Projective twists and the Hopf correspondence

BRUNELLA CHARLOTTE TORRICELLI

Given Lagrangian (real, complex) projective spaces $K_1, \ldots, K_m$ in a Liouville manifold (X, ω) satisfying a certain cohomological condition, we show there is a Lagrangian correspondence (in the sense of Wehrheim and Woodward (2012)) that assigns a Lagrangian sphere $L_i \subset K$ of another Liouville manifold (Y, Ω) to any given projective Lagrangian $K_i \subset X$ for $i = 1, \ldots, m$.

We use the Hopf correspondence to study *projective twists*, a class of symplectomorphisms akin to Dehn twists, but defined starting from Lagrangian projective spaces. When this correspondence can be established, we show that it intertwines the autoequivalences of the compact Fukaya category $\mathcal{F}\mathrm{uk}(X)$ induced by the projective twists $\tau_{K_i} \in \pi_0(\mathrm{Symp}_{\mathrm{ct}}(X))$ with the autoequivalences of $\mathcal{F}\mathrm{uk}(Y)$ induced by the Dehn twists $\tau_{L_i} \in \pi_0(\mathrm{Symp}_{\mathrm{ct}}(Y))$ for $i = 1, \ldots, m$.

Using the Hopf correspondence, we obtain a free generation result for projective twists in a *clean plumbing* of projective spaces and various results about products of positive powers of Dehn/projective twists in Liouville manifolds.

The same techniques are also used to show that the Hamiltonian isotopy class of the projective twist (along the zero section in $T^*\mathbb{CP}^n$) in $\mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$ does depend on a choice of framing for $n \geqslant 19$. Another application of the Hopf correspondence delivers smooth homotopy complex projective spaces $K \simeq \mathbb{CP}^n$ that do not admit Lagrangian embeddings into $(T^*\mathbb{CP}^n, d\lambda_{\mathbb{CP}^n})$ for $n = 4, 7$.

1 Introduction

1.1 Questions

Given a symplectic manifold (M, ω) with contact boundary, an interesting object of study is the group $\mathrm{Symp}_{\mathrm{ct}}(M)$ of compactly supported symplectomorphisms that are the identity in a neighbourhood of the boundary. Its quotient $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(M))$ by the relation of symplectic isotopy is the *symplectic mapping class group*, and is already a highly nontrivial object. When $H^1(M; \mathbb{R}) = 0$, a symplectic isotopy is automatically Hamiltonian, and $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(M))$ coincides with the quotient $\mathrm{Symp}_{\mathrm{ct}}(M)/\mathrm{Ham}_{\mathrm{ct}}(M)$ by the subgroup $\mathrm{Ham}_{\mathrm{ct}}(M) \subset \mathrm{Symp}_{\mathrm{ct}}(M)$ of (compactly supported) Hamiltonian symplectomorphisms (namely time-1 maps of compactly supported Hamiltonian flows).

The symplectic mapping class group carries a (forgetful) comparison map

$$(1) \qquad\qquad\qquad c \colon \pi_0(\mathrm{Symp}_{\mathrm{ct}}(M)) \to \pi_0(\mathrm{Diff}_{\mathrm{ct}}^+(M))$$

to the (compactly supported and orientation-preserving) smooth mapping class group of M. In general, the map is neither injective nor surjective. Its kernel is of particular interest as it captures phenomena which are exclusively symplectic and not visible in the smooth mapping class group. The question of whether a symplectomorphism $\varphi \in \mathrm{Symp}_{\mathrm{ct}}(M)$ is a nontrivial element of the kernel of c (ie is smoothly isotopic to the identity but not symplectically so) is called the *symplectic isotopy problem*.

In dimension two, the kernel of c is always trivial, and the symplectic mapping class group is isomorphic to the smooth mapping class group $\pi_0(\mathrm{Diff}_{\mathrm{ct}}^+(M))$; this is a consequence of *Moser's argument* [1965].

Dehn twists often provide examples of nontrivial symplectomorphisms that lie in the kernel of (1). Given a sphere L (and a choice of parametrisation, called a *framing*; see Definition 2.7), the periodicity of the (co)geodesic flow can be used to construct a compactly supported symplectomorphism of the cotangent bundle $\tau_L \in \mathrm{Symp}_{\mathrm{ct}}(T^*L)$ (see Definition 2.8), called a standard Dehn twist.

The standard Dehn twist has infinite symplectic order, ie infinite order in $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(T^*S^n))$ [Seidel 2000] and, for $n = 2$, it generates the entire mapping class group $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(T^*S^2))$ [Seidel 1998].

Given a general symplectic manifold (M, ω) and an embedded Lagrangian sphere $L \subset M$, the local construction of the standard Dehn twist can be implanted in a neighbourhood of L via Weinstein's neighbourhood theorem, to yield a compactly supported symplectomorphism $\tau_L \in \mathrm{Symp}_{\mathrm{ct}}(M)$. When $\dim(L)$ is even, the Dehn twist has finite order in $\mathrm{Diff}_{\mathrm{ct}}^+(M)$ but often has infinite order in $\mathrm{Symp}_{\mathrm{ct}}(M)$. Seidel's early investigations provided the first global examples of (symplectically) nontrivial Dehn twists, in particular nontrivial elements of the kernel of the comparison map (1). For example, for a $K3$–surface (M, ω) containing two disjoint Lagrangian spheres $L_1, L_2 \subset M$, the class of τ_{L_1} has infinite order in $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(M))$, and hence in that case c has infinite kernel [Seidel 2000]. Other important examples in which the kernel of c is large include Dehn twists in Milnor fibres of any isolated hypersurface singularity

[Keating 2014] and Dehn twists in projective hypersurfaces of degree $d > 2$ (and more general divisors [Tonkonog 2015]).

One of the widely employed methodologies used in these investigations is symplectic Picard–Lefschetz theory. In this context, Dehn twists are regarded as the class of symplectomorphisms that encode symplectic monodromy maps associated to nodal degenerations, ie monodromies of *Lefschetz fibrations* (see Section 2).

For an exact symplectic manifold (M, ω), any Dehn twist τ_L along a Lagrangian sphere $L \subset (M, \omega)$ can be realised as the local monodromy of an exact Lefschetz fibration (with exact smooth fibre (M, ω) and exact base). One important result that has been proved recently in [Barth et al. 2019] (an alternative proof of which can be found in this paper) is that the global monodromy of such Lefschetz fibrations can never be isotopic to the identity in the symplectic mapping class group.

Theorem A [Barth et al. 2019, Theorem 1.4]   *Let (M, ω) be a Liouville manifold, and let $L_1, \ldots, L_m \subset M$ be Lagrangian spheres. Let $\phi = \prod_{i=1}^{k} \tau_{L_{j_i}} \in \mathrm{Symp}_{\mathrm{ct}}(M)$, $j_i \in \{1, \ldots, m\}$ be a positive word of Dehn twists. Then ϕ is not compactly supported isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(M)$.*

As a result, Dehn twists represent an extremely important source of (symplectically) nontrivial symplectic automorphisms of exact symplectic manifolds.

In a more general setting, we can consider both positive as well as negative powers of Dehn twists. In this case, the intersection pattern of the Lagrangians generating the twists determines the behaviour of a product of such twists. For example, if L and L' are two Lagrangian spheres of a Liouville manifold (M, ω) which intersect in a single point, then the corresponding twists $\tau_L, \tau_{L'} \in \mathrm{Symp}_{\mathrm{ct}}(M)$ satisfy the *braid relation*, an isotopy $\tau_L \tau_{L'} \tau_L \simeq \tau_{L'} \tau_L \tau_{L'}$ in $\mathrm{Symp}_{\mathrm{ct}}(M)$ [Seidel 1999; Seidel and Thomas 2001]. In a general situation, Keating showed that the suitable quantifier that obstructs the possibility of a nontrivial relation between the twists τ_L and $\tau_{L'}$ is the rank of the Floer cohomology group $\mathrm{HF}(L, L')$, as follows:

Theorem 1.1 [Keating 2014, Theorems 1.1 and 1.2]   *Let (M, ω) be a Liouville manifold of dimension greater than 2, and $L, L' \subset M$ be two Lagrangian spheres satisfying $\mathrm{rank}\,\mathrm{HF}(L, L') \geqslant 2$ and that are such that L and L' are not quasi-isomorphic in the Fukaya category. Then the Dehn twists τ_L and $\tau_{L'}$ generate a free subgroup of $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(M))$, and the associated functors T_L and T'_L generate a free subgroup of $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(M))$.*

Note that the two-dimensional case holds via a result due to Ishida [1996].

Seidel [2000] introduced a class of symplectomorphisms defined from Lagrangian submanifolds with periodic geodesic flow. This type of Lagrangian includes spheres — in which case the symplecto-morphisms are squared Dehn twists — and projective spaces. This paper focuses on the latter class of symplectomorphisms, which we call *projective twists* (the appellation *Dehn* will be associated exclusively to Dehn twists along spheres). The complex projective analogues of Dehn twists are *always* contained in

the kernel of the comparison map (1) [Seidel 2000, Proposition 4.6], which means that they are a class of symplectomorphisms which are never detectable by the smooth structure. Similarly to Dehn twists, projective twists arise as local monodromies of fibration-like structures [Perutz 2007]; these fibrations are called *Morse–Bott–Lefschetz* fibrations and their singularities are Morse–Bott degenerations.

Unlike their spherical counterparts, projective twists have not been in the spotlight of research in symplectic topology, and this is for a number of reasons. The definition of projective twist requires the existence of a Lagrangian embedding of a projective space in the ambient symplectic manifold, which can result in strong topological restrictions. Moreover, the symplectic Picard–Lefschetz theory of [Seidel 2008a] does not have such immediate applications as for Dehn twists.

Nevertheless, a series of recent results indicates that projective twists do have interesting properties of the calibre of Dehn twists [Evans 2011; Harris 2011; Mak and Wu 2018].

The results of the present research are driven by the following questions, which in the existing literature have been considered for Dehn twists exclusively:

Questions 1   Let (M, ω) be a Liouville manifold.

(a) Can a reduced word of projective twists be symplectically isotopic to the identity (ie are there twists satisfying any nontrivial relations) in $\mathrm{Symp}_{\mathrm{ct}}(M)$?

(b) Can a reduced *positive* word (ie a product of positive powers) of projective twists be symplectically isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(M)$?

1.2   Methods: the Hopf correspondence

How can we study projective twists? Because much of the scholarship that emerged from the study of Dehn twists is the result of successful applications of symplectic Picard–Lefschetz theory, a first intuitive move is to approach the study of projective twists by means of their presentation as monodromies of Morse–Bott–Lefschetz fibrations. One strategy could be to adapt some of the arguments originally tailored for Dehn twists to a more general Picard–Morse–Bott–Lefschetz theory, as developed in [Wehrheim and Woodward 2016]. However, this setting presents serious complications related to a potential loss of compactness of the moduli spaces of pseudoholomorphic curves of these fibrations (the total space of Morse–Bott–Lefschetz fibrations is in general not exact, and the singular locus — a smooth manifold of the singular fibre — often admits rational curves).

To examine the properties of these symplectomorphisms, in this paper we adopt a strategy that allows to reduce the study of projective twists to that of Dehn twists in an auxiliary Liouville manifold; this is made possible via the theoretical device of *Lagrangian correspondences*.

A Lagrangian correspondence between two symplectic manifolds (W, ω) and (Y, Ω) is a Lagrangian submanifold of the product $W^- \times Y := (W \times Y, -\omega \oplus \Omega)$. By [Wehrheim and Woodward 2012; 2010a; 2010b; Gao 2017a; 2017b], under suitable conditions, a Lagrangian correspondence induces a functor which associates a Lagrangian in Y to a Lagrangian in W.

In a first stage, we define an appropriate Lagrangian correspondence that relates a set of Lagrangian projective spaces in a given Liouville manifold (W, ω) to a set of Lagrangian spheres in an auxiliary manifold (Y, Ω) expressly built under some cohomological conditions. Fix a tuple $(\mathbb{A}, k, *, R) \in \{(\mathbb{R}, 0, 1, \mathbb{Z}/2\mathbb{Z}), (\mathbb{C}, 1, 2, \mathbb{Z})\}$. Assume there are Lagrangian projective spaces $\mathbb{A}\mathbb{P}^n \cong K_1, \ldots, K_m \subset W$ and a nontrivial class $\alpha \in H^*(W; R)$ such that $\alpha|_{K_i}$ generates $H^*(\mathbb{A}\mathbb{P}^n; R)$. Then there is a Liouville manifold (Y, Ω), realised as a T^*S^k–bundle $q \colon Y \to W$, which contains an S^k–fibred coisotropic submanifold $V \to W$, defining a Lagrangian correspondence $\Gamma := \{(q(y), y) \mid y \in V\} \subset W^- \times Y$ in the sense of [Perutz 2008]. Over each projective Lagrangian $K_i \subset W$, the correspondence yields a Lagrangian sphere $L_i \subset Y$ for $i = 1, \ldots, m$ (Sections 3.1 and 3.2). We name Γ the *Hopf correspondence*.

Once the Hopf correspondence is constructed, we use Ma'u–Wehrheim–Woodward theory and Gao's extension for nonclosed correspondences to show that there is an induced functor $\Theta_\Gamma \colon \mathcal{F}\mathrm{uk}(W) \to \mathcal{F}\mathrm{uk}(Y)$ between the compact Fukaya categories (see Section 4.2). We then prove the existence of a commuting diagram (Section 4.4)

(2)
$$
\begin{array}{ccc}
\mathcal{F}\mathrm{uk}(Y) & \xrightarrow{\;\;T_{L_i}\;\;} & \mathcal{F}\mathrm{uk}(Y) \\
{\scriptstyle \Theta_\Gamma}\big\uparrow & & {\scriptstyle \Theta_\Gamma}\big\uparrow \\
\mathcal{F}\mathrm{uk}(W) & \xrightarrow{\;\;T_{K_i}\;\;} & \mathcal{F}\mathrm{uk}(W)
\end{array}
$$

where $T_{K_i} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ and $T_{L_i} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$ are the twist functors induced by the graphs of the respective twists $\tau_{K_i} \in \mathrm{Symp}_{\mathrm{ct}}(W)$ and $\tau_{L_i} \in \mathrm{Symp}_{\mathrm{ct}}(Y)$.

1.3 Results

1.3.1 Free groups generated by projective twists In Section 5 we consider Question 1(b) and give a first answer to it. We consider a *clean plumbing* (see Definition 5.1) of Lagrangian projective spaces: a symplectic construction in which two copies of cotangent bundles $T^*\mathbb{A}\mathbb{P}^n$ are glued along a common submanifold of the zero sections, and prove the following:

Theorem B *Let* $W := T^*\mathbb{A}\mathbb{P}^n \#_{\mathbb{A}\mathbb{P}^l} T^*\mathbb{A}\mathbb{P}^n$ *be a clean plumbing of (real, complex) projective spaces along a linearly embedded subprojective space* $\mathbb{A}\mathbb{P}^l \subset W$, $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}\}$. *Let* $K_1, K_2 \cong \mathbb{A}\mathbb{P}^n \subset W$ *denote the Lagrangian core components of the plumbing. Then the projective twists* τ_{K_1} *and* τ_{K_2} *generate a free group inside* $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(W))$, *and the associated functors* T_{K_1} *and* T_{K_2} *generate a free subgroup of* $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$.

In the complex case, this theorem yields a new criterion for projective twists to generate a free subgroup of the kernel of the comparison map (1).

We prove Theorem B using the Hopf correspondence to relate the functors $T_{K_1}, T_{K_2} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ to functors $T_{L_1}, T_{L_2} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$ induced by Dehn twists in a Liouville manifold (Y, Ω) constructed as a T^*S^k–bundle over W for $k \in \{0, 1\}$. This is made possible via the commuting diagram (2).

We can then apply Keating's result (Theorem 1.1) to our setting to obtain a free generation result for T_{L_1} and T_{L_2}, which we translate into a free generation result for T_{K_1} and T_{K_2} via the Hopf correspondence.

Remark 1.2 The case $W := T^*\mathbb{CP}^1_1 \#_{\mathrm{pt}} T^*\mathbb{CP}^1_2$ can be obtained with the current literature [Seidel 1999; Seidel and Thomas 2001; Khovanov and Seidel 2002], by considering W as an A_2–configuration and using the isotopies $\tau_{\mathbb{CP}^1_i} \simeq \tau^2_{S^2_i}$ (see Remark 5.2). ◁

1.3.2 Positive products of twists in Liouville manifolds In Section 6.3, we restrict our attention to products of positive powers of twists, ie Question 1(b). In a first instance, we analyse this question for Dehn twists, and we present an alternative proof of Theorem A, which was originally proved (by Barth, Geiges and Zehmisch [Barth et al. 2019]) via techniques involving open book decompositions. Our proof is implemented using Picard–Lefschetz theory. The idea is to build a Lefschetz fibration $\pi \colon E \to \mathbb{C}$ with smooth fibre the Liouville manifold (M, ω) and vanishing cycles the given Lagrangian spheres involved in the product $\phi \in \mathrm{Symp}_{\mathrm{ct}}(M)$. In that way, the monodromy of π is given by ϕ. Assuming that there exists an isotopy $\phi \simeq \mathrm{Id}$ as in the statement, we extend π to a fibration over $\mathbb{CP}^1$, and, by analysing the moduli space of pseudoholomorphic sections (following [Seidel 2003]), obtain a contradictory statement.

It can happen that a product of Dehn twists, despite being necessarily not (compactly supported symplectically) isotopic to the identity, preserves some Lagrangian submanifolds of M. The question arises whether one can find a Lagrangian $T \subset M$ such that there can be no compactly supported symplectic isotopy $\phi(T) \simeq T$. The existence of such a Lagrangian would result in a stronger version of Theorem A. In Section 6.2, we address this question. We find one possible candidate Lagrangian $T \subset M$ with the above properties, but unfortunately cannot prove that such a Lagrangian always exist.

A Lagrangian $T \subset M$ is called conical if it is an exact, properly embedded Lagrangian that is preserved by the Liouville flow over the cylindrical ends of M.

Theorem C *Let (M^{2n}, ω) be a Liouville manifold containing embedded Lagrangian spheres $L_1, \ldots, L_m$ and a conical Lagrangian disc T intersecting one of the spheres L_j transversely in a point. Let $\phi := \prod_{i=1}^k \tau_{L_{j_i}} \in \mathrm{Symp}_{\mathrm{ct}}(M)$ with $j_i \in \{1, \ldots, m\}$ be a positive word of Dehn twists involving τ_{L_j}. Then the Lagrangians T and $\phi(T)$ are not isotopic via a compactly supported Lagrangian isotopy.*

Example 1.3 For $m > 0$, consider an iterated transverse plumbing

$$M := T^* S^m \#_{\mathrm{pt}} T^* S^m \#_{\mathrm{pt}} T^* S^m \#_{\mathrm{pt}} \cdots \#_{\mathrm{pt}} T^* S^m$$

(see Section 5.1 for the definition of plumbing). Let $\phi \in \mathrm{Symp}_{\mathrm{ct}}(M)$ be a product of Dehn twists along the Lagrangian spheres of M such that ϕ contains the Dehn twist along the jth sphere. In this case, there is at least one conical Lagrangian disc $T \subset M$ as in Theorem C; any cotangent fibre of the jth summand will do. ◁

The arguments we use in the proof of Theorem C are centred around the same principles as the method used for Theorem A, with some necessary adjustments due to the noncompactness of the Lagrangian $T \subset M$.

At last, in Section 6.3, we turn to applications related to projective twists. Using the Hopf correspondence, we prove a result that can be considered the (real) projective counterpart to Theorem A.

Theorem D *Let* (W^{2n}, ω) *be a Liouville manifold containing Lagrangian real projective spaces* $K_1, \ldots, K_m$ *with* $K_i \cong \mathbb{RP}^n$. *Suppose that there is a class* $\alpha \in H^1(W; \mathbb{Z}/2\mathbb{Z})$ *such that, for every* $i = 1, \ldots, m, \alpha|_{K_i}$ *generates* $H^*(\mathbb{RP}^n; \mathbb{Z}/2\mathbb{Z})$. *Let* $\varphi \in \mathrm{Symp}_{ct}(W)$ *be a positive word in the subset of projective twists* $\{\tau_{K_i}\}_{i \in \{1,\ldots,m\}}$. *Then* φ *is not isotopic to the identity in* $\pi_0(\mathrm{Symp}_{ct}(W))$.

Using the cohomological assumption of the theorem, we establish the Hopf correspondence and prove the theorem by contradiction. The idea is that, in these circumstances, there exists a product of Dehn twists $\phi \in \mathrm{Symp}_{ct}(\widetilde{W})$ in the symplectic double cover $q : (\widetilde{W}, \widetilde{\omega}) \to (W, \omega)$ such that $q \circ \phi = \varphi \circ q$. Then an isotopy $\varphi \simeq \mathrm{Id}$ in $\mathrm{Symp}_{ct}(W)$ can be lifted to an isotopy $\phi \simeq \mathrm{Id}$ in $\mathrm{Symp}_{ct}(\widetilde{W})$, contradicting Theorem A.

Unfortunately, the same techniques do not yield a result for complex projective twists; in that case, the auxiliary manifold (Y, Ω) defines a $\mathbb{C}^*$–bundle $q : Y \to W$ and a compactly supported isotopy on W does not lift to a compactly supported isotopy on Y, so that the arguments used for Theorem D do not apply here.

1.3.3 Framing of projective twists and Lagrangian embeddings of homotopy projective spaces

The last section uses the Hopf correspondence to examine the ways in which the symplectic structure interferes with the underlying topological structures, such as diffeomorphism and homeomorphism class, of Lagrangian homotopy projective spaces. In this paper, a manifold that is homeomorphic but not diffeomorphic to a standard (real or complex) projective space is called an AD projective space. A manifold that is homotopy equivalent but not homeomorphic to a standard (real or complex) projective space is called an AT projective space. Similarly, an AD sphere is a sphere that is homeomorphic but not diffeomorphic to the standard sphere (we decide to drop the usual epithet *exotic*; see Definition 7.6).

A notorious conjecture, the *nearby Lagrangian conjecture*, states that, given a closed smooth manifold Q, any closed exact Lagrangian submanifold of $(T^*Q, d\lambda_Q)$ is Hamiltonian isotopic to the zero section. This conjecture has generated a lot of interest in the symplectic community, but its statement is currently confirmed only up to simple homotopy equivalence [Abouzaid 2012b; Kragh 2013; Abouzaid and Kragh 2018]; in Section 7.1 we summarise the state of the art of this conjecture. For a homotopy sphere L, it is known that the choice of smooth structure can be an obstruction to the existence of a Lagrangian embedding $L \hookrightarrow T^*S^n$. Namely, for $n > 4$ odd, AD spheres which do not bound parallelisable manifolds admit no Lagrangian embedding into T^*S^n [Abouzaid 2012a; Ekholm et al. 2016].

Using the existing literature about S^1–actions on AD spheres [Bredon 1967; James 1980; Kasilingam 2016], we find, in Section 7.1, examples of nonstandard homotopy complex projective spaces which do not admit Lagrangian embeddings into $T^*\mathbb{CP}^n$. These results are compatible with the predictions derived from the nearby Lagrangian conjecture.

Theorem E   *There is a manifold P homotopy equivalent to $\mathbb{CP}^4$ and with the same first Pontryagin class such that neither P nor $P \# \Sigma^8$ admits an exact Lagrangian embedding into $T^*\mathbb{CP}^4$.*

Theorem F   *There is an element Σ^{14} in the group of homotopy 14–spheres Θ_{14} such that $\mathbb{CP}^7 \# \Sigma^{14}$ does not admit an exact Lagrangian embedding into $T^*\mathbb{CP}^7$.*

On the other hand, in Section 7.2, we present new results which prove that, in general, the Hamiltonian isotopy class of projective twists does depend on a choice of framing, ie a choice of smooth parametrisation $f : \mathbb{CP}^n \to L$ (see Definition 2.7). It was proved by Dimitroglou Rizell and Evans [2015] that a nonstandard parametrisation $S^n \to L$ of a Lagrangian sphere can give rise to a Dehn twist that is not isotopic to the standard Dehn twist τ_{S^n}.

We use classical homotopy theory and the Hopf correspondence to transpose the existence of nonstandard parametrisations of Dehn twists of [Dimitroglou Rizell and Evans 2015] into instances of projective twists depending on their framing.

Theorem G   *The $\mathbb{CP}^n$–twist depends on the framing when $n = 19, 23, 25, 29$.*

This shows that, in general, $\mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$ is not generated by the standard projective twist along the zero section $\tau_{\mathbb{CP}^n}$ (see Corollary 7.26). Moreover, we also note that the use of advanced topological technology (*topological modular forms*) can prove the existence of infinitely many nonstandardly framed (complex) projective twists (Proposition 7.24).

Organisation of the paper

The rest of the paper is organised as follows.

Sections 3 and 4 are the two theoretical cores that support the arguments throughout the paper. After recalling the principal properties of twists in Section 2, in Section 3 we prove commutative diagrams involving Dehn twists, the Hopf map and projective twists in the geometric setting. In Section 4, we define the Hopf correspondence and its applications for diagrams of functors of the Fukaya category induced by Dehn/projective twists.

The central body of the paper is divided in three parts, in which we apply the methods developed. We prove a free group generation criterion for projective twists in plumbings in Section 5, we study positive products of twists in general Liouville manifolds in Section 6, and we study framings of projective twists as well as Lagrangian embeddings of homotopy projective spaces in Section 7.

Acknowledgements

This is my final contribution to a mathematical practice held captive by the institution — as a monolithic body of knowledge and a machine producing universal truths. My journey has led me to honour relational, living, and life-affirming mathematical praxes that exist beyond Cartesian binaries and the necropolitics of technoscience.

May we remember that mathematics is a verb, conjugated plurally.

2 Twists

This section provides the contextualisation necessary for studying Dehn twists and projective twists in symplectic topology; it can be skipped by the expert reader. We summarise the constructions of twists from a geodesic flow perspective (Section 2.1), and as local monodromies of fibrations (Section 2.2).

2.1 Twists from geodesic flow

In this section we recall the definitions of Dehn and projective twists that employ the periodicity of the geodesic flow of spheres and projective spaces (the main references are [Seidel 2003, Section 1.2; 2000, Section 4.b; Mak and Wu 2018, Section 2.1]). Let (K, g) be a closed connected Riemannian manifold admitting a periodic cogeodesic flow $\Phi_K^t : T^*K \to T^*K$ on its cotangent bundle $(T^*K, d\lambda_{T^*K})$ such that each geodesic of length 2π is closed (so that the shortest period of a unit-speed geodesic is 2π).

Let $\|\cdot\|_K$ be the norm associated to the given Riemannian metric g. The normalised cogeodesic flow satisfies $\Phi_K^{2\pi} = \mathrm{Id}$ and can be extended to a Hamiltonian S^1–action σ_t^H on $T^*K \setminus K$, with moment map $H : T^*K \setminus K \to \mathbb{R}$, $H(v) = \|v\|_K$.

