduli space.

37E35

1 Introduction

We consider the problem of the existence of regular closed geodesics in dilation surfaces. Our main theorem is the following:

Theorem 1.1 *For any closed dilation surface Σ, there is a dense set of directions θ such that the directional foliation $\mathcal{F}_\theta$ has a periodic leaf. Equivalently, the set of directions covered by a cylinder is dense in $\mathbb{RP}^1$.*

In particular, any dilation surface Σ carries at least one closed geodesic. This generalizes to the context of dilation surfaces a celebrated theorem of Masur [4] for translation surfaces.

As the two equivalent formulations of Theorem 1.1 suggest, it can be viewed from either a dynamical or geometric perspective. From the geometric point of view, it guarantees that every dilation surface contains the simplest building block that can be imagined, a cylinder, thus giving valuable insight into the geometric structure of the arbitrary dilation surface.

On the dynamical side, this theorem guarantees the ubiquity of periodic orbits in some particular (but very natural) one-parameter families of one-dimensional dynamical systems, in the form of the following corollary:

Corollary 1.2 *For every affine interval exchange transformation $T_0 : [0,1] \to [0,1]$, the set of parameters t such that the map $x \mapsto T_0(x) + t \mod 1$ has a periodic orbit is dense in $\mathbb{R}$.*

Results about particular families of dynamical systems of this type are usually difficult to prove; a result analogous to Corollary 1.2 where T_0 is an arbitrary generalized interval exchange map seems out of reach of current methods.

1.1 Affine structures on surfaces

The question of the existence of closed geodesics can be considered in the wider context of affine (complex or real) structures on surfaces.[1] For Riemannian structures, the existence of closed geodesics has been known for a long time (see for example Gromoll and Meyer [2]). The case of translation surfaces, which lies in the intersection of the affine and Riemannian world, is now very well understood. On the contrary, for general affine structures very little is known. We therefore pose the following problem.

Problem 1.3 Characterize the affine structures on closed surfaces which carry a regular[2] closed geodesic.

Note that a complete solution to this problem is likely to be very difficult, as it contains as a particular case the notoriously hard question of determining whether the billiard flow of every polygonal table has a periodic orbit.

1.2 Dilation surfaces vs general affine surfaces

Dilation surfaces are particular complex affine surfaces whose structural group is the set of transformations of the form $z \mapsto az + b$ where a is a positive real number and $b \in \mathbb{C}$. Although it is expected that generic complex affine surfaces do not have any closed geodesics, our main theorem predicts that any dilation surface does.

We explain what the condition on the structural group defining dilation surfaces implies at the dynamical level. Every (complex or real) affine structure induces a geodesic foliation on $T^1\Sigma$ the unit tangent bundle of the surface. $T^1\Sigma$ is a three-dimensional manifold, thus the dynamical system induced by the foliation is essentially two-dimensional. Indeed, for a given Poincaré section, the first return map may change both the direction and the position of the intersection of the leaf with the interval.

[1] A real (resp. complex) affine structure on a surface is an atlas of charts taking values in $\mathbb{R}^2$ (resp. $\mathbb{C}$) such that transition maps lie in the group of real affine transformations $\mathrm{GL}^+(2,\mathbb{R}) \ltimes \mathbb{R}^2$ (resp. complex affine transformations $\mathbb{C}^* \ltimes \mathbb{C}$), with possibly finitely many cone-type singularities.

[2] A regular closed geodesic is a closed geodesic that does not contain any singularity of the affine structure.

In the particular case of dilation surfaces, $T^1\Sigma$ decomposes into a one-parameter family of invariant surfaces for the foliation. While this gives no indication as to which affine structures always have periodic leaves, it explains why dilation surfaces are essentially different from the general case:

- the problem for dilation surfaces is about finding periodic orbits in *one-parameter families* of one-dimensional dynamical systems,
- the problem for the generic affine surface is about finding a periodic orbit for a *given* two-dimensional dynamical system.

The analysis of two-dimensional dynamical systems is far more intricate than that of their one-dimensional counterparts; furthermore with dilation surfaces we have an entire one-parameter family of one-dimensional dynamical systems (which are easier to analyze) to find a periodic orbit. This discussion also explains why, despite the fact that in principle it is plausible that a lot of real affine surfaces carry closed geodesics, the dilation case is of a different nature and probably easier to analyze.

1.3 The action of $\mathrm{SL}(2,\mathbb{R})$ and strategy of proof

We now explain the ideas behind the proof of Theorem 1.1. It is very much inspired by the translation case, and we remind the reader of the general structure of its proof. We refer to Masur [4] for the original proof in the translation case.

Both moduli spaces of dilation and translation surfaces carry an action of the group $\mathrm{SL}(2,\mathbb{R})$. This action is naturally defined by the postcomposition of the charts defining the dilation/translation structure. It has the following important property: two surfaces are on the same $\mathrm{SL}(2,\mathbb{R})$-orbit if and only if they define the same underlying real affine structure. In particular, if a surface has a closed geodesic, it is the case for every surface in its $\mathrm{SL}(2,\mathbb{R})$-orbit.

In the translation case, the proof goes by induction on the combinatorial complexity of the surface.[3]

(1) It is easy to check that translation surfaces of lowest complexity (flat tori) always carry closed geodesics.

(2) Assume that we know that all surfaces of complexity less than k do carry closed geodesics, and consider a translation surface Σ of complexity k. It is not hard to find a sequence $(\Sigma_n)_{n\in\mathbb{N}}$ of translation surfaces in the $\mathrm{SL}(2,\mathbb{R})$-orbit of Σ which diverges, ie leaves any compact subset in the moduli space of surfaces of complexity k.

(3) Geometric tools building on the Riemannian structure of translation surfaces allow us to show the following dichotomy: either $(\Sigma_n)_{n\in\mathbb{N}}$ Gromov–Hausdorff converges (up to passing to a subsequence) towards a translation surface of less complexity, or the Riemannian diameter of Σ_n tends to infinity.

[3] We define the complexity to be the number of triangles in a triangulation whose set of vertices is the set of singular points of the surface.

(4) In the first case, having a cylinder is a property that is open in parameter space, and by the induction hypothesis, for n large enough, Σ_n has a closed geodesic. Since Σ_n has the same real affine structure as Σ, so does Σ.

(5) In the second case, an elegant lemma due to Masur and Smillie [5, Corollary 5.5] ensures that a translation surface of large diameter contains a long flat cylinder and thus contains a closed geodesic, which concludes the proof.

This strategy relies heavily on the Riemannian nature of translation surfaces to get a rather simple analysis of the ways a sequence of translation surfaces can degenerate; this part of the proof breaks down when trying to generalize it to the case of dilation surfaces. Most of the work done here is to replace the last three points of the strategy outlined above by a suitable analysis of the different ways a sequence of dilation surfaces can degenerate. We will give a precise roadmap of the proof in Section 4. The three key technical steps of the proof (Propositions 4.1, 4.3 and 4.4) are proved respectively in Sections 5, 6 and 7.

1.4 An important shortcoming and an open problem

We prove that every dilation surface contains a closed geodesic, but unfortunately we were not able to infer anything concerning the nature of the cylinder carrying this closed geodesic. In particular, our proof does not preclude the existence of a dilation surface which is not a translation surface all of whose cylinders are flat (although the existence of such a surface seems highly unlikely).

Problem 1.4 Show that a dilation surface whose cylinders are all flat is a translation surface.

Acknowledgements Ghazouani is greatly indebted to Bertrand Deroin for introducing him to the topic of affine structures on surfaces and asking him the question that lead to the present article. Tahar would like to thank Dmitry Novikov for interesting feedback. The authors are grateful to the referee for valuable remarks and discussions.

2 Dilation surfaces

The symbol Σ will always stand for a compact surface of genus $g \geq 0$ with a finite number of boundary components.

2.1 Dilation cones

Singularities of dilation surfaces are modeled on singularities of dilation cones.

For any $k \in \mathbb{N}^*$, a *flat cone* of angle $2k\pi$ is obtained as the cyclic cover of $\mathbb{C}$ of degree k ramified at 0. The vertex 0 is a *cone point* of angle $2k\pi$ in the flat cone.

For any $k \in \mathbb{N}^*$ and any $\lambda \in \mathbb{R}^*$, a *dilation cone* of angle $2k\pi$ and multiplier λ is obtained from a flat cone of angle $2k\pi$ by cutting a slit along a half-line starting from the vertex 0 and identifying the left side with the right side by a homothety of multiplier λ. The vertex 0 is then a *cone point* of angle $2k\pi$ and dilation multiplier λ.

In particular, for the affine structure induced by the gluing, the holonomy of any closed simple loop around the vertex is a homothety of dilation multiplier λ.

2.2 Generalities

The main objects we will deal with are dilation structures, defined as follows:

Definition 2.1 A *marked topological surface* is a topological surface Σ — possibly with boundary — with a nonempty finite set $S \subset \Sigma$ of *marked points* such that each boundary component contains an element of S.

A *dilation structure* on a marked topological surface (Σ, S) is an atlas of charts $\mathcal{A} = (U_i, \varphi_i)_{i \in I}$ on $\Sigma \setminus S$ such that

- the transition maps are locally restrictions of elements of $\mathrm{Aff}_{\mathbb{R}_+^*}(\mathbb{C}) = \{z \mapsto az + b \mid a \in \mathbb{R}_+^*, b \in \mathbb{C}\}$,
- each marked point in the interior of Σ has a punctured neighborhood which is affinely equivalent to a punctured neighborhood of the cone point of a dilation cone,
- each marked point on the boundary of Σ has a punctured neighborhood which is affinely equivalent to a neighborhood of the center of a Euclidean angular sector of arbitrary angle,
- unless it is a marked point, each point of the boundary of Σ has a punctured neighborhood which is affinely equivalent to a neighborhood of the center of a Euclidean angular sector of angle π.

Elements of S are the *singularities* of the dilation structure.

A particularly simple way of constructing a dilation surface is to glue planar polygons together by using translations and dilations as illustrated in Figure 1. We will see that, up to addition of finitely many singularities with an angle of 2π and a trivial dilation multiplier, every dilation surface can be constructed in this way.

Note that the notion of a straight line on the surface is well defined, since changes of coordinates are affine maps. Moreover, in any direction $\theta \in \mathbb{RP}^1$, the foliation by straight lines of $\mathbb{C}$ in the direction defined by θ being invariant by dilation maps, it gives rise to a well-defined oriented foliation $\mathcal{F}_\theta$ on any dilation surface. Such a foliation is called a *directional foliation*. We call the resulting family of foliations the *directional foliations*; it is indexed by $\mathbb{RP}^1$ and denoted by $(\mathcal{F}_\theta)_{\theta \in \mathbb{RP}^1}$. We shall call any oriented leaf of one of these foliations a *trajectory*.

Definition 2.2 Let Σ be a dilation surface.

- A *closed geodesic* in Σ is a periodic leaf of a directional foliation.

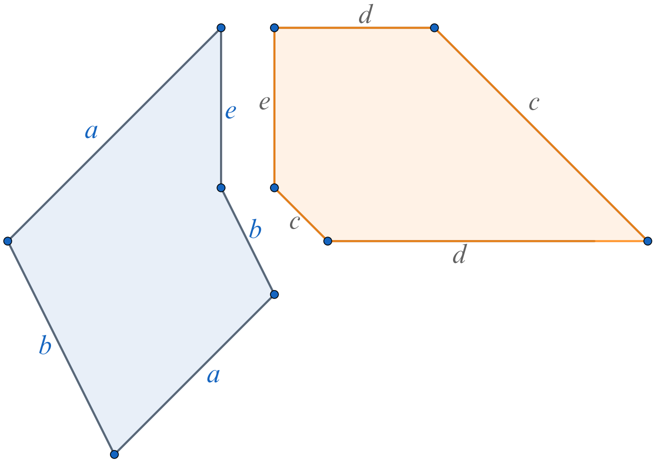


Figure 1: The sides of the two polygons are glued according to their names. Topologically, the resulting surface has genus two and has only one singularity, which corresponds to the extremal points of these two polygons.

- A *saddle connection* is a topological segment on the surface $\Sigma \setminus S$ which is also a straight line (a piece of leaf of a directional foliation) and whose boundary consists of two singularities (possibly identical).

We conclude this subsection with the following definition, which we will use to measure the complexity of a dilation surface:

Lemma 2.3 *We consider a compact topological surface X of genus g with b boundary components, n_i marked points in its interior and n_b marked points on its boundary.*

Assuming $n_b + n_i \geq 1$ and that every boundary component contains at least one marked point, any topological triangulation of X whose set of vertices coincides with the marked points of X is formed by exactly $4g + 2n_i + 2b + n_b - 4$ topological triangles.