Definition 2.1 Let K be diffeomorphic to S^n. For $\varepsilon > 0$, define an auxiliary smooth cut-off function $r_\varepsilon : \mathbb{R}^+ \to \mathbb{R}$ such that $0 < r_\varepsilon(t) < \pi$ for all $t < \varepsilon$ and

$$
(3) \qquad r_\varepsilon(t) = \begin{cases} \pi - t & \text{if } t \ll \varepsilon, \\ 0 & \text{if } t \geqslant \varepsilon. \end{cases}
$$

The model Dehn twist $\tau_K^{\mathrm{loc}} : T^*K \to T^*K$ is defined as

$$
(4) \qquad \tau_K^{\mathrm{loc}}(\xi) = \begin{cases} \sigma_{r_\varepsilon(\|\xi\|_K)}^H(\xi) & \text{if } \xi \notin K, \\ -\xi & \text{if } \xi \in K. \end{cases} \qquad \lhd
$$

Definition 2.2 Let K be diffeomorphic to $\mathbb{AP}^n$ for $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$. For $\varepsilon > 0$, let $r_\varepsilon : \mathbb{R}^+ \to \mathbb{R}$ be a smooth cut-off function such that $0 < r_\varepsilon(t) < 2\pi$ for all $t < \varepsilon$ and

$$
(5) \qquad r_\varepsilon(t) = \begin{cases} 2\pi - t & \text{if } t \ll \varepsilon, \\ 0 & \text{if } t \geqslant \varepsilon. \end{cases}
$$

The model projective twist $\tau_K^{\mathrm{loc}} : T^*K \to T^*K$ is defined as

$$
(6) \qquad \tau_K^{\mathrm{loc}}(\xi) = \begin{cases} \sigma_{r_\varepsilon(\|\xi\|_K)}^H(\xi) & \text{if } \xi \notin K, \\ \xi & \text{if } \xi \in K. \end{cases} \qquad \lhd
$$

Remark 2.3 Our choice of cut-off functions r_ε follows [Mak and Wu 2018, Section 2.1], but the construction is independent of such choices, up to suitable isotopy [Seidel 2000]. $\lhd$

Theorem 2.4 [Seidel 2000, Corollary 4.5] *Let (K, g) be a Riemannian manifold admitting a periodic (co)geodesic flow and satisfying $H^1(K; \mathbb{R}) = 0$. Then the symplectomorphisms τ_K^{loc} have infinite order in $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(T^*K))$.*

Theorem 2.5 [Seidel 2000, Proposition 4.6] *The symplectomorphism $\tau_{\mathbb{CP}^n}^{\mathrm{loc}}$ of Definition 2.1 is isotopic to the identity in $\mathrm{Diff}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$.*

Remark 2.6 We will often denote the standard twists by $\tau_{S^n} := \tau_{S^n}^{\mathrm{loc}}$ or, for $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$, $\tau_{\mathbb{AP}^n} := \tau_{\mathbb{AP}^n}^{\mathrm{loc}}$. With the conventions (3) and (5), the isomorphisms $S^1 \cong \mathbb{RP}^1$, $S^2 \cong \mathbb{CP}^1$ and $S^4 \cong \mathbb{HP}^1$ induce isotopies $\tau_{S^1}^2 \simeq \tau_{\mathbb{RP}^1}$, $\tau_{S^2}^2 \simeq \tau_{\mathbb{CP}^1}$ and $\tau_{S^4}^2 \simeq \tau_{\mathbb{HP}^1}$, respectively (see [Seidel 2000; Harris 2011]). $\lhd$

Now suppose (L, g) is a Riemannian manifold admitting a Lagrangian embedding $L \subset M$ into a general symplectic manifold (M, ω).

Definition 2.7 Let $K \in \{S^n, \mathbb{RP}^n, \mathbb{CP}^n, \mathbb{HP}^n\}$. A *framed Lagrangian* sphere/projective space is a Lagrangian submanifold $L \subset M$ together with an equivalence class $[f]$ of diffeomorphisms $f : K \to L$, where $f_1 \sim f_2$ if and only if $f_2^{-1} f_1$ is isotopic, in $\mathrm{Diff}(K)$, to an element of the isometry group $\mathrm{Iso}(K, g)$. An equivalence class $[f]$ as above is called a *framing*. ◁

Definition 2.8 Let $(L, [f])$ be a framed Lagrangian sphere/projective space in (M, ω). Using Weinstein's neighbourhood theorem, extend a framing representative $f : K \to L$ to a symplectic embedding $\iota : D_s T^* K \to M$, where $D_s T^* K := \{v \in T^* K \mid \|v\|_K < s\}$ for $s > 0$. There is a model twist τ_K^{loc}, supported in the interior of $D_s T^* K$, and we define

$$\tau_L \cong \begin{cases} \iota \circ \tau_K^{\mathrm{loc}} \circ \iota^{-1} & \text{on } \mathrm{Im}(\iota), \\ \mathrm{Id} & \text{elsewhere.} \end{cases}$$

In the case where L is a sphere, the map τ_L is the well-known *Dehn* twist. When L is a projective space, the resulting map is called a *projective twist*. In this paper, the term *Dehn* twist is exclusively reserved for twists that are constructed from a Lagrangian sphere. ◁

Remark 2.9 (1) A Dehn twist along an exact Lagrangian sphere, or a projective twist along an exact projective Lagrangian in an exact symplectic manifold, is an exact symplectomorphism in the sense of Definition 2.11. The same holds for products of such twists. This follows by construction (for direct computations, see for example [Barth et al. 2019, Lemma 4.4; Chiang et al. 2016, Lemma 2.1]).

(2) Theorem 2.5 implies that, given a symplectic manifold (M, ω), any Lagrangian $L \cong \mathbb{CP}^n \subset M$ will define an element $\tau_L \in \mathrm{Symp}_{\mathrm{ct}}(M)$ that is isotopic to the identity in $\mathrm{Diff}_{\mathrm{ct}}(M)$. ◁

As shown by Dimitroglou Rizell and Evans [2015], the choice of framing does play a role in determining the symplectic isotopy class of a spherical Dehn twist. In Section 7, we prove that this is also the case for projective twists. Before then, any given Lagrangian submanifold involved in the construction of a twist is assumed to be endowed with a choice of framing and we omit mentioning this datum, as the results of this paper, up to the last section, are independent of such choices. This is because the autoequivalence of the Fukaya category induced by a Dehn twist (see Section 2.3) is independent of the choice of framing (as a consequence of the shape of the functor; see [Seidel 2008a, Corollary 17.17]). The same is true for the functor induced by the projective twist [Mak and Wu 2018, Theorem 6.10].

2.2 Twists as monodromies

This section approaches twists from a different perspective, one that presents these symplectomorphisms as monodromy maps of fibration-like structures. Dehn twists occur as (local) monodromies of Lefschetz

fibrations, and this is one of the features that has made the study of Dehn twists particularly productive. On the other hand (it is a lesser known fact that) projective twists can be modelled as local monodromies of *Morse–Bott–Lefschetz* fibrations, another class of fibrations admitting more degenerate singularities, which we will not discuss here.

Below we give a brief review of Lefschetz fibrations (mainly following [Seidel 2008a; Maydanskiy and Seidel 2010]) on Liouville manifolds, aimed at setting the notation for future sections, and recall the well-known *Picard–Lefschetz theorem*.

Definition 2.10 A Liouville manifold of finite type is an exact symplectic manifold $(W, \omega = d\lambda_W)$, where $\lambda_W \in \Omega^1(W)$ is called the Liouville form, such that there exists a proper function $h_W : W \to [0, \infty)$ and $c_0 > 0$ with the following property: for all $c \in (c_0, \infty)$ and $x \in h_W^{-1}(c)$, the vector field Z_W dual to λ_W, called the Liouville vector field, satisfies $dh_W(Z_W)(x) > 0$.

For a regular value c of h_W, a closed sublevel set $M := h_W^{-1}([0, c])$ of a Liouville manifold $(W, d\lambda_W)$ is a compact symplectic manifold with contact type boundary $(\Sigma := h_W^{-1}(c), \lambda_W|_\Sigma)$, and it is called a Liouville domain. ◁

Definition 2.11 An exact symplectomorphism between two Liouville manifolds $(W_1, d\lambda_1)$ and $(W_2, d\lambda_2)$ is a diffeomorphism $\psi : W_1 \to W_2$ satisfying $\psi^* \lambda_2 - \lambda_1 = df$ for a compactly supported function $f : W_1 \to \mathbb{R}$. ◁

Definition 2.12 Now let $(M, d\lambda)$ be a Liouville domain with contact boundary $(\Sigma = \partial M, \alpha = \lambda|_\Sigma)$. The negative Liouville flow identifies a collar neighbourhood $C(\Sigma)$ of the boundary with $(-\varepsilon, 0] \times \partial M$, so that $\lambda|_{C(\Sigma)} = e^t \alpha$. An almost complex structure J of *contact type near the boundary* is one that satisfies $de^t \circ J = -\lambda$. ◁

Definition 2.13 Given a Liouville domain $(M, d\lambda)$ as above, we can use the identification of the collar neighbourhood $C(\Sigma)$ to glue an infinite cone and define the symplectic completion of M,

$$(7) \qquad\qquad (W, \omega_W) := \big(M \cup [0, \infty) \times \partial M, d(e^t \alpha)\big),$$

where t is the coordinate on $(0, \infty)$, such that the Liouville flow extends to Z_W with $Z_W|_{[0,\infty) \times \partial M} = \partial_t$.

An almost complex structure J of contact type extends to an almost complex structure J_W on the completion satisfying:

- $J_W(\partial/\partial t) = R_\alpha$, where R_α is the Reeb vector field associated to α.
- J_W is invariant under translations in the t–direction.
- $J_W|_M = J$.

This kind of almost complex structure will be called cylindrical. ◁

We will only consider Liouville manifolds that are complete (ie with complete Liouville vector field) and of finite type, which we can identify as the union of a Liouville domain with a cylindrical noncompact end, equipped with an almost complex structure cylindrical at infinity.

Let $(E^{2n+2}, \Omega_E, \lambda_E)$ be a Liouville manifold, with a compatible almost complex structure J_E, and consider the complex plane with its standard symplectic form and complex structure $j_{\mathbb{C}}$. Let $\pi : E \to \mathbb{C}$ be a map with finitely many critical points, which are all nondegenerate, and contained in a compact set of E. Denote by $\mathrm{Crit}(\pi) := \{x \in E \mid D_x\pi = 0\}$ the set of critical points, and by $\mathrm{Crit}\, v(\pi) := \pi(\mathrm{Crit}(\pi))$ the set of critical values.

Definition 2.14 A Lefschetz fibration on (the Liouville manifold) E is a $(J_E, j_{\mathbb{C}})$–holomorphic map π, ie $D\pi \circ J_E = j_{\mathbb{C}} \circ D\pi$, with the above properties and the following additional features:

(1) For all $x \in E \setminus \mathrm{Crit}(\pi)$, $\ker(D_x\pi) \subset T_xE$ is symplectic.

(2) Every smooth fibre is symplectomorphic to the completion of a Liouville domain $(M, d\lambda_M)$.

(3) There is an open neighbourhood $U^h \subset E$ such that $\pi : E \setminus U^h \to \mathbb{C}$ is proper and $\pi|_{U^h}$ can be trivialised via an isomorphism $f : U^h \cong \mathbb{C} \times ([0, \infty) \times \partial M)$ such that

$$(8) \qquad\qquad f^*(\lambda_E) = \lambda_{\mathbb{C}} + e^t \lambda_M. \qquad\qquad \triangleleft$$

For more details about how this fibration is modelled outside of a neighbourhood of the critical points, see [Maydanskiy and Seidel 2010, (2.1)].

By the first point above, there is a symplectic splitting

$$(9) \qquad\qquad T_xE = \ker(D_x\pi) \oplus T_xE^h,$$

where T_xE^h is the symplectic complement of $\ker(D_x\pi)$ with respect to Ω_E. The decomposition in (9) defines a canonical connection over $\mathbb{C} \setminus \mathrm{Crit}\, v(\pi)$. By the triviality condition (3), for every path $\gamma : [0, 1] \to D \setminus \mathrm{Crit}(\pi)$, there are well-defined parallel transport maps $h_\gamma : E_{\gamma(0)} \to E_{\gamma(1)}$ which yield symplectomorphisms between smooth fibres.

Definition 2.15 A pair $(J_E, j_{\mathbb{C}})$ is said to be *compatible* with π if the following holds:

- $D\pi \circ J_E = j_{\mathbb{C}} \circ D\pi$.

- There is a local Kähler structure J_0 such that $J_E = J_0$ in a neighbourhood of $\mathrm{Crit}(\pi)$.

- On the neighbourhood U^h, J_E is a product, $f^*(J_E) = (j_{\mathbb{C}}, J^{vv})$, where J^{vv} is a cylindrical almost complex structure compatible with $d(e^t\lambda_M)$.

- $\Omega_E(\cdot, J_E\cdot)$ is symmetric and positive definite. $\qquad\qquad \triangleleft$

Remark 2.16 This choice of almost complex structure is not generic. However, the space of compatible almost complex structures on the total space of an exact Lefschetz fibration is contractible [Seidel 2003, Section 2.1], and the moduli spaces we will consider still meet the usual regularity requirements [Seidel 2003, Section 2.2]. ◁

For a Lefschetz fibration on a Liouville manifold (E, Ω_E), the proper fibration obtained as $E \setminus U^h \to \mathbb{C}$ for an open neighbourhood $U^h \subset E$ as above carries the same symplectic information as π with the difference that its fibres are Liouville domains, and as a result the total space admits a nontrivial *horizontal boundary*, given by the union of the boundaries of all fibres.

In most of the paper we will employ this latter type of Lefschetz fibration (for notational simplicity), and, unless specified, an *exact* Lefschetz fibration will denote a fibration obtained in this way.

Now let $\pi \colon E \to \mathbb{C}$ be an exact Lefschetz fibration, with smooth fibre given by the Liouville domain $(M, d\lambda)$. By the triviality assumption of Definition 2.14, there is a neighbourhood of $U^\partial \subset E$ of the horizontal boundary $\partial^h E$ that is isomorphic to an open neighbourhood of the trivial bundle $\mathbb{C} \times \partial M$,

$$\tag{10} U^\partial \cong \mathbb{C} \times M^{\mathrm{out}} \subset \mathbb{C} \times M,$$

where $M^{\mathrm{out}} \subset M$ is a collar neighbourhood of ∂M. The isomorphism is compatible with the Liouville forms and the almost complex structures.

Let $\pi \colon E \to \mathbb{C}$ be a Lefschetz fibration with exact compact fibre (M, ω) and distinct critical values $\mathrm{Crit}\, v(\pi) = \{w_0, \dots, w_m\} \subset D_R$, where $D_R \subset \mathbb{C}$ is a disc of radius R. Fix a basepoint $z_* \in \mathbb{R}$ such that $z_* \gg R$, and an identification $\pi^{-1}(z_*) \cong M$. In what follows we will frequently use the fact that, via parallel transport, any fibre $\pi^{-1}(z)$ for $z \in \mathbb{C}$ with $\mathrm{Re}(z) > R$ can be symplectically identified with the smooth fixed fibre M via parallel transport.

Definition 2.17   (1) A vanishing path associated to a critical value $w_i \in \mathrm{Crit}\, v(\pi)$ is a properly embedded path $\gamma_i \colon \mathbb{R}^+ \to \mathbb{C}$ with $\gamma_i^{-1}(\mathrm{Crit}\, v(\pi)) = \{0\}$, $\gamma_i(0) = w_i$ and $\lim_{t \to \infty} \mathrm{Re}(\gamma(t)) = \infty$ such that, outside of a compact set containing the critical values, the image of γ_i is a horizontal half ray at height $a_i \in \mathbb{R}$:

$$\tag{11} \exists T > 0 \;\; \forall t > T \quad \mathrm{Re}(\gamma_i(t)) > R \;\; \text{and} \;\; \mathrm{Im}(\gamma_i(t)) = a_i.$$

   (2) A distinguished basis of vanishing paths for π is a collection of $m+1$ disjoint paths $(\gamma_0, \dots, \gamma_m) \subset \mathbb{C}$ defined as above, with pairwise distinct heights satisfying $a_0 < a_1 < \cdots < a_m$.

   (3) The corresponding basis of Lefschetz thimbles is the unique set of Lagrangian discs $(\Delta_{\gamma_0}, \dots, \Delta_{\gamma_m})$ in E, where Δ_{γ_i} is defined as the set of points which under the limit $t \to 0$ of the parallel transport maps over γ_i are mapped to the critical point in $\pi^1(w_i)$ (the proof of uniqueness can be found in [Seidel 2008a, (16b)]). Given a general Lefschetz thimble $\mathcal{L}$, define its height $a(\mathcal{L})$ as the value defined in (11). For a pair of thimbles $(\mathcal{L}_0, \mathcal{L}_1)$, set $\mathcal{L}_0 > \mathcal{L}_1$ if $a(\mathcal{L}_0) > a(\mathcal{L}_1)$.
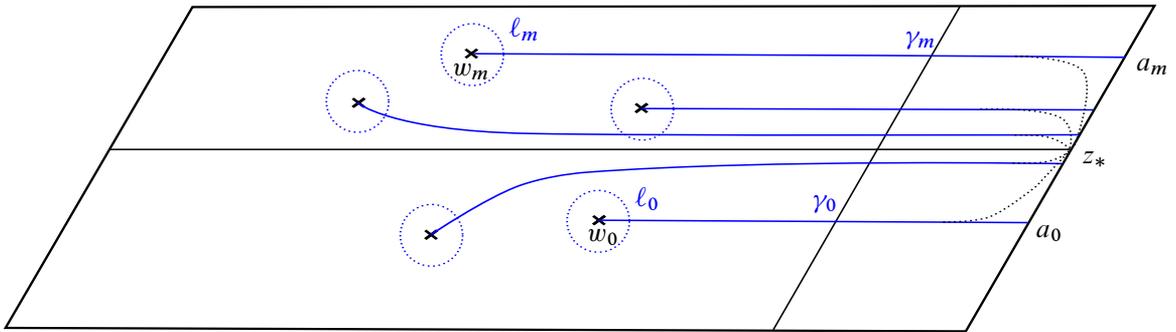
Figure 1: A distinguished basis of vanishing paths $(\gamma_0, \ldots, \gamma_m)$.

(4) There is an associated basis of vanishing cycles $(V_0, \ldots, V_m)$ where, for all $i = 0, \ldots, m$,

$$V_i = \partial \Delta_{\gamma_i} = \Delta_{\gamma_i} \cap M \subset M$$

(using the above identification for smooth fibres). Every vanishing cycle $V_i \subset M$ is an exact Lagrangian sphere which comes with an equivalence class in of diffeomorphisms $S^n \to V_i$ defined up to the action of $O(n+1)$ (called a *framing*). This is induced by the restriction of a diffeomorphism $D^{n+1} \to \Delta_i$ (which is canonical; see [Seidel 2003, Lemma 1.14]). ◁

Definition 2.18 The global monodromy is the symplectomorphism $\phi \in \mathrm{Symp}_{\mathrm{ct}}(M)$ whose Hamiltonian isotopy class is defined by the anticlockwise parallel transport map around a loop through the basepoint z_* encircling all the critical values of the fibration. (Typically, this loop is defined as the smoothing of the concatenation of the loops centred at z_* going around a single critical value as in Figure 1.) ◁

The symplectic Picard–Lefschetz theorem [Arnold 1995] states that the global monodromy ϕ is isotopic to the product of the Dehn twists along the vanishing cycles $(V_0, \ldots, V_m)$,

$$\phi \simeq \tau_{V_0} \cdots \tau_{V_m} \in \mathrm{Symp}_{\mathrm{ct}}(M), \tag{12}$$

and the Hamiltonian isotopy class is independent of the choice of basis of vanishing paths.

On the other hand, given the data $\{(M, \omega), (V_0, \ldots, V_m)\}$, there is an exact Lefschetz fibration $\pi : E \to \mathbb{C}$ with fibre (M, ω), and vanishing cycles $(V_0, \ldots, V_m) \subset M$, unique up to exact symplectomorphism [Seidel 2008a, (16e)].

Remark 2.19 Lefschetz fibrations can be viewed as a special case of *Morse–Bott–Lefschetz* (abbreviated MBL) fibrations, a class of fibrations which allows nonisolated singularities. The monodromies of such fibrations are symplectomorphisms called *fibred twists* [Perutz 2007], which naturally generalise Dehn twists. Projective twists are a special type of fibred twists, and therefore also admit a presentation as local monodromies. However, in this paper, we won't study projective twists from this perspective. ◁

2.3 Functor twists

This section only contains the notation (and the general notions involved) that we will use for the functors of the Fukaya category that are induced by twists.

Let (M, ω) be a Liouville manifold and let k be a field of characteristic 2. Given two closed exact Lagrangian submanifolds $L_0, L_1 \subset M$ the Floer complex is freely generated as a vector space by the intersection points of the (perturbed) Lagrangians $CF(L_0, L_1; k) := \bigoplus_{x \in L_0 \cap L_1} k \langle x \rangle$. The boundary operator $\partial \colon CF(L_0, L_1; k) \to CF(L_0, L_1; k)$ counts J_M–holomorphic strips with boundary conditions on (L_0, L_1) and asymptotic conditions on intersection points. For a compatible cylindrical almost complex structure J_M, the moduli spaces of such curves are compact oriented manifolds [Seidel 2008a, Sections 8–9] and the operator ∂ squares to zero [Seidel 2008a, (9e)], so that $(CF(L_0, L_1; k), \partial)$ is a well-defined cochain complex whose cohomology is the Floer cohomology ring $HF(L_0, L_1; k)$. Floer cohomology is designed to be invariant under Hamiltonian isotopies; if ϕ is the flow of a Hamiltonian vector field, then $HF(L_0, \phi(L_1)) \cong HF(L_0, L_1)$.

Very simply put, the compact Fukaya category, $\mathcal{F}uk(M)$, is an A_∞–category whose objects are closed exact Lagrangian *branes*, which are Lagrangian submanifolds with additional algebraic data, and morphisms the Floer cochain groups between transversely intersecting Lagrangians [Seidel 2008a, (9j) and (12g)]. This category encodes intersection data associated to all its objects, including the Floer differential $\partial = \mu^1$, the Floer cup product μ^2 and higher-order products μ^k (see eg [Seidel 2008a, (9j) and (12g)]). It is well defined for any Liouville manifold (see [Seidel 2008a]).

Two Lagrangians that are Hamiltonian isotopic are quasi-isomorphic objects in the Fukaya category, which means they are isomorphic objects of the associated cohomological category, which we denote by $H(\mathcal{F}uk(M))$. We denote the automorphisms of $H(\mathcal{F}uk(M))$ (ie the automorphisms of the Fukaya category up to quasi-isomorphism) by $\text{Auteq}(\mathcal{F}uk(M))$.

Let $\text{Tw}(\mathcal{F}uk(M))$ be the category of twisted complexes in $\mathcal{F}uk(M)$ (see [Seidel 2008a, (3k)]), and $D^\flat \mathcal{F}uk(M) := H(\text{Tw}\,\mathcal{F}uk(M))$ the cohomology category of $\text{Tw}(\mathcal{F}uk(M))$.

There is a map

$$(13) \qquad\qquad \Phi \colon \text{Symp}_{ct}(M) \to \text{Auteq}(D^\flat \mathcal{F}uk(M))$$

to the group of auto-equivalences of the Fukaya category (modulo quasi-isomorphism) such that, given $\phi \in \text{Symp}_{ct}(M)$, $\Phi(\phi)$ sends a Lagrangian $L \subset M$ to another Lagrangian $\phi(L) \subset M$ (we avoid discussing a graded situation in this context). The map factors through the quotient by the subgroup $\text{Ham}_{ct}(M) \subset \text{Symp}_{ct}(M)$ of compactly supported Hamiltonian diffeomorphisms, so, given an exact Lagrangian sphere/projective space L and its associated twist τ_L, $\Phi(\tau_L)$ defines a well-defined element of $\text{Aut}(D^\flat \mathcal{F}uk(M))$, which we denote by T_L.

Seidel [2003] showed that, for a Dehn twist τ_L, the induced functor $T_L \in \mathrm{Aut}(D^\flat \mathcal{F}\mathrm{uk}(M))$ fits into an exact triangle (see [Seidel 2008a, (17j)]).

Recently, there have been generalisations of Seidel's triangle for a wider class of symplectomorphisms, achieved through a range of different techniques. Wehrheim and Woodward [2016] proved the existence of an exact triangle for fibred twists using quilt theory adapted to Morse–Bott Lefschetz fibrations.

Mak and Wu [2018] treated the case of projective twists, using Lagrangian cobordism theory as developed in [Biran and Cornea 2013; 2014]. They proved that the autoequivalence induced by a (real, complex, quaternionic) projective twist is isomorphic to a double cone of functors in $\mathrm{Aut}(\mathrm{Tw}\,\mathcal{F}\mathrm{uk}(M))$ [Mak and Wu 2018, Theorem 6.10].

Under the appropriate circumstances, the mirror symmetry conjecture gives conjectural descriptions of such functors. If a symplectic manifold (M, ω) has a mirror complex manifold (X, J), there are autoequivalences of the Fukaya category of M that are induced by autoequivalences of the derived category of coherent sheaves of X (we call such autoequivalences *algebraic twist functors*, and will only refer to them in Remark 4.15).

3 Commuting diagrams of twists

In this section we introduce the geometric ideas underpinning the philosophy of the Hopf correspondence. We prove a criterion for relating projective twists in a Liouville manifold (W, ω) to Dehn twists in another Liouville manifold (Y, Ω).

3.1 Complex projective Lagrangians

We begin by considering Lagrangian complex projective spaces.

Fix the round metric on S^{2n+1}, with norm $\|\cdot\|_S$, and consider the free S^1–action on S^{2n+1} by complex multiplication. The orbits of the action are great circles ("Hopf circles"), hence geodesics, and the action is isometric.

Consider the quotient map $h: S^{2n+1} \to \mathbb{CP}^n$, which is the (generalised) Hopf fibration. It is a Riemannian submersion that uniquely defines the Fubini–Study metric g_P on $\mathbb{CP}^n$. Identify the tangent bundles with their corresponding cotangent bundles $TS^{2n+1} \cong T^*S^{2n+1}$ and $T\mathbb{CP}^n \cong T^*\mathbb{CP}^n$ via the canonical isomorphism induced by the metrics.

The Hopf action on S^{2n+1} lifts to a Hamiltonian S^1–action on the cotangent bundle $(T^*S^{2n+1}, \omega_{T^*S^n})$ [Guillemin and Sternberg 1990]. Let $\mu: T^*S^{2n+1} \to \mathbb{R}$ be the moment map of this action. Assume 0 is a regular value of μ and consider the level set $V := \mu^{-1}(0) \subset T^*S^{2n+1}$, which has the structure of a principal S^1–bundle $p: V \to T^*\mathbb{CP}^n$ over the symplectic quotient $T^*S^{2n+1} /\!\!/ S^1 := V/S^1 \cong T^*\mathbb{CP}^n$.

Lemma 3.1 *Let $\tau_{S^{2n+1}} \in \mathrm{Symp}_{\mathrm{ct}}(T^*S^{2n+1})$ and $\tau_{\mathbb{CP}^n} \in \mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$ be the model Dehn and projective twists, respectively. Let $p \colon V := \mu^{-1}(0) \to T^*\mathbb{CP}^n$ be the symplectic quotient map as above. There is a commuting diagram*

(14)
$$
\begin{array}{ccc}
V & \xrightarrow{\ \tau_{S^{2n+1}}|_V\ } & V \\
\Big\downarrow{\scriptstyle p} & & \Big\downarrow{\scriptstyle p} \\
T^*\mathbb{CP}^n & \xrightarrow{\ \tau_{\mathbb{CP}^n}\ } & T^*\mathbb{CP}^n
\end{array}
$$

Proof The Hopf action is isometric, ie for any $g \in S^1$, the induced map $\psi_g \in \mathrm{Diff}(S^{2n+1})$ is an isometry. This implies that the differential maps on the tangent bundles $D_x\psi_g \colon T_x S^{2n+1} \to T_{\psi_g(x)} S^{2n+1}$ commute (for any $x \in S^{2n+1}$) with the geodesic flow.

The cogeodesic flow $\Phi_{\tilde{H}}^t$ on T^*S^{2n+1} is induced by the Hamiltonian function

(15)
$$
\tilde{H} \colon T^*S^{2n+1} \to \mathbb{R}, \quad (x, \xi) \mapsto \|\xi\|_S.
$$

This is S^1–invariant, so there is a Hamiltonian function $H \colon T^*\mathbb{CP}^n \to \mathbb{R}$ defined on the quotient with respect to the submersion metric g_P, which induces the (co)geodesic flow on $T^*\mathbb{CP}^n$. Since p is induced by a Riemannian submersion, we have the relation $p \circ \Phi_{\tilde{H}}^t|_V = \Phi_H^t \circ p|_V$, and, for any choice of cut-off function r_ε as in Section 2.1,

(16)
$$
p \circ \sigma_{r_\varepsilon(\|\xi\|_S)}^{\tilde{H}}(\xi) = \sigma_{r_\varepsilon(\|p(\xi)\|_P)}^{H} \circ p(\xi), \quad \xi \in V \subset T^*S^{2n+1},
$$

where σ_t^H and $\sigma_t^{\tilde{H}}$ are the Hamiltonian S^1–actions induced by H and $\tilde{H}$, respectively, as in Section 2.1.

Any geodesic connecting a point on S^{2n+1} to its antipode projects, under p, to a closed geodesic of minimal period on $\mathbb{CP}^n$ (it cannot collapse to a point since the Hopf action is isometric), so the definitions of the twists in Section 2.1 imply that $p \circ \tau_{S^{2n+1}}|_V = \tau_{\mathbb{CP}^n} \circ p|_V$. $\square$

We now extend the above discussion to a more global situation; in order to do that it is necessary to set the following assumption:

Assumption (CX) Let (W, ω) be a $4n$–dimensional Liouville manifold with a homology class $\alpha \in H^2(W; \mathbb{Z})$ and Lagrangian complex projective spaces $K_1, \ldots, K_m \subset W$ such that

$$
\alpha|_{K_i} = x \in H^2(\mathbb{CP}^n; \mathbb{Z}) \quad \text{for all } i,
$$

where $x = c_1(\mathcal{O}_{\mathbb{CP}^n}(-1))$ is the generator of the cohomology ring $H^*(\mathbb{CP}^n; \mathbb{Z}) \cong \mathbb{Z}[x]/x^{n+1}$.

Proposition 3.2 *Let (W, ω) be a $4n$–dimensional Liouville manifold containing embedded Lagrangian complex projective spaces $K_1, \ldots, K_m \subset W$. Assume there exists a class $\alpha \in H^2(W; \mathbb{Z})$ satisfying Assumption (CX). Then there is a $(4n+2)$–dimensional Liouville manifold (Y, Ω) with Lagrangian*

spheres $L_1, \ldots, L_m \subset Y$, a coisotropic submanifold $V \subset Y$ with the structure of an S^1–fibre bundle $p \colon V \to W$ such that, for each $i \in \{1, \ldots, m\}$, $L_i \subset V$, and there is a commuting diagram

(17)

$$
\begin{array}{ccc}
V & \xrightarrow{\;\tau_{L_i}|_V\;} & V \\
\downarrow{\scriptstyle p} & & \downarrow{\scriptstyle p} \\
W & \xrightarrow[\;\tau_{K_i}\;]{} & W
\end{array}
$$

The class $\alpha \in H^2(W; \mathbb{Z})$ restricts to a generator $x \in H^2(\mathbb{CP}^n; \mathbb{Z})$ on each Lagrangian K_i, so there is a complex line bundle $\mathcal{L} \to W$ satisfying $c_1(\mathcal{L}) = \alpha$ which is modelled on the tautological line bundle $\mathcal{O}_{\mathbb{CP}^n}(-1)$ over K_i for $i = 1, \ldots, m$. Fix a metric $\|\cdot\|_{\mathcal{L}}$ on $\mathcal{L}$, and for $u \in \mathcal{L}$ define a function $r(u) := \|u\|_{\mathcal{L}}$. Set $V := \{u \in \mathcal{L}, \ r(u) = 1\}$. Over K_i, V defines a sphere $L_i := V|_{K_i}$.

Lemma 3.3 *The $\mathbb{C}^*$–bundle associated to $\mathcal{L}$ is a Liouville domain, where the spheres L_i are embedded as Lagrangian submanifolds.*

Proof enote this bundle by $q \colon Y \to W$. Following [Ritter 2014, Section 7.2], we build a symplectic form Ω on Y, making the spheres L_i Lagrangian, and find the appropriate vector field which will be Liouville with respect to Ω.

The metric induces a connection one-form γ^∇ on $\mathcal{L} \setminus 0$ satisfying

(18) $\quad \gamma^\nabla|_{H_u^\nabla} = 0, \quad \gamma^\nabla|_{T_u^v \mathcal{L}} = \gamma \quad$ for all $u \in \mathcal{L} \setminus 0$, $\qquad [d\gamma^\nabla] = -q^*(c_1(\mathcal{L})) = -q^*(\alpha),$

where $H_u^\nabla \mathcal{L}$ is the horizontal distribution associated to the connection ∇ at u, $T_u^v \mathcal{L}$ the vertical distribution, and γ the fibrewise angular form defined by the metric. Let $\Omega := q^*\omega + d(f(r)\gamma^\nabla)$ for a function $f \in C^\infty(\mathbb{R})$ with

$$
f(1) = 0, \qquad f'(r) > 0 \quad \text{for all } r \in \mathbb{R}.
$$

Then Ω defines a symplectic form in a neighbourhood of $\{r = 1\}$, and L_i is Lagrangian with respect to Ω. Let λ be the Liouville one-form on W with $d\lambda = \omega$. Define $\lambda_Y := q^*\lambda + f(r)\gamma^\nabla$, so that $d(\lambda_Y) = \Omega$. Then (λ_Y, Ω) defines a Liouville structure near $\{r = 1\}$ (the symplectic dual to λ_Y points outwards along a small neighbourhood of $\{r = 1\}$). Therefore, a symplectic completion along this neighbourhood yields a Liouville manifold that is diffeomorphic to Y, containing the Lagrangian spheres $L_1, \ldots, L_m$. $\qquad\square$

Proof of Proposition 3.2 Let $\mathcal{L} \to W$ be the complex line bundle we have constructed above with $c_1(\mathcal{L}) = \alpha$. For each Lagrangian projective space $K_i \subset W$, the restriction of the bundle $\mathcal{L}|_{K_i}$ is modelled on the tautological line bundle, which implies that $L_i \to K_i$ is modelled on the Hopf quotient map $h \colon S^{2n+1} \to \mathbb{CP}^n$. The commutativity of (17) follows by the local commuting diagram of cotangent bundles (14). $\qquad\square$

Example 3.4 Without Assumption (CX), Proposition 3.2 is in general not true, as the following example illustrates. Consider the manifold W obtained by attaching a 3–handle to the contact boundary of $D_s T^* \mathbb{CP}^2$ ($s > 0$) such that $H^2(W; \mathbb{Z}) = 0$. On one hand, W contains a nontrivial Lagrangian $K = \mathbb{CP}^2 \subset W$ coming from the zero section (which is preserved by the handle attachment, since it is disjoint from the boundary; note that the handle attachment is subcritical, so in fact the whole wrapped Fukaya category is preserved; see [Ganatra et al. 2024]). However, as there is no nontrivial 2–cohomology class on W, there is no nontrivial S^1–bundle over W that can be used to build a sphere over K. ◁

3.2 Real projective Lagrangians

A similar procedure can be applied to a Liouville manifold containing real projective Lagrangians with an appropriate cohomology criterion. First recall the following.

Let $S^0 \cong \mathbb{Z}/2\mathbb{Z}$ act on the sphere S^n by the antipodal map. The quotient map $h \colon S^n \to \mathbb{RP}^n$ is in this case a covering map, and induces a symplectic double cover $q \colon T^* S^n \to T^* \mathbb{RP}^n$ with $q^* \omega_{T^* \mathbb{RP}^n} = \omega_{T^* S^n}$.

Lemma 3.5 [Mak and Wu 2018, Lemma 2.4] *Let* $\tau_{\mathbb{RP}^n} \in \mathrm{Symp}_{\mathrm{ct}}(T^* \mathbb{RP}^n)$ *be the* $\mathbb{RP}^n$–*twist defined as in Section 2.1. Then the diagram*

(19)
$$
\begin{array}{ccc}
T^* S^n & \xrightarrow{\ \tau_{S^n}\ } & T^* S^n \\
\downarrow{\scriptstyle q} & & \downarrow{\scriptstyle q} \\
T^* \mathbb{RP}^n & \xrightarrow{\ \tau_{\mathbb{RP}^n}\ } & T^* \mathbb{RP}^n
\end{array}
$$

commutes.

Assumption (RE) Let (W, ω) be a $2n$–dimensional Liouville manifold with a homology class $\alpha \in H^1(W; \mathbb{Z}/2\mathbb{Z})$ and Lagrangian real projective spaces $K_1, \ldots, K_m \subset W$ such that

$$
\alpha|_{K_i} = x \in H^1(\mathbb{RP}^n; \mathbb{Z}/2\mathbb{Z}) \quad \text{for all } i,
$$

where $x = e(\gamma_{\mathbb{R}}^{1,n+1})$ is the Euler class of the real tautological bundle $\gamma_{\mathbb{R}}^{1,n+1} \to \mathbb{RP}^n$, and generator of the cohomology ring $H^*(\mathbb{RP}^n; \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}[x]/x^{n+1}$.

Proposition 3.6 *Let* (W, ω) *be a* $2n$–*dimensional Liouville manifold containing embedded Lagrangian real projective spaces* $K_1, \ldots, K_m \subset W$. *Assume there is a class* $\alpha \in H^1(\mathbb{RP}^n; \mathbb{Z}/2\mathbb{Z})$ *satisfying Assumption (RE). Then there is a* $2n$–*dimensional Liouville manifold* $(\widetilde{W}, \widetilde{\omega})$ *containing Lagrangian spheres* $L_1, \ldots, L_m \subset \widetilde{W}$ *and a commuting diagram*

(20)
$$
\begin{array}{ccc}
\widetilde{W} & \xrightarrow{\ \tau_{L_i}\ } & \widetilde{W} \\
\downarrow{\scriptstyle q} & & \downarrow{\scriptstyle q} \\
W & \xrightarrow{\ \tau_{K_i}\ } & W
\end{array}
$$

Proof In this case, the class $\alpha \in H^1(W; \mathbb{Z}/2\mathbb{Z})$ defines a symplectic double cover $q \colon (\widetilde{W}, \widetilde{\omega}) \to (W, \omega)$. Each Lagrangian $K_i \cong \mathbb{RP}^n$ then lifts to its double cover L_i, which is a sphere $S^n \subset \widetilde{W}$. Let λ be the Liouville form on W. As q is symplectic, $\widetilde{\omega} = q^*(\omega) = q^*(d\lambda) = d(q^*(\lambda))$, and $\widetilde{\lambda} := q^*(\lambda)$ defines a Liouville form on $\widetilde{W}$, which gives $\widetilde{W}$ the structure of a Liouville manifold. Then the result follows by the local case, illustrated by Lemma 3.5. $\qquad\square$

Remark 3.7 It is possible to obtain an analogous diagram for the quaternionic twist as follows. Consider the free $S^3 \simeq \mathrm{Sp}(1)$–action on S^{4n+3} inducing the quotient map $h \colon S^{4n+3} \to \mathbb{HP}^n$. This is a submersion as in the complex case, and the same arguments (with the natural metrics) yield the local commuting diagram

(21)
$$
\begin{array}{ccc}
\mu^{-1}(0) & \xrightarrow{\ \tau_{S^{4n+3}}|_{\mu^{-1}(0)}\ } & \mu^{-1}(0) \\
\downarrow{\scriptstyle p} & & \downarrow{\scriptstyle p} \\
T^*\mathbb{HP}^n & \xrightarrow{\ \tau_{\mathbb{HP}^n}\ } & T^*\mathbb{HP}^n
\end{array}
$$

where $p \colon \mu^{-1}(0) \to T^*\mathbb{HP}^n$ is the S^3–fibre bundle induced given by the symplectic quotient map of the Hamiltonian action induced on T^*S^{4n+3}.

Given an $8n$–dimensional symplectic manifold (W, ω) containing quaternionic projective Lagrangians, one would hope to find a cohomological condition to ensure the existence of a symplectic $(8n+6)$–dimensional manifold (Y, Ω) with corresponding Lagrangian spheres, as we did for the real and complex cases. However, homotopy classes of maps $W \to \mathbb{HP}^\infty \cong \mathrm{Sp}(1)$ do not classify quaternionic line bundles over W, so there is no analogue of Assumptions (CX) and (RE) to ensure the existence of such a manifold and a commuting diagram of the form of (17). $\qquad\triangleleft$

4 The Hopf correspondence

In this section we discuss the main theoretical device in action in this paper; Lagrangian correspondences. We begin by reviewing the main concepts from Wehrheim–Woodward Lagrangian correspondence theory (Section 4.1). The rest of the section is then focussed on the correspondence that will be used in our applications, the *Hopf correspondence*. Given a real/complex projective Lagrangian $K \subset W$ in a Liouville manifold (W, ω) satisfying (RE)/(CX), the Hopf correspondence associates to it a Lagrangian sphere $L \subset Y$ in an auxiliary Liouville manifold (Y, Ω). The key use of the Hopf correspondence in this section is aimed at achieving a categorical version of the commuting diagrams of the previous section. To do this, we first show that the Hopf correspondence $\Gamma \subset W^- \times Y$ induces a well-defined functor $\Theta_\Gamma \colon \mathcal{F}\mathrm{uk}(W) \to \mathcal{F}\mathrm{uk}(Y)$ (Sections 4.2 and 4.3). We then show that the functors of $\mathcal{F}\mathrm{uk}(W)$ induced by projective twists are entwined, via the correspondence, with the functors of $\mathcal{F}\mathrm{uk}(Y)$ induced by the Dehn twists (Section 4.4). In Section 4.5, we show that the Hopf correspondence can be used to build a symplectic Gysin sequence as established in [Perutz 2008].

4.1 Lagrangian correspondences

We summarise the basic definitions and results associated to Lagrangian correspondences in the setting of [Wehrheim and Woodward 2012; 2010a; 2010b; Ma'u et al. 2018]. For the entire section we let k be a coefficient field of characteristic two.

Definition 4.1 [Wehrheim and Woodward 2010b] A *Lagrangian correspondence* between two symplectic manifolds (M_k, ω_k) and (M_{k+1}, ω_{k+1}) ("from M_k to M_{k+1}") is a Lagrangian submanifold $L_{k,k+1} \subset (M_k^- \times M_{k+1}) := (M_k \times M_{k+1}, -\omega_k \oplus \omega_{k+1})$. A *cycle of Lagrangian correspondences* of length $r \geq 1$ is a sequence of symplectic manifolds $(M_0, \ldots, M_{r+1} = M_0)$ together with a sequence of Lagrangian correspondences $\underline{L} := (L_{01}, L_{12}, \ldots, L_{(r-1)r}, L_{r0})$ such that $L_{k(k+1)} \subset M_k^- \times M_{k+1}$ for $k = 0, \ldots, r$. ◁

For example, a Lagrangian submanifold L of a symplectic manifold (M, ω) is a trivial example of Lagrangian correspondence, seen as $L \subset \{\text{pt}\}^- \times M = M$ (see other examples below).

Definition 4.2 [Wehrheim and Woodward 2010a, Definition 2.0.4] Let (M_i, ω_i) for $i = 0, 1, 2$ be symplectic manifolds and $L_{01} \subset M_0^- \times M_1$ and $L_{12} \subset M_1^- \times M_2$ be Lagrangian correspondences.

(1) The correspondence transpose to L_{01} is defined as $L_{01}^t := \{(m_1, m_0) \mid (m_0, m_1) \in L_{01}\} \subset M_1^- \times M_0$. Note that, for a simple Lagrangian $L \subset M$ of a single symplectic manifold M, we won't distinguish L from its conjugate.

(2) The composition of L_{01} and L_{12} is defined as

(22) $L_{01} \circ L_{12} := \{(m_0, m_2) \in M_0^- \times M_2 \mid (m_0, m_1) \in L_{01}, (m_1, m_2) \in L_{12} \text{ for some } m_1 \in M_1\}$
$\subset M_0^- \times M_2$

and it is called embedded if it defines an embedded Lagrangian submanifold of $M_0^- \times M_2$. ◁

Example 4.3 [Perutz 2008, 1.1] Let (M^{2n}, ω_M) be a symplectic manifold with a coisotropic embedding $\iota : V \hookrightarrow M$. If the foliation defined by the integrable distribution TV^ω is a fibration $p : V \to B$, then the leaf space is a symplectic manifold (B, ω_B) satisfying $p^* \omega_B = \iota^*(\omega_M)$. The (transpose) graph of p,

$$\Gamma := \{(p(v), v) \mid v \in V\} \subset (B \times M, -\omega_B \oplus \omega_M),$$

is a Lagrangian correspondence. ◁

A special case of Example 4.3 is when the coisotropic submanifold is obtained as a level set of a moment map induced by a Hamiltonian action:

Example 4.4 [Wehrheim and Woodward 2010b, Example 2.0.2(e)] Let (M, ω_M) be a symplectic manifold. Let G be a compact Lie group acting on M Hamiltonianly with moment map $\mu : M \to \mathfrak{g}^*$. If G acts freely on $\mu^{-1}(0)$, the latter is a smooth G–fibred coisotropic over the symplectic quotient $W := M /\!\!/ G = \mu^{-1}(0)/G$. W is a symplectic manifold with symplectic structure $\omega_{M /\!\!/ G}$ given by the

Marsden–Weinstein theorem (see for example [McDuff and Salamon 2017, Section 5.4]). The graph of the quotient map $p\colon \mu^{-1}(0) \to W$ is a Lagrangian submanifold of $(M \times W, -\omega_M \oplus \omega_W)$ and defines a Lagrangian correspondence, relating Lagrangians of M with Lagrangians of its symplectic quotient. ◁

4.2 Induced functors

Wehrheim and Woodward [2010a; 2010b] introduced a Floer cohomology theory adapted to cycles of closed Lagrangian correspondences $\underline{L} := (L_{01}, \ldots, L_{r0})$, called *quilted Floer cohomology* and denoted by $\mathrm{HF}(\underline{L}; k)$. Pseudoholomorphic quilts are a generalisation of the usual pseudoholomorphic strips used in standard Lagrangian Floer theory, and the quilted invariant is defined by counting pseudoholomorphic quilts with boundary constraints defined by the Lagrangian correspondences [Wehrheim and Woodward 2010b, Section 5]. It can be viewed as a Floer theory in product symplectic manifolds (we refer to [Wehrheim and Woodward 2010b, Section 4.3] for definitions).

One of the main features is that, given a cycle $\underline{L}$ of Lagrangian correspondences, quilted Floer cohomology is invariant under embedded composition (as in Definition 4.2) of subsequent Lagrangians in $\underline{L}$.

Theorem 4.5 [Wehrheim and Woodward 2010b, Theorem 5.4.1] *Let $\underline{L} = (L_{01}, \ldots, L_{r0})$ be a cyclic sequence of closed, exact embedded and oriented Lagrangian correspondences between Liouville manifolds $(M_0, \ldots, M_{r+1} = M_0)$ such that $L_{(i-1)i} \circ L_{i(i+1)}$ is embedded for each i. Then, for $\underline{L}' := (L_{01}, \ldots, L_{(j-1)j} \circ L_{j(j+1)}, \ldots, L_{r0})$, there is an isomorphism $\mathrm{HF}(\underline{L}; k) \cong \mathrm{HF}(\underline{L}'; k)$.*

Ma'u et al. [2018] proved that, under certain assumptions, a Lagrangian correspondence Γ_{01} between given symplectic manifolds (M_0, ω_0) and (M_1, ω_1) defines an A_∞–functor $\Theta_{\Gamma_{01}}$ between $\mathcal{F}\mathrm{uk}(M_0)$ and the dg-category of A_∞–modules over $\mathcal{F}\mathrm{uk}(M_1)$. The functor is realised as the geometric composition $(\cdot) \circ \Gamma_{01}$ of Lagrangians submanifolds of M_0 with the correspondence, and this important result relies on the invariance of Theorem 4.5. If for every Lagrangian in M_0 the composition outputs an embedded Lagrangian of M_1, the induced functor is between Fukaya categories.

Theorem 4.6 [Ma'u et al. 2018, Theorem 1.1] *Assume M_0 and M_1 are Liouville manifolds, and let $\Gamma_{01} \subset M_0^- \times M_1$ be a closed, exact and embedded correspondence such that, for any closed embedded Lagrangian $K_0 \subset M_0$, the geometric composition*

(23) $$L_1 := K_0 \circ \Gamma_{01} = \{m_1 \in M_1 \mid (m_0, m_1) \in \Gamma_{01} \text{ for some } m_0 \in K_0\} \subset M_1$$

is a closed embedded Lagrangian in M_1. This assignment defines an A_∞–functor

(24) $$\Theta_{\Gamma_{01}} \colon \mathcal{F}\mathrm{uk}(M_0) \to \mathcal{F}\mathrm{uk}(M_1), \quad \Theta_{\Gamma_{01}}(K_0) = L_1.$$

In the above theorem, the correspondences are required to be closed, exact (or satisfy suitable monotonicity conditions) and embedded. Gao [2017a; 2017b] developed noncompact generalisations of Theorem 4.6, including noncompact Lagrangian correspondences, in the setting of wrapped Fukaya categories.

In both cases, the main theoretical device at work behind a result such as Theorem 4.6 (or Gao's equivalent) is quilted Floer theory, which, in [Gao 2017b], was adapted to a version suitable for noncompact correspondences. In this work we focus on a Lagrangian correspondence in a setting that features some properties of both theories. Before introducing our setting (see below), we review the types of Lagrangians that are admitted in a Gao's setting.

Let (M_0, ω_0) and (M_1, ω_1) be Liouville manifolds with cylindrical almost complex structures J_0 and J_1 and Liouville flows Z_0 and Z_1, respectively. The product manifold $(M_0 \times M_1, -\omega_0 \times \omega_1)$ is a Liouville manifold with respect to the product almost complex structure $J_{01} := -J_0 \times J_1$ and Liouville flow $Z_{01} := \pi_0^*(Z_0) + \pi_1^*(Z_1)$ for the projections $\pi_i \colon M_0 \times M_1 \to M_i$ for $i = 1, 2$.

Let $\Sigma \subset M_0 \times M_1$ be the contact hypersurface given in [Gao 2017b, Section 2.2], so that we can fix a choice of cylindrical end that is compatible with the choices above. In other words, there is a compact set $U \subset M_0 \times M_1$ bounded by Σ such that there is a diffeomorphism $M_0 \times M_1 \setminus U \cong [0, \infty) \times \Sigma$ [Gao 2017b, (2.5)].

Definition 4.7 A Lagrangian submanifold is said to be *conical* if it is an exact, properly embedded Lagrangian which is preserved by the Liouville vector field over the cylindrical end. ◁

Definition 4.8 [Gao 2017b, Definition 3.9] A Lagrangian submanifold $\Gamma_{01} \subset M_0^- \times M_1$ is called admissible if it is

(1) either a product of conical Lagrangian submanifolds of M_0^- and M_1,

(2) or a Lagrangian that is conical with respect to the cylindrical end $\Sigma \times [0, +\infty)$. ◁

Gao [2017b, Theorem 1.5] defines geometric composition for this type of Lagrangian correspondences and proves the analogue of Theorem 4.5. Moreover, he shows the open version of Theorem 4.6, namely that such a Lagrangian correspondence induces a functor of wrapped Fukaya categories [Gao 2017a, Theorem 1.2].

Below we focus on the type of correspondences we consider in this paper, which arises as a special case of Example 4.3 for a noncompact coisotropic. It is a class of exact, embedded, but not closed correspondences between Liouville manifolds.

Setting Let (M_0, ω_0) and (M_1, ω_1) be Liouville manifolds such that there is a fibration $q \colon M_1 \to M_0$ with Liouville fibres.

Let $\Gamma_{01} \subset M_0^- \times M_1$ be a Lagrangian correspondence obtained as the (transpose) graph of a proper fibration $p \colon V \to M_0$, where $V \subset M_1$ is a fibred coisotropic as in Example 4.3, and $q|_V = p$.

On $M_0^- \times M_1$ set the product almost complex structure $J_{01} = -J_0 \times J_1$ for cylindrical almost complex structures on M_0 and M_1, so that the fibration $(\mathrm{id}, q) \colon M_0^- \times M_1 \to M_0^- \times M_0$ is (J_{01}, J_{00})–holomorphic for $J_{00} = -J_0 \times J_0$.

Then $\Gamma_{01} = \{(p(v), v) \mid v \in V\}$ is properly fibred over the diagonal $\Delta_{M_0} := \{(p(v), p(v)) \mid v \in V\} \subset M_0^- \times M_0$, which is a conical Lagrangian correspondence in $M_0^- \times M_0$. However, the original correspondence Γ_{01} is not conical, or more generally admissible in the sense of Definition 4.8.

Consequently, the above setting doesn't exactly fit either the compact or the open quilted theories, but it is a combination of the two: it depicts a class of noncompact correspondences which nevertheless induces a functor of compact Fukaya categories.

Axiom 1 The type of Lagrangian correspondence $\Gamma_{01} \subset M_0^- \times M_1$ defined in the above setting induces a functor

$$(25) \qquad \Theta_{\Gamma_{01}} : \mathcal{F}uk(M_0) \to \mathcal{F}uk(M_1), \quad \Theta_{\Gamma_{01}}(K_0) = L_1.$$

Experts will recognise the validity of the above statement that we have set as an axiom. Proving it as a theorem would require a lengthy digression necessary to fill in all details covered in [Wehrheim and Woodward 2010a; Gao 2017b; 2017a]. In Lemma 4.9, we restrict to proving the invariance of quilted Floer cohomology under Lagrangian correspondences. Given invariance, the results of [Wehrheim and Woodward 2010a] yield a functor on the cohomological category. The extension to an A_∞–functor, which would turn Axiom 1 into a theorem, can then be obtained by considering higher A_∞–products, which we omit here.

Lemma 4.9 *Let $K \subset M_0$ and $L' \subset M_1$ be closed exact Lagrangians and consider the cycle of correspondences $(K, \Gamma_{01}, L') \subset (\mathrm{pt}, M_0, M_1)$. Then the quilted Floer cohomology group $\mathrm{HF}(K, \Gamma_{01}, L')$ is well defined and satisfies the invariance property*

$$(26) \qquad \mathrm{HF}(K, \Gamma_{01}, L') \cong \mathrm{HF}(K \circ \Gamma_{01}, L') = \mathrm{HF}(L, L').$$

Proof By definition (see [Wehrheim and Woodward 2010b, Section 4.3]), the generators of the cochain complex $\mathrm{CF}(K, \Gamma_{01}, L')$ are given by the generators of $\mathrm{CF}(K \times L, \Gamma_{01})$. These intersection points must be contained in a compact region, since $K \subset M_0$ and $L' \subset M_1$ are closed Lagrangians. By [Wehrheim and Woodward 2012, Proposition 2.2.1], the cochain groups $\mathrm{CF}(K \circ \Gamma_{01}, L') = \mathrm{CF}(L, L')$ and $\mathrm{CF}(K \times L, \Gamma_{01})$ are isomorphic.

We now analyse the Floer trajectories involved in the computation of $\mathrm{HF}(K, \Gamma_{01}, L')$.

By the maximum principle, the only noncompactness phenomenon that could occur would be a J_{01}–holomorphic curve escaping a compact set on the noncompact boundary condition Γ_{01}. However, all such curves, and any Floer trajectory of interest, are contained in a compact set, as we now explain.

By assumption, J_{01}–holomorphic curves with boundary conditions on (K, Γ_{01}, L') project under (id, q) to J_{00}–holomorphic curves involved in the complex for the tuple (K, Δ_{M_0}, K'), where $K' \circ \Gamma_{01} = L'$ and $(\mathrm{id}, q)(\Gamma_{01}) = \Delta_{M_0}$.