Proof The Euler characteristic $\chi(X)$ of surface X is $2-2g-b$. For any such topological triangulation, the number of vertices is $n_i + n_b$. Thus $2-2g-b = T - A + n_i + n_b$, where T is the number of triangles in the triangulation and A is the number of arcs.

Connected components of the boundary are loops. Thus the number of boundary arcs is exactly n_b. Every arc has two sides (excepted the boundary arcs). Thus $3T = 2A - n_b$. We have $4-4g-2b = 2T - 2A + 2n_i + 2n_b$. It follows that $4-4g-2b = -T + 2n_i + n_b$ and thus $T = 4g + 2n_i + 2b + n_b - 4$. □

Definition 2.4 The *complexity* of a marked topological surface is the number of triangles of any topological triangulation whose set of vertices is exactly the set of marked points. By convention, we define the complexity of the empty set to be zero.

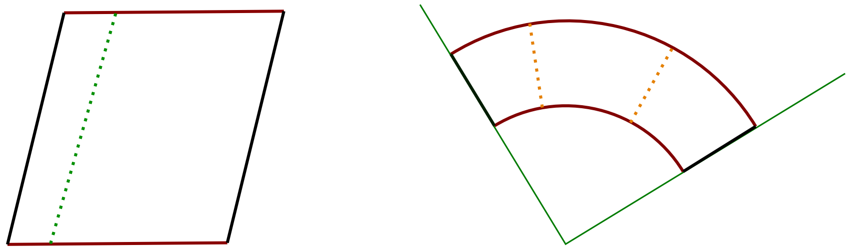


Figure 2: Left: a flat cylinder with a closed geodesic represented by the dashed line (corresponding to the only direction in which there is a closed geodesic). Right: a dilation cylinder with two closed geodesics of two different directions.

We define the complexity of a dilation surface as the complexity of the underlying marked topological surface.

2.3 Cylinders

Cylinders are the geometric counterpart of the periodic leaves of the directional foliations as, in particular, each cylinder contains a closed geodesic. Conversely, any neighborhood of a closed geodesic contains a portion of cylinder. Here we always understand cylinders as maximal: we say that a cylinder is maximal if it is not included in any cylinder but itself.

A *flat cylinder* is a dilation surface with boundary obtained by gluing a pair of opposite sides of a parallelogram embedded in $\mathbb{R}^2$.

A *dilation cylinder* is a dilation surface (with boundary) obtained by cutting a sector C_θ of angle θ in the universal cover of $\mathbb{C}^*$. The quotient of C_θ by the dilation $z \mapsto \lambda z$ with $\lambda > 1$ real is called a *dilation cylinder* (see Figure 2).

2.4 Moduli of cylinders

In this subsection we give an interpretation of conformal moduli of cylinders in dilation structures.

- Recall that the modulus of a flat cylinder obtained from a rectangle of base $(z_1, z_2) \in \mathbb{C}^2$ where the sides glued together are those corresponding to z_2 is by definition $|z_2|/|z_1|$.
- A dilation cylinder of angle θ and dilation multiplier $\lambda > 1$ is biholomorphic (using the exponential map) to the flat cylinder obtained from a rectangle of base $(\ln(\lambda), i\theta)$. Its conformal modulus is thus $\theta/\ln(\lambda)$.

We call a closed geodesic within a cylinder a waist curve of this cylinder.

Lemma 2.5 *There is an absolute constant $M > 0$ such that for any pair of cylinders C_1 and C_2 of conformal modulus at least M in a dilation surface Σ, either C_1 and C_2 are disjoint, or their waist curves are in the same homotopy class.*

Proof A consequence of the Margulis lemma is the existence of a universal constant ϵ such that in any hyperbolic surface of unit area, closed geodesics of length smaller than ϵ are automatically disjoint (see [3, Section 4.2.4] for a reference).

In the conformal class of Σ (punctured at the singularities), we consider the unique hyperbolic metric of unit area. In the homotopy class of waist curves of cylinder C_1 (resp. C_2), there is a unique simple closed geodesic γ_1 (resp. γ_2). We denote by $l(\gamma_1)$ and $l(\gamma_2)$ the lengths of geodesics γ_1 and γ_2. Assuming that waist curves of C_1 and C_2 do not belong to the same homotopy class, γ_1 and γ_2 are distinct.

Following the interpretation of conformal modulus in terms of extremal length, if the conformal modulus of C_1 is strictly bigger than $M = \epsilon^{-2}$, then $l(\gamma_1) < \epsilon$. The same holds for C_2 and $l(\gamma_2)$. Thus γ_1 and γ_2 are disjoint, and waist curves of C_1 and C_2 do not intersect. □

Corollary 2.6 *For any dilation surface Σ, there exists a constant $M(\Sigma) > 0$ such that any cylinder in Σ has conformal modulus smaller than $M(\Sigma)$.*

Proof We assume for contradiction that Σ contains an infinite family of cylinders of arbitrarily large moduli. Since two different cylinders always define two different free homotopy classes, we can always find intersecting cylinders with arbitrarily large moduli. This contradicts Lemma 2.5. □

2.5 Pencils

Pencils were studied in [7] to make explicit the geometric properties of strict dilation surfaces in comparison with translation surfaces. We gather the needed results in this section.

Definition 2.7 A *pencil* is a continuous family of oriented trajectories starting from the same point. Let x be a (possibly singular) point of a dilation surface Σ, and I an open interval of $\mathbb{RP}^1$. The notation $P(x, I)$ will refer to a pencil of trajectories starting at x and covering directions of I.

It should be noted that there are usually several pencils for a given pair (x, I).

The following statement provides a geometric criterion for the existence of dilation cylinders:

Lemma 2.8 [7, Lemma 3.3] *Let x be a point in a dilation surface Σ (possibly with boundary) and I be an open interval of $\mathbb{RP}^1$. For a given pencil $P(x, I)$, at least one of the following statements must hold*:

(1) *a trajectory of $P(x, I)$ hits a singularity*,

(2) *there exists a closed geodesic whose direction belongs to the interval I*,

(3) *there is an open subset $J \subset I$ such that trajectories of the restricted pencil $P(x, J)$ cross the interior of a boundary component of Σ.*

Note that in the case where Σ is without boundary then only the two first items can hold.

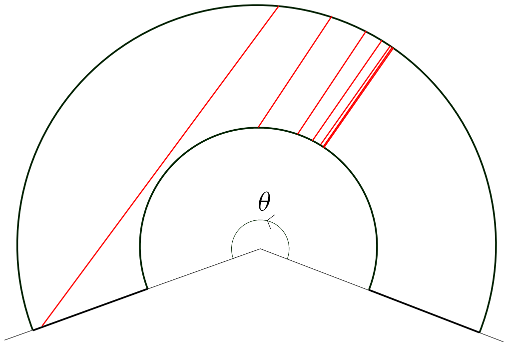


Figure 3: A fundamental domain for the action of $z \mapsto 2z$ on a cone of angle $\theta > \pi$. Any trajectory entering the cylinder is trapped within it forever regardless of the direction of the trajectory, as the one represented here. This property prevents a polygonation from "connecting" both sides of the cylinder.

A second result proves the existence of dilation cylinders in dilation surfaces with nonempty boundary where trajectories of a pencil avoid the boundary:

Proposition 2.9 [7, Corollary 4.6] *Let Σ be a connected dilation surface with a nonempty boundary, a point $x \in \Sigma$ and an open interval I in $\mathbb{RP}^1$. Then at least one of following statements holds:*

(1) *there is an open interval $J \subset I$ such that every trajectory of the restricted pencil $P(x, J)$ accumulates on a closed geodesic of a dilation cylinder of Σ,*

(2) *there is an open interval $J \subset I$ such that every trajectory of $P(x, J)$ crosses the interior of a boundary saddle connection of Σ.*

2.6 Nonpolygonable surfaces

Definition 2.10 A polygonation of a dilation surface Σ is family of saddle connections $\gamma_1, \dots, \gamma_k$ such that

(i) interiors of saddle connections $\gamma_1, \dots, \gamma_k$ are disjoint,

(ii) connected components of $\Sigma \setminus \bigcup_{i=1}^{k} \gamma_i$ are flat polygons without any interior singularity.

A surface Σ is polygonable if it admits a polygonation.

Veech's criterion provides a geometric characterization of polygonable surfaces. This theorem is optimal since cylinders of angle at least π are not polygonable, as shown in Figure 3.

Theorem 2.11 (Veech's criterion [1]) *For a closed dilation surface Σ containing at least one singularity, the three following propositions are equivalent:*

- *Σ is polygonable,*
- *Σ does not contain a dilation cylinder of angle at least π,*
- *every affine immersion of the open unit disk $\mathbb{D} \subset \mathbb{C}$ in Σ extends continuously to its closure $\overline{\mathbb{D}}$.*

Remark 2.12 Up to adding enough singularities of angle 2π and trivial dilation multiplier, we can nevertheless decompose cylinders of angle at least π into smaller cylinders and then into polygons.

For our purpose, Theorem 2.11 proves in particular that every dilation surface that is not polygonable carries cylinders, and one can focus on polygonable surfaces.

2.7 The action of $\mathrm{SL}(2,\mathbb{R})$

We now define a natural action of $\mathrm{SL}(2,\mathbb{R})$ on the space of dilation surfaces.

Let Σ be a dilation surface and consider $A \in \mathrm{SL}(2,\mathbb{R})$. Let $(U_i, \varphi_i)_{i\in I}$ be a maximal atlas defining the dilation structure of Σ. Define $A\Sigma$ to be the dilation structure defined by the maximal atlas $(U_i, A\circ\varphi_i)_{i\in I}$ where A acts on $\mathbb{C}$ via the standard identification $\mathbb{C} \simeq \mathbb{R}^2$. This new atlas indeed defines a dilation structure, as $\mathrm{SL}(2,\mathbb{R})$ centralizes the group formed by maps $z \mapsto az+b$ where $a \in \mathbb{R}_+^*$ and $b \in \mathbb{C}$.

If the dilation surface was given by gluing a bunch of polygons together, the new surface is also polygonable. Indeed, the image of the initial set of polygons with edges identified is mapped by the linear action of the matrix A to another set of polygons. Since A is linear, the sides of the polygon that were parallel are still parallel after applying the matrix A, so that one can still glue them using dilations of the plane. The resulting dilation surface is the image of Σ under the matrix A.

A remarkable subgroup of $\mathrm{SL}(2,\mathbb{R})$ is formed by matrices

$$g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$$

for $t \in \mathbb{R}$. The flow expands the horizontal direction and contracts the vertical direction.

3 Delaunay polygonations

3.1 Delaunay polygonation

The goal of this subsection is to define the Delaunay polygonation of a (polygonable) dilation surface. The construction we will give actually is Veech's proof of Theorem 2.11. To show that surfaces that do not carry cylinders of angle larger than π are polygonable, he proved that the following construction defines a polygonation. We refer to [1] for the full proof and will only describe it here.

The vertices of this polygonation are by definition the singularities of Σ. The edges of the polygonation are saddle connections: a given saddle connection between singularities s_1 and s_2 belongs to the edges of the Delaunay triangulation if there is a closed disk immersed in Σ such that s_1 and s_2 belong to the boundary circle of this disk and such that there are no other singularities in its interior. A disk in Σ is said to be Delaunay if it does not contain any singularities in its interior but at least three on its boundary.

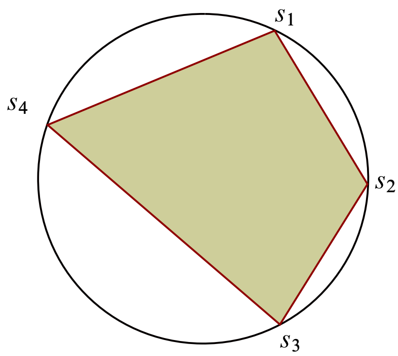


Figure 4: A Delaunay disk with four boundary singularities. Their convex hull in the disk is a face of the Delaunay polygonation.

The faces correspond to what is left after suppressing the edges and the vertices: they are convex polygons whose extremal points all belong to the same Delaunay disks. Figure 4 illustrates the construction: the disk is a Delaunay disk whose boundary contains singularities s_1, s_2, s_3 and s_4. The quadrilateral is one of the faces of the polygonation while its four sides are edges.

Note that the Delaunay polygonation gives you a way to recover from an "abstract" polygonable dilation surface a concrete set of polygons that defines it.

Remark 3.1 Here, even if surfaces with boundary may appear, we will only consider Delaunay polygonations of dilation surfaces without boundary.

3.2 Polygons up to dilation and their limits

In this subsection we consider the space of polygons with exactly $p \geq 3$ vertices arising from Delaunay polygonation. We consider these polygons *as marked and up to dilation*, which means that

- we think of Delaunay polygons as within the unit circle, as we can use a dilation to map the Delaunay circle to the unit one,
- we keep track of the role of each side and each vertex, which is what we mean by marked,
- a polygon and its image under a rotation are considered to be different (because polygons are considered up to dilation and not similarity).

We denote the set described above by $\mathcal{P}_p$. Each polygon is characterized by a p-tuple $(\theta^1, \dots, \theta^p) \in (\mathbb{R}/2\pi\mathbb{Z})^p$, where θ^i is the angle of vertex i in the Delaunay circle and $\theta^i \neq \theta^j$ for $i \neq j$.

Definition 3.2 Let p be fixed. Consider a sequence of polygons $(P_n)_{n\in\mathbb{N}}$ in $\mathcal{P}_p$. We say that this sequence is *Delaunay-convergent* if the following conditions hold:

- the cyclic ordering of the vertices in the circle is constant,
- each vertex $(\theta^i_n)_{n\in\mathbb{N}}$ converges in the circle.

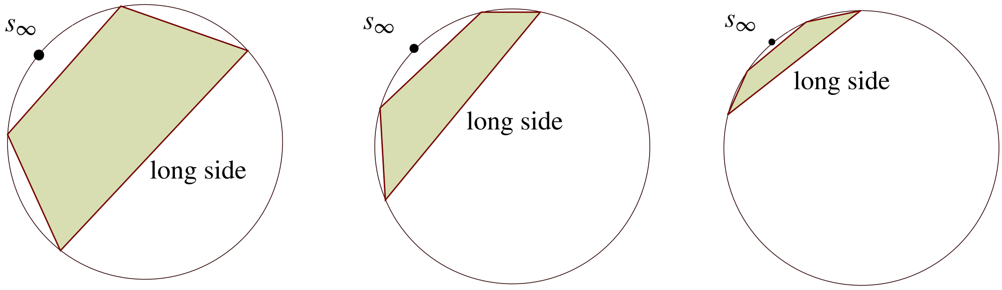


Figure 5: The three first polygons of a degenerating sequence of type 1 whose vertices all converge toward s_∞.

Besides, a Delaunay-convergent polygon is

- a *polygon of type* 1 if all the vertices of $(P_n)_{n\in\mathbb{N}}$ converge towards a given point s_∞ of the circle containing all vertices of P_n (see Figure 5),
- a *polygon of type* 2 if the set of vertices of $(P_n)_{n\in\mathbb{N}}$ converges towards a set of exactly two points s^1_∞ and s^2_∞ of the Delaunay circle (see Figure 6),
- a *polygon of type* 3 if the set of vertices of $(P_n)_{n\in\mathbb{N}}$ converges towards a set of at least three vertices.

In the second case, the slope of the limit edge in $\mathbb{RP}^1$, relating the two remaining vertices, is called the *limit slope*.

By compactness, one can from any sequence of polygons $(P_n)_{n\in\mathbb{N}}$ extract a Delaunay-convergent subsequence.

We now introduce the following terminology which will be useful when proving our main theorem:

Definition 3.3 If $(P_n)_{n\in\mathbb{N}}$ is of type 1, the longest side of the polygon, corresponding to the closest one from the center of the circle in which it is inscribed, is called the *long side*, while the other sides will be called *short*. The *direction* of the polygon in $\mathbb{RP}^1$ is the direction given by its long side

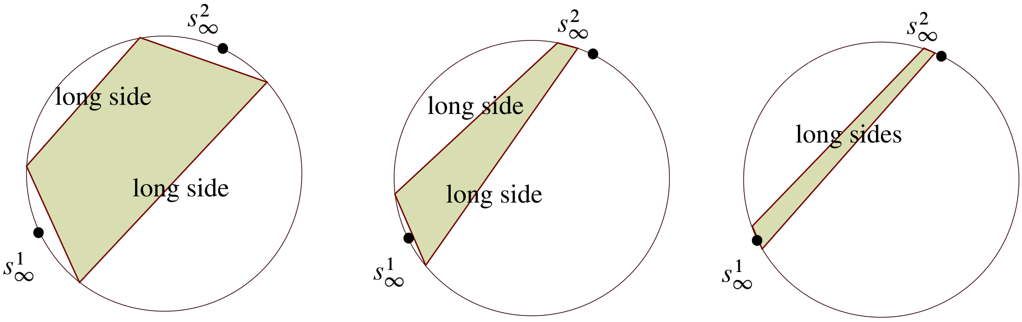


Figure 6: The three first polygons of a degenerating sequence of type 2 whose vertices all converge toward either s^1_∞ or s^2_∞.

If $(P_n)_{n\in\mathbb{N}}$ is of type 2 or 3, the longest sides whose vertices converge to different limit point will be called the *long sides* while the others sides are called *short*. In the type-2 case, the directions tangent to the circle at the two remaining vertices are called the *short side limit slopes* as the short sides are asymptotic to this direction.

Be aware that the terminology is about sequences of polygons, and more precisely about their asymptotic behavior: one can change finitely many polygons of the sequence without changing its long or short sides.

By a harmless abuse of notation, we will refer to a sequence of polygons $(P_n)_{n\in\mathbb{N}}$ in a Delaunay-convergent sequence as a *polygon*. We will use the terms of degenerating polygons. The terms long sides and short sides for these polygons will refer similarly to sequences of edges.

3.3 Delaunay-convergent sequences of dilation surfaces

In this subsection we consider sequences of dilation surfaces of fixed topological type (the underlying marked topological surfaces are isomorphic).

Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of dilation surfaces of same topological type. Up to extracting a subsequence, we can assume that their Delaunay polygonations are all combinatorially equivalent. Precisely, this means that for any $n\in\mathbb{N}$ their Delaunay polygonations have the same pattern. We label for each n the set I of polygons $(P_{i,n})_{i\in I}$ in such a way that

- the sequence $(P_{i,n})_{n\in\mathbb{N}}$ has always the same numbers of sides,
- one can mark the sides of the polygons so that the gluing pattern of the sides of the marked polygons $(P_{i,n})_{n\in\mathbb{N}}$ is constant with respect to the marking.

In that case we say that sequence of surfaces $(\Sigma_n)_{n\in\mathbb{N}}$ has *constant Delaunay pattern*.

Definition 3.4 A sequence $(\Sigma_n)_{n\in\mathbb{N}}$ is said to be *Delaunay-convergent* if

(1) the sequence $(\Sigma_n)_{n\in\mathbb{N}}$ has constant Delaunay pattern,

(2) every polygon $(P_{i,n})_{n\in\mathbb{N}}$ is Delaunay convergent (see Definition 3.2).

We refer to the edges of these polygons as the *Delaunay edges* of the pattern.

For a given sequence of dilation surfaces of fixed topological type there are finitely many Delaunay patterns, so we can always extract a Delaunay-convergent subsequence.

3.4 Maximal domains of type 1

Properties of Delaunay polygonations induce constraints on the Delaunay patterns involving polygons of type 1.

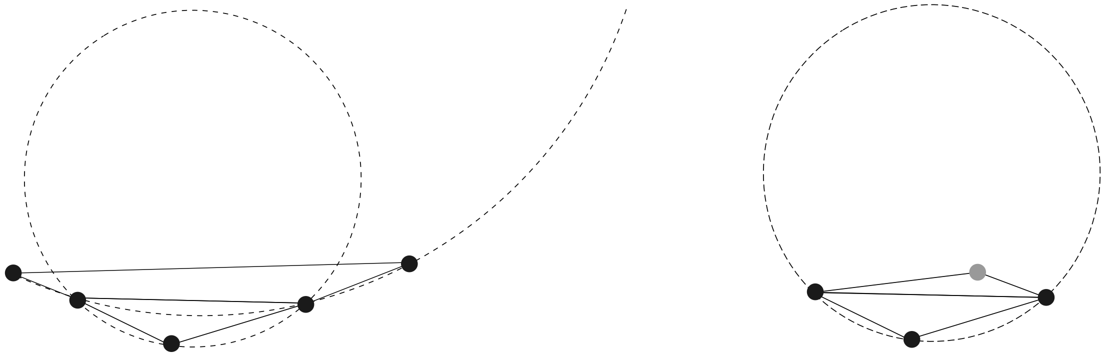


Figure 7: Left: A configuration of two polygons of type 1. None of the singularities lie inside the two Delaunay disks. As the polygons shrink, the Delaunay disks tend to cover half a plane. Right: A forbidden configuration. The gray dot lies inside a large Delaunay disk.

Proposition 3.5 *In a Delaunay-convergent sequence* $(\Sigma_n)_{n\in\mathbb{N}}$ *of polygonable dilation surfaces, the long side* $(L_n)_{n\in\mathbb{N}}$ *of a polygon* $(P_n)_{n\in\mathbb{N}}$ *of type* 1 *can only be incident to a short side of a polygon* (*of any type*).

Proof We assume that in addition to being a side of $(P_n)_{n\in\mathbb{N}}$, the edge $(L_n)_{n\in\mathbb{N}}$ is a long side of a polygon $(Q_n)_{n\in\mathbb{N}}$. In these cases, vertices of $(P_n)_{n\in\mathbb{N}}$ distinct from the ends of $(L_n)_{n\in\mathbb{N}}$ are included for n large enough in the interior of the Delaunay disk of polygons $(Q_n)_{n\in\mathbb{N}}$; see Figure 7. This is a contradiction. □

Following Proposition 3.5, the long side of a polygon of type 1 is always incident to the short side of another polygon. We say that two polygons of type 1 belong to the same *domain of type* 1 if the long side of the first is incident to a short side of the second. The equivalence relation generated by these relations defines classes. This way, each polygon of type 1 belongs to a unique *maximal domain of type* 1.

Besides, in a maximal domain of type 1, any internal edge is a long side of a polygon while being a short side of another (it may happen that the polygons coincide). Thus the incidence graph of a maximal domain is actually an oriented graph with a unique oriented edge leaving each vertex (since each polygon of type 1 has a unique long side). It follows from that there are two types of maximal domains of type 1:

- *noncyclic domains*, where the incidence graph is a rooted tree (the edges being oriented towards the root),
- *cyclic domains*, where the oriented incidence graph contains a unique (oriented) cycle.

Since a maximal domain of type 1 is connected, there is no other type of graphs of incidence.

3.5 Maximal domains of type 2

Polygons of type 2 have two long sides. Two polygons of type 2 glued along an edge that is a long side for each of them belong to a same domain of type 2. This way, each polygon of type 2 belongs to a

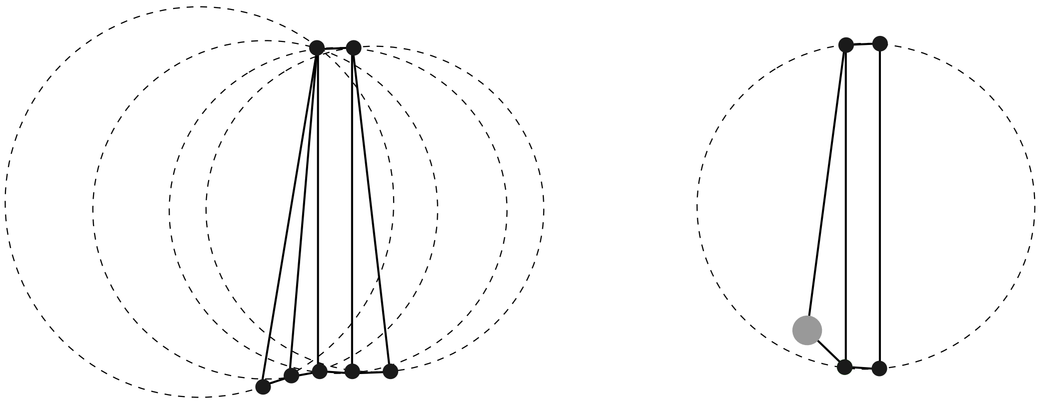


Figure 8: Left: An admissible configuration of four degenerating polygons of type 2. None of the singularities lie inside the Delaunay disks. As the polygons shrink, the Delaunay disks tend to cover a half-plane. Right: A forbidden configuration. The gray dot lies inside a large Delaunay disk.

unique *maximal domain of type* 2. These domains can be *cyclic* or not. In the noncyclic case we call the two long edges that are not glued with another polygon of type 2 the *extremal edges*.