The (quilted) Floer cohomology group for the cycle of Lagrangian correspondences $(K, \Delta_{M_0}, K') \subset$ (pt, M_0, M_0) can be defined as the Floer cohomology group $\mathrm{HF}(K, \Delta_{M_0}, K') := \mathrm{HF}^*(K \times K', \Delta_{M_0})$ [Gao 2017b, Lemma 4.8]. Moreover, by [Gao 2017a, Theorem 1.2], Δ_{M_0} induces the identity functor, so clearly all the J_{00}–holomorphic strips involved in the complex $\mathrm{CF}(K, \Delta_{M_0}, K')$ are well behaved, and moreover we have $\mathrm{HF}(K, \Delta_{M_0}, K') \cong \mathrm{HF}(K, K')$.

Because of properness of $(q, \mathrm{id})|_{\Gamma_{01}} \colon \Gamma_{01} \to M_0 \times M_0$, if there were any J_{01}–holomorphic curve escaping to infinity at the boundary condition Γ_{01}, then it would project to a J_{00}–holomorphic curve escaping to infinity at the boundary condition on Δ_{M_0}, which cannot happen. $\square$

Remark 4.10 Let $K, K' \subset (M_0, \omega_0)$ be closed exact Lagrangians. For any conical correspondence (not just the diagonal) $\Gamma_{00} \subset M_0^- \times M_0$, compactness of moduli spaces of curves involved in the quilted complex $\mathrm{CF}(K, \Gamma_{00}, K')$ (for compact Lagrangians $K, K' \subset M_0$) is preserved. Namely, all intersection points lie in a compact region, so, by exactness, both energy and symplectic area are bounded. We can apply a reverse isoperimetric inequality, according to which the length of the boundary of such a curve is bounded by a quantity proportional to its area [Groman and Solomon 2014, Theorem 1.4].

This ensures that the boundary of all pseudoholomorphic curves is contained in a compact set, which can then be determined by using a monotonicity lemma in the likes of [Seidel and Smith 2005, Lemma 13]. Again, by exactness there is no bubbling, so the moduli spaces of such curves are compact. $\triangleleft$

4.3 The Hopf correspondence

We can finally introduce the correspondence of interest, the *Hopf correspondence*. This is a Lagrangian correspondence obtained as the graph of a spherically fibred coisotropic submanifold as in Example 4.3.

We use the discussions of Sections 3.1 and 3.2 to explain how, for each type of Lagrangian projective space $K \cong \mathbb{A}\mathbb{P}^n \subset W$ with $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}\}$ in a Liouville manifold (W, ω) satisfying the appropriate cohomology assumption (RE) or (CX), there is a Lagrangian correspondence relating K to a Lagrangian sphere L in an auxiliary Liouville manifold (Y, Ω).

4.3.1 Lagrangian $\mathbb{C}\mathbb{P}^n$ Let (W^{4n}, ω) be a Liouville manifold admitting Lagrangian submanifolds $K_i \cong \mathbb{C}\mathbb{P}^n \hookrightarrow W$ for $i = 1, \ldots, m$. Assume there is a class $\alpha \in H^2(W; \mathbb{Z})$ satisfying Assumption (CX). The discussion of Section 3.1 delivers a $\mathbb{C}^*$–bundle $q \colon Y \to W$ (associated to the complex line bundle $\mathcal{L} \to W$ with $c_1(\mathcal{L}) = \alpha$), whose total space is a Liouville manifold (Y, Ω) (proof of Proposition 3.2). Set $V := Y|_{\{r=1\}}$, the unit length bundle (determined by the metric on Y induced by a choice of hermitian metric on $\mathcal{L}$). If $V \hookrightarrow Y$ is the inclusion, then, by construction, $\iota^*\Omega = q^*\omega|_V$, so the symplectic reduction of V by S^1 is given by (W, ω), and V is a fibred coisotropic submanifold of (Y, Ω) with S^1–fibre bundle structure $p = q|_V \colon V \to W$.

For any Lagrangian projective space $K_i \subset W$, the restriction $V|_{K_i} \to K_i$ is a Lagrangian sphere $L_i \cong S^{2n+1} \subset Y$.

Definition 4.11 The (transpose) graph

$$\Gamma := \{(p(y), y), y \in V\} \subset W^- \times Y \tag{27}$$

defines a Lagrangian correspondence [Perutz 2008, Proposition 1.1], which we call the *Hopf correspondence*. By construction, for $K_i \cong \mathbb{CP}^n \subset \{\text{pt}\} \times W$, the correspondence maps K_i to the embedded Lagrangian sphere $L_i := K_i \circ \Gamma \cong S^{2n+1} \subset \{\text{pt}\} \times Y \cong Y$ for $i = 1, \ldots, m$ via geometric composition. ◁

Remark 4.12 This Lagrangian correspondence can equivalently be thought of as a correspondence of the type of Example 4.4, where the coisotropic V is a regular level set of a Hamiltonian S^1 "Hopf" action, and (W, ω) its symplectic quotient (note that the local models (14), (19) and (21) are obtained from this perspective). This explains the choice of name for the correspondence. ◁

4.3.2 Lagrangian $\mathbb{RP}^n$ Let (W^{2n}, ω) be a Liouville manifold admitting Lagrangian embeddings $K_i \cong \mathbb{RP}^n \hookrightarrow W$ for $i = 1, \ldots, m$. Assume there is a cohomology class $\alpha \in H^1(W; \mathbb{Z}/2\mathbb{Z})$ satisfying Assumption (RE).

Then there is a Liouville manifold $(Y, \Omega) = (\widetilde{W}^{2n}, \widetilde{\omega})$ obtained as the symplectic double cover of W and containing Lagrangian spheres $L_1, \ldots, L_m \subset \widetilde{W}$. The double cover $q \colon \widetilde{W} \to W$ defines an S^0–fibration over W, and in this case the "coisotropic submanifold" is the total space itself. As above, we define the Hopf correspondence as $\Gamma := \{(q(y), y) \mid y \in \widetilde{W}\} \subset W^- \times \widetilde{W}$.

4.4 Commuting diagrams of functors

Let (W, ω) and (Y, Ω) be Liouville manifolds and $K_1, \ldots, K_m \subset W$ be real/complex projective Lagrangians satisfying (RE)/(CX). Let $q \colon Y \to W$ be the fibration we constructed in the previous subsections, and $\Gamma \subset W^- \times Y$ be the Hopf correspondence obtained as the graph $\Gamma = \{(p(v), v) \mid v \in V\}$ of the spherically fibred coisotropic $p = q|_V \colon V \to W$. This correspondence is properly fibred over the diagonal $\Delta_W = \{(p(v), p(v)) \mid v \in V\} \subset W^- \times W$, via $(\text{id}, q) \colon W \times Y \to W \times W$, and satisfies the conditions of Axiom 1. Therefore, there is a well-defined functor

$$\Theta_\Gamma \colon \mathcal{F}\text{uk}(W) \to \mathcal{F}\text{uk}(Y), \quad \Theta_\Gamma(K) = K \circ \Gamma =: L. \tag{28}$$

Let $L_1, \ldots, L_m \subset Y$ be the Lagrangian spheres associated to $K_1, \ldots, K_m$ through the correspondence. For each $i = 1, \ldots, m$, let $T_{K_i} \in \text{Auteq}(\mathcal{F}\text{uk}(W))$ and $T_{L_i} \in \text{Auteq}(\mathcal{F}\text{uk}(Y))$ be the (geometric) twist functors induced by the graphs of the respective twists $\tau_{K_i} \in \text{Symp}_{\text{ct}}(W)$ and $\tau_{L_i} \in \text{Symp}_{\text{ct}}(Y)$.

Corollary 4.13 *There is a commuting diagram at the level of compact Fukaya categories*

$$(29) \qquad \begin{array}{ccc} \mathcal{F}\text{uk}(Y) & \xrightarrow{T_{L_i}} & \mathcal{F}\text{uk}(Y) \\ {\scriptstyle \Theta_\Gamma} \uparrow & & \uparrow {\scriptstyle \Theta_\Gamma} \\ \mathcal{F}\text{uk}(W) & \xrightarrow{T_{K_i}} & \mathcal{F}\text{uk}(W) \end{array}$$

In particular, iterative applications of this diagram yield

$$\Theta_\Gamma \circ \prod T_{K_i}^{k_i} = \prod T_{L_i}^{k_i} \circ \Theta_\Gamma. \tag{30}$$

Proof   Consider the functors T_{K_i} and T_{L_i} as correspondences induced by the graphs of the respective twists $\tau_{K_i} \in \mathrm{Symp}_{\mathrm{ct}}(W)$ and $\tau_{L_i} \in \mathrm{Symp}_{\mathrm{ct}}(Y)$. Then we have to check that the compositions of correspondences $\Theta_\Gamma \circ T_{K_i} = T_{L_i} \circ \Theta_\Gamma$, as Lagrangians in $W^- \times Y$, coincide. By construction, this equality amounts to the commutativity of the diagram (17) or (20), respectively.   □

Remark 4.14   For a coefficient field of characteristic zero, the functor associated to the real projective twist has a different shape which produces a different diagram [Mak and Wu 2019, Corollary 1.3].   ◁

Remark 4.15   Given a hypothetical mirror pair (X, M) for a symplectic manifold (M, ω) with $c_1(M) = 0$ and complex manifold (X, J), we can make the following observation.

Huybrechts and Thomas [2006] conjectured that the functors induced by projective twists on the derived Fukaya category $D^\flat(\mathcal{F}uk(M))$ should be mirror to a class of autoequivalences of $D^\flat(X)$, induced by so-called $\mathbb{P}$–objects (see [Huybrechts and Thomas 2006, Definition 1.1]). This is the analogue of the statement proved by Seidel that autoequivalences of $D^\flat(\mathcal{F}uk(M))$ induced by Dehn twists should be mirror to autoequivalences of $D^\flat(X)$ induced by "spherical objects" (see [Seidel and Thomas 2001, Definition 1.1]).

From this perspective, we can view the diagram 4.13 as a conjectural mirror to the following situation.

By [Huybrechts and Thomas 2006, Proposition 1.4], a $\mathbb{P}$–object $\mathcal{P} \in D^\flat(X)$ in the central fibre of an algebraic deformation $j \colon X \hookrightarrow \mathcal{X}$ and satisfying $0 \neq A(\mathcal{P}) \cdot \kappa(\mathcal{X}) \in \mathrm{Ext}^2(\mathcal{P}, \mathcal{P})$ has an associated spherical object given by $j_*(\mathcal{P}) \in D^\flat(\mathcal{X})$. Here, $A(\mathcal{P}) \in \mathrm{Ext}^1(\mathcal{P}, \mathcal{P} \otimes \Omega^1_X)$ is the Atiyah class of $\mathcal{P}$ and $\kappa(\mathcal{X}) \in H^1(X, \mathcal{T}_X)$ the Kodaira–Spencer class of the family $\mathcal{X}$. Furthermore, the autoequivalences associated to each object (also called "twists"), $T_\mathcal{P}$ and $T_{j_*\mathcal{P}}$, are related by a commutative diagram [Huybrechts and Thomas 2006, Proposition 2.7]

(31)
$$
\begin{array}{ccc}
D^\flat(X) & \xrightarrow{\ \ j_*\ \ } & D^\flat(\mathcal{X}) \\
\downarrow{\scriptstyle T_\mathcal{P}} & & \downarrow{\scriptstyle T_{j_*\mathcal{P}}} \\
D^\flat(X) & \xrightarrow{\ \ j_*\ \ } & D^\flat(\mathcal{X})
\end{array}
$$
                                                                                    ◁

4.5   Lagrangian Gysin sequence

Let $\Gamma \subset W^- \times Y$ be the Hopf correspondence. Given real/complex projective Lagrangian submanifolds $K, K' \subset W$ and their corresponding spherical lifts $L, L' \subset Y$ through the functor Θ_Γ, a version of Perutz's Gysin sequence [2008] can be used to establish a relationship between the ranks of the Floer cohomology groups $\mathrm{HF}(K, K')$ and $\mathrm{HF}(L, L')$. We will need this relation in the next section for the proof of Theorem B.

Let $V \to W$ be the S^k–fibred coisotropic defining the correspondence, $k \in \{0, 1\}$, with Euler class $\alpha \in H^{k+1}(W; R)$, $R \in \{\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}\}$ and Lagrangian projective spaces $K, K' \subset W$ satisfying (RE)/(CX), respectively.

Let $L = \Theta_\Gamma(K) = K \circ \Gamma \subset Y$ and $L' = \Theta_\Gamma(K') = K' \circ \Gamma \subset Y$ be the associated Lagrangian spheres given by the correspondence.

Lemma 4.16 *There is an exact triangle of the shape*

(32)

Proof This exact sequence follows from the Gysin triangle proved by Perutz [2008, Theorem 1], which has the more general form

(33) $$\cdots \to \mathrm{HF}^*(K, K') \xrightarrow{e(V)\cup\cdot} \mathrm{HF}^{*+k+1}(K, K') \xrightarrow{\Gamma_*} \mathrm{HF}^{*+k+1}(K, \Gamma^t, \Gamma, K') \to \cdots,$$

where the last group is the quilted Floer cohomology group of the cycle of Lagrangian correspondences $\underline{L} := (K, \Gamma, \Gamma^t, K') \subset (\mathrm{pt}, W, Y, W)$, satisfying $\mathrm{HF}(K, \Gamma, \Gamma^t, K') \cong \mathrm{HF}^*(K \circ \Gamma, K' \circ \Gamma) \cong \mathrm{HF}(L, L')$.

The isomorphism follows from Axiom 1 (in particular Lemma 4.9 applied to a sequence of four Lagrangian correspondences). The compositions $L = K \circ \Gamma$ and $L' = \Gamma^t \circ K' = K' \circ \Gamma$ are embedded, and coincide with the spheres (which are Lagrangian in W) in the sphere bundle V over K and K', respectively.

The first map in the original exact sequence (32) is the quantum cup product with the Euler class $e(V) \in QH^*(W)$. In this case the exactness assumptions on the ambient symplectic manifold W ensure the well-definedness of the operation and, as $QH^*(W) \cong H^*(W)$ is a ring isomorphism, there is no quantum deformation involved and we obviously have $e(V) = \alpha \in H^*(W)$. The second map, Γ_*, is induced by the Lagrangian correspondence, and needs to be understood in the context of quilted Floer theory. We refer the reader to [Perutz 2008, Section 4.1] for a more refined description of the maps (in the setting of Hamiltonian Floer theory). $\square$

Corollary 4.17 *The Gysin sequence produces the rank inequality*

(34) $$\mathrm{hf}(L, L') := \mathrm{rank}\,\mathrm{HF}(L, L') \leqslant 2\,\mathrm{rank}\,\mathrm{HF}(K, K').$$

In Section 5, we will need to compare functors induced by (projective and Dehn) twists to the identity functor. In particular, it will be necessary to distinguish objects of $\mathcal{F}\mathrm{uk}(W)$ with their image under the twist functors. The following lemma gives a helpful criterion:

Lemma 4.18 *Let K' and $\overline{K}'$ be quasi-isomorphic objects in $\mathcal{F}\mathrm{uk}(W)$. Then the maps*

$$f_1 \colon \mathrm{CF}^*(K, K') \xrightarrow{\alpha\cup\cdot} \mathrm{CF}^{*+k+1}(K, K') \quad and \quad f_2 \colon \mathrm{CF}^*(K, \overline{K}') \xrightarrow{\alpha\cup\cdot} \mathrm{CF}^{*+k+1}(K, \overline{K}')$$

have quasi-isomorphic mapping cones.

Proof Consider the long exact sequences associated to the mapping cones of the cup product maps $f_1 \colon \mathrm{CF}^*(K, K') \to \mathrm{CF}^{*+k+1}(K, K')$ and $f_2 \colon \mathrm{CF}^*(K, \overline{K}') \to \mathrm{CF}^{*+k+1}(K, \overline{K}')$.

These sequences fit in a diagram of the shape

$$
\begin{array}{ccccccc}
\mathrm{CF}^*(K, K') & \xrightarrow{f_1 = \alpha \cup \cdot} & \mathrm{CF}^{*+k+1}(K, K') & \longrightarrow & \mathrm{Cone}(f_1) & \longrightarrow & \mathrm{CF}^{*+k+1}(K, K') \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
\mathrm{CF}^*(K, \overline{K}') & \xrightarrow{f_2 = \alpha \cup \cdot} & \mathrm{CF}^{*+k+1}(K, \overline{K}') & \longrightarrow & \mathrm{Cone}(f_2) & \longrightarrow & \mathrm{CF}^{*+k+1}(K, \overline{K}')
\end{array}
$$

Since K' and $\overline{K}'$ are quasi-isomorphic objects in $\mathcal{F}uk(W)$, there is a characteristic element $\eta \in \mathrm{HF}(K', \overline{K}')$ which induces an isomorphism $\mathrm{HF}(K, K') \to \mathrm{HF}(K, \overline{K}')$ (the Floer product with η; see [Seidel 2008a, (8k)]). Therefore, the vertical maps $\mathrm{CF}^*(K, K') \to \mathrm{CF}^*(K, \overline{K}')$ are well defined, and they are quasi-isomorphisms. By the five lemma, the mapping cones $\mathrm{Cone}(f_1)$ and $\mathrm{Cone}(f_2)$ are also quasi-isomorphic. $\square$

5 Free groups generated by projective twists

In this section we apply the Hopf correspondence to prove our first result about products of projective twists.

Consider a transverse plumbing $W := T^*\mathbb{AP}^n \#_{\mathrm{pt}} T^*\mathbb{AP}^n$ of cotangent bundles of projective spaces for $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}\}$. Then the main result of this section (Theorem B) shows that the Lagrangian cores of the plumbing define two projective twists which generate a free subgroup of $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(W))$. In fact, Theorem B is a stronger statement, which holds not only for transverse plumbings but also, more generally, for *clean* plumbings along subprojective spaces (see Definition 5.1).

For the proof, we use the Hopf correspondence to reduce the statement of Theorem B to a statement about Dehn twists, and apply *Keating's free generation result* [2014] (Theorem 5.3) for Dehn twists.

As a corollary, we show that there are infinitely many Lagrangian isotopy classes of embeddings $\mathbb{CP}^n \hookrightarrow W$ which are smoothly isotopic, but pairwise not Lagrangian isotopic.

5.1 Clean Lagrangian plumbing

We first recall a construction from [Abouzaid 2011, Appendix A] of clean *Lagrangian plumbing* of two Riemannian manifolds Q_1 and Q_2 along a submanifold $B \subset Q_i$ for $i = 1, 2$. Fix three closed smooth manifolds B, Q_1 and Q_2, for each $i = 1, 2$ an embedding $B \hookrightarrow Q_i$, and an isomorphism $\varrho \colon \nu_{B/Q_1} \to \nu^*_{B/Q_2}$ from the normal bundle ν_{B/Q_1} to the conormal bundle ν^*_{B/Q_2}.

Pick a Riemannian metric on B, an inner product and a connection on $\nu_{B/Q_1} \cong \nu^*_{B/Q_2}$ (which induces an inner product and connection on $\nu_{B/Q_2} \cong \nu^*_{B/Q_1}$). This data induces a metric on the total spaces ν_{B/Q_i}, and a neighbourhood U_i of $B \subset Q_i$ can be identified with a disc subbundle $D_\varepsilon \nu_{B/Q_i}$ of radius $\varepsilon > 0$. With this identification we write $x \in U_i$ as $x = (a, b)$ for $b \in B$ and $a \in D_\varepsilon(\nu_{B/Q_i})_b$ (the fibre over b).

For each $x = (a, b) \in U_i$, the connection gives a decomposition of the fibres $T_x^* Q_i \cong T_b^* B \oplus (\nu_{B/Q_i}^*)_b$. We get an identification of a neighbourhood of $B \subset T^* Q_i$ as

$$(35) \qquad D_\varepsilon \nu_{B/Q_i} \oplus D_\varepsilon T^* B \oplus D_\varepsilon \nu_{B/Q_i}^*.$$

Let V_i be a neighbourhood of $Q_i \subset T^* Q_i$ which in (35) coincides with $D_\varepsilon T^* B \oplus D_\varepsilon \nu_{B/Q_i}^*$ over $U_i \cong D_\varepsilon \nu_{B/Q_i}$.

Definition 5.1 (1) As a smooth manifold, the clean plumbing of Q_1 and Q_2 along B, denoted by $M := D_\varepsilon(T^* Q_1 \#_B T^* Q_2)$, is defined by gluing V_1 to V_2 along $D_\varepsilon \nu_{B/Q_1} \oplus D_\varepsilon T^* B \oplus D_\varepsilon \nu_{B/Q_1}^* \subset V_1$ identified with $D_\varepsilon \nu_{B/Q_2}^* \oplus D_\varepsilon T^* B \oplus D_\varepsilon \nu_{B/Q_2}$ via $(\varrho, \mathrm{id}_{T^*B}, -\varrho^*)$. Its Liouville completion will be denoted by $T^* Q_1 \#_B T^* Q_2$.

(2) The plumbing construction inherits an exact symplectic structure, since the identification maps of (1) preserve the canonical structures on $D^* Q_i$. Let Z_i be the standard radial Liouville vector field on V_i. We define a Liouville vector field Z on the plumbing by letting $Z = \rho_1 Z_1 + \rho_2 Z_2$ for smooth functions $\rho_i \colon M \to [0, 1]$ supported on V_i such that $\rho_1 + \rho_2 = 1$. This endows M with the structure of an exact symplectic manifold. ◁

In the next sections, we will apply this plumbing construction to cotangent bundles of projective spaces and spheres. We will work with (ungraded) Floer cohomology groups $\mathrm{HF}(Q_1, Q_2, k)$, where k is a coefficient field of characteristic two. Note that, by exactness of the Lagrangians and the manifold W, the Floer differentials in $\mathrm{CF}(Q_i, Q_i)$ vanish, and, as Q_1 and Q_2 intersect cleanly along B, there is an isomorphism $\mathrm{HF}(Q_1, Q_2) \cong H^*(B)$ [Poźniak 1994].

5.2 Proof of Theorem B

We now prove the main theorem of this section.

Theorem B *Let $W := T^* \mathbb{A}\mathbb{P}^n \#_{\mathbb{A}\mathbb{P}^l} T^* \mathbb{A}\mathbb{P}^n$ be a clean plumbing of (real, complex) projective spaces along a linearly embedded subprojective space $\mathbb{A}\mathbb{P}^l \subset W$, $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}\}$. Let $K_1, K_2 \cong \mathbb{A}\mathbb{P}^n \subset W$ denote the Lagrangian core components of the plumbing. Then the projective twists τ_{K_1} and τ_{K_2} generate a free group inside $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(W))$, and the associated functors T_{K_1} and T_{K_2} generate a free subgroup of $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$.*

Remark 5.2 The case $W := T^* \mathbb{C}\mathbb{P}_1^1 \#_{\mathrm{pt}} T^* \mathbb{C}\mathbb{P}_2^1$ can be deduced from the existing literature by considering X as an A_2–configuration and the isotopies $\tau_{\mathbb{C}\mathbb{P}_i^1} \simeq \tau_{S_i^2}^2$ of Remark 2.6. There is a homomorphism [Seidel 1999, Proposition 8.4] $\rho \colon \mathrm{Br}_3 \to \pi_0(\mathrm{Symp}_{\mathrm{ct}}(W))$ sending the generators of the braid group σ_i to $\rho(\sigma_i) = \tau_{S_i^2}$ for $i = 1, 2$. The associated homomorphism $\hat{\rho} \colon \mathrm{Br}_3 \to \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ fits in the diagram

(36)

$$\begin{array}{ccc} \mathrm{Br}_3 & \xrightarrow{\ \rho\ } & \pi_0(\mathrm{Symp}_{\mathrm{ct}}(W)) \\ & \searrow{\scriptstyle \hat{\rho}} & \downarrow \\ & & \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W)) \end{array}$$

and its injectivity [Seidel and Thomas 2001] implies the injectivity of ρ. Then, as $\langle \sigma_1^2, \sigma_2^2 \rangle \cong \mathrm{Free}_2$, it follows that $\langle \tau_{K_1}, \tau_{K_2} \rangle \cong \mathrm{Free}_2$.

Also note that ρ is in fact an isomorphism [Wu 2014], so $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(T^*S^2 \#_{\mathrm{pt}} T^*S^2)) = \mathrm{Br}_3$.   ◁

Theorem B takes inspiration from Keating's free generation result for Dehn twists in Liouville manifolds.

Theorem 5.3 [Keating 2014, Theorem 1.1 and 1.2] *Let (Y, Ω) be a Liouville manifold of dimension greater than 2, and $L, L' \subset Y$ be two Lagrangian spheres satisfying* $\mathrm{rank}\, \mathrm{HF}(L, L') \geq 2$ *and such that L and L' are not quasi-isomorphic in $\mathcal{F}\mathrm{uk}(Y)$. The Dehn twists $\tau_L, \tau_{L'}$ generate a free subgroup of $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(Y))$, and the associated functors $T_L, T_{L'} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$ generate a free subgroup of $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$.*

Keating proves the geometric part of Theorem B by making a categorical detour, first proving that the associated functors $T_L, T_{L'} \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$ induced by the Dehn twists generate a free subgroup of $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$, so that the composition

(37) $$\mathrm{Free}_2 \to \pi_0 \mathrm{Symp}_{\mathrm{ct}}(Y) \to \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$$

is injective.

By identifying a Dehn twist with its associated functor, Keating exploits the algebraic properties of the latter to arrive at the following rank inequalities (which are central in her final proof):

Lemma 5.4 [Keating 2014, Lemma 8.1] *Let $\tilde{L}, L, L' \subset Y$ be Lagrangians such that $\tilde{L}$ is a sphere, $\tilde{L} \not\cong L$ in the Fukaya category, and* $\mathrm{hf}(\tilde{L}, L) := \mathrm{rank}(\mathrm{HF}(\tilde{L}, L')) \geq 2$. *Then, for all $n \neq 0$,*

(38) $$\mathrm{hf}(\tilde{L}, L') > \mathrm{hf}(L, L') \implies \mathrm{hf}(\tilde{L}, \tau_{\tilde{L}}^n(L')) < \mathrm{hf}(L, \tau_{\tilde{L}}^n(L')).$$

Lemma 5.5 [Keating 2014, Claim 8.2] *Let $L, L' \subset Y$ be two Lagrangian spheres in an exact symplectic manifold as in Theorem 5.3 satisfying* $\mathrm{hf}(L, L') = 2$. *Then, for all $m \neq 0$,*

(39) $$\mathrm{hf}(L', L) = \mathrm{hf}(L', \tau_{L'}^m L) < \mathrm{hf}(L, \tau_{L'}^m L).$$

We will apply these inequalities to Lagrangian spheres obtained from the Hopf correspondence, to produce similar results for projective twists and prove Theorem B.

5.2.1 Strategy The plumbing (W, ω) and its real/complex projective Lagrangian cores $K_1, K_2 \subset W$ satisfy the cohomological conditions (RE)/ (CX).

In the case in which W is a transverse plumbing (which retracts to the wedge sum of the two spheres), there is a ring isomorphism (with $\mathbb{A} \in \{\mathbb{R}, \mathbb{C}\}$ and $R \in \{\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}\}$)

$$\tilde{H}^*(W; R) \cong \tilde{H}^*(K_1; R) \oplus \tilde{H}^*(K_2; R) \cong \tilde{H}^*(\mathbb{A}\mathbb{P}^n; R) \oplus \tilde{H}^*(\mathbb{A}\mathbb{P}^n; R),$$

so it is immediate to see the existence of a class $\alpha = (\alpha_1, \alpha_2) \in H^k(W; R)$ restricting in K_1 and K_2 to the generator of $H^*(\mathbb{AP}^n; R)$ for $k \in \{1, 2\}$. For a clean plumbing along a linearly embedded subprojective space $\mathbb{AP}^l$, this restriction property still holds because the "difference" map of the Mayer–Vietoris sequence is always zero.

By Propositions 3.2 and 3.6, the cohomological condition ensures the existence of a Liouville manifold $(Y, \Omega) \to (W, \omega)$ and a Hopf correspondence $\Gamma \subset W^- \times Y$ that gives rise to associated Lagrangian spheres $S^m \cong L_i = K_i \circ \Gamma \subset Y$ for $i = 1, 2$ and commuting diagrams of twist functors (29). Then, given a product (a word in τ_{K_1} and τ_{K_2}) $\varphi \in \mathrm{Symp}_{\mathrm{ct}}(W)$ of projective twists, the Hopf correspondence yields a corresponding product of Dehn twists (a word in τ_{L_1} and τ_{L_2}) $\phi \in \mathrm{Symp}_{\mathrm{ct}}(Y)$.

In the real projective case, the geometric statement of Theorem B can be obtained by an isotopy-lifting argument using the geometric diagrams of Section 3 (the strategy adopted in Section 6.3). Assuming the projective twists do satisfy a relation, this procedure lifts the isotopy to $\mathrm{Symp}_{\mathrm{ct}}(Y)$, producing a relation between Dehn twists, which cannot hold, by Keating's theorem. However, this geometric argument does not give a statement at the level of Fukaya categories, for which the use of the Hopf correspondence at the level of functors in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ is necessary (Sections 4.4 and 4.5).

The spheres $L_1, L_2 \cong S^m$ intersect cleanly along a subsphere S^r for a tuple $(\mathbb{A}, m, r)$ that is one of $(\mathbb{R}, n, l)$ or $(\mathbb{C}, 2n + 1, 2l + 1)\}$ $(n, l \in \mathbb{N}^*)$, and, as noted before, $\mathrm{HF}(L_1, L_2; k) \cong H^*(S^r; k)$. Since $L_i \subset Y$ are exact spheres, $\mathrm{HF}(L_i, L_i; k) \cong H^*(S^m; k)$ [Floer 1988] and therefore

$$(40) \qquad \mathrm{rank}\, \mathrm{HF}(L_1, L_2) = \mathrm{rank}\, \mathrm{HF}(L_i, L_i) = 2 \quad \text{for } i = 1, 2.$$

In the following sections we will study the ranks of the Floer cohomology groups $\mathrm{HF}(\cdot, \varphi(\cdot))$ and show that there is always a Lagrangian $\hat{K} \subset W$ such that

$$(41) \qquad \mathrm{HF}(\hat{K}, \hat{K}) \not\cong \mathrm{HF}(\hat{K}, \varphi(\hat{K})).$$

As a result, $\hat{K}$ and $\varphi(\hat{K})$ are not quasi-isomorphic objects in $\mathcal{F}\mathrm{uk}(W)$, and therefore the functor induced by φ cannot be isomorphic to the identity in $\mathrm{Auteq}\,\mathcal{F}\mathrm{uk}(W)$. This will also rule out the possibility of φ being isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(W)$.

We prove (41) by applying the rank inequalities (38) and (39) to Lagrangian spheres in (Y, Ω) obtained via the correspondence Γ, in combination with the symplectic Gysin sequence associated to Γ (Corollary 4.17).

In the following section, we rederive a part of Keating's proof of Theorem 5.3 using the rank inequalities (38) and (39), which hold for the word of Dehn twists $\phi \in \langle \tau_{L_1}, \tau_{L_2} \rangle$ associated to φ via the Hopf correspondence. This will clarify the methods used in proving the analogous statement for projective twists (namely Theorem B) in Section 5.2.3.

5.2.2 Associated word of Dehn twists Let $\varphi \in \langle \tau_{L_1}, \tau_{L_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ be a word of projective twists as in the statement of Theorem B. Consider the Hopf correspondence $\Gamma \subset W^- \times Y$ and the associated

word of Dehn twists $\phi \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ as above. In this section we replicate the last steps in Keating's proof of the injectivity of the homomorphism

$$\mathrm{Free}_2 \to \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W)).$$

We first make the following observation about a word of twists and its conjugates:

Lemma 5.6 Let $\phi \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ be a symplectomorphism which has the shape of a global conjugate, ie $\phi = \psi^{-1}\phi'\psi$ for $\psi, \phi' \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ not isotopic to the identity. Then there is a closed Lagrangian $\widetilde{L} \subset Y$ such that $\mathrm{HF}(\widetilde{L}, \phi(\widetilde{L})) \not\cong \mathrm{HF}(\widetilde{L}, \widetilde{L})$ if and only if ϕ' satisfies $\mathrm{HF}(\widehat{L}, \phi'(\widehat{L})) \not\cong \mathrm{HF}(\widehat{L}, \widehat{L})$ for some closed Lagrangian $\widehat{L} \subset Y$.

Proof Assume there is a Lagrangian $\widehat{L} \subset Y$ such that $\mathrm{HF}(\widehat{L}, \phi'(\widehat{L})) \not\cong \mathrm{HF}(\widehat{L}, \widehat{L})$. Then, by invariance of Floer cohomology under symplectomorphisms, $\mathrm{HF}(\psi^{-1}\widehat{L}, \psi^{-1}\phi'\psi(\psi^{-1}\widehat{L})) \cong \mathrm{HF}(\psi^{-1}\widehat{L}, \psi^{-1}(\phi'\widehat{L}) \cong \mathrm{HF}(\widehat{L}, \phi'\widehat{L}) \not\cong \mathrm{HF}(\widehat{L}, \widehat{L})$, so, for $\widetilde{L} := \psi^{-1}(\widehat{L})$, we have $\mathrm{HF}(\widetilde{L}, \phi(\widetilde{L})) \not\cong \mathrm{HF}(\widetilde{L}, \widetilde{L})$. The other direction is similar. $\square$

We will apply the above lemma to a word of Dehn twists $\phi \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ which in its reduced has the shape of a global conjugate, ie a word $\phi = \psi^{-1}\phi'\psi$ for two reduced words $\psi, \phi' \in \langle \tau_{L_1}, \tau_{L_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(Y)$ not isotopic to the identity. Then the lemma shows that it is always possible to switch between ϕ and its conjugate, as the correct choice of Lagrangian keeps track of the Floer cohomological action of the original word.

Without loss of generality, we can therefore use this conjugation argument to restrict the focus on reduced words $\phi \in \langle \tau_{L_1}, \tau_{L_2} \rangle$ which are either (for $i, j \in \{1, 2\}$)

(1) a power of a single Dehn twist, ie $\phi = \tau_{L_i}^s$, $s \in \mathbb{Z}^*$ (Lemma 5.7); or

(2) a word starting with a power of τ_{L_i} and ending in a power of τ_{L_j} with $i \neq j$ (Lemma 5.9).

Lemma 5.7 Let $\phi = \tau_{L_i}^s \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ be a reduced word of Dehn twists which is a power of a single Dehn twist, with $i \in \{1, 2\}$ and $s \in \mathbb{Z}^*$. The associated functor is not isomorphic to the identity in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$, so, in particular, ϕ cannot be isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(Y)$.

Proof We show that there exists a closed Lagrangian $\widehat{L} \subset Y$ such that

$$\mathrm{HF}(\widehat{L}, \phi(\widehat{L})) \not\cong \mathrm{HF}(\widehat{L}, \widehat{L}).$$

For $\phi = \tau_{L_i}^s$, a possible candidate is given by $\widehat{L} = L_j$ for $i, j \in \{1, 2\}$ and $i \neq j$.

Namely, the rank inequality stated by Lemma 5.5, gives

$$2 = \mathrm{hf}(L_j, L_j) = \mathrm{hf}(L_i, L_j) = \mathrm{hf}(L_i, \tau_{L_i}^s L_j) < \mathrm{hf}(L_j, \tau_{L_i}^s L_j). \qquad \square$$

Remark 5.8 The geometric result of the above lemma can also be proven independently from Keating's results, as a corollary to Theorem A (see Section 6.1).   ◁

Lemma 5.9 Let $\phi \in \langle \tau_{L_1}, \tau_{L_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(Y)$ *be a reduced word of Dehn twists around the Lagrangian (spherical) cores which is a product where the first and last factors are powers of distinct Dehn twists. Then the functor associated to ϕ is not isomorphic to the identity in* $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$, *so, in particular, ϕ cannot be isotopic to the identity in* $\mathrm{Symp}_{\mathrm{ct}}(Y)$.

Proof We show that there is a closed Lagrangian $\hat{L} \subset Y$ such that $\mathrm{HF}(\hat{L}, \phi\hat{L}) \not\cong \mathrm{HF}(\hat{L}, \hat{L})$.

We can assume without loss of generality that the first factor of ϕ is a power of τ_{L_2} and the last is a power of τ_{L_1} (otherwise consider ϕ^{-1}), so that we have a word of shape

$$(42) \qquad \phi = \tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1}, \quad a_i, b_i \in \mathbb{Z}^* \text{ for } 1 \leqslant i \leqslant k.$$

In the case we are considering, we have $\mathrm{hf}(L_1, L_2) = 2$. Apply Lemma 5.5 to get

$$2 = \mathrm{hf}(L_1, L_1) = \mathrm{hf}(L_2, L_1) = \mathrm{hf}(L_2, \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) < \mathrm{hf}(L_1, \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1).$$

Now apply Lemma 5.4 (with $n = a_2$, $\tilde{L} = L_1$, $L = L_2$ and $L' = \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1$) and get

$$\mathrm{hf}(L_1, \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) = \mathrm{hf}(L_1, \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) < \mathrm{hf}(L_2, \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1).$$

Apply Lemma 5.4 again (with $n = b_2$, $\tilde{L} = L_2$, $L = L_1$ and $L' = \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1$)

$$\mathrm{hf}(L_2, \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) = \mathrm{hf}(L_2, \tau_{L_2}^{b_2} \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) < \mathrm{hf}(L_1, \tau_{L_2}^{b_2} \tau_{L_1}^{a_2} \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1).$$

Continue to apply Lemma 5.4 iteratively until the final step

$$\mathrm{hf}(L_2, \tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) < \mathrm{hf}(L_1, \tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1).$$

Then

$$\mathrm{hf}(L_1, \tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1} L_1) > 2 + 2k - 1 = 2k + 1.$$

So, setting $\hat{L} = L_1$, we have $\mathrm{HF}(\hat{L}, \phi(\hat{L})) \not\cong \mathrm{HF}(\hat{L}, \hat{L})$. $\qquad \square$

Corollary 5.10 Let $\phi \in \langle \tau_{L_1}, \tau_{L_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(Y)$ *be a word of Dehn twists that is a product of the shape* (42). *Then there is a Lagrangian $\hat{L} \subset Y$ such that*

$$\lim_{s \to \infty} \mathrm{rank}\, \mathrm{HF}^*(\hat{L}, \phi^s(\hat{L})) = \infty.$$

Proof Let ϕ be of the shape (42). Then

$$\phi^s = (\tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1})(\cdots)(\cdots)(\tau_{L_2}^{b_k} \tau_{L_1}^{a_k} \cdots \tau_{L_2}^{b_1} \tau_{L_1}^{a_1})$$

has "factor length" $k \cdot s$ (in the sense of (42)). By the proof of Lemma 5.9, the rank of $\mathrm{HF}(L_1, \phi(L_1))$ depends on the number $k \in \mathbb{N}$ appearing in the factor decomposition of ϕ. Therefore,

$$\mathrm{hf}(L_1, \phi^s(L_1)) > 2ks + 1,$$

so we can set $\hat{L} := L_1$. $\qquad \square$

5.2.3 Proof We now go back to the original word $\varphi \in \langle \tau_{K_1}, \tau_{K_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ of projective twists in the statement of Theorem B, and we show that it cannot induce the identity functor in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$. Lemma 5.6 holds for any symplectomorphism, so, by the same conjugation argument explained before, we can focus the attention on words that are either (for $i, j \in \{1, 2\}$)

(1) a power of a single twist $\varphi = \tau_{K_i}^s$, $s \in \mathbb{Z}^*$, or

(2) a mixed product of the shape $\varphi := \tau_{K_i}^{b_k} \tau_{K_j}^{a_k} \cdots \tau_{K_i}^{b_1} \tau_{K_j}^{a_1} \in \mathrm{Symp}_{\mathrm{ct}}(W)$ with $i \neq j$ and $a_m, b_m \in \mathbb{Z}^*$ for $1 \leqslant m \leqslant k$.

Proposition 5.11 *Let $\varphi = \tau_{K_i}^s \in \langle \tau_{K_1}, \tau_{K_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ be a reduced word of projective twists which is a power of a single twist with $i \in \{1, 2\}$ and $s \in \mathbb{Z}^*$. Then the functor induced by φ is not isomorphic to the identity in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$, and in particular φ cannot be isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(W)$.*

Proof Let $\varphi = \tau_{K_i}^s \in \mathrm{Symp}_{\mathrm{ct}}(W)$ with $s \in \mathbb{Z}^*$. Assume by contradiction that the functor induced by φ (still denoted by φ) is isomorphic to the identity, so that any Lagrangian $\widehat{K} \subset W$ is quasi-isomorphic, as an object of $\mathcal{F}\mathrm{uk}(W)$, to $\varphi(\widehat{K})$.

By Lemma 4.18, there is a quasi-isomorphism of the mapping cones of the cup product maps

$$f_1 \colon \mathrm{CF}^*(\widehat{K}, \widehat{K}) \to \mathrm{CF}^{*+k+1}(\widehat{K}, \widehat{K}) \quad \text{and} \quad f_2 \colon \mathrm{CF}^*(\widehat{K}, \varphi(\widehat{K})) \to \mathrm{CF}^{*+k+1}(\widehat{K}, \varphi(\widehat{K}))$$

(we are considering ungraded Floer cohomology groups, so technically the degrees are irrelevant here). Therefore, by the exact triangle of Lemma 4.16, if $\widehat{L} \subset Y$ is the Lagrangian lift of $\widehat{K}$ through the correspondence Γ and $\phi \in \mathrm{Symp}_{\mathrm{ct}}(Y)$ the symplectomorphism associated to φ, then $\mathrm{HF}(\widehat{L}, \widehat{L}) \cong \mathrm{HF}(\widehat{L}, \phi(\widehat{L}))$.

So, if we set $\widehat{K} := K_j$ with $j \neq i$, by assumption we have $\mathrm{HF}(K_j, \varphi(K_j)) \cong \mathrm{HF}(K_j, K_j)$ and the above argument yields $\mathrm{HF}(L_j, \phi(L_j)) = \mathrm{HF}(L_j, \tau_i^s(L_j)) \cong \mathrm{HF}(L_j, L_j)$, which is clearly in contradiction to (the proof of) Lemma 5.7 (according to which these two groups have distinct ranks). Hence, φ cannot be isomorphic to the identity functor in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$. $\qquad\square$

Proposition 5.12 *Let $\varphi \in \langle \tau_{K_1}, \tau_{K_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ be a reduced word of projective twists around the Lagrangian cores which is a product where the first and last factors are powers of distinct projective twists. Then the functor induced by φ is not isomorphic to the identity in $\mathrm{Auteq}\,\mathcal{F}\mathrm{uk}(W)$; so, in particular, φ is not isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(W)$.*

Proof By the analogous discussion in the proof of Lemma 5.9, it is enough to prove the statement for a word whose reduced form is of the shape

$$(43) \qquad \varphi := \tau_{K_2}^{b_k} \tau_{K_1}^{a_k} \cdots \tau_{K_2}^{b_1} \tau_{K_1}^{a_1} \in \mathrm{Symp}_{\mathrm{ct}}(W), \quad a_m, b_m \in \mathbb{Z}^*, \ 1 \leqslant m \leqslant k.$$

Denote the product of twist functors induced by (43) also by $\varphi \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$. By iteratively using commutativity of the functors in diagram (29), one can define the corresponding composition of (Dehn) twist functors $\phi \in \mathrm{Auteq}(\mathcal{F}\mathrm{uk}(Y))$, which, by Theorem 5.3, cannot be isomorphic to the identity functor.

Moreover, Corollary 5.10 shows that, not only is $\mathrm{HF}(L_1, \phi(L_1))$ nonisomorphic to $\mathrm{HF}(L_1, L_1)$, but also $\lim_{s \to \infty} \mathrm{hf}(L_1, \phi^s(L_1)) \to \infty$.

The Lagrangian $L_1 = \Gamma \circ K_1 \subset Y$ is the Lagrangian associated to K_1 via the Hopf correspondence, and the symplectic Gysin exact sequence (Corollary 4.17) applied to the Hopf correspondence gives the inequality

$$\mathrm{hf}(L_1, \phi(L_1)) \leqslant 2\mathrm{hf}(K_1, \varphi(K_1)), \tag{44}$$

which implies that rank $\mathrm{hf}(K_1, \varphi^s(K_1))$ also grows at least linearly with s. $\qquad\square$

Corollary 5.13 *Let $\varphi \in \langle \tau_{K_1}, \tau_{K_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ be a word of projective twists of the shape (43). Then there is a Lagrangian $\hat{K} \subset W$ such that*

$$\lim_{s \to \infty} \mathrm{rank}\, \mathrm{HF}^*(\hat{K}, \varphi^s(\hat{K})) = \infty. \qquad\square$$

Finally, we can summarise the proof of Theorem B.

Proof of Theorem B Let $\varphi \in \langle \tau_{K_1}, \tau_{K_2} \rangle \subset \mathrm{Symp}_{\mathrm{ct}}(W)$ be a word in the projective twists along the Lagrangian cores of W.

(1) If the word has the shape $\varphi = \tau_{K_i}^s \in \mathrm{Symp}_{\mathrm{ct}}(W)$ with $i \in \{1, 2\}$ and $s \in \mathbb{Z}^*$, then its induced functor is not isomorphic to the identity in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ by Proposition 5.11.

(2) If the word has the shape $\varphi := \tau_{K_i}^{b_k} \tau_{K_j}^{a_k} \cdots \tau_{K_i}^{b_1} \tau_{K_j}^{a_1} \in \mathrm{Symp}_{\mathrm{ct}}(W)$ with $i, j \in \{1, 2\}$ and $i \neq j$, and $a_m, b_m \in \mathbb{Z}^*$ for $1 \leqslant m \leqslant k$, then its induced functor is not isomorphic to the identity in $\mathrm{Auteq}(\mathcal{F}\mathrm{uk}(W))$ by Proposition 5.12.

(3) If φ has any other form, then it must be a conjugate of a word of shape (1) or (2) and hence the induced functor is not isomorphic to the identity by Lemma 5.6. $\qquad\square$

5.3 Knotted Lagrangian projective spaces

The phenomenon that a single (smooth) isotopy class of submanifolds contains infinitely many Lagrangian isotopy classes is called Lagrangian "knottedness" [Seidel 1999; Evans 2010; Hind 2012; Li and Wu 2012; Wu 2014]. Often, the quest for knottedness is intimately related to the study of isotopy classes of Dehn twists.

In the plumbing of spheres $L_i \cong S^2$ and $Y := T^*L_1 \#_{\mathrm{pt}} T^*L_2$, we know that, for any $r \in \mathbb{Z}$, $\tau_{L_2}^{2r}(L_1)$ is smoothly isotopic to the identity, but not symplectically; as first shown by Seidel [1999, Theorem 1.1], none of the powers $\tau_{L_2}^{2r}(L_1)$ are Hamiltonian isotopic. Our results yield the analogue for plumbing of complex projective spaces (of any dimension):

Corollary 5.14 *Let $W := T^*\mathbb{CP}^n \#_{\mathbb{CP}^l} T^*\mathbb{CP}^n$ be a clean plumbing along a projective subspace $\mathbb{CP}^l \subset \mathbb{CP}^n$. Each Lagrangian core $K_i \cong \mathbb{CP}^n$ of W defines a smooth isotopy class which contains infinitely many symplectic isotopy classes of Lagrangian projective spaces.*

Proof Let $K_1, K_2 \cong \mathbb{C}\mathbb{P}^n \subset W$ be the two Lagrangian cores of the plumbing. For $i, j \in \{1, 2\}$ with $i \neq j$, define the element $\varphi := \tau_{K_i} \tau_{K_j}$. Then, by Proposition 5.12,

$$\lim_{s \to \infty} \operatorname{rank} \operatorname{HF}(K_j, \varphi^s(K_j)) = \infty, \tag{45}$$

which, in particular, means that $\varphi^{s_a}(K_j)$ is not Lagrangian isotopic to $\varphi^{,s_b}(K_j)$ for any $s_a \neq s_b$, despite being smoothly isotopic (by Theorem 2.5). $\qquad \square$

Remark 5.15 The low-dimensional case $n = 1$ corresponds to a transverse plumbing of spheres $W :=$ $T^*L_1 \#_{\mathrm{pt}} T^*L_2$ and $L_i \cong S^2$. In that case, the symplectic mapping class group $\pi_0(\operatorname{Symp}_{\mathrm{ct}}(W))$ is generated by the Dehn twists τ_{L_1} and τ_{L_2} (see Remark 5.2). Moreover, Hind [2012] proved that, for any Lagrangian sphere $L \subset W$, there is a word $\tau \in \langle \tau_{L_1}, \tau_{L_2} \rangle$ such that $\tau(L)$ is isotopic to one of the cores L_1 or L_2. $\qquad \triangleleft$

6 Positive products of twists in Liouville manifolds

The present section covers our results about products of positive powers of Dehn and projective twists.

In the first part, Section 6.1, we analyse products of (positive powers of) Dehn twists. We reprove a theorem by Barth, Geiges and Zehmisch (Theorem A) asserting that, in a Liouville manifold (M, ω), no product $\phi \in \operatorname{Symp}_{\mathrm{ct}}(M)$ of positive powers of Dehn twists can be symplectically isotopic to the identity. We provide an alternative proof that was suggested by Paul Seidel. Based on symplectic Picard–Lefschetz theory, the argument for the proof relies on a count of pseudoholomorphic sections of a Lefschetz fibration constructed from the data given by ϕ and the Lagrangian spheres associated to the Dehn twists.

Using similar tools, we then prove Theorem C (Section 6.2), which can be interpreted as a relative version of Theorem A. This states that a Liouville manifold (M, ω) containing Lagrangian spheres and a conical Lagrangian disc T (Definition 4.7) intersecting one of the spheres transversely at a point cannot admit a positive product of Dehn twists preserving T up to compactly supported symplectic isotopy.

In Section 6.3, we explore the analogous questions for projective twists, by means of the tools developed in Section 4. After setting the necessary conditions to ensure the existence of the Hopf correspondence, we use Theorem A to prove a comparable result for real projective twists.

6.1 Alternative proof of Theorem A

In this section, we reprove the following theorem:

Theorem A [Barth et al. 2019, Theorem 1.4] *Let (M, ω) be a Liouville manifold, and let $L_1, \ldots, L_m \subset M$ be Lagrangian spheres. Let $\phi = \prod_{i=1}^k \tau_{L_{j_i}} \in \operatorname{Symp}_{\mathrm{ct}}(M)$, $j_i \in \{1, \ldots, m\}$ be a positive word of Dehn twists. Then ϕ is not compactly supported isotopic to the identity in $\operatorname{Symp}_{\mathrm{ct}}(M)$.*

Example 6.1 The exactness condition of Theorem A is necessary, as the following examples show:

(a) Consider the 2–torus $M := T^2$, and let $a, b \subset M$ represent the longitude and meridian of M. Then the associated Dehn twists satisfy $(\tau_a \tau_b)^6 = \text{Id}$ in $\pi_0(\text{Symp}_{\text{ct}}(M))$. This is a classical result; see for example [Farb and Margalit 2012] (see [Auroux 2003, Section 3.1] for the same example in a symplectic setting).

(b) Let $(M := S^2 \times S^2, \omega_{S^2} \oplus \omega_{S^2})$, and consider the antidiagonal $\overline{\Delta} := \{(x, y) \in S^2 \times S^2 \mid x + y = 0\} \subset M$. Then the Dehn twist $\tau_{\overline{\Delta}}$ is symplectically isotopic to an involution $(x, y) \mapsto (y, x)$, which implies $\tau_{\overline{\Delta}}^2 = \text{Id}$ in $\pi_0(\text{Symp}_{\text{ct}}(M))$ (see [Seidel 2008b, Example 2.9]). ◁

Remark 6.2 (1) The two-dimensional case of Theorem A (for a product of Dehn twists in a Riemann surface) is a consequence of [Smith 2001, Theorem 1.3].

(2) The outcome of Theorem A is strictly geometric, and may not hold for the compact Fukaya category: we are not able to obtain information about the functors associated to the Dehn twists. Consider a punctured torus $M := T^2 \setminus \{*\}$ (the same applies to a punctured genus g surface), and the two (Lagrangian) circles a and b, representatives of the homological generators. In the closed case, the composition $(\tau_a \tau_b)^6$ is isotopic to the identity by the example above. In the punctured torus, there is an isotopy $(\tau_a \tau_b)^6 \simeq \tau_d$, where τ_d is the Dehn twist along the boundary curve d encircling the puncture (this is a consequence of the *chain relation*; see [Farb and Margalit 2012, Proposition 4.12]). But, since the support of τ_d is disjoint from any exact compact circle in M, the product $(\tau_a \tau_b)^6$ still acts as the identity on objects of the *compact* Fukaya category $\mathcal{F}\text{uk}(M)$. ◁

The original proof of [Barth et al. 2019] relies on the theory of open book decompositions, whereas the proof below uses Picard–Lefschetz theory. To simplify notation we prove the version of the theorem where $(M, \omega = d\lambda_M)$ is a Liouville domain.

Let $\phi = \prod_{i=1}^{k} \tau_{L_{j_i}} \in \text{Symp}_{\text{ct}}(M)$ with $j_i \in \{1, \ldots, m\}$ be the word in positive powers of Dehn twists in the given collection, and assume by contradiction that this product is compactly supported Hamiltonianly isotopic to the identity (recall ϕ is an exact symplectomorphism).

Let $\pi : (E, \Omega_E, \lambda_E) \to (\mathbb{C}, \lambda_{\mathbb{C}})$ be the exact Lefschetz fibration determined by the data

$$\{(M, \lambda_M), (L_{j_1}, \ldots, L_{j_k})\}$$

as in Section 2.2. Let $z_* \in \mathbb{C}$ be the basepoint, so that $\pi^{-1}(z_*) \cong M$, and $\phi \in \text{Symp}_{\text{ct}}(M)$ be the total monodromy of π.

Let $j_{\mathbb{C}}$ be the standard complex structure on $\mathbb{C}$.

By assumption, the monodromy of π is isotopic to the identity via a compactly supported Hamiltonian isotopy $(\phi_t)_{t \in [0,1]}$ with $\phi_0 = \phi$ and $\phi_1 = \text{Id}$. Then π can be extended to a fibration $\hat{\pi} : \widehat{E} \to \mathbb{CP}^1$ as follows. Let $D_R \subset \mathbb{C}$ be a large circle of radius $R > 0$ passing through z_* and containing all the critical

values. Define a fibration $E' \to D_{R+1}$ by extending $E|_{D_R}$ to a larger disc D_{R+1} such that, for $t \in [0, 1]$, the monodromy around D_{R+t} is $\phi_t \in \mathrm{Symp}_{\mathrm{ct}}(M)$. Then $\widehat{E}$ is obtained after gluing E' to a trivial fibration with fibre (M, ω) over a disc neighbourhood of the point at "infinity", $\hat{z} \in \mathbb{C} \cup \{\infty\} \simeq \mathbb{CP}^1$.

Moreover, as the symplectic connection around the fibre $\hat{\pi}^{-1}(\hat{z})$ is trivial, $\hat{\pi} \colon \widehat{E} \to \mathbb{CP}^1$ has the following properties:

(1) There is a closed (possibly degenerate) two-form $\widehat{\Omega}_{\widehat{E}}$ on $\widehat{E}$ satisfying $\widehat{\Omega}_{\widehat{E}}|_{\hat{\pi}^{-1}(z)} = \Omega_E|_{\pi^{-1}(z)}$ for all $z \in \mathbb{CP}^1 \setminus \hat{z}$,

(2) A neighbourhood of the horizontal boundary $V \supset \partial^h \widehat{E}$ can be trivialised as $V \cong \mathbb{CP}^1 \times M^{\mathrm{out}}$, where $M^{\mathrm{out}} \subset M$ is an open neighbourhood of the boundary of the smooth fibre.