Remark 3.6 The case of a long side of a polygon of type 2 glued along a short side of another polygon of type 2 can happen. We can obtain such a configuration in a variant of the degeneration presented in Figure 9. If the modulus of the connecting flat cylinder decreases to zero (instead of going to infinity), the cylinder is a maximal domain of type 2 and its upper extremal edge is glued on the short side of a polygon of type 2.

An observation that will be needed in the proof of Theorem 1.1 is that short sides of maximal domains of type 2 form "concavely shaped" curves, as shown in Figure 8. It proceeds from the following statement:

Proposition 3.7 *We consider a Delaunay-convergent sequence* $(\Sigma_n)_{n\in\mathbb{N}}$ *of polygonable dilation surfaces. In the polygon formed by two degenerating polygons of type* 2 *or* 3 *glued along a common long side, the magnitude of the limit inner angle between two consecutive short sides is at least* π.

Proof Two consecutive short sides belonging to the same polygon of type 2 or 3 have the same limit slope because their endpoints converge to the same limit point in the Delaunay circle (see Definition 3.3). Therefore, the limit inner angle between them is equal to π.

Now we consider the case of two consecutive short sides $[A_n, B_n]_{n\in\mathbb{N}}$ and $[B_n, C_n]_{n\in\mathbb{N}}$ belonging respectively to two distinct incident degenerating polygons $(P_n^1)_{n\in\mathbb{N}}$ and $(P_n^2)_{n\in\mathbb{N}}$ of type 2 or 3. These two sides have well-defined limit slopes (corresponding to the slope of the tangent line at their limit point in their Delaunay circle). We will assume for contradiction that the limit inner angle θ between these sides at B_n is strictly smaller than π.

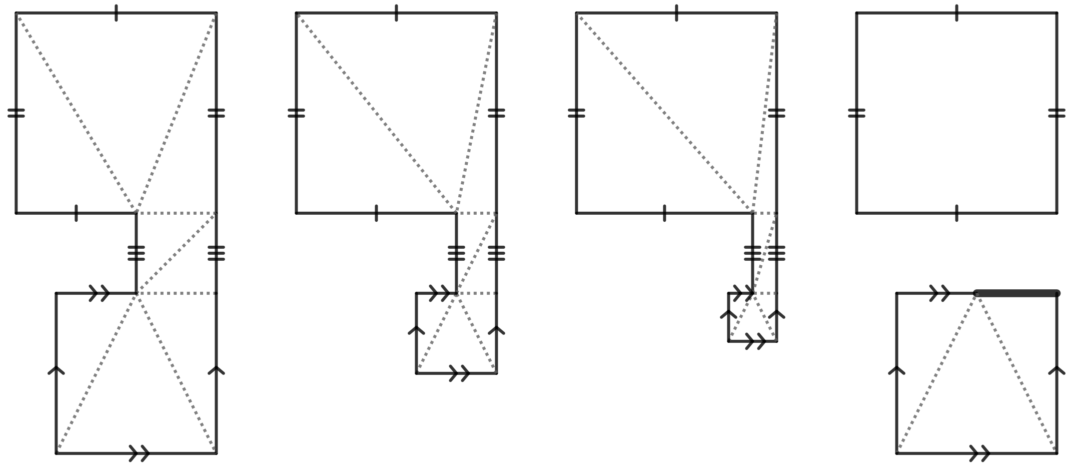


Figure 9: The three first drawings represent some terms of a degenerating dilation surface of genus two with one singularity whose "connecting" flat cylinder degenerates as its modulus goes to infinity. At the level of the Delaunay polygonation, the two parts left after removing the connecting cylinder converge to: (i) a flat torus which is the "Delaunay limit" of the upper part, (ii) a torus with one horizontal boundary component (in bold).

It follows that there is $N > 0$ such that for any $n \geq N$, the segment $[B_n C_n]$ intersects the Delaunay disk $\mathcal{D}_n^1$ of P_n^1. Since by hypothesis C_n cannot belong to P_n^1, the segment $[B_n C_n]$ intersects the boundary of P_n^1 in some point C_n'. The triangle $A_n B_n C_n'$ is inscribed in the Delaunay circle that bounds $\mathcal{D}_n^1$.

Since $P_n^1 \cup P_n^2$ is contractible, it can be endowed with a flat metric in such a way that $[A_n, B_n]_{n\in\mathbb{N}}$ and $[B_n, C_n]_{n\in\mathbb{N}}$ have meaningful lengths. The latter metric is normalized by fixing the radius of the Delaunay disk $\mathcal{D}_n^1$ to 1. As $n \to +\infty$, the length of $[A_n, B_n]$ shrinks to zero. Since the inner angle at B_n converges to $\theta < \pi$, the length of $[B_n C_n']$ converges to some nonzero limit. It follows that the length of $[B_n C_n]$ cannot decrease to zero as n tends to infinity. In $(P_n^2)_{n\in\mathbb{N}}$, the length of $[B_n C_n]$ does not become negligible in comparison with the length of the common edge between $(P_n^1)_{n\in\mathbb{N}}$ and $(P_n^2)_{n\in\mathbb{N}}$. In other words, $[B_n C_n]$ is not a short side of $(P_n^2)_{n\in\mathbb{N}}$, and we get a contradiction. □

3.6 Delaunay limits

For any Delaunay-convergent sequence $(\Sigma_n)_{n\in\mathbb{N}}$ of closed dilation surfaces, we can define a *Delaunay limit* Σ_∞ formed by the polygons that do not completely degenerate; see Figure 9 for an example.

The limit surface Σ_∞ will be a polygonable dilation surface. However, we should be careful. The limit surface can have several connected components. It can also have a boundary, and it can even be empty.

Definition 3.8 Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a Delaunay-convergent sequence of closed dilation surfaces. We define the *Delaunay limit* Σ_∞ in the following way:

- Σ_∞ is the union of limits of polygons of type 3 in $(\Sigma_n)_{n\in\mathbb{N}}$,

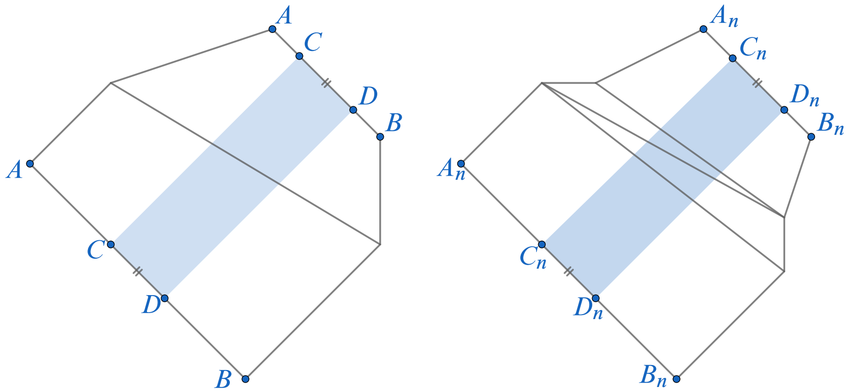


Figure 10: The bands here are shaded. Left: the neighborhood of some periodic orbit. Right: the corresponding band, defined by using the affine identification $\mathcal{A}_n$ of the edges.

- two sides of limit polygons of $(\Sigma_n)_{n\in\mathbb{N}}$ are identified to each other if they are incident in the Delaunay pattern of the sequence,
- the sides of limit polygons are also identified to each other if they are connected by a noncyclic maximal domain of type 2 (see Section 3.5).

As they degenerate to polygons with empty interior, polygons of type 1 and 2 completely disappear in the Delaunay limit Σ_∞ (see Figures 9 and 11).

The simple, but key, feature about this notion of limit is that "carrying a cylinder" is an open property. Namely, if a sequence of dilation surfaces has a Delaunay limit which carries a cylinder, then for any large enough index of the sequence the corresponding surface also carries a cylinder.

Proposition 3.9 *We consider a Delaunay-convergent sequence $(\Sigma_n)_{n\in\mathbb{N}}$. If its Delaunay limit Σ_∞ is nonempty and contains a closed geodesic in some direction $\theta \in \mathbb{RP}^1$, then for any $\epsilon > 0$, there is $N > 0$ such that for any $n \geq N$, Σ_n contains a closed geodesic in a direction of $]\theta-\epsilon, \theta+\epsilon[$.*

Proof We denote by γ a closed geodesic of slope θ in the limit surface Σ_∞. Such a geodesic must cross an edge of the Delaunay polygonation as there is no closed geodesic contained in the interior of a Delaunay polygon. Let us denote by $[A, B]$ an edge crossed by γ and by $[A, B]_\gamma$ the intersection of $[A, B]$ with γ. By definition, closed geodesics do not contain any singularity. It follows that $[A, B]_\gamma$ belongs to the interior of $[A, B]$ (it is not a singularity). Moreover, as $[A, B]_\gamma$ belongs to a periodic leaf of $\mathcal{F}_\theta$, the foliation $\mathcal{F}_\theta$ on Σ_∞ has a well-defined first return map on a neighborhood $[C, D]$ of $[A, B]_\gamma$; see Figure 10.

By definition of the Delaunay limit, the edge $[A, B]$ is the limit edge of a sequence of long sides $([A_n, B_n])_{n\in\mathbb{N}}$ of a polygon of type 3 in the Delaunay polygonation of $(\Sigma_n)_{n\in\mathbb{N}}$. As the polygons converge, the unique (up to translation) complex affine mapping $\mathcal{A}_n$ of the plane that maps $[A_n, B_n]$ to $[A, B]$ converges to the identity as $n \to \infty$. We set $x_n := \mathcal{A}_n^{-1}([A, B]_\gamma)$. Note that $x_n \to [A, B]_\gamma$ as $n \to \infty$.

For n large enough, the leaf of the foliation $\mathcal{F}_\theta$ in Σ_n starting at x_n crosses the edges corresponding to the edges crossed by γ in Σ_∞ (several edges of Σ_n can correspond to the same edge of Σ_∞ if they are long sides of the same noncyclic maximal domain of type 2) and then crosses back $[A_n, B_n]$ at some point.

As the finitely many polygons encountered converge toward nondegenerate polygons or degenerates toward "edges", there is a bound $N > 0$ such that for any $n \geq N$, the first return map T_n of $[A_n, B_n]$ is well defined on a neighborhood $[C_n, D_n] := \mathcal{A}_n^{-1}([C, D])$.

All the oriented leaves of $\mathcal{F}_\theta$ starting from $[C_n, D_n]$, taken up to their first return on $[A_n, B_n]$, give rise to a band $\mathcal{B}_n$ (a parallelogram) whose sides contained in $[A_n, B_n]$ partially coincide. In particular, $\mathcal{B}_n$ contains a closed geodesic. Similarly, we define $\mathcal{B}_\infty$ in Σ_∞.

For an arbitrarily small $\eta > 0$, we can choose a neighborhood $[C, D]$ of $[A, B]_\gamma$ such that the slopes of the two diagonals of $\mathcal{B}_\infty$ are contained in $]\theta - \eta, \theta + \eta[$. Then, provided n is large enough, the slopes of the diagonals of $\mathcal{B}_n$ can be made arbitrarily close to slopes of the diagonals of $\mathcal{B}_\infty$. The slope of a closed geodesic contained in $\mathcal{B}_n$ belongs to an interval whose ends are the slopes of the diagonals of $\mathcal{B}_n$. It follows that for any $\epsilon > 0$, there is $N > 0$ such that for any $n \geq N$, $\mathcal{B}_n$ contains a closed geodesic whose slope is contained in $]\theta - \epsilon, \theta + \epsilon[$. □

We also need to keep track of the polygons involved in the Delaunay limit. To this purpose, we introduce the notion of *core sequence*:

Definition 3.10 For a given Delaunay-convergent sequence $(\Sigma_n)_{n\in\mathbb{N}}$, the *core sequence* $(C\Sigma_n)_{n\in\mathbb{N}}$ is defined for each n as the union of

- polygons of type 3,
- noncyclic maximal domains of type 2 in which at least one extremal edge is incident to a long side of a polygon of type 3.

Remark 3.11 It follows from Definitions 3.8 and 3.10 that each connected component of Σ_∞ corresponds to a unique connected component of $C\Sigma_n$. Boundary saddle connections of Σ_∞ correspond to long boundary edges of the core.

Maximal domains of type 1, maximal domains of type 2 and connected components of the core are the fundamental pieces of the decomposition we will use in the proof of Theorem 1.1.

Definition 3.12 Polygons of $(\Sigma_n)_{n\in\mathbb{N}}$ are grouped into *Delaunay pieces* that are:

- the connected components of the core $(C\Sigma_n)_{n\in\mathbb{N}}$,
- the maximal domains of type 1,
- the maximal domains of type 2 that do not belong to the core.

It follows from Definition 3.10 that every polygon of $(\Sigma_n)_{n\in\mathbb{N}}$ belongs to a unique Delaunay piece.

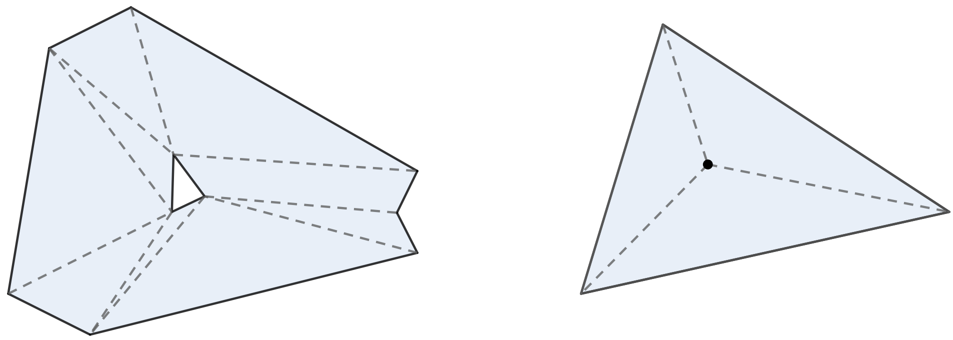


Figure 11: Left: The surface $C\Sigma_n$ of the core sequence with two boundary components. The exterior boundary component contains three long sides, while the interior boundary component is formed by short sides only. Right: The Delaunay limit Σ_∞ with an exterior boundary formed by three sides and a singularity at the center.

The important feature of the decomposition into Delaunay pieces is that their boundary edges cannot be long sides for both of their incident polygons.

Lemma 3.13 *If a Delaunay edge $(L_n)_{n\in\mathbb{N}}$ of $(\Sigma_n)_{n\in\mathbb{N}}$ is a long side for its two incident polygons, then it cannot belong to the boundary of a Delaunay piece.*

Proof It follows from Proposition 3.5 that neither of the incident polygons of $(L_n)_{n\in\mathbb{N}}$ can be polygons of type 1. If both of them are polygons of type 2, then they belong to the same maximal domain and $(L_n)_{n\in\mathbb{N}}$ is not a boundary edge of a Delaunay piece. If at least one of the two polygons incident to $(L_n)_{n\in\mathbb{N}}$ is of type 3, then it follows from Definition 3.10 that $(L_n)_{n\in\mathbb{N}}$ is an interior edge of some connected component of the core $(C\Sigma_n)_{n\in\mathbb{N}}$. □

4 Overview of the proof of Theorem 1.1

As mentioned above, we will argue by induction on the complexity of the closed surface Σ (see Definition 2.4). It is easy to deal with the case of smallest complexity: the flat tori formed by a pair of triangles. The induction assumption is that any closed dilation surface of complexity lower than k carries cylinders in a dense set of directions. We want to show that it is still the case for surfaces of complexity $k+1$.

It follows from Lemma 2.8 that for a dilation surface Σ without boundary, an open set of $\mathbb{RP}^1$ that does not contain any direction of saddle connection contains the direction of a closed geodesic. Therefore, it remains to prove that any direction $\theta \in \mathbb{RP}^1$ that is approached by directions of saddle connections is also approached by directions of closed geodesics. In other words, any open subset of $\mathbb{RP}^1$ containing the direction of a saddle connection should also contain the direction of a closed geodesic. Up to the action of an element of $SL_2(\mathbb{R})$, we can assume that Σ contains a vertical saddle connection γ and U is an open subset of $\mathbb{RP}^1$ containing the vertical direction.

To rely on the induction assumption, we will use the Teichmüller flow g_t that contracts the vertical direction, and therefore the saddle connection γ, and that expands the horizontal direction. We have seen in Section 3.3 that one can extract a sequence of times $t_n \to +\infty$ such that the sequence $(\Sigma_n)_{n\in\mathbb{N}} := (g_{t_n}\Sigma)_{n\in\mathbb{N}}$ is Delaunay convergent. We are doing induction on closed surfaces, and this is why the case where Σ_∞ has a boundary will need to be handled separately. We first rule out the case in which $(\Sigma_n)_{n\in\mathbb{N}}$ Delaunay-converges toward a closed surface Σ_∞ of the same complexity (see Section 3.6 for a precise definition of Delaunay limits).

Proposition 4.1 *Let Σ be a closed dilation surface that carries a vertical saddle connection and a sequence of times $t_n \to +\infty$ such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges toward a surface Σ_∞. Then one of the following statements holds*:

- *for any $\epsilon > 0$, Σ carries a cylinder whose direction belongs to $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$,*
- *Σ_∞ is of strictly smaller complexity than Σ.*

Proposition 4.1 is proved in Section 5. Note that in the first case there is nothing more to be proven. In the second case, as soon as Σ_∞ is nonempty and contains a component without boundary, we can also conclude. Indeed, the induction assumption guarantees that directions of closed geodesics of Σ_∞ are dense in $\mathbb{RP}^1$. In particular, Σ_∞ contains closed geodesics whose directions are arbitrarily close to the vertical direction.

One can now conclude using Proposition 3.9 that provided n is large enough, Σ_n contains a closed geodesic whose direction is arbitrarily close to the vertical direction. It follows that $\Sigma = g_{t_n}\Sigma_n$ contains a closed geodesic whose direction is even closer to the vertical direction.

It then remains to deal with two cases:

- Σ_∞ is empty,
- every connected component of Σ_∞ has a nonempty boundary.

We will deal with these two cases at once by thoroughly examining how the Delaunay polygonations, and especially the Delaunay pieces, degenerate under the Teichmüller flow.

The key notion here is that of *short and long boundary*. We will say that a Delaunay piece (see Definition 3.12) *has a long boundary side* if one of its boundary sides is the long boundary (in the sense of Definition 3.3) of a Delaunay polygon that belongs to the Delaunay piece.

Recall that a Delaunay piece is either a component of the core, or a maximal domain of type 1 or 2. A boundary side of the core can be short or long, but the short ones disappear in Σ_∞ by construction. In particular, if the limit of a Delaunay piece that belongs to the core has a boundary side, then the Delaunay piece in question must have a long boundary side. A maximal domain of type 1 or 2 must have, by construction, a long boundary side, except in the case where it is cyclic. We summarize the content of this discussion within the following structural lemma:

Lemma 4.2 *Let Σ be a dilation surface and $t_n \to +\infty$ be such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges toward a surface Σ_∞ of strictly smaller complexity than Σ. Then at least one the following conditions holds*:

(1) *one of the Delaunay pieces converges toward a closed nonempty dilation surface of smaller complexity*,

(2) *one of the Delaunay pieces is a cyclic maximal domain of type* 1 *or of type* 2,

(3) *all the Delaunay pieces have at least one long boundary side.*

As mentioned above, the first case is dealt with using the induction assumption. The second case will follow from the next result, which will be proven in Section 6.

Proposition 4.3 *Let Σ be a dilation surface, and $t_n \to +\infty$ be such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges and such that at least one of its Delaunay pieces is a cyclic maximal domain of type* 1 *or of type* 2*. Then for any $\epsilon > 0$, Σ carries a cylinder whose direction belongs to $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$.*

One is then left to analyze the last case, in which all the Delaunay pieces have at least one long boundary side. This is the most subtle part of the article.

Proposition 4.4 *Let Σ be a dilation surface, and $t_n \to +\infty$ be such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges and such that all the Delaunay pieces in $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ have at least one long boundary side.*

Then, for any open set $U \subset \mathbb{RP}^1$, there is $N > 0$ such that for any $n \geq N$, Σ_n contains closed geodesics whose directions belong to U.

The above proposition shows in particular that Σ carries closed geodesics whose directions are as close as we want to the vertical one as, as usual, nonhorizontal closed geodesics of Σ_n are images of almost vertical ones of Σ under the Teichmüller flow. This proves that any open set of $\mathbb{RP}^1$ containing the vertical direction contains the direction of a closed geodesic of Σ. The three cases of Lemma 4.2 are settled, and Theorem 1.1 is now proven.

The proof of Proposition 4.4 will be given in Section 7.

5 The nondegenerating case (proof of Proposition 4.1)

In this section, we apply the Teichmüller flow to a closed dilation surface Σ containing a vertical saddle connection. For a sequence of times $t_n \to +\infty$ such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges to a limit surface Σ_∞ with the same complexity as Σ, we prove that Σ_∞ (and subsequently Σ) contains closed geodesics whose directions are arbitrarily close to $\frac{1}{2}\pi$.

Remark 5.1 Before entering the proof of the above proposition, let us mention that there is a class of dilation surfaces, called quasi-Hopf surfaces (see [6] for details), having a vertical saddle connection and whose Teichmüller orbit is periodic. These surfaces decompose into disjoint dilation cylinders whose boundary saddle connections are either horizontal or vertical. We cannot extract from the Teichmüller orbit of these surfaces a sequence that Delaunay-converges to a surface of smaller complexity.

We recall that for positive times, the Teichmüller flow expands the horizontal direction and contracts the vertical direction. We first prove the existence of a vertical saddle connection γ_∞ in Σ_∞.

Lemma 5.2 *Let Σ be a closed dilation surface that carries a vertical saddle connection and a sequence of times $t_n \to +\infty$ such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges toward a surface Σ_∞ having the same complexity as Σ. Then the limit surface Σ_∞ carries at least one vertical saddle connection.*

Proof We first prove that for n large enough the vertical saddle connection of $g_{t_n}\Sigma$ belongs to an edge of the Delaunay polygonation. Indeed, in the dilation surface Σ, there is an affine immersion of an elliptic domain D of eccentricity $e < 1$ such that vertical saddle connection γ coincides with the image of the major axis. The ratio of lengths between the (horizontal) semiminor axis and the (vertical) semimajor axis is $\sqrt{1-e^2}$. Teichmüller flow will deform the ellipse. For $T = -\frac{1}{4}\ln(1-e^2) > 0$, the saddle connection $g_T\gamma$ in the surface $g_T\Sigma$ is the vertical diameter of the immersed disk $g_T D$. Consequently, for any $t \geq T$, $g_t\gamma$ is an edge of the Delaunay polygonation of the dilation surface $g_t\Sigma$.

We now prove that the limit surface carries indeed a vertical saddle connection (that belongs to the Delaunay polygonation). By definition of being Delaunay-convergent, all the Delaunay polygons of the sequence $g_{t_n}\Sigma$ converge toward a limit polygon of Σ_∞. A polygon cannot converge toward a polygon with more sides, so the complexity can only decrease. Assuming that the complexity of Σ_∞ and Σ are the same, all the limit polygons keep the same number of sides. In particular, the side corresponding to the vertical connection does not vanish, and the limit surface carries a vertical saddle connection as well. □

On a dilation surface, vertical saddle connections have a top and a bottom endpoint. For any vertical saddle connection γ, we denote by $\mathcal{R}(\gamma)$ the vertical ray satisfying the following properties:

- The starting point M of $\mathcal{R}(\gamma)$ is the top endpoint of the saddle connection γ.
- At M, $\mathcal{R}(\gamma)$ and γ form an angular sector of amplitude π contained in the right half-plane (by convention). In other words, $\mathcal{R}(\gamma)$ is obtained from γ by turning counterclockwise around M by an angle of π.

We will prove that Σ_∞ contains a cylinder with vertical boundary saddle connection by exhibiting a cyclic sequence of vertical saddle connections.