Definition 6.3 The set of almost complex structures compatible with $\hat{\pi}$, denoted by $\mathcal{J}(\widehat{E}, \hat{\pi}, j)$, is defined as follows. An element $\widehat{J} \in \mathcal{J}(\widehat{E}, \hat{\pi}, j)$ satisfies:

- $D\hat{\pi} \circ \widehat{J} = j \circ D\hat{\pi}$, where j is the standard complex structure on $\mathbb{CP}^1$.

- There is an integrable almost complex structure J_0 such that $\widehat{J} = J_0$ in a neighbourhood of $\mathrm{Crit}(\hat{\pi})$.

- For all $z \in \mathbb{CP}^1$, the restriction $J^{vv} := \widehat{J}|_{\hat{\pi}^{-1}(z)}$ is an almost complex structure of contact type compatible with the Liouville form λ_M, and its restriction to V is isomorphic to a product $j \times J^{vv}$.

- $\widehat{\Omega}_{\widehat{E}}(\cdot, \widehat{J}\cdot)$ is symmetric and positive definite. $\triangleleft$

The form $\widehat{\Omega}_{\widehat{E}}$ can be modified to a symplectic form $\widehat{\Omega} := \widehat{\Omega}_E + \hat{\pi}^*(\beta)$ that tames $\widehat{J}$ for $\beta \in \Omega^2(\mathbb{CP}^1)$ (similar to [Seidel 2003, Lemma 2.1; McDuff and Salamon 2017, Theorem 6.1.4]).

From now onwards, we fix a generic element $\widehat{J} \in \mathcal{J}(\widehat{E}, \hat{\pi}, j)$, so that, by the same arguments as in [Seidel 2003, Lemma 2.4], all the moduli spaces we encounter satisfy the necessary regularity conditions.

Consider the moduli space of closed $(\widehat{J}, j)$–holomorphic sections

$$(46) \qquad \mathcal{M}_{\widehat{J}} = \{u \colon \mathbb{CP}^1 \to \widehat{E} \mid \hat{\pi} \circ u = \mathrm{id}_{\mathbb{CP}^1}, \widehat{J} \circ Du = Du \circ j\}.$$

The moduli space has a nonempty boundary, but, as we explain below, this does not cause compactness issues, as the only sections reaching the boundary must be trivial.

Lemma 6.4 *The space $\mathcal{M}_{\widehat{J}}$ is not empty. Moreover, there is a compact subset $K \subset \widehat{E} \setminus \partial^h \widehat{E}$ such that, for all $u \in \mathcal{M}_{\widehat{J}}$, either $\mathrm{Im}(u) \subset K$ or u is a trivial (constant) section.*

Proof Let $q \in V$ a point in a neighbourhood $\partial^h \widehat{E} \subset V$ of the horizontal boundary as in (2) above. Via the trivialisation of this neighbourhood, one obtains a trivial section $s \colon \mathbb{CP}^1 \to \widehat{E}$ with $s(z) = q$ for all $z \in \mathbb{CP}^1$, which is a regular $(\widehat{J}, j)$–holomorphic section, so $\mathcal{M}_{\widehat{J}}$ is not empty. The rest of the proof follows from a maximum principle as in [Seidel 2003, Lemma 2.2]. $\square$

We can adapt the argument of [Seidel 2003, Lemma 2.3] to the case of closed curves to show that, for our choice of almost complex structure $\hat{J}$, the moduli space $\mathcal{M}_{\hat{J}}$ is a compact smooth manifold with boundary. The only issue that could possibly occur is a loss of compactness for the component containing sections outside the compact part K, which, by Lemma 6.4, can only be trivial sections. These elements have bounded energy, as they are all in the same homology class. By the Gromov compactness theorem, the only noncompact phenomenon that can occur in this case is sphere bubbling. The next lemma shows how to discard bubbles.

Lemma 6.5 *Let u_∞ be the limit of a (sub)sequence of pseudoholomorphic sections $(u_n)_{n \in \mathbb{N}}$ of the Lefschetz fibration $\hat{\pi}$. A component of u_∞ is either an element in the class $[u_i]$ (for $i \in \mathbb{N}$) or is contained in a single fibre. In the latter case, the component is a bubble.*

Proof Let $v_1, v_2, \ldots, v_k$ be the components of u_∞. The limiting curve u_∞ is assumed to be $(\hat{J}, j)$–holomorphic and nonconstant, so it has to have degree one, as $\sum_{j=1}^{k}[\pi \circ v_i] = [\pi \circ u_\infty] = [\mathbb{CP}^1]$. It follows that the degrees of its components sum up to one. All degrees are nonnegative, so there is only one component with degree one. If in addition there were a bubble, it would be represented in a degree zero component and therefore would have to be entirely contained in a fibre (note that, by positivity of intersections, the bubble cannot intersect other fibres).

Since the fibres are exact, there can be no bubbling of the type of Lemma 6.5, so the moduli space $\mathcal{M}_{\hat{J}}$ is compact. $\square$

Lemma 6.6 *Through each point of the smooth fibre M there is at least one holomorphic section $s \in \mathcal{M}_{\hat{J}}$.*

Proof As in the proof of Lemma 6.4, we consider a neighbourhood of the horizontal boundary $V \supset \partial^h \hat{E}$ and $q \in V$ such that $\hat{\pi}(q) =: z_{\text{gen}} \in \mathbb{CP}^1 \setminus \text{Crit} \, v(\hat{\pi})$ and the trivial section through q is $s \colon \mathbb{CP}^1 \to \hat{E}$. Consider

$$\mathcal{M}(\hat{J}, q) := \{u \in \mathcal{M}_{\hat{J}} \mid q \in \text{Im}(u)\} \subset \mathcal{M}_{\hat{J}}.$$

It is a smooth compact manifold (by the same arguments as for $\mathcal{M}_{\hat{J}}$). Moreover, by Lemma 6.4, the only element in $\mathcal{M}(\hat{J}, q)$ is the trivial $\hat{J}$–holomorphic section $s \colon \mathbb{CP}^1 \to \hat{E}$ through q.

Let $p \in \hat{\pi}^{-1}(z_{\text{gen}})$ be any other point in the fibre of q, and consider a path $\alpha \colon [0, 1] \to M$ with $\alpha(0) = p$ and $\alpha(1) = q$. For every point $\alpha(t)$, $t \in [0, 1]$, define $\mathcal{M}(\hat{J}, \alpha(t), [s]) := \{u \in \mathcal{M}_{\hat{J}} \mid \alpha(t) \in \text{Im}(u) \text{ and } [u] = [s]\}$. Clearly, $\mathcal{M}(\hat{J}, \alpha(1), [s]) = \mathcal{M}(\hat{J}, q)$.

Consider

(47) $$\mathcal{M}_{\text{cob}} := \bigcup_{t \in [0,1]} \mathcal{M}(\hat{J}, \alpha(t), [s]) \subset \mathcal{M}_{\hat{J}}.$$

The boundary components of (47) are given by $\partial \mathcal{M}_{\text{cob}} = \mathcal{M}(\hat{J}, p, [s]) \sqcup \mathcal{M}(\hat{J}, q)$. We want to show that the space $\mathcal{M}_{\text{cob}}$ is compact, so that it defines a one-dimensional cobordism between $\mathcal{M}(\hat{J}, p, [s])$ and $\mathcal{M}(\hat{J}, q)$. As before, since $\hat{J} \in \mathcal{J}(\hat{E}, \hat{\pi}, j)$ is chosen to be generic, $\mathcal{M}_{\text{cob}}$ is a smooth manifold.

To show that $\mathcal{M}_{\mathrm{cob}}$ is compact, the same strategy applies as in the case of $\mathcal{M}_{\hat{J}}$. In particular, consider a sequence $t_i \subset [0, 1]$ and for each i a section $u_i \in \mathcal{M}(\hat{J}, \alpha(t_i), [s]) \subset \mathcal{M}_{\mathrm{cob}}$.

All the sections of the sequence we are considering belong to the same homology class, by definition. In particular, they have the same area, so Gromov's theorem applies. Consequently, as t_i tends to a limit value t_∞, the sequence u_i converges to a stable map u_∞. As before (in the proof of Lemma 6.5), if sphere bubbling occurred, bubbles would have to be "vertical" (meaning entirely contained in the fibres), which is impossible by the exactness of the fibres.

It follows that $\mathcal{M}_{\mathrm{cob}}$ is compact, and hence $\mathcal{M}(\hat{J}, p, [s])$ and $\mathcal{M}(\hat{J}, q)$ are indeed cobordant. Since the signed count of the boundary components of a one-dimensional compact manifold is zero, the zero-dimensional components of the two spaces have the same cardinality. In particular, for any $p \in \hat{\pi}^{-1}(z_{\mathrm{gen}})$, $\mathcal{M}(\hat{J}, p, [s])$ is not empty, which means there is at least one element in $\mathcal{M}_{\hat{J}}$ that passes through p. $\qquad\square$

Corollary 6.7 *The map induced by the evaluation map*

$$(48) \qquad\qquad \mathrm{ev} \colon \mathcal{M}_{\hat{J}} \times \mathbb{CP}^1 \to \hat{E}, \quad (u, z) \mapsto \mathrm{ev}_z(u) = u(z),$$

is surjective.

Proof By Lemma 6.6, the image of the map (48) is dense, since each point on a smooth fibre has a preimage. As $\mathcal{M}_{\hat{J}}$ is compact and the mapping is continuous, the result extends to all points of $\hat{E}$ and hence (48) is surjective. $\qquad\square$

Proof of Theorem A Assume by contradiction that the product $\phi = \tau_{L_{j_1}} \cdots \tau_{L_{j_k}}$ is isotopic to the identity, and build the fibration $\hat{\pi} \colon \hat{E} \to \mathbb{CP}^1$ and the moduli space $\mathcal{M}_{\hat{J}}$ as above. By Corollary 6.7, the evaluation map $\mathrm{ev} \colon \mathcal{M}_{\hat{J}} \times \mathbb{CP}^1 \to \hat{E}$ is surjective. Consider the commuting diagram

$$
\begin{array}{ccc}
\mathcal{M}_{\hat{J}} \times \mathbb{CP}^1 & \xrightarrow{\ \mathrm{ev}\ } & \hat{E} \\
& \searrow^{\mathrm{pr}_2} & \downarrow^{\hat{\pi}} \\
& & \mathbb{CP}^1
\end{array}
$$

where $\mathrm{pr}_2 \colon \mathcal{M}_{\hat{J}} \times \mathbb{CP}^1 \to \mathbb{CP}^1$ is the projection to the second factor.

Let $x \in \mathrm{Crit}(\hat{\pi}) \subset \hat{E}$ be any point in the critical set. By the surjectivity of ev, there is a pair $(u, w) \in \mathcal{M}_{\hat{J}} \times \mathbb{CP}^1$ such that $u(w) = x$, so that $w \in \mathbb{CP}^1$ is the critical value associated to x. From the diagram, we obtain

$$D_{(u,w)}(\mathrm{pr}_2) = D_{(u,w)}(\hat{\pi} \circ \mathrm{ev}), \quad D_{(u,w)}(\mathrm{pr}_2) = D_x \hat{\pi} \, D_{(u,w)}(\mathrm{ev}).$$

As x is a critical point, $D_x \hat{\pi} = 0$, which forces $D_{(u,w)}(\mathrm{pr}_2)$ to be the zero map. But this is in contradiction with $D_{(u,w)}(\mathrm{pr}_2)$ being surjective. $\qquad\square$

Corollary 6.8 *There is no exact Lefschetz fibration with global monodromy symplectically isotopic to the identity, except for the trivial fibration.* $\qquad\square$

6.2 Relative version

Let

(49) $$M := T^*S^m \#_{\mathrm{pt}} T^*S^m \#_{\mathrm{pt}} T^*S^m \#_{\mathrm{pt}} \cdots \#_{\mathrm{pt}} T^*S^m$$

be a "multiplumbing" of m spheres (an iterated construction of transverse plumbing of spheres; see Section 5.1 for the definition of plumbing). By Theorem A, we know that no product $\phi \in \mathrm{Symp}_{\mathrm{ct}}(M)$ of Dehn twists along the core spheres can be compactly supported symplectically isotopic to the identity. However, the theorem, a priori, doesn't prevent such a product from acting trivially on some Lagrangian submanifolds of M. Is it possible to tell whether there are Lagrangians that detect the nontriviality of ϕ? Let T be a cotangent fibre of the jth T^*S^n–summand for $j \in \{1, \ldots, m\}$. The theorem we prove in this section shows that any product of positive Dehn twists along Lagrangian cores and involving the jth sphere does not preserve T up to compactly supported symplectic isotopy.

Theorem C *Let (M^{2n}, ω) be a Liouville manifold containing embedded Lagrangian spheres $L_1, \ldots, L_m$ and a conical Lagrangian disc T intersecting one of the spheres L_j transversely in a point. Let $\phi := \prod_{i=1}^k \tau_{L_{j_i}} \in \mathrm{Symp}_{\mathrm{ct}}(M)$ with $j_i \in \{1, \ldots, m\}$ be a positive word of Dehn twists involving τ_{L_j}. Then the Lagrangians T and $\phi(T)$ are not isotopic via a compactly supported Lagrangian isotopy.*

We prove the statement of Theorem C in the equivalent version where $(M, \omega = d\lambda_M)$ is a Liouville domain and $T \subset M$ is a Lagrangian disc preserved by the Liouville flow near the boundary ∂M (so that $\partial T \subset \partial M$). This is only chosen so that the Lefschetz fibrations involved have exact compact fibres.

As in the statement, write $\phi = \prod_{i=1}^k \tau_{L_{j_i}}$ with $j_i \in \{1, \ldots, m\}$. By assumption, there is at least one index $l \in \{1, \ldots, k\}$ such that $j_l = j$. Assuming $\phi(T) \simeq T$ via a compactly supported isotopy, we arrive at the contradictory statement $j \notin \{j_1, \ldots, j_k\}$.

From the data $(M, (L_{j_1}, \ldots, L_{j_l}, \ldots, L_{j_k}))$, build an exact Lefschetz fibration $\pi': (E', \Omega_{E'}, \lambda_{E'}) \to (\mathbb{C}, \lambda_{\mathbb{C}})$ with smooth fibre the Liouville domain $(M, d\lambda)$, basepoint $z_* \in \mathbb{R}$ with $z_* \gg 0$ such that the k critical values $\mathrm{Crit}\, v(\pi') = \{w_{j_1}, \ldots, w_{j_l}, \ldots, w_{j_k}\}$ are ordered vertically on the imaginary line, $\mathrm{Crit}\, v(\pi) \subset i\mathbb{R}$ with a basis of vanishing paths $(\gamma_{j_1}, \ldots, \gamma_{j_k})$ [Seidel 2008a, (16e)].

Let $(\Delta_{\gamma_{j_1}}, \ldots, \Delta_{j_k})$ be the corresponding basis of Lefschetz thimbles and $V_{j_i} := \pi^{-1}(z_*) \cap \Delta_{\gamma_{j_i}}$ for $i = 1, \ldots, k$, be the associated vanishing cycles, which, under the identification $\pi^{-1}(z_*) = M$, correspond to L_{j_i}. Let $\sigma: S^1 \to \mathbb{C}$ be a loop encircling all critical values.

Build a new exact fibration $\pi: (E, \Omega_E, \lambda_E) \to (\mathbb{C}, \lambda_{\mathbb{C}})$ associated to the data

$$(M, (V_{j_1}, \ldots, V_{j_l}, \ldots, V_{j_k}, V_{j_l})),$$

with basepoint $z_* \in \mathbb{C}$, an extra critical value $w_{j+1} \in \mathrm{Crit}\, v(\pi) \subset i\mathbb{R}$ and an extra vanishing path $\gamma_{j_{k+1}}$ such that $\mathrm{Im}(\gamma_{j_{k+1}}) \cap \mathrm{Im}(\sigma) = \varnothing$ (all the other choices are the same as for π').
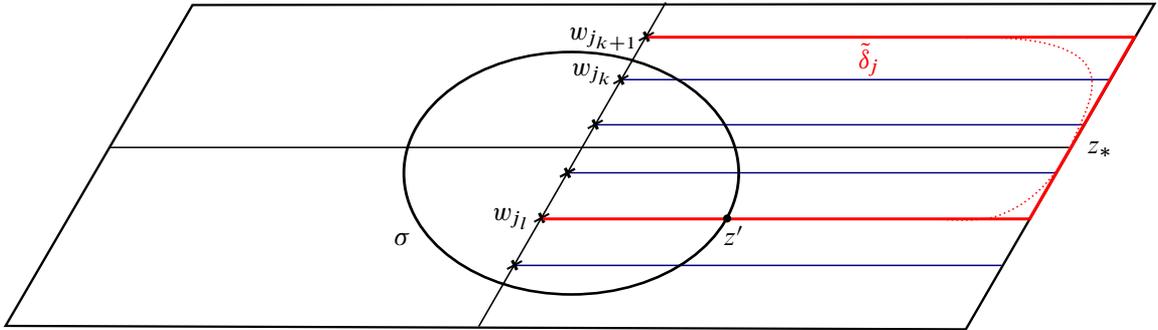
Figure 2: The new fibration π has an extra critical value $w_{j_{k+1}}$ and a matching sphere that fibres over the smoothing δ_j of the red arc $\tilde{\delta}_j$.

Compared to π', there are now two critical points w_{j_l} and $w_{j_{k+1}}$ associated to the same vanishing cycle V_{j_l}. Therefore, there is a matching path $\delta_j : [0,1] \to \mathbb{C}$ with $\delta_j(0) = w_{j_l}$ and $\delta(\frac{1}{2}) = z_*$, $\delta_j(1) = w_{j_{k+1}}$ whose parallel transport is a Lagrangian matching sphere $S_j \cong S^{n+1} \subset E$ (Section 2.2 and [Seidel 2008a, (16g)]) fibred by Lagrangians isomorphic to L_j (see Figure 2). Let $z' \in \mathrm{Im}(\delta_j) \cap \mathrm{Im}(\sigma)$, and, via parallel transport, identify $T \subset \pi^{-1}(z_*)$ with a copy of the Lagrangian in $\pi^{-1}(z')$.

By construction, the monodromy around σ is given by the product ϕ. By assumption, there is an isotopy $\phi(T) \simeq T$, so parallel transport of T along σ yields a well-defined Lagrangian $P_\sigma \subset E$. For $z \in \mathrm{Im}(\sigma)$, let $T_z \subset \pi^{-1}(z)$ be the exact fibres of P_σ. Then $\Omega_E|_{P_\sigma} = df_\sigma + \pi^*(\kappa_\sigma)$ for a function $f_\sigma \in C^\infty(P_\sigma, \mathbb{R})$ such that, for every $z \in \mathrm{Im}(\sigma)$, $f_\sigma|_{\pi^{-1}(z)}$ makes T_z exact and $\kappa_\sigma \in \Omega^1(\mathrm{Im}(\sigma))$ [Seidel 2003, Lemma 1.3].

Lemma 6.9 *The Lagrangian P_σ defines a nontrivial class in $H_{n+1}(E, \partial E; \mathbb{Z})$.*

Proof The matching sphere S_j and the disc bundle P_σ are properly embedded Lagrangian submanifolds meeting transversely at the point $y \in L_j$ lying over the intersection between σ and the matching path associated to S_j. Their homological intersection, which is the image of a nondegenerate pairing

$$H_{n+1}(E; \mathbb{Z}) \times H_{n+1}(E, \partial E; \mathbb{Z}) \to \mathbb{Z},$$

is one, so, in particular, P_σ represents a nontrivial homology class in $H_{n+1}(E, \partial E; \mathbb{Z})$. □

6.2.1 Proof of Theorem C Let $D \subseteq \mathbb{C}$ be the disc bounded by the loop σ in the base of π. The idea for the proof of Theorem C is based on a section count which follows the same principles as Section 6. In this context, however, we consider pseudoholomorphic sections defining boundary conditions for $E|_D$ on P_σ.

Let $\mathcal{J}(\pi, E, j_\mathbb{C})$ be the set of almost complex structures compatible with π (see Definition 2.15), where $j_\mathbb{C}$ is the standard complex structure on $\mathbb{C}$. For a generic element $J \in \mathcal{J}(\pi, E, j_\mathbb{C})$, let

(50) $$\mathcal{M}(J, P_\sigma) := \{u : (D, \partial D) \to (E, P_\sigma) \mid \pi \circ u = \mathrm{id}_D, \, J \circ Du = Du \circ j_\mathbb{C}|_D\}$$

be the moduli space of pseudoholomorphic sections with boundary condition on P_σ.

The Lagrangian P_σ is fibred by copies of the exact Lagrangian $T \subset M$, and therefore $P_\sigma \cap \partial^h E \neq \varnothing$, ie it is not disjoint from the horizontal boundary. As a result, the moduli space $\mathcal{M}(J, P_\sigma)$ is not compact, but fortunately its noncompact ends are very well behaved.

Below, we show that, for a generic almost complex structure in $\mathcal{J}(\pi, E, j_\mathbb{C})$, the "noncompact" elements (those sections reaching the horizontal boundary) of the moduli space (50) are regular. We do this by showing that such sections must be trivial — and the trivial section can be made regular, as the almost complex structure is product-like near $\partial^h E$. For all the other holomorphic sections, which are entirely contained in the compact region, the same regularity arguments as in [Seidel 2003] apply.

Lemma 6.10 *There is $\hat{J} \in \mathcal{J}(\pi, E, j_\mathbb{C})$ with the following property: there is no $\hat{J}$–holomorphic section $v : (D, \partial D) \to (E, P_\sigma)$ with boundary condition on P_σ such that there are $z_1, z_2 \in \partial D$ with $v(z_1) \in P_\sigma \setminus (P_\sigma \cap \partial^h E)$ and $v(z_2) \in P_\sigma \cap \partial^h E$.*

Proof We show that any generic element $J \in \mathcal{J}(\pi, E, j_\mathbb{C})$ can be deformed to an almost complex structure $\hat{J} \in \mathcal{J}(\pi, E, j_\mathbb{C})$ as in the statement. To do that we use a *reverse isoperimetric inequality* from [Groman and Solomon 2014] that applies to the Liouville completion of E.

Identify a collar neighbourhood of $\partial^h E$ with $C(\partial^h E) := \mathbb{C} \times ((-\varepsilon, 0] \times \partial M))$, and consider the Liouville completion of E, $(\bar{E}, \omega_{\bar{E}})$, obtained by gluing a cylindrical end $U^h := \mathbb{C} \times ([0, \infty) \times \partial M)$ along a collar neighbourhood of the horizontal boundary $\partial^h E$, such that $\omega_{\bar{E}}|_{U^h} = d(\lambda_\mathbb{C} + e^t \lambda_M|_{\partial M})$ for the coordinate t on $[0, \infty)$.

Let $(\bar{M}, \bar{\omega})$ be the generic smooth fibre of $\bar{E}$, and $\bar{T} \subset \bar{M}$ the Lagrangian obtained from T by gluing a conical end at the boundary. Accordingly, let $\bar{P}_\sigma \subset \bar{E}$ be the "completion" of $P_\sigma \subset E$ in $\bar{E}$. This Lagrangian can be trivialised outside of a compact set as $\partial D \times U^\infty \subset \mathbb{C} \times U^\infty \subset \bar{E}$, where $U^\infty \subset \bar{M}$ is a neighbourhood of the cylindrical end of $\bar{T}$. Extend J to a cylindrical almost complex structure $\bar{J}$ on $\bar{E}$ (see Definition 2.12).

By [Ganatra et al. 2020, Lemma 2.43], $(\bar{E}, \omega_{\bar{E}})$ has *bounded geometry* in the sense of [Ganatra et al. 2020, Definition 2.42], which is equivalent to the notion of bounded geometry of [Groman and Solomon 2014, Section 1.4]; see [Ganatra et al. 2020, page 104]. The same holds for the Lagrangian $\bar{P}_\sigma \subset \bar{E}$, as it is compact in the base direction, and conical in the fibre direction (see the proof of [Ganatra et al. 2020, Lemma 2.43]). Bounded geometry implies that for any $\bar{J}$–holomorphic section $u : (D, \partial D) \to (\bar{E}, \bar{P}_\sigma)$ there is a reverse isoperimetric inequality [Groman and Solomon 2014, Theorem 1.4]

$$(51) \qquad \qquad \ell(u|_{\partial D}) \leqslant a(u) \cdot C,$$

where ℓ is the length function associated to a $\bar{J}$–compatible metric $g_{\bar{J}}$, $C > 0$ is a constant depending on $\bar{E}$, and $a(u)$ is the area of the curve.

Let $A := \int_{\partial D} \kappa_\sigma$ (for $\kappa_\sigma \in \Omega^1(\mathrm{Im}(\sigma))$ as above), and set $R := A \cdot C$. For $R > 0$, consider a piece of symplectisation $(E_{R+1} := E \cup \mathbb{C} \times ([0, R+1] \times \partial M), \omega_{R+1})$ with $\omega_{R+1}|_E = \omega_E$ and $\omega_{R+1}|_{\mathbb{C} \times ([0, R+1] \times \partial M)} = d(\lambda_\mathbb{C} + e^t \lambda_M)$, and a compatible almost complex structure $J_{R+1} = \bar{J}|_{E_{R+1}}$ of contact type. Clearly

$E \subset E_{R+1} \subset \overline{E}$, and there is a diffeomorphism $\psi \colon E_{R+1} \to E$ that is the identity on $E \setminus C(\partial^h E)$ and compresses $\mathbb{C} \times ((-\varepsilon, R+1] \times \partial M)$ to $\mathbb{C} \times ((-\varepsilon, 0] \times \partial M)$ via the negative Liouville flow.

Every $\overline{J}$–holomorphic curve $u \colon (D, \partial D) \to (\overline{E}, \overline{P}_\sigma)$ such that there are $z_1, z_2 \in \partial D$ with $u(z_1) \in \mathrm{Int}(E)$ and $u(z_2) \in \overline{E} \setminus E_{R+1}$ satisfies $d(u(z_1), u(z_2)) > A \cdot C$ and the inequality (51)

Now set $\widehat{J} := \psi_*(J_{R+1})$. This satisfies the requirements of the lemma.

Namely, let $v \colon (D, \partial D) \to (E, P_\sigma)$ be a $\widehat{J}$–holomorphic section as in the statement, ie such that there are $z_1, z_2 \in \partial D$ with $v(z_1) \in P_\sigma \setminus (P_\sigma \cap \partial^h E)$ and $v(z_2) \in P_\sigma \cap \partial^h E$. Then we certainly have $d(v(z_1), v(z_2)) < \ell(v|_{\partial D})$ for the distance function d and the length ℓ associated to a compatible metric $g_{\widehat{J}}$. On the other hand, the area of v is bounded by a fixed upper bound since $a(v) = \int_D v^* \Omega_E = \int_D d(v^* \lambda_E) = \int_{\partial D} v^*(\lambda_E) = \int_{\partial D} \kappa_\sigma = A$ by exactness of Ω_E and fibrewise exactness of P_σ.

By stretching the neck in a neighbourhood of the boundary of E to E_{R+1}, the pullback $\psi^*(v)$ produces a contradiction, since $d(\psi^*(v(z_1)), \psi^*(v(z_2))) < \ell(\psi^*(v|_{\partial D})) < a(\psi^*(v)) \cdot C = A \cdot C$, but also $d(\psi^*(v(z_1)), \psi^*(v(z_2))) > A \cdot C$ by construction of E_{R+1}. □

From now onwards, fix an almost complex structure $\widehat{J} \in \mathcal{J}(\pi, E, j_\mathbb{C})$ as in Lemma 6.10. The above results imply that the only possible scenario left to consider in the case of a nonconstant section with boundary condition on P_σ intersecting $\partial^h E$ is to be entirely contained in the horizontal boundary of the fibration.