Lemma 5.3 *For any vertical saddle connection γ in the limit surface Σ_∞, $\mathcal{R}(\gamma)$ is a vertical saddle connection too.*

Proof We argue by contradiction, assuming that Σ_∞ contains a vertical saddle connection γ such that $\mathcal{R}(\gamma)$ is not a saddle connection. We denote by s_{top} the top singularity of γ and choose a continuous

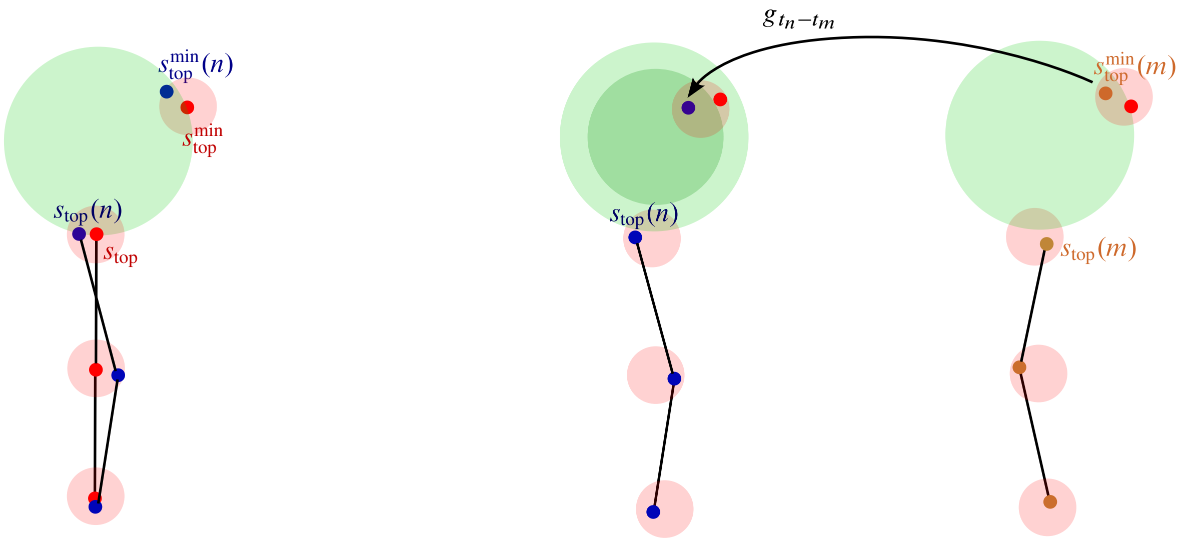


Figure 12: Left: In green, the maximal disk defining the singularity s_{top}^{min}. The red dots correspond to the singularities encountered on an $\mathcal{R}$-orbit in Σ_∞. The blue dots correspond to the associated singularities in Σ_n. Right: The left part corresponds to the surface Σ_n as the right one to Σ_m. The shadow of the closest singularity on Σ_m is mapped within a 2ϵ shorter disk of Σ_n.

parametrization $\Gamma(t)$ of the ray $\mathcal{R}(\gamma)$ with $\Gamma(0) = s_{top}$; see Figure 12. For t small enough, we define $\mathcal{D}_t$ to be the unique immersed disk whose center is $\Gamma(t)$ and whose radius is given by the segment $[s_{top}, \Gamma(t)]$. Note that for any $t > t'$ we have $\mathcal{D}_{t'} \subset \mathcal{D}_t$. We then define the "closest" singularity s_{top}^{min} to γ_∞ as the first singularity encountered when considering the increasing sequence of disks $(\mathcal{D}_t)_{t\geq 0}$. Note that such a sequence must actually encounter a singularity because of Theorem 2.11.

By hypothesis, s_{top}^{min} does not belong to $\mathcal{R}(\gamma)$. We will reach our contradiction by showing that there is a closer singularity to s_{top} than s_{top}^{min}. In order to do so, we rely on the assumption that Σ_∞ is the Delaunay limit of $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$. All polygons of the Delaunay triangulation converge toward their limit polygon, and all the quantities indexed by n must converge toward their ∞-indexed corresponding quantity. For $N > 0$ large enough, any surface $g_{t_n}\Sigma$ contains a well-defined saddle connection γ_n corresponding to γ_∞ in Σ_∞. Analogously we denote by $s_{top}(n)$ and $s_{top}^{min}(n)$ the top singularity of such a sequence and the singularity in $g_{t_n}\Sigma$ corresponding to s_{top}^{min}; see Figure 12.

Then let $m > n$ such that $t_m - t_n > 0$. By construction $g_{t_n-t_m}\Sigma_m = \Sigma_n$. Note that $t_n - t_m < 0$ so that the Teichmüller flow is now expanding (by a definite amount independent of n, m and ϵ) in the vertical direction and contracting the horizontal one. Note also that the image of $s_{top}^{min}(m)$ under $g_{t_n-t_m}$ must be a singularity of Σ_n. Since ϵ is arbitrary, one can take it as small as for the singularity $g_{t_n-t_m}(s_{top}^{min}(m))$ to be inside the disk centered at the same point as the maximal disk defining s_{top}^{min} but of radius 2ϵ shorter. This contradicts our initial assumption: as $g_{t_n-t_m}(s_{top}^{min}(m))$ must be ϵ close to a singularity of Σ_∞, this new singularity would be inside the maximal disk defining s_{top}^{min}; see Figure 12. □

Proof of Proposition 4.1 Assuming that Σ_∞ and Σ have the same complexity, Lemma 5.2 proves that Σ_∞ contains a vertical saddle connection γ_∞. Using repeatedly Lemma 5.3, we get that γ_∞ belongs to a periodic sequence of vertical saddle connections. In other words, two consecutive saddle connections in this sequence differ by an angle of π. The curve formed by these saddle connections becomes simple if moved slightly toward the right. Such a simple curve must be the boundary of a cylinder. Consequently, the vertical saddle connection γ_∞ belongs to the boundary of some cylinder of Σ_∞. Proposition 3.9 guarantees that for any $\epsilon > 0$ there is $N > 0$ such that for any $n \geq N$, $g_{t_n}\Sigma$ contains a closed geodesic whose direction belongs to $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$. For positive times, the Teichmüller flow g_t expands the horizontal direction and shrinks the vertical direction. Thus, a fortiori, the same holds for Σ. □

6 Cyclic maximal domains of type 1 and 2 (proof of Proposition 4.3)

We will show that the existence of a cyclic maximal domain of type 1 or 2 (see Sections 3.4 and 3.5) in the Delaunay limit Σ_∞ of a Delaunay-convergent subsequence $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ of positive Teichmüller orbit of a dilation surface Σ implies the existence of closed geodesic in Σ whose direction is arbitrarily close to the vertical direction.

We will prove Proposition 4.3 by contradiction. We first give estimates on the moduli and directions of cylinders in surfaces of the positive Teichmüller orbit of a dilation surface without closed geodesics in a neighborhood of the vertical direction.

Lemma 6.1 *For $\delta > 0$, we consider a closed dilation surface Σ that does not contain any closed geodesic whose direction belongs to the interval $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$.*

Then there is a positive constant $C_\delta > 0$ such that for any $t \geq 0$ the modulus of any cylinder of the surface $g_t\Sigma$ is bounded above by C_δ.

Besides, for any $\epsilon > 0$, there is a time T_ϵ such that for any $t \geq T_\epsilon$ directions of closed geodesics of $g_t\Sigma$ are contained in $[\pi - \epsilon, \pi]$.

Proof The second claim follows immediately from the action of the Teichmüller flow on the interval $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$ in $\mathbb{RP}^1$.

Recall that Corollary 2.6 asserts that every cylinder of Σ is of modulus at most M for some $M > 0$. We will prove that the moduli of cylinders of surfaces $(g_t\Sigma)_{t\in\mathbb{R}^+}$ satisfy a global upper bound $C_\delta = M/\sin^2(\delta)$. Note that cylinders of $g_t\Sigma$ correspond to cylinders of Σ.

We first consider a flat cylinder C of Σ. Normalizing its area to 1, its modulus is equal to h^{-2} where h is the (normalized) length of its closed geodesics. These geodesics have a direction θ which is δ far away from $\frac{1}{2}\pi$ by assumption. Now we consider the images of C under the action of the Teichmüller flow. The normalized area remains identical while the lengths h_t of closed geodesics of g_tC satisfy

$$\frac{h_t}{h} = \sqrt{e^{2t}\cos^2(\theta) + e^{-2t}\sin^2(\theta)}.$$

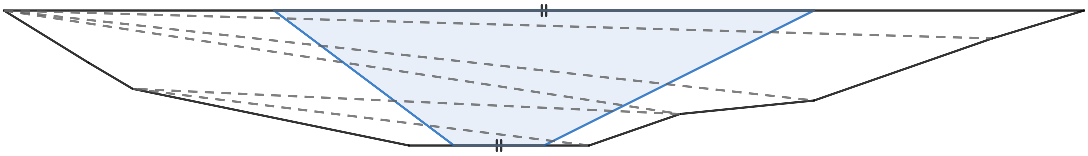


Figure 13: A maximal domain of type 1 containing a dilation cylinder.

It follows that
$$\frac{1}{h_t} \le \frac{1}{\sin^2(\delta)h},$$
which is desired result.

It then remains to deal with the case where the cylinder C is a dilation cylinder of Σ whose directions of closed geodesics is the interval $]\theta_1, \theta_2[\subset \left]-\frac{1}{2}\pi + \delta, \frac{1}{2}\pi - \delta\right[$. We denote by $\lambda > 1$ the dilation multiplier of C. Recall that the modulus M of such a cylinder is given by the relation
$$M = \frac{|\theta_2 - \theta_1|}{\ln(\lambda)}.$$
The action of the Teichmüller flow preserves the dilation multiplier. On the other hand, g_t transforms any slope θ into
$$\arctan(e^{-2t}\tan(\theta)).$$
Using that for any $x, y < 0$, we have $|\arctan(y) - \arctan(x)| \le |y - x|$, we get that the size $|\theta_1(t) - \theta_2(t)|$ of the interval $g_t(](\theta_1, \theta_2[)$ satisfies
$$|\theta_1(t) - \theta_2(t)| \le e^{-2t}|\tan(\theta_1) - \tan(\theta_2)| \le |\theta_1 - \theta_2| \sup_{\theta \in]-\pi/2+\delta, \pi/2-\delta[} |\tan'(\theta)|.$$
Actually, $\sup_{\theta \in]-\pi/2+\delta, \pi/2-\delta[} |\tan'(\theta)| = 1/\sin^2(\delta)$, so we have
$$|\theta_1(t) - \theta_2(t)| \le \frac{|\theta_1 - \theta_2|}{\sin^2(\delta)}.$$
Thus the modulus of the image cylinder is bounded above by $M/\sin^2(\delta)$. □

We split the proof of Proposition 4.3 in two statements, corresponding to the maximal domains of type 1 and type 2.

6.1 Cyclic maximal domains of type 1

Polygons of type 1 assemble into maximal domains of type 1 (see Section 3.4). Following Proposition 3.5, the (unique) long side of a degenerating polygon of type 1 must be glued to the short side of any other polygon. A cyclic maximal domain is formed by polygons of type 1 glued long side on short side, as in Figure 13.

We prove that cyclic maximal domains of type 1 contain dilation cylinders whose angular amplitude is bounded below.

Proposition 6.2 *Let Σ be a dilation surface, and $t_n \to +\infty$ be such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges and such that at least one of its Delaunay pieces is a cyclic maximal domain of type* 1*. Then for any $\epsilon > 0$, Σ carries a cylinder whose direction belongs to $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$.*

Proof Assuming for contradiction that for some $\epsilon > 0$, the interval $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi + \epsilon\right[$ does not contain any direction of a closed geodesic of Σ, we use Lemma 6.1 to prove that the maximal angular amplitude of dilation cylinders of $\Sigma_n = g_{t_n}\Sigma$ becomes arbitrarily small as n tends to infinity. We will obtain a contradiction by proving that the cyclic maximal domain of type 2 $(X_n)_{n\in\mathbb{N}}$ in $(\Sigma_n)_{n\geq\mathbb{N}}$ contains a dilation cylinder whose angular amplitude is bounded below provided n is large enough.

The incidence graph of $(X_n)_{n\in\mathbb{N}}$ is connected and contains a unique (oriented) cycle C (see Section 3.4 for details). For any $n \in \mathbb{N}$, X_n is a topological cylinder. We consider an edge $(L_n)_{n\in\mathbb{N}}$ between two polygons of the cycle C.

Cutting along the edge $(L_n)_{n\in\mathbb{N}}$ in $(X_n)_{n\in\mathbb{N}}$, we obtain a sequence of simply connected flat surfaces $(P_n)_{n\in\mathbb{N}}$ with a unique boundary component. It is formed by the gluing of polygons of type 1 according to an incidence graph which is a tree.

By definition of a polygon of type 1 (see Definition 3.2), all the sides of $(P_n)_{n\in\mathbb{N}}$ have the same limit direction in $\mathbb{RP}^1$. We normalize $(P_n)_{n\in\mathbb{N}}$ in such a way that all the sides tend to be horizontal and every surface P_n has unit area. In particular, provided n is large enough, P_n is a planar polygon in the classical sense.