Lemma 6.11 *Let $u \colon D \to E$ be a $\widehat{J}$–holomorphic section such that $\mathrm{Im}(u) \subset \partial^h E$. Then u is a constant section.*

Proof Assume there is a nonconstant section $u \colon D \to E$ such that $\mathrm{Im}(u) \subset \partial^h E$. Identify (via a trivialisation as in (8)) a neighbourhood of $\partial^h E$ as $U^\partial \cong \mathbb{C} \times M^{\mathrm{out}} \subset \mathbb{C} \times M$ for a collar neighbourhood $M^{\mathrm{out}} \subset M$ of ∂M. Then the projection of $\mathrm{Im}(u)$ to M defines a nonconstant $\widehat{J}|_M$–holomorphic disc $u \colon (D, \partial D) \to (M, T)$, which, by the exactness assumptions on M, cannot exist. Therefore, u must be a constant section. □

We now prove that there are no compactness issues. The moduli space $\mathcal{M}(\widehat{J}, P_\sigma)$ has noncompact end, but by the regularity discussion above, the only sections reaching it are the constant ones, and all elements of $\mathcal{M}(\widehat{J}, P_\sigma)$ have bounded energy so that the Gromov compactness theorem applies. The bubbles in the Gromov limit of a sequence of $(\widehat{J}, j_\mathbb{C})$–holomorphic sections in $\mathcal{M}(\widehat{J}, P_\sigma)$ are either spheres in the fibres over D, or discs in the fibres $\pi^{-1}(z)$ for $z \in \mathrm{Im}(\sigma)$ with boundary condition on T_z. Both options can be discarded by exactness of E and fibrewise exactness of P_σ.

Lemma 6.12 *The evaluation map*

$$(52) \qquad \mathrm{ev} \colon \mathcal{M}(\widehat{J}, P_\sigma) \times D \to E, \quad (u, z) \mapsto u(z),$$

(i) *is proper;*

(ii) *restricts to a surjective map $\mathcal{M}(\widehat{J}, P_\sigma) \times \partial D \to P_\sigma$ of degree one.*

Proof (i) To prove this property is enough to show that every sequence of sections $\{u_k\}_{k\in\mathbb{N}}$ in $\mathcal{M}(\hat{J}, P_\sigma)$ whose image under ev lies in a relatively compact set of E has a convergent subsequence. Consider such a sequence. If its image under (52) lie in a compact set, then, by exactness, there is an upper bound to the energy of all elements in the sequence (which is bounded by a finite value determined by the maximum among all areas of the curves). Then, by the Gromov compactness theorem, $\{u_k\}_k$ admits a subsequence converging to a stable map, which, in the absence of bubbles, can only be another section.

(ii) To prove the second point, we show that the algebraic count of sections through every point of P_σ is one. Let $U^\partial \supset \partial^h E$ be a neighbourhood of the horizontal boundary as in the proof of the previous lemma and $q \in U^\partial \cap \pi^{-1}(\sigma)$. Since ϕ is compactly supported in a neighbourhood of the vanishing cycles, the monodromy around σ preserves q. By Lemmas 6.10 and 6.11, the moduli space $\mathcal{M}(\hat{J}, q) := \{u \in \mathcal{M}(\hat{J}, P_\sigma) \mid q \in \text{Im}(u)\} \subset \mathcal{M}(\hat{J}, P_\sigma)$ is compact and only contains the constant section $s: D \to E$ through q.

Given another point $p \in P_\sigma$, consider the path $\alpha: [0, 1] \to M$ with $\alpha(0) = p$ and $\alpha(1) = q$, and define

$$\mathcal{M}(\hat{J}, P_\sigma, \alpha(t), [s]) := \{u \in \mathcal{M}(\hat{J}, P_\sigma) \mid \alpha(t) \in \text{Im}(u) \text{ and } [u] = [s]\}.$$

Clearly, $\mathcal{M}(\hat{J}, P_\sigma, \alpha(1), [s]) = \mathcal{M}(\hat{J}, q)$.

Consider

$$(53) \qquad \mathcal{M}_{\text{cob}} := \bigcup_{t \in [0,1]} \mathcal{M}(\hat{J}, P_\sigma, \alpha(t), [s]) \subset \mathcal{M}(\hat{J}, P_\sigma).$$

All elements in $\mathcal{M}_{\text{cob}}$ are in the same homology class so that the same compactness arguments apply as above. Compactness implies that, for every $p \in P_\sigma$, the moduli space $\mathcal{M}(\hat{J}, P_\sigma, p, [s])$ is cobordant to the moduli space $\mathcal{M}(\hat{J}, q)$. Therefore, by the same reasoning as in the proof of Lemma 6.6, through each point of P_σ there is algebraically a unique section in $\mathcal{M}(\hat{J}, P_\sigma)$, so the restriction $\mathcal{M}(\hat{J}, P_\sigma) \times \partial D \to P_\sigma$ is surjective and of degree one. □

Proof of Theorem C Under the assumption that $\phi(T) \simeq T$, we have proved that P_σ represents a nontrivial class in $H_{n+1}(E, \partial E)$ (Lemma 6.9). The same assumption however also yields Lemma 6.12, which in particular implies that $\text{ev}_*(\mathcal{M}_{\hat{J}}(E, P_\sigma) \times \partial D) = [P_\sigma] \in H_{n+1}(E, \partial E)$ is realised as the boundary of the chain $\text{ev}_*(\mathcal{M}_{\hat{J}}(E, P_\sigma) \times D) \in C_{n+2}(E, \partial E)$. This is a contradiction. □

6.3 Product of projective twists

We continue the investigation on positive products of twists in Liouville manifolds, this time focussing on projective twists. Ideally, one would try to generalise as many results from the previous sections to this situation.

The previous section heavily relied on the link between Dehn twists and Lefschetz fibrations, and many constructions we used depended on section-count invariants of Lefschetz fibrations.

Perutz [2007] showed that any *fibred twist* admits a representation as the local monodromy of a Morse–Bott–Lefschetz (MBL) fibration. Projective twists can be thought of as an example of S^1–fibred twists, so we could envisage extending the mechanisms behind the proof for the spherical case to the setting of MBL fibrations (following [Perutz 2007; Wehrheim and Woodward 2016]) to show the analogous statement for projective twists.

Question 2 Let $\varphi \in \mathrm{Symp}_{\mathrm{ct}}(W)$ be a nonempty composition of positive powers of projective twists on a Liouville manifold (W, ω) of dimension at least four. Can φ be isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(W)$?

Unfortunately, the section-count strategy presents a route filled with obstacles, the central problem being the lack of compactness of moduli spaces of sections of MBL fibrations. The critical locus $\mathrm{Crit}(\pi)$ of such a fibration is a compact symplectic submanifold of the total space, and in general contains rational curves. The total space of a MBL fibration $\pi \colon E \to \mathbb{C}$ associated to a projective twist cannot be made into an exact symplectic manifold, so bubbling phenomena can become an issue when considering moduli spaces of pseudoholomorphic sections.

Instead, the idea remains, as in Section 5, to use the Hopf correspondence to translate a situation involving projective twists into one involving Dehn twists.

Theorem D Let (W^{2n}, ω) be a Liouville manifold containing Lagrangian real projective spaces $K_1, \ldots, K_m$ with $K_i \cong \mathbb{RP}^n$. Suppose that there is a class $\alpha \in H^1(W; \mathbb{Z}/2\mathbb{Z})$ such that, for every $i = 1, \ldots, m, \alpha|_{K_i}$ generates $H^*(\mathbb{RP}^n; \mathbb{Z}/2\mathbb{Z})$. Let $\varphi \in \mathrm{Symp}_{\mathrm{ct}}(W)$ be a positive word in the subset of projective twists $\{\tau_{K_i}\}_{i \in \{1, \ldots, m\}}$. Then φ is not isotopic to the identity in $\pi_0(\mathrm{Symp}_{\mathrm{ct}}(W))$.

Proof As in Section 4.3.2, let $q \colon (\widetilde{W}, \widetilde{\omega}) \to (W, \omega)$ be the symplectic double cover given by the class α and $L_1, \ldots, L_m \subset \widetilde{W}$ Lagrangian spheres obtained as double cover of $K_1, \ldots, K_m \subset W$. The composition of projective twists $\varphi \in \mathrm{Symp}_{\mathrm{ct}}(W)$ lifts to a composition of spherical Dehn twists $\phi \in \mathrm{Symp}_{\mathrm{ct}}(\widetilde{W})$. Assume there is an isotopy $(\varphi_t)_{0 \leqslant t \leqslant 1}$ connecting the composition of projective twists $\varphi_0 = \varphi$ to the identity $\varphi_1 = \mathrm{Id}$. The isotopy lifts to a family of compactly supported maps $(\phi_t)_{0 \leqslant t \leqslant 1}$ in the double cover $\widetilde{W}$, where $\phi_0 = \phi$ is the lift of φ. Then ϕ_1 covers the identity and can therefore only be either the identity or a deck transformation. The latter type would define a noncompactly supported symplectomorphism, hence ϕ_1 must be the identity. It follows that $\phi \in \mathrm{Symp}_{\mathrm{ct}}(\widetilde{W})$ is a composition of Dehn twists in a Liouville domain which is isotopic to the identity, contradicting Theorem A. $\square$

Remark 6.13 A similar argument fails when applied to complex projective twists. Let (W^{4n}, ω) be a symplectic manifold with complex projective Lagrangians $K_1, \ldots, K_m$ satisfying Assumption (CX). The fibration $(Y, \Omega) \to (W, \omega)$ constructed from the cohomological condition is not proper, so an isotopy in $\mathrm{Symp}_{\mathrm{ct}}(W)$ cannot be lifted to an isotopy in $\mathrm{Symp}_{\mathrm{ct}}(Y)$.   ◁

7 Epilogue: framings of projective twists, homotopy projective Lagrangians

As a last application of the Hopf correspondence, we examine homotopy projective Lagrangians. We prove two nonembedding results for Lagrangian projective spaces in nonstandard homeomorphism/diffeomorphism classes (Theorems E and F), and, for $n \geq 19$, the existence of projective twists obtained from a nonstandard choice of framing that are not Hamiltonian isotopic to the standard $\tau_{\mathbb{CP}^n} \in \mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$:

Theorem G *The $\mathbb{CP}^n$–twist depends on the framing when $n = 19, 23, 25, 29$.*

Embedding theorems are obtained in Section 7.1 using homotopy theory results combined with the existing state of the art of the nearby Lagrangian conjecture, and the use of the Hopf correspondence.

We subsequently investigate the question of framings for projective twists in Section 7.2. For that purpose, we utilise the current literature on framing of Dehn twists, a pairing constructed by Bredon, and the Hopf correspondence. This enables us to obtain instances in which the (Hamiltonian isotopy class of the) local projective twist does depend on a choice of framing of the associated Lagrangian projective space. With the additional use of topological modular forms, we explain why there should be infinitely many such examples.

7.1 Lagrangian nonembeddings of projective spaces

The *nearby Lagrangian conjecture* states that, given a closed smooth manifold Q, any closed exact Lagrangian submanifold of $(T^*Q, d\lambda_Q)$ is Hamiltonian isotopic to the zero section. If this conjecture were true, the existence of another closed exact Lagrangian embedding $L \hookrightarrow T^*Q$ would yield a diffeomorphism $L \cong Q$. By Weinstein's neighbourhood theorem, the latter version of the statement can also be read as: if $(T^*L, d\lambda_{T^*L})$ is symplectomorphic to $(T^*Q, d\lambda_{T^*Q})$, then L is diffeomorphic to Q.

The conjecture has been verified for some specific examples (T^*S^2 and $T^*\mathbb{RP}^2$ by [Hind 2012; Li and Wu 2012] and T^*T^2 by Dimitroglou Rizell, Goodman and Ivrii [Dimitroglou Rizell et al. 2016]), and weaker versions of it have been proved. Currently, the most general feature one can deduce from an exact Lagrangian embedding in $(T^*Q, d\lambda_{T^*Q})$ is (simple) homotopy equivalence:

Theorem 7.1 [Abouzaid 2012b; Kragh 2013; Abouzaid and Kragh 2018] *If $L \subset T^*Q$ is a closed, exact Lagrangian embedding, then the projection $L \subset T^*Q \xrightarrow{p} Q$ is a (simple) homotopy equivalence.*

Remark 7.2 If $L \subset T^*Q \xrightarrow{p} Q$ is a Lagrangian as in the above statement, then $TL \otimes \mathbb{C} \cong p^*(TQ \otimes \mathbb{C})$. The Pontryagin classes $p_i \in H^{4k}(\cdot)$ satisfy $2p_i(L) = 2p_i(Q)$. Moreover, the (rational) Pontryagin classes p_i are homeomorphism invariants [Novikov 1965]. ◁

Equipped with the connected sum operation, the set of h–cobordism classes of homotopy m–spheres Θ_m has an abelian group structure (where the standard sphere plays the role of neutral element). We will always assume $m > 5$, in which case the elements of Θ_m correspond to diffeomorphism classes of m–spheres.

The group Θ_m fits in an exact sequence [Kervaire and Milnor 1963]

$$(54) \qquad\qquad 0 \to \mathrm{bP}_{m+1} \to \Theta_m \xrightarrow{\;\psi\;} \mathrm{coker}(J_m) \to \mathrm{bP}_m.$$

Here $\mathrm{bP}_{m+1} = \ker(\psi) \subset \Theta_m$ denotes the subgroup of homotopy m–spheres bounding an $(m+1)$–dimensional parallelisable manifold, and $J_m \colon \pi_m(O) \to \pi_m(S)$ is a map from the mth stable homotopy group $\pi_m(O) = \lim_{l\to\infty} \pi_m(\mathrm{SO}(l))$ to the mth stable homotopy group of spheres $\pi_m(S) := \lim_{l\to\infty} \pi_{m+l}(S^l)$ (see eg [Levine 1985, Section 3]). This group is also called the mth *stable stem*.

Throughout the section, we will repeatedly use the following fact about the sequence (54):

Theorem 7.3 [Kervaire and Milnor 1963, Theorem 5.1] *If m is an odd integer, $\mathrm{bP}_m = 0$. Consequently, for any odd m, $\psi \colon \Theta_m \to \mathrm{coker}(J_m)$ is surjective.*

In the symplectic setting, homotopy spheres are good candidates to test the nearby Lagrangian conjecture.

Theorem 7.4 [Abouzaid 2012a] (extended by [Ekholm et al. 2016]) *Let $m > 4$ odd. If $\Sigma, \Sigma' \in \Theta_m$ and $T^*\Sigma$ is symplectomorphic to $T^*\Sigma'$, then $[\Sigma] = \pm[\Sigma'] \in \Theta_m/\mathrm{bP}_{m+1}$.*

It will be practical to paraphrase the above theorem as follows:

Corollary 7.5 *If $m > 4$ is odd and $\Sigma \in \Theta_m \setminus \mathrm{bP}_{m+1}$, then Σ does not admit a Lagrangian embedding into T^*S^m.*

Definition 7.6 We choose to depart from the classic terminology of *exotic* manifolds. Instead, we will call a smooth manifold that is homeomorphic, but not diffeomorphic, to the standard sphere an *AD sphere* (AD stands for alternative differentiable structure). Correspondingly, a smooth manifold that is homeomorphic, but not diffeomorphic, to the standard $\mathbb{CP}^n$ will be called an *AD projective space*. Finally, a smooth manifold that is homotopy equivalent, but not homeomorphic, to the standard projective space will be called an *AT projective space* (where AT stands for alternative topological structure).   ◁

7.1.1 Results The results of this section hinge on the existence of homotopy projective spaces that are obtained as the reduced space of a circle action on an AD sphere. It is not always possible to relate an n–dimensional AD/AT projective space to a $(2n+1)$–dimensional AD sphere in this way. Below, we start by exploring a few facts about AD/AT projective spaces, after which we can discuss three interesting examples where the desired phenomenon is observed (the spaces of Theorems E and F).

Definition 7.7 [Kawakubo 1969] The inertia group $I(M)$ of an oriented closed smooth manifold M is the subgroup of Θ_m consisting of homotopy spheres $S \in \Theta_m$ such that the connected sum $M \# S$ is in the same diffeomorphism class as M.   ◁

If $I(\mathbb{CP}^n) = 0$ and Θ_{2n} is nontrivial, one can build an AD projective space as follows. Given an AD sphere $\Sigma \in \Theta_{2n}$ the connected sum $\mathbb{CP}^n \# \Sigma$ (a zero-dimensional surgery) is another manifold homeomorphic to $\mathbb{CP}^n$ but not diffeomorphic to it. For $n \geq 8$, there are examples for which the inertia group $I(\mathbb{CP}^n)$ is nontrivial (see [Kawakubo 1969]); in those cases the smooth structure of the resulting manifold is not automatically distinct from the standard smooth structure on $\mathbb{CP}^n$. In dimension four, we know:

Theorem 7.8 [Kasilingam 2016] *There are two possible distinct smooth structures on a manifold homeomorphic to $\mathbb{CP}^4$: the standard $\mathbb{CP}^4$–structure, and the one on $\mathbb{CP}^4 \# \Sigma^8$, where $\Sigma^8 \in \Theta_8$ is the unique AD 8–sphere.*

In contrast, it is known that there is an abundance of AT projective spaces: for even integers $n \geq 4$, there are infinitely many AT projective spaces, distinguished by the first Pontryagin class $p_1 \in H^4(\mathbb{CP}^n; \mathbb{Z})$ [Hsiang 1966].

Is there a way to associate an AD sphere to an AD/AT projective space? Given an AD/AT projective space K, the unit bundle of the line bundle $\mathcal{L} \to K$ satisfying $c_1(\mathcal{L}) = \alpha_K$ (where $\alpha_K \in H^2(K; \mathbb{Z})$ is the cohomology generator) could still be diffeomorphic to a standard sphere. Note that, in the special case where the projective space is a surgery of the form $K = \mathbb{CP}^n \# \Sigma$, for an AD sphere $\Sigma \in \Theta_{2n}$, the $(2n+1)$–sphere obtained as the unit bundle of $\mathcal{L} \to K$ is given by $\mathrm{stab}(\Sigma) \in \Theta_{2n+1}$ (where stab is the map constructed in Section 7.2; see Remark 7.17).

On the other hand, one could examine S^1–quotients of AD spheres $\widetilde{S} \in \Theta_{2n+1}$. A priori this is not always a successful strategy, as not all homotopy spheres admit a smooth free circle action. But, if such an action exists, then the quotient $P := \widetilde{S}/S^1$ resulting from it is an AD or AT projective space. Namely, this reduced space is necessarily homotopy equivalent to a projective space [Hsiang 1966], but it is at least not diffeomorphic to the standard $\mathbb{CP}^n$ (since circle bundles over P are classified by elements of $H^2(P; \mathbb{Z})$, and, if P were the standard projective space, then the total space of the line bundle would have to be a standard sphere).

Theorem 7.9 [James 1980, Sections 2–3] *There is a homotopy 9–sphere $\widetilde{S}$ such that*:

(i) $\widetilde{S} \notin \mathrm{bP}_{10} \cong \mathbb{Z}/2\mathbb{Z}$.

(ii) $\widetilde{S}$ *admits a free action of S^1.*

(iii) *The quotient $P := \widetilde{S}/S^1$ is not homeomorphic to $\mathbb{CP}^4$.*

(iv) *P and the standard $\mathbb{CP}^4$ have the same tangent bundles.*

Remark 7.10 James [1980, Section 3] notes that there is another S^1–action on $\widetilde{S}$ with quotient space $P \# \Sigma^8$. The latter is an AT projective space that is not diffeomorphic to P. ◁

We now have enough material to state and prove the results of this section.

Theorem E *There is a manifold P homotopy equivalent to $\mathbb{CP}^4$ and with the same first Pontryagin class such that neither P nor $P \# \Sigma^8$ admits an exact Lagrangian embedding into $T^*\mathbb{CP}^4$.*

Proof Consider the homotopy 9–sphere $\widetilde{S}$ admitting a free S^1–action of Theorem 7.9. We let $P := \widetilde{S}/S^1$ and prove it is the right candidate to satisfy the claim. The quotient P is homotopy equivalent to $\mathbb{CP}^4$, but, by Theorem 7.9(iii), it is not homeomorphic to it. The first Pontryagin classes of P and $\mathbb{CP}^4$ coincide by Theorem 7.9(iv). Assume there is a Lagrangian embedding $P \hookrightarrow T^*\mathbb{CP}^4$. The Hopf correspondence (see Lemma 3.1) lifts P to $\widetilde{S}$, giving an exact Lagrangian embedding $\widetilde{S} \hookrightarrow T^*S^9$. However, by Theorem 7.9, $\widetilde{S} \in \Theta_9 \setminus \mathrm{bP}_{10}$, so the existence of the Lagrangian embedding contradicts Corollary 7.5.

The same argument applies to prove that $P \# \Sigma^8$ does not embed as Lagrangian into $T^*\mathbb{CP}^4$. Namely, the Hopf correspondence would, in that case too, lift (via the S^1–action of Remark 7.10) $P \# \Sigma^8$ to $\widetilde{S}$ [James 1980, Section 3]. □

Remark 7.11 Our techniques do not allow to prove whether the AD projective space $\mathbb{CP}^4 \# \Sigma^8$ of Theorem 7.8 does admit a Lagrangian embedding into $T^*\mathbb{CP}^4$ or not.   ◁

Theorem F *There is an element Σ^{14} in the group of homotopy 14–spheres Θ_{14} such that $\mathbb{CP}^7 \# \Sigma^{14}$ does not admit an exact Lagrangian embedding into $T^*\mathbb{CP}^7$.*

Proof First note that $\Theta_{14} \cong \mathbb{Z}/2\mathbb{Z}$ and $\mathrm{bP}_{15} = 0$ [Kervaire and Milnor 1963], so there is a unique AD 14–sphere. We define Σ^{14} to be this AD 14–sphere and prove it is the right candidate to satisfy the claim. By [Bredon 1967, Theorem 4.6], there is an AD sphere $\Sigma^{15} \in \Theta_{15} \setminus \mathrm{bP}_{16}$ admitting a free S^1–action, with quotient $P := \mathbb{CP}^7 \# \Sigma^{14}$. If P admitted a Lagrangian embedding in $T^*\mathbb{CP}^7$, the Hopf correspondence would yield a Lagrangian embedding $\Sigma^{15} \hookrightarrow T^*S^{15}$. But $\Sigma^{15} \notin \mathrm{bP}_{16}$, which would contradict Corollary 7.5 (for the same reasons as in the proof of Theorem E).   □

7.2 Framing of projective twists

The background material that we use to examine the question of framing of projective twists is based on [Dimitroglou Rizell and Evans 2015], in which it is proved that the Hamiltonian isotopy class of a Dehn twist does in general depend on a choice of framing.

Let (M, ω) be a symplectic manifold. Given a framing of a Lagrangian sphere $L \subset M$, ie a diffeomorphism $S^n \to L$ (see Section 2.1), the precomposition with an element $F \in \mathrm{Diff}(S^m)$ yields another framing.

Consider the symplectomorphism $F^* \colon T^*S^m \to T^*S^m$ induced by the lift of F to the cotangent bundle T^*S^m. The standard model twist $\tau_{S^m}^{\mathrm{loc}} \in \mathrm{Symp}_{\mathrm{ct}}(T^*S^m)$ can be replaced by $F^* \circ \tau_{S^m}^{\mathrm{loc}} \circ (F^{-1})^*$, and the latter can be implanted in a Weinstein neighbourhood as in Definition 2.8 to produce a new element in $\mathrm{Symp}_{\mathrm{ct}}(M)$. To study framings of twists, we can then restrict to these parametrisations of the standard model twist $\tau_{S^m} := \tau_{S^m}^{\mathrm{loc}} \in \mathrm{Symp}_{\mathrm{ct}}(T^*S^m)$.

A core fact for the study of parametrisations of twists is the isomorphism $\pi_0(\text{Diff}^+(S^m)) \cong \Theta_{m+1}$ [Kervaire and Milnor 1963; Cerf 1970]. In particular, given a nontrivial diffeomorphism $F \in \text{Diff}^+(S^m)$, there is an $(m+1)$–dimensional AD sphere constructed as follows:

Definition 7.12 Let $F \in \text{Diff}(S^m)$ be a diffeomorphism not isotopic to the identity. Then $\Sigma_F := D^{m+1} \cup_F D^{m+1} \in \Sigma_{m+1}$ is an $(m+1)$–dimensional homotopy sphere obtained by gluing two $(m+1)$–discs along their boundary S^m twisted by F. In the notation of [Dimitroglou Rizell and Evans 2015, Definition 1.4] (which is more apt to visualise the Lagrangian suspension we utilise in Section 7.2.2), this is equivalent to

$$\Sigma_F := (D^{m+1} \times S^0) \cup_\Phi S^m \times [0, 1]$$

glued along $S^m \times S^0$ via $\Phi \colon S^m \times S^0 \to S^m \times S^0$, $\Phi(x, y) = (F(x), y)$. ◁

Also recall there is an isomorphism $\pi_0(\text{Diff}^+(S^m)) \cong \pi_0(\text{Diff}^+_{\text{ct}}(D^m))$ induced by a map $\text{Diff}^+_{\text{ct}}(D^m) \to \text{Diff}^+(S^m)$ which extends all elements of $\text{Diff}^+_{\text{ct}}(D^n)$ over a capping disc.

Dimitroglou Rizell and Evans proved the existence of Dehn twists, whose Hamiltonian isotopy class depends on the choice of framing.

Definition 7.13 [Dimitroglou Rizell and Evans 2015, Definition 1.1] Fix a cotangent fibre $\Lambda \subset T^*S^m$ and let $\mathcal{L}_m \subset \Theta_m$ be the subset of homotopy spheres which admit a Lagrangian embedding into T^*S^m with the additional requirement that the embedding intersects Λ transversely in exactly one point. ◁

Theorem 7.14 [Dimitroglou Rizell and Evans 2015, Theorem A] *Let* $F \in \text{Diff}^+(S^m)$ *be such that* $\Sigma_F \notin \mathcal{L}_{m+1}$. *Then* $\tau_{S^m}^{-1} \circ (F^* \circ \tau_{S^m} \circ (F^{-1})^*)$ *is not trivial in* $\pi_0(\text{Symp}_{\text{ct}}(T^*S^m))$.

In the rest of the section, we analyse the analogous problem for reparametrisations $f \in \text{Diff}(\mathbb{CP}^n)$ of projective twists. We prove that there exist $n \in \mathbb{N}$ such that the twist $\tau_f := f^* \circ \tau_{\mathbb{CP}^n} \circ (f^{-1})^*$ is not isotopic to the standard projective twist in $\pi_0(\text{Symp}_{\text{ct}}(T^*\mathbb{CP}^n))$, where $f^* \colon T^*\mathbb{CP}^n \to T^*\mathbb{CP}^n$ is the symplectomorphism induced by the lift of f to the cotangent bundle. We will not directly use Theorem 7.14 but an intermediary result (Proposition 7.15 below) that Dimitroglou Rizell and Evans proved (using [Abouzaid 2012a; Abouzaid and Kragh 2018; Ekholm and Smith 2014]) to support their arguments.

Proposition 7.15 [Dimitroglou Rizell and Evans 2015, Proposition 1.2] *There is an inclusion* $\mathcal{L}_m \subset \text{bP}_{m+1}$.

Remark 7.16 There is a slight abuse of terminology in the entirety of the section. A *framing* will be employed (as in the rest of the paper) in the nonstandard sense à la Seidel to signify a smooth parametrisation of a sphere. The classical topological notion of framing (as a trivialisation of the normal bundle) is also needed in this section, and, in order to avoid a conflict of nomenclature, we call the latter a *normal framing*. ◁

7.2.1 Bredon's pairing We begin by introducing an essential component of the arguments of this section: a map

(55) $$\mathrm{stab}\colon \Theta_m \to \Theta_{m+1}$$

obtained as a special case of a homomorphism $\Theta_m \otimes \pi_l(S) \to \Theta_{m+l}$ studied in [Bredon 1967].

Consider the linear action of $\mathrm{SO}(2) \simeq S^1$ on $S^{m+1} \subset \mathbb{R}^{m+2}$ via the representation

$$\mathrm{SO}(2) \to \mathrm{SO}(m+2), \quad A \mapsto \varphi(A) = \begin{bmatrix} A & & \\ & A & \\ & & \ddots \end{bmatrix},$$

with 1 in the right-hand bottom corner if m is odd. This is the linear S^1–action on S^{m+1}, which is free if m even (in which case it is the standard Hopf action), and whose fixed-point set is S^0 if m odd.

For $\Sigma \in \Theta_m$, Bredon's construction [1967, Sections 1 and 4] yields a homotopy $(m+1)$–sphere as follows. Let $V \subset \Sigma$ be an open neighbourhood of a point $p \in \Sigma$, and $g\colon (V, p) \to (\mathbb{R}^m, 0)$ an orientation-reversing diffeomorphism. Let $B := g^{-1}(D^m) \subset V \subset \Sigma$, where $D^m \subset \mathbb{R}^m$ is the unit disc.

Let $\mathcal{C} \cong S^1 \subset S^{m+1}$ be a principal orbit of the $\mathrm{SO}(2)$–action on S^{m+1}, equipped with a normal framing $\mathcal{F}\colon \mathcal{C} \times \mathbb{R}^m \to S^{m+1}$.

Define

(56) $$\mathrm{stab}(\Sigma) := S^{m+1} \setminus (\mathcal{F}(\mathcal{C} \times D^m)) \cup \mathcal{C} \times (\Sigma \setminus B),$$

where the two pieces are glued along their boundaries, which can be identified via a diffeomorphism $\mathcal{F}(\mathcal{C} \times (\mathbb{R}^m \setminus \{0\})) \cong \mathcal{C} \times (V \setminus \{p\})$ as in [Bredon 1967, page 435].

The normally framed orbit $(\mathcal{C}, \mathcal{F})$ represents an element $\gamma \in \pi_{m+1}(S^m) \cong \pi_4(S^3) \cong \mathbb{Z}/2\mathbb{Z}$ via the Thom–Pontryagin construction (see [Milnor 1965, Section 7]). With this identification in mind, the map stab is derived from a pairing $\Theta_m \times \pi_{m+1}(S^m) \to \Theta_{m+1}$, $(\Sigma, \gamma) \mapsto \mathrm{stab}(\Sigma) = \langle \Sigma, \gamma \rangle$ (see [Bredon 1967, (1)]). The latter induces a homomorphism [Bredon 1967, (2)]

(57) $$\Theta_m \otimes \pi_1(S) \to \Theta_{m+1}.$$

To determine the class γ, we follow [Bredon 1967, (4.1)] and find that $\gamma = \eta^j$, where $\eta \in \pi_1(S) := \pi_4(S^3) \cong \mathbb{Z}/2\mathbb{Z}$ is the nontrivial element in the stable stem $\pi_1(S)$ and

$$j = \begin{cases} \frac{1}{2}m & \text{if } m \text{ even (ie the action on } S^{m+1} \text{ is free)}, \\ \frac{1}{2}(m-1) & \text{if } m \text{ odd (ie the action } S^{m+1} \text{ is not free)}. \end{cases}$$

Intuitively, if a normally framed Hopf circle in S^3 represents the class $\eta \in \pi_4(S^3)$, then γ is determined by the number of times (mod 2) that this normal framing fits in the normal bundle to $\mathcal{C} \subset S^{m+1}$.

For $m + 1 = 2n + 1$ and $m + 1 = 2n + 2$, we have $j = n$ and

(58) $$\gamma = \eta^n = \begin{cases} \eta & \text{if } n \text{ odd}, \\ 0 & \text{if } n \text{ even}. \end{cases}$$

Remark 7.17 For an even-dimensional homotopy sphere $\Sigma \in \Theta_{2n}$, the image stab(Σ) can also be described as follows (this remark is relevant for Section 7.1). Consider the surgery $\mathbb{CP}^n \# \Sigma$ and the complex line bundle $\mathcal{L} \to \mathbb{CP}^n \# \Sigma$ associated to the generator of $H^2(\mathbb{CP}^n \# \Sigma; \mathbb{Z})$. Then stab($\Sigma$) is the homotopy sphere obtained as the unit circle bundle of $\mathcal{L}$. ◁

We now focus on the case $m + 1 = 2n + 2$.

Lemma 7.18 The map $\Theta_{2n+1} \to \Theta_{2n+2}/\mathrm{bP}_{2n+3}$ is nontrivial for $n = 19, 23, 25, 29$.

Proof There is a commuting diagram (see [Bredon 1967, Corollary 2.2]) obtained from the exact sequence (54),

(59)
$$
\begin{array}{ccc}
\Theta_{2n+1} & \xrightarrow{\ \ \mathrm{stab}\ \ } & \Theta_{2n+2} \\
\downarrow{\scriptstyle\psi} & & \downarrow{\scriptstyle\psi} \\
\mathrm{coker}(J_{2n+1}) & \xrightarrow{\ (-)\cdot\eta^n\ } & \mathrm{coker}(J_{2n+2})
\end{array}
$$

where $(-) \cdot \eta^n : \mathrm{coker}(J_{2n+1}) \to \mathrm{coker}(J_{2n+2})$, is a map descending from the multiplication $\pi_{2n+1}(S) \times \pi_1(S) \to \pi_{2n+2}(S)$ with the class $\eta \in \pi_1(S) \cong \mathbb{Z}/2\mathbb{Z}$, which is well defined since, for $l + 1 < m$, $\mathrm{Im}(J_m) \cdot \mathrm{Im}(J_l) \subseteq \mathrm{Im}(J_{m+l})$, the image of the J–homomorphism is preserved under multiplication with elements of the stable stems.

By (58), we know that a necessary requirement for the map stab to be nontrivial is to have $n = 2k + 1$ for some $k \in \mathbb{N}$ so that $\eta^n = \eta$ is nontrivial. In that case, we get

(60)
$$
\begin{array}{ccc}
\Theta_{4k+3} & \xrightarrow{\ \ \mathrm{stab}\ \ } & \Theta_{4(k+1)} \\
\downarrow{\scriptstyle\psi} & & \downarrow{\scriptstyle\psi} \\
\mathrm{coker}(J_{4k+3}) & \xrightarrow{\ (-)\cdot\eta\ } & \mathrm{coker}(J_{4(k+1)})
\end{array}
$$

The vertical maps are both surjective since ψ is always surjective in odd dimensions and when $m \equiv 0 \mod 4$ (see [Levine 1985, Theorem 5.4]).

The exact sequence (54) implies that $\mathrm{coker}(J_{4k+4}) \cong \Theta_{4k+4}/\ker(\psi) \cong \Theta_{4k+4}$, and the nontriviality of the composition $\psi \circ \mathrm{stab} : \Theta_{4k+3} \to \Theta_{4k+4}$ is equivalent to the nontriviality of the multiplication $(-) \cdot \eta : \mathrm{coker}(J_{4k+3}) \to \mathrm{coker}(J_{4k+4})$. This amounts to looking for elements in the stable stems whose η–multiples are not in the image of J. As η is of order two, this information can be found in the "two-primary part" of the stable stems, the subgroups obtained after quotienting all elements of odd order. These are tabulated in a diagram in [Hatcher 2002, page 385], where the elements of interest appear to be in degrees $2n + 1 \in \{39, 47, 51, 59\}$, which means that $n \in \{19, 23, 25, 29\}$. □

The rest of the section is dedicated to explaining how to relate a parametrisation $f \in \mathrm{Diff}^+(\mathbb{CP}^n)$ of the standard projective twist to a parametrisation $F \in \mathrm{Diff}^+(S^{2n+1})$ of the standard Dehn twist.

Lemma 7.19 *Let n be an odd integer and $f \in \mathrm{Diff}^+(\mathbb{CP}^n)$ an orientation-preserving diffeomorphism. There exists a diffeomorphism $F \in \mathrm{Diff}^+(S^{2n+1})$ satisfying $h \circ F = f \circ h$, ie F is the lift of f by the Hopf bundle map.*

Proof Let $h \colon S^{2n+1} \to \mathbb{CP}^n$ the Hopf bundle map. A diffeomorphism $f \colon \mathbb{CP}^n \to \mathbb{CP}^n$ induces a continuous function $F \colon f^*(S^{2n+1}) \to S^{2n+1}$ covering f such that the diagram

(61)
$$
\begin{array}{ccc}
f^*(S^{2n+1}) & \xrightarrow{\quad F \quad} & S^{2n+1} \\
\downarrow{\scriptstyle h'} & & \downarrow{\scriptstyle h} \\
\mathbb{CP}^n & \xrightarrow{\quad f \quad} & \mathbb{CP}^n
\end{array}
$$

commutes, where $h' \colon f^*(S^{2n+1}) \to \mathbb{CP}^n$ is the pullback bundle of h by f. The map induced by f on the second cohomology $\bar{f} \colon H^2(\mathbb{CP}^n) \to H^2(\mathbb{CP}^n)$ is $\pm\mathrm{Id}$. Therefore, the Euler classes of $h \colon S^{2n+1} \to \mathbb{CP}^n$ and $h' \colon f^*(S^{2n+1}) \to \mathbb{CP}^n$ coincide up to sign, so these principal S^1–bundles must have diffeomorphic total spaces. It follows that $F \colon f^*(S^{2n+1}) \to S^{2n+1}$ is in fact a diffeomorphism $F \colon S^{2n+1} \to S^{2n+1}$ covering f satisfying $h \circ F = f \circ h$. $\qquad\square$

Lemma 7.20 *Let n be an odd integer and $f \in \mathrm{Diff}^+(\mathbb{CP}^n)$ be an orientation-preserving diffeomorphism supported in an open chart, ie f is induced by an element of $\mathrm{Diff}_{\mathrm{ct}}^+(D^{2n})$, and let $\Sigma_f \in \Theta_{2n+1}$ be the homotopy $(2n+1)$–sphere associated to f.*

Let $F \in \mathrm{Diff}^+(S^{2n+1})$ be the S^1–equivariant lift of f of Lemma 7.19 and $\Sigma_F \in \Theta_{2n+2}$ the corresponding homotopy $(2n+2)$–sphere. Then $\mathrm{stab}(\Sigma_f) = \Sigma_F$.

Proof The lift $F \in \mathrm{Diff}^+(S^{2n+1})$ is supported in a tubular neighbourhood of a Hopf circle in S^{2n+1}. To build Σ_F, identify S^{2n+1} with an equator in S^{2n+2}, and consider the above circle as a normally framed circle $\mathcal{C} \subset S^{2n+2}$. That requires a choice of trivialisation of the normal bundle to a Hopf circle, a normal framing $\mathcal{F} \colon \mathcal{C} \times \mathbb{R}^{2n+1} \to S^{2n+2}$ that defines the support of the gluing map for the construction of Σ_F: on $\mathcal{F}(\mathcal{C} \times D^{2n+1})$, F acts as $\mathrm{id} \times f$. By the same arguments as in the beginning of this section, the normal framing of this Hopf circle corresponds to the class $\eta^n \in \pi_1(S)$, so $\Sigma_F = \mathrm{stab}(\Sigma_f)$ by the construction (56). $\qquad\square$

7.2.2 Results We first mention an auxiliary result from [Dimitroglou Rizell and Evans 2015] we will need in the proof of Theorem G.

Lemma 7.21 [Dimitroglou Rizell and Evans 2015, Proposition 2.5] *Consider $(T^*S^{2n+1}, d\lambda_{T^*S^{2n+1}})$ equipped with the well-known structure of a Lefschetz fibration $T^*S^{2n+1} \to \mathbb{C}$ with smooth fibre $(T^*S^{2n}, d\lambda_{T^*S^{2n}})$ and two singular fibres. Let $L \subset T^*S^{2n+1}$ be the standard Lagrangian embedding of the zero section. There is an open symplectic embedding*

(62)
$$
e \colon T^*S^{2n+1} \times T^*[0, 1] \to T^*S^{2n+2}
$$

such that:

- $L \times [0, 1]$ *is sent to a subset of the zero section $S^{2n+2} \subset T^*S^{2n+2}$* (*the matching sphere*).

- *The image of the embedding is disjoint from a particular cotangent fibre $\Lambda \subset T^*S^{2n+2}$* (*a Lefschetz thimble*).

Proposition 7.22 *If the map* $\mathrm{stab} \colon \Theta_{2n+1} \to \Theta_{2n+2}$ *is nontrivial and n is odd, then the $\mathbb{CP}^n$–twist depends on a choice of framing.*

Proof Choose a framing $f \in \mathrm{Diff}^+(\mathbb{CP}^n)$, coming from an element of $\mathrm{Diff}^+_{\mathrm{ct}}(D^{2n})$ extended by the identity on the projective space. Let $F \in \mathrm{Diff}^+(S^{2n+1})$ be the S^1–equivariant lift of f as in Lemma 7.19, supported in a tubular neighbourhood of a Hopf circle $\mathcal{F} \colon \cong S^1 \times D^{2n} \subset S^{2n+1}$. Let $\Sigma_f \in \Theta_{2n+1}$ be the sphere associated to f, and $\Sigma_F \in \Theta_{2n+2}$ that associated to F. By Lemma 7.20, $\Sigma_F = \mathrm{stab}(\Sigma_f) = \langle \Sigma_f, \eta^n \rangle \in \Theta_{2n+2}$. Since n is odd, $\eta^n = \eta$ and $\Sigma_F = \mathrm{stab}(\Sigma_f) \in \Theta_{2n+2}$ is nontrivial.

The map f^* induced by f on the cotangent bundle is not compactly supported, but can be used to define the compactly supported conjugation

$$(63) \qquad \tau_f := f^* \circ \tau_{\mathbb{CP}^n} \circ (f^{-1})^* \colon T^*\mathbb{CP}^n \to T^*\mathbb{CP}^n$$

of the projective twist $\tau_{\mathbb{CP}^n} \in \mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$.

We next show below that τ_f belongs to a Hamiltonian class distinct from that of the standard projective twist:

Lemma 7.23 *The twist τ_f defined in (63) is not isotopic to the standard twist $\tau_{\mathbb{CP}^n} \in \mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$.*

Proof Assume by contradiction that $\tau_{\mathbb{CP}^n}^{-1} \circ \tau_f$ is (Hamiltonian) isotopic to the identity in $\mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$. Let $(\phi_t)_{t \in [0,1]}$ be an isotopy connecting the two symplectomorphisms in $\mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$ such that there are $s' > s \in (0, 1)$ with

$$(64) \qquad \phi_t = \begin{cases} \tau_{\mathbb{CP}^n}^{-1} \circ \tau_f & \text{if } t \leqslant s, \\ \mathrm{Id} & \text{if } t \geqslant s'. \end{cases}$$

Let $H \colon T^*\mathbb{CP}^n \times [0, 1] \to \mathbb{R}$ be the generating Hamiltonian function. Define the Lagrangian embedding

$$(65) \qquad \psi' \colon K \times [0, 1] \to T^*\mathbb{CP}^n \times T^*[0, 1], \quad (x, t) \mapsto \big(\phi_t(x), t, -H(\phi_t(x), t)\big),$$

where $K \subset T^*\mathbb{CP}^n$ is the standard Lagrangian embedding of the zero section. By construction, near the ends of the interval, $K \times [0, 1]$ is preserved by (65) in the sense that $\psi'(K \times [0, s]) \subset K \times [0, s]$ and $\psi'(K \times [s', 1]) \subset K \times [s', 1]$.

On each fibre $T^*\mathbb{CP}^n$ of $T^*\mathbb{CP}^n \times T^*[0, 1]$, apply the Hopf correspondence to lift the image of (65) to a Lagrangian embedding

$$(66) \qquad \Psi \colon L \times [0, 1] \to T^*S^{2n+1} \times T^*[0, 1]$$

(where $L \subset T^*S^{2n+1}$ is the standard Lagrangian embedding of the zero section) such that $L \times I$ is preserved by Ψ near the ends of the interval.

By Lemma 7.21, we can replace $e(L \times [0,1]) \subset T^*S^{2n+2}$ by the Lagrangian suspension $\Psi(L \times [0,1])$, so that the ends of $\Psi(L \times [0,1])$ are "capped" into a $(2n+2)$–dimensional sphere diffeomorphic to $\Sigma_F \in \Theta_{2n+2}$ (see [Dimitroglou Rizell and Evans 2015, Section 3.3]) which intersects a cotangent fibre once transversely and is therefore contained in $\mathcal{L}_{2n+2}$. By Proposition 7.15, $\mathcal{L}_{2n+2} \subset \mathrm{bP}_{2n+3}$ and since $\mathrm{bP}_{2n+3} = 0$ (this holds for all odd integers; see [Kervaire and Milnor 1963]), Σ_F has to be the standard sphere. However, as we have seen above, $\Sigma_F = \langle \eta^n, \Sigma_f \rangle \in \Theta_{2n+2}$ is nontrivial as n is odd. This is a contradiction, which proves Lemma 7.23; τ_f cannot be isotopic to the standard projective twist in $\mathrm{Symp}_{\mathrm{ct}}(T^*\mathbb{CP}^n)$.   □

This also concludes the proof of Proposition 7.22.   □

The above results are sufficient to prove the following:

Theorem G *The $\mathbb{CP}^n$–twist depends on the framing when $n = 19, 23, 25, 29$.*

Proof The statement is proved by combining Lemma 7.18 with Proposition 7.22.   □

Proposition 7.24 *The $\mathbb{CP}^n$–twist depends on the choice of framing for infinitely many dimensions n.*

Proof One way to obtain infinite families of nontrivial multiples of η which are not contained in the image of J is by detecting them in topological modular forms, denoted by tmf (we refer to [Henriques 2014] for a survey on the subject). There is a "Hurewicz homomorphism" $\pi_*(\mathbb{S}) \to \pi_*(\mathrm{tmf})$ between the ring of stable homotopy groups of spheres and the homotopy ring of tmf, and the two primary components of the ring of homotopy groups have a certain kind of periodicity of degree 192. Therefore, if we can identify an element in one of the homotopy groups $\pi_{4k+3}(\mathrm{tmf})$ that is also in the image of the Hurewicz homomorphism and arises as a product of η, we obtain a periodic family of elements to which the argument of Lemma 7.18 applies.

A (partially conjectural) diagram depicting the two-primary components can be found in [Henriques 2014] and it is helpful to first identify a potential candidate. Degree $39 = 4 \cdot 9 + 3$ presents an element which has been confirmed to be the image of a nontrivial multiple of η (see [Hopkins and Mahowald 2014, Corollary 11.2], there the element in question is called u and arises as image of a product of $\bar{\kappa}, \nu, \eta$ and κ; all of these are standard names of generators of stable homotopy groups stems). It follows that, in every dimension $m \equiv 39 \bmod 192$, there is an element for which the map $(-) \cdot \eta \colon \mathrm{coker}(J_m) \to \mathrm{coker}(J_{m+1})$ and hence $\mathrm{stab} \colon \Theta_m \to \Theta_{m+1}$ are not trivial. Recall that $m = 4k+3 = 2n+1$, so that, by Proposition 7.22, the projective twist depends on the framing for $n \equiv 19 \bmod 96$. Further scrutiny of the literature would provide other such elements, eg for $m = 59$ ($n = 29$).   □

Remark 7.25 It is very likely that a version of Theorem G holds for $\mathbb{HP}^n$–twists as well. Bredon [1967, page 446] computes the class that would be associated to a framing of $S^3 \subset S^{4n+3}$, which is a power of $\nu \in \pi_3(S) = \lim_m \pi_{m+3}(S^m) \cong \pi_8(S^5) \cong \mathbb{Z}_{24}$. Nontriviality results for the map stab in this case would not only depend on the parity of n, so a nonvanishing criterion would be harder to obtain.

But such a criterion could then be combined with the existence of smooth semifree actions of S^3 on homotopy $(4k+3)$–spheres explicitly computed in [Bredon 1967, Theorems 4.4 and 4.7] (also note that there are infinitely many inequivalent free S^3–actions on homotopy S^{4k+3}–spheres, by [Hsiang 1966, Theorem 3]). Then the above strategy could be applied to obtain infinitely many dimensions in which the $\mathbb{HP}^n$–twist would depend on the framing. ◁

Corollary 7.26 *In the above dimensions,* $\mathrm{Symp}_{ct}(T^*\mathbb{CP}^n) \not\simeq \mathbb{Z}$.

Proof If $\tau_{\mathbb{CP}^n} \in \pi_0(\mathrm{Symp}_{ct}(T^*\mathbb{CP}^n))$ is the standardly framed twist along the zero section, then we claim that $\mathbb{Z}\langle\tau_{\mathbb{CP}^n}\rangle \subsetneq \pi_0(\mathrm{Symp}_{ct}(T^*\mathbb{CP}^n))$. Let $f \in \mathrm{Diff}_{ct}^+(\mathbb{CP}^n)$ be a framing such that the projective twist $\tau_f \in \mathrm{Symp}_{ct}(T^*\mathbb{CP}^n)$ defined using f is not isotopic to $\tau_{\mathbb{CP}^n}$, as in Theorem G. Then $\tau_f^{-1} \circ \tau_{\mathbb{CP}^n}$ cannot be isotopic to any power $\tau_{\mathbb{CP}^n}^k$ for any $k \in \mathbb{Z}$. This is because $\tau_{\mathbb{CP}^n}$, viewed as a graded symplectomorphism, acts nontrivially on the grading of the zero section, viewed as a graded Lagrangian (see [Seidel 2000, Lemma 5.7]), whereas $\tau_f^{-1} \circ \tau_{\mathbb{CP}^n}$ acts trivially on the grading (see also [Dimitroglou Rizell and Evans 2015, Remark 1.5]). □

References

[Abouzaid 2011] **M Abouzaid**, *A topological model for the Fukaya categories of plumbings*, J. Differential Geom. 87 (2011) 1–80 MR Zbl

[Abouzaid 2012a] **M Abouzaid**, *Framed bordism and Lagrangian embeddings of exotic spheres*, Ann. of Math. 175 (2012) 71–185 MR Zbl

[Abouzaid 2012b] **M Abouzaid**, *Nearby Lagrangians with vanishing Maslov class are homotopy equivalent*, Invent. Math. 189 (2012) 251–313 MR Zbl

[Abouzaid and Kragh 2018] **M Abouzaid**, **T Kragh**, *Simple homotopy equivalence of nearby Lagrangians*, Acta Math. 220 (2018) 207–237 MR Zbl

[Arnold 1995] **V I Arnold**, *Some remarks on symplectic monodromy of Milnor fibrations*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 99–103 MR Zbl

[Auroux 2003] **D Auroux**, *Monodromy invariants in symplectic topology*, preprint (2003) arXiv math/0304113

[Barth et al. 2019] **K Barth**, **H Geiges**, **K Zehmisch**, *The diffeomorphism type of symplectic fillings*, J. Symplectic Geom. 17 (2019) 929–971 MR Zbl

[Biran and Cornea 2013] **P Biran**, **O Cornea**, *Lagrangian cobordism, I*, J. Amer. Math. Soc. 26 (2013) 295–340 MR Zbl

[Biran and Cornea 2014] **P Biran**, **O Cornea**, *Lagrangian cobordism and Fukaya categories*, Geom. Funct. Anal. 24 (2014) 1731–1830 MR Zbl

[Bredon 1967] **G E Bredon**, *A Π_*–module structure for Θ_* and applications to transformation groups*, Ann. of Math. 86 (1967) 434–448 MR Zbl

[Cerf 1970] **J Cerf**, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. 39 (1970) 5–173 MR Zbl

[Chiang et al. 2016] **R Chiang**, **F Ding**, **O van Koert**, *Non-fillable invariant contact structures on principal circle bundles and left-handed twists*, Internat. J. Math. 27 (2016) art. id. 1650024 MR Zbl

[Dimitroglou Rizell and Evans 2015] **G Dimitroglou Rizell**, **J D Evans**, *Exotic spheres and the topology of symplectomorphism groups*, J. Topol. 8 (2015) 586–602 MR Zbl

[Dimitroglou Rizell et al. 2016] **G Dimitroglou Rizell**, **E Goodman**, **A Ivrii**, *Lagrangian isotopy of tori in $S^2 \times S^2$ and $\mathbb{C}P^2$*, Geom. Funct. Anal. 26 (2016) 1297–1358 MR Zbl

[Ekholm and Smith 2014] **T Ekholm**, **I Smith**, *Exact Lagrangian immersions with one double point revisited*, Math. Ann. 358 (2014) 195–240 MR Zbl

[Ekholm et al. 2016] **T Ekholm**, **T Kragh**, **I Smith**, *Lagrangian exotic spheres*, J. Topol. Anal. 8 (2016) 375–397 MR Zbl

[Evans 2010] **J D Evans**, *Lagrangian spheres in del Pezzo surfaces*, J. Topol. 3 (2010) 181–227 MR Zbl

[Evans 2011] **J D Evans**, *Symplectic mapping class groups of some Stein and rational surfaces*, J. Symplectic Geom. 9 (2011) 45–82 MR Zbl

[Farb and Margalit 2012] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR Zbl

[Floer 1988] **A Floer**, *Morse theory for Lagrangian intersections*, J. Differential Geom. 28 (1988) 513–547 MR Zbl

[Ganatra et al. 2020] **S Ganatra**, **J Pardon**, **V Shende**, *Covariantly functorial wrapped Floer theory on Liouville sectors*, Publ. Math. Inst. Hautes Études Sci. 131 (2020) 73–200 MR Zbl

[Ganatra et al. 2024] **S Ganatra**, **J Pardon**, **V Shende**, *Sectorial descent for wrapped Fukaya categories*, J. Amer. Math. Soc. 37 (2024) 499–635 MR Zbl

[Gao 2017a] **Y Gao**, *Functors of wrapped Fukaya categories from Lagrangian correspondences*, preprint (2017) arXiv 1712.00225

[Gao 2017b] **Y Gao**, *Wrapped Floer cohomology and Lagrangian correspondences*, preprint (2017) arXiv 1703.04032

[Groman and Solomon 2014] **Y Groman**, **J P Solomon**, *A reverse isoperimetric inequality for J–holomorphic curves*, Geom. Funct. Anal. 24 (2014) 1448–1515 MR Zbl

[Guillemin and Sternberg 1990] **V Guillemin**, **S Sternberg**, *Symplectic techniques in physics*, 2nd edition, Cambridge Univ. Press (1990) MR Zbl

[Harris 2011] **R M Harris**, *Projective twists in $\mathcal{A}$–infinity categories*, preprint (2011) arXiv 1111.0538

[Hatcher 2002] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[Henriques 2014] **A Henriques**, *The homotopy groups of tmf and of its localizations*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) 189–205 MR Zbl

[Hind 2012] **R Hind**, *Lagrangian unknottedness in Stein surfaces*, Asian J. Math. 16 (2012) 1–36 MR Zbl

[Hopkins and Mahowald 2014] **M J Hopkins**, **M Mahowald**, *From elliptic curves to homotopy theory*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Math. Surveys Monogr. 201, Amer. Math. Soc., Providence, RI (2014) 261–285 MR Zbl

[Hsiang 1966] **W-c Hsiang**, *A note on free differentiable actions of S^1 and S^3 on homotopy spheres*, Ann. of Math. 83 (1966) 266–272 MR Zbl

[Huybrechts and Thomas 2006] **D Huybrechts**, **R Thomas**, ℙ*–objects and autoequivalences of derived categories*, Math. Res. Lett. 13 (2006) 87–98 MR Zbl

[Ishida 1996] **A Ishida**, *The structure of subgroup of mapping class groups generated by two Dehn twists*, Proc. Japan Acad. Ser. A Math. Sci. 72 (1996) 240–241 MR Zbl

[James 1980] **D M James**, *Free circle actions on homotopy nine spheres*, Illinois J. Math. 24 (1980) 681–688 MR Zbl

[Kasilingam 2016] **R Kasilingam**, *Classification of smooth structures on a homotopy complex projective space*, Proc. Indian Acad. Sci. Math. Sci. 