Since every polygon of type 1 has a unique long side (see Definition 3.3), provided that n is large enough, P_n has a unique upper side S_n (or a unique lower side, depending on the normalization) and several lower sides $T_n^1, \dots, T_n^{p-1}$ (where p is the number of sides of P_n for any n).

The edge $(L_n)_{n\in\mathbb{N}}$ corresponds to the identification of the unique upper side $(S_n)_{n\in\mathbb{N}}$ with some lower side $(T_n^{i_0})_{n\in\mathbb{N}}$. Since S_n and $T_n^{i_0}$ have the same slope, they cannot be adjacent in the boundary of P_n. Thus the corner angles and the ends of $T_n^{i_0}$ tend to π as n tends to infinity. Provided that n is large enough, rays starting from the ends of the side $T_n^{i_0}$ in directions $\frac{1}{4}\pi$ and $\frac{3}{4}\pi$ intersect S_n and there exists a trapezoid M_n in P_n formed by $T_n^{i_0}$ (the lower side of M_n), a side of slope $\frac{1}{4}\pi$, a portion of S_n (the upper side of M_n) and a side of slope $\frac{3}{4}\pi$.

Since any point of the upper side of M_n is identified with a point of $T_n^{i_0}$, the lift of the trapezoid M_n in X_n contains a family of closed geodesics whose slopes sweep an interval of length at least $\frac{1}{2}\pi$ in $\mathbb{RP}^1$ (see Figure 13). Thus, for any large enough n, the surface Σ_n contains a dilation cylinder of angle at least $\frac{1}{2}\pi$. This is the desired contradiction. □

6.2 Cyclic maximal domains of type 2

In Section 3.5, we defined maximal domains of type 2 as collections of polygons of type 2 glued along their long boundary sides. Such a maximal domain is cyclic if the polygons are glued according to a cyclic graph, as in Figure 14.

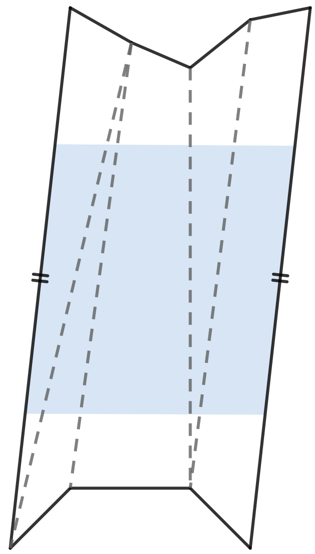


Figure 14: A maximal domain of type 2 containing a cylinder of large modulus (shaded).

We prove that cyclic maximal domains of type 2 contain cylinders of arbitrarily large modulus.

Proposition 6.3 *Let Σ be a dilation surface, and $t_n \to +\infty$ be such that $(g_{t_n}\Sigma)_{n\in\mathbb{N}}$ Delaunay-converges and such that at least one of its Delaunay pieces is a cyclic maximal domain of type* 2. *Then for any $\epsilon > 0$, Σ carries a cylinder whose direction belongs to $\left]\frac{1}{2}\pi-\epsilon, \frac{1}{2}\pi+\epsilon\right[$.*

Proof We proceed similarly as when proving Proposition 6.2. Assuming for contradiction that for some $\epsilon > 0$, the interval $\left]\frac{1}{2}\pi-\epsilon, \frac{1}{2}\pi+\epsilon\right[$ does not contain any direction of a closed geodesic of Σ, we use Lemma 6.1 to obtain an upper bound $M > 0$ on the modulus of any cylinder in any dilation surface $\Sigma_n = g_{t_n}\Sigma$. We will prove that the cyclic maximal domain of type 2 $(X_n)_{n\in\mathbb{N}}$ in $(\Sigma_n)_{n\in\mathbb{N}}$ contains cylinders of arbitrarily large modulus as n tends to infinity.

In particular, for any $n \in \mathbb{N}$, X_n is a topological cylinder. We cut along some edge $(L_n)_{n\in\mathbb{N}}$ and obtain a sequence of polygons $(P_n)_{n\in\mathbb{N}}$. We normalize each polygon P_n in such a way that the two sides corresponding to edge L_n are vertical and P_n has unit area. These two vertical sides will be referred to as S_n (for the left side) and T_n (for the right side).

In a polygon of type 2 (see Definition 3.2), the ratio between the length $|S_n|$ of S_n and the length $|T_n|$ of T_n converges to 1, and the length of any other side of P_n becomes negligible in comparison with $|S_n|$ and $|T_n|$ (see Figure 14). Since P_n has unit area for any $n \in \mathbb{N}$, $|S_n|$ and $|T_n|$ tend to infinity while the distance between S_n and T_n tends to zero as $n \to +\infty$.

It follows that for any $\epsilon > 0$, there is $N > 0$ such that for any $n \geq N$, the polygon P_n contains a rectangle R_n satisfying the following conditions:

- sides of R_n are either vertical or horizontal,

- the vertical left and right sides are portions of S_n and T_n,
- the length of the vertical sides of R_n is at least $(1-\epsilon)|S_n|$,
- the length of the vertical sides of R_n tend to infinity as $n \to +\infty$,
- the length of the horizontal sides of R_n tend to zero as $n \to +\infty$.

Since the sides S_n and T_n are identified, the lift of the rectangle R_n in X_n contains a family of closed geodesics covering most of the rectangle R_n (the complement of a part of arbitrarily small relative area). Therefore, provided that n is large enough, X_n contains a cylinder of arbitrarily large modulus (see Figure 14). □

7 Long sides and short sides (proof of Proposition 4.4)

We will actually prove the following slightly stronger version of Proposition 4.4, which does not involve the Teichmüller flow:

Proposition 4.4 *Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a Delaunay-convergent sequence of dilation surfaces such that all its Delaunay pieces have at least one long boundary side. Then, for any open set $U \subset \mathbb{RP}^1$, there is $N > 0$ such that for any $n \geq N$, Σ_n contains closed geodesics whose directions belong to U.*

The proof is based on Proposition 2.9. This proposition asserts that, in the case of a dilation surface with boundary, either a given open set of directions contains a cylinder, or a set of trajectories having these directions, a pencil to be precise, must leave across a boundary component of the dilation surface. This proposition then shows that we can concentrate on the case where each trajectory of a Delaunay piece (see Definition 3.12) leaves it by hitting the boundary. It can do it by crossing either a long edge or a short one. We will actually rule out the short edge case in Sections 7.1 and 7.2. Indeed, these boundaries are by definition very small compared to the long edges and it will be unlucky to leave the piece through such a short side. The trajectories of the pencil will then have to leave the Delaunay piece through a long side and then enter a new Delaunay piece through a short side (as by construction Delaunay pieces are glued to one another short side to long side; see Lemma 3.13). If the pencil does not enter in a cylinder, one can repeat the argument to get a sequence of Delaunay pieces such that the pencil enters them by short sides and leaves them by long sides. As there are only finitely many boundary components, such a pencil will cross a given edge twice. The first return map on such an edge is a very dilating mapping, as going from short sides to long sides induces a huge contraction. This concludes the argument, as contracting mappings have periodic orbits.

7.1 Trajectories inside maximal domains of type 1 or 2

For a trajectory whose slope is far enough from the limit directions of the (finitely many) Delaunay edges, we have some control on its behavior in Delaunay pieces formed by degenerating polygons.

Definition 7.1 For any $\epsilon > 0$, $\Theta_\epsilon \subset \mathbb{RP}^1$ is the open subset of slopes whose distance to any limit direction of a Delaunay edge of $(\Sigma_n)_{n\in\mathbb{N}}$ is strictly bigger than ϵ.

The following proposition asserts that a for a given direction in Θ_ϵ a trajectory entering a maximal domain of type 1 by a small edge must exit it through a long one, provided that n is large enough.

Proposition 7.2 *Let $(E_n)_{n\in\mathbb{N}}$ be a short boundary edge of a maximal domain of type 1 $(X_n)_{n\in\mathbb{N}}$ that has at least one boundary long edge.*

For any $\epsilon \in \left]0, \frac{1}{2}\pi\right[$, there is a long boundary edge $(M_n)_{n\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$ and $N > 0$ such that for any $n \geq N$, any trajectory of X_n whose direction belongs to Θ_ϵ starting from E_n eventually leaves X_n through the interior of M_n.

Proof Since a maximal domain of type 1 is formed by polygons of type 1, Delaunay edges have the same limit slope. Without loss of generality, we will assume that this unique limit slope is horizontal.

We start by discussing the case of a polygon of type 1. Note that for any $\delta > 0$, there is $N_\delta \in \mathbb{N}$ such that for any $n \geq N_\delta$, every (convex) Delaunay polygon P_n of X_n satisfies the following properties:

- the slope of every Delaunay edge belongs to $]-\delta, \delta[\subset \mathbb{RP}^1$,
- the inner angle between two short sides of P_n is at least $\pi - \delta$,
- the inner angle between a short side and a long side of P_n is at most δ.

If a trajectory of P_n starts from a short side and leaves P_n through another short side, then it cuts out P_n into two polygons. Computing the sum of the inner angles in each of them, we deduce that the slope of t belongs to $[-p\delta, p\delta]$, where p is the number of sides of P_n. Thus, by choosing $\delta \geq \epsilon/q$ where q is the number of Delaunay edges of P_n, one makes sure that for $n \geq N_\delta$, a trajectory of polygon P_n whose slope belongs to Θ_ϵ starting from a short side of P_n leaves it through the interior of its unique long side.

The proof of the noncyclic domain of type 1 follows the exact same line. The key remark being that type-1 polygons piled up long side to short side form a polygon that satisfies the three points above (see Figure 13). Therefore, if we set $\delta = \epsilon/m$ where m is the total number of edges that are short sides of at least one polygon of X_n, following the argumentation above, we see that there is N_δ large enough that for $n \geq N_\delta$ the trajectory visits finitely many long boundaries of X_n and exits X_n, as otherwise the domain would be cyclic. □

We now address the case of maximal domains of type 2.

Proposition 7.3 *Let $(E_n)_{n\in\mathbb{N}}$ be a short boundary edge of a maximal domain of type 2 $(X_n)_{n\in\mathbb{N}}$ that has at least one boundary long edge. We also consider a nonempty open interval $\mathcal{I} \subset \Theta_\epsilon$ for some $\epsilon > 0$.*

There is a long boundary edge $(M_n)_{n\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$ and $N > 0$ such that for any $n \geq N$, any trajectory of X_n whose direction belongs to $\mathcal{I}$ starting from E_n eventually leaves X_n through the interior of M_n.

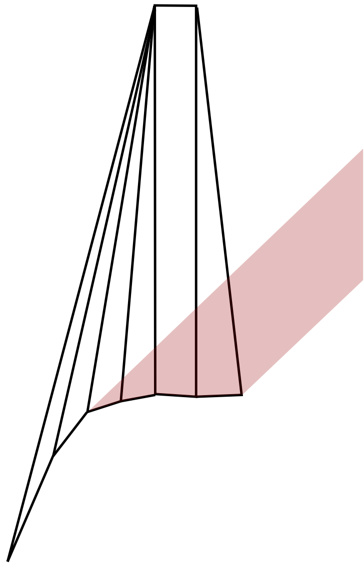


Figure 15: A maximal domain of type 2 with trajectories starting from short sides in a direction far from that of the long sides.

Proof A maximal domain of type 2 with at least one boundary edge actually has two long boundary edges, since its graph of incidence is linear (see Section 3.5). Without loss of generality, we assume that the limit slope of the long Delaunay edges of $(X_n)_{n\in\mathbb{N}}$ is vertical. Therefore, we will refer to the boundary edges of X_n (which is a polygon) as the long left side, the long right side, the short upper sides and the short lower sides.

There is $N > 0$ such that for any $n \geq N$, the slope of any straight segment joining an upper vertex and a lower vertex of X_n is contained in $\left]\frac{1}{2}\pi - \epsilon, \frac{1}{2}\pi\right] + \epsilon\left[\right.$. It follows that the slope of a trajectory of X_n joining a short upper side and a short lower side cannot belong to Θ_ϵ for $n \geq N$. It remains to consider the case of a trajectory t joining two upper (or lower) short sides of X_n.

It follows from Proposition 3.7 that for any $\delta > 0$, there is N_δ such that for any $n \geq N_\delta$, the inner angle between two consecutive upper (or lower) short sides of X_n is at least $\pi - \delta$. Any trajectory t cuts out X_n into two polygons. Computing the sum of inner angles in each of them, we deduce that the trajectory t forms an angle of magnitude smaller than $p\delta$ with one of the short sides of X_n (here p is the number of sides of X_n). Since δ can be made arbitrarily small, there exists a bound $N' > 0$ such that for any $n \geq N'$, a trajectory joining two upper (or lower) short sides of X_n cannot belong to Θ_ϵ.

Consequently, for any n satisfying $n \geq \max(N, N')$, any trajectory starting from a short boundary edge E_n of X_n eventually leaves X_n through the interior of one of its two extremal edges (see Figure 15). If we restrict ourselves to trajectories whose slope belongs to a connected open subset U of Θ_ϵ, a continuity argument proves that two trajectories starting from E_n leave X_n through the same extremal edge. □

7.2 Trajectories inside connected components of the core

The case of Delaunay pieces that are connected components of the core $(C\Sigma_n)_{n\in\mathbb{N}}$ is a bit more complicated. In order to find an open set of directions where trajectories starting from the same short boundary edge leave a component of $C\Sigma_n$ through the same long boundary edge, we first prove the analogous result for connected components of the limit surface Σ_∞, which is an easy consequence of Proposition 2.9.

Lemma 7.4 *For any nonempty open subset $U \subset \mathbb{RP}^1$ and any connected component X_∞ of Σ_∞ with a nonempty boundary, one of the following statements holds:*

- *there exists a closed geodesic in X_∞ whose slope is contained in U,*
- *there is a nonempty open subset $V \subset U$ such that every trajectory starting from a singularity x of X_∞ in a direction of V eventually leaves X_∞ through the interior of a boundary saddle connection.*

Proof Let $S_{x,U}$ be the set of (oriented) trajectories starting from the singularity x with a slope in U. The topology of $S_{x,U}$ is induced by the canonical projection π_x to $\mathbb{RP}^1$. Assuming that no closed geodesic of X_∞ belongs to a direction of U, it has been proved in Proposition 2.9 that trajectories of $S_{x,U}$ leaving X_∞ through the interior of a boundary saddle connection form an open dense subset of $S_{x,U}$. Since there are finitely many such singularities in Σ_∞ and projections π_x have finitely many preimages, there is an open dense subset V of U such that every trajectory starting from such a singularity x in a direction of V leaves its component through the interior of a boundary saddle connection. □

Since short boundary edges of connected components of the core degenerate to singular points in the Delaunay limit, we obtain a result about trajectories in the connected components of the core:

Proposition 7.5 *Let $(X_n)_{n\in\mathbb{N}}$ be a connected component of the core $(C\Sigma_n)_{n\in\mathbb{N}}$ with at least one long boundary edge. Let $(E_n)_{n\in\mathbb{N}}$ be a short boundary edge of $(X_n)_{n\in\mathbb{N}}$. For any nonempty open subset $U \subset \mathbb{RP}^1$, one of the following statements holds:*

- *there is a bound $N > 0$ such that for any $n \geq N$, there exists a closed geodesic in X_n whose slope is contained in U,*
- *there is a nonempty open subset $V \subset U$, a bound $N > 0$ and a long boundary edge $(M_n)_{n\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$ such that for any $n \geq N$, any trajectory of X_n in a direction $\theta \in V$ starting from E_n eventually leaves X_n through the interior of M_n.*

Proof We first decompose the proof into two subcases, depending whether Σ_∞ contains a closed geodesic whose direction belongs to U or not. In the first case, we deduce from Proposition 3.9 that there exists $N > 0$ such that for any $n \geq N$, Σ_n contains a closed geodesic whose direction belongs to U.

In the second case, we fix ϵ small enough that $U \cap \Theta_\epsilon$ is nonempty. Lemma 7.4 then proves the existence of a nonempty open subset V of $U \cap \Theta_\epsilon$ such that any trajectory of the Delaunay limit X_∞ of $(X_n)_{n\in\mathbb{N}}$

whose direction belongs to V and that starts from a singularity leaves X_∞ through the interior of a boundary saddle connection.

Let $(E_n)_{n\in\mathbb{N}}$ be a short boundary edge of $(X_n)_{n\in\mathbb{N}}$. By construction this short boundary edge converges toward a point x of X_∞ in the limit. For any interval I in V, we consider the two-parameter family $P(E_n, I)$ of trajectories starting from the edge E_n and whose directions belong to I. This family of trajectories accumulates on a pencil $P(x, I)$ of X_∞ as n tends to infinity.

The edge $(E_n)_{n\in\mathbb{N}}$ is a short edge of a chain of polygons of type 2 belonging to $(X_n)_{n\in\mathbb{N}}$ (see Figure 11). Using Proposition 3.7 as in the proof of Proposition 6.3, we deduce that provided that n is large enough, trajectories of $P(E_n, I)$ leave each of these polygons of type 2 through one of its long side (the hypothesis that I is disjoint from Θ_ϵ is crucial here). Then these trajectories finally enter a polygon of type 3 of $(X_n)_{n\in\mathbb{N}}$. A continuity argument proves that trajectories of the pencil $P(x, I)$ leave X_∞ through the interior of the same boundary saddle connection M_∞ which is the limit of a long boundary edge $(M_n)_{n\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$. Up to replacing I by a smaller open interval, we can assume that the intersection of trajectories of $P(x, I)$ with M_∞ is disjoint from a neighborhood of the endpoints of M_∞. We deduce that, provided n is large enough, trajectories of $P(E_n, I)$ leave X_n through the interior of M_n. □

We combine the previous results to exhibit a set of directions and a lower bound that hold for every Delaunay piece of $(\Sigma_n)_{n\in\mathbb{N}}$:

Corollary 7.6 *For any nonempty open subset $U \subset \mathbb{RP}^1$, one of the following statements holds*:

- *there is a bound $N > 0$ such that for any $n \geq N$, there exists a closed geodesic in Σ_n whose slope is contained in U*,
- *there is a nonempty open subset $V \subset U$ and a bound $N > 0$ such that for any $n \geq N$ and any short boundary edge E_n in any Delaunay piece X_n of Σ_n having a long boundary edge*, *there is a long boundary edge M_n such that every trajectory of X_n starting from E_n and whose slope belongs to U eventually leaves X_n through the interior of M_n.*

Proof Provided ϵ is small enough, $\Theta_\epsilon \cap U$ is nonempty. For such a small ϵ, we consider an open interval $I \subset \Theta_\epsilon \cap U$. Since there are finitely many Delaunay pieces and Delaunay edges in $(\Sigma_n)_{n\in\mathbb{N}}$, there is a global bound N_0 such that the second statement holds for trajectories whose slope is in the interval I for any short boundary edge in any Delaunay piece X_n that is a maximal domain of type 1 or 2 (see Propositions 7.2 and 7.3).

Then we apply Proposition 7.5 to a boundary short edge $(E_n)_{n\in\mathbb{N}}$ in a connected component $(X_n)_{n\in\mathbb{N}}$ of the core. If X_n contains a closed geodesic, provided n is large enough, then the first statement of our proposition holds. Otherwise, the second statement of Proposition 7.5 provides a nonempty open subset I' of I and a new bound such that the property also holds for this edge. After finitely many steps, we

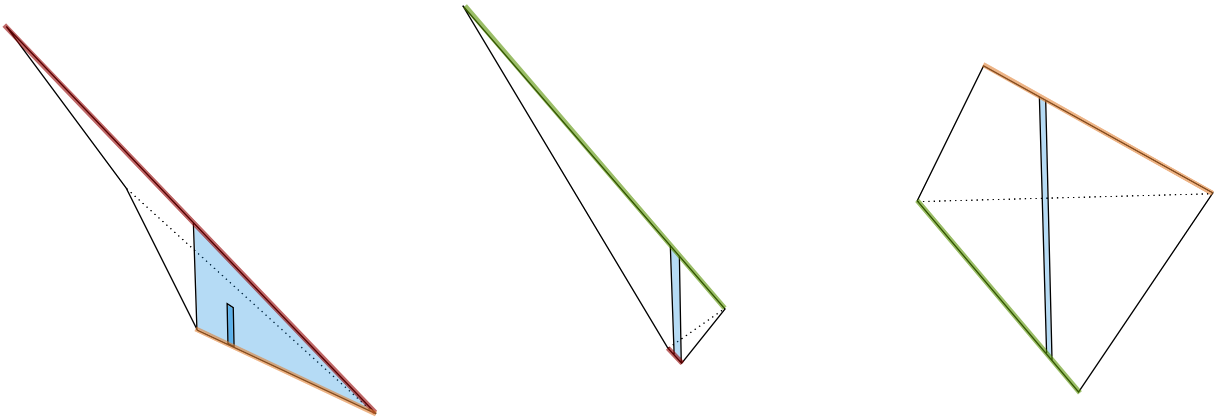


Figure 16: From left to right: a maximal domain D_1 of type 1, a maximal domain D_2 of type 2 and a component C of the core. The orange edge on the left corresponds to E_n. The ribbon of parallel trajectories leaves D_1 to enter D_2 through its short side which goes to C from a long side of D_2. When the ribbon enters back into D_1 it has been contracted by an amount that goes to infinity when $n \to +\infty$.

obtain a nonempty open subset V of $\mathbb{RP}^1$ and a bound $N > 0$ such that the property holds of trajectories whose slope belongs to V for every short boundary edge in every Delaunay piece having a long boundary edge (provided that $n \geq N$). □

Proof of Proposition 4.4 Corollary 7.6 shows that it is enough to prove that any direction d such that for n large enough any trajectory starting from any short edge of any Delaunay piece exits the Delaunay piece through one of its long sides carries a cylinder. As we only have finitely many Delaunay pieces, any such trajectory will have to cross twice some boundary edge $(E_n)_{n\in\mathbb{N}}$ of two Delaunay pieces (one for which it is a short side and one for which it is a long side).

Given a direction d and a short edge $(E_n)_{n\in\mathbb{N}}$, we denote by $P(E_n, d)$ the set of trajectories starting from a point of $(E_n)_{n\in\mathbb{N}}$ of direction d pointing inside the Delaunay piece for which $(E_n)_{n\in\mathbb{N}}$ is the short side.

By definition of E_n, there is a trajectory t of $P(E_n, d)$ which crosses back E_n for n large enough. We claim that all the trajectories of $P(E_n, d)$ cross E_n alongside t. Indeed, as any trajectory of $P(E_n, d)$ only exits a Delaunay piece by its long side and enters one by its short side, the contraction ratio of the first return map on a neighborhood of E_n converges to 0 as $n \to +\infty$. In particular, for n large enough, the image of E_n must be fully contained in E_n, which implies that it has a periodic point. This periodic point of the first return map corresponds to a closed geodesic. □

References

[1] **E Duryev**, **C Fougeron**, **S Ghazouani**, *Dilation surfaces and their Veech groups*, J. Mod. Dyn. 14 (2019) 121–151 MR Zbl

[2] **D Gromoll**, **W Meyer**, *Periodic geodesics on compact riemannian manifolds*, J. Differential Geometry 3 (1969) 493–510 MR Zbl

[3] **B Martelli**, *An introduction to geometric topology*, CreateSpace, Scotts Valley, CA (2016)

[4] **H Masur**, *Closed trajectories for quadratic differentials with an application to billiards*, Duke Math. J. 53 (1986) 307–314 MR Zbl

[5] **H Masur**, **J Smillie**, *Hausdorff dimension of sets of nonergodic measured foliations*, Ann. of Math. 134 (1991) 455–543 MR Zbl

[6] **G Tahar**, *Horizon saddle connections, quasi-Hopf surfaces and Veech groups of dilation surfaces*, Geom. Dedicata 212 (2021) 1–16 MR Zbl

[7] **G Tahar**, *Horizon saddle connections and Morse–Smale dynamics of dilation surfaces*, J. Mod. Dyn. 19 (2023) 417–431 MR Zbl

Institut Mathématique de Marseille, CNRS, Aix-Marseille Université
Marseille, France

Department of Mathematics, University College London
London, United Kingdom

Faculty of Mathematics and Computer Science, Beijing Institute of Mathematical Sciences and Applications
Beijing, China

adrien.boulanger@univ-amu.fr, s.ghazouani@ucl.ac.uk, guillaume.tahar@bimsa.cn

Proposed: Benson Farb — Received: 24 August 2023
Seconded: Mladen Bestvina, David Fisher — Revised: 16 May 2024

Guidelines for Authors

Submitting a paper to Geometry & Topology

Papers must be submitted using the upload page at the GT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Geometry & Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in GT is available on the GT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

GEOMETRY & TOPOLOGY

Volume 29 Issue 4 (pages 1693–2250) 2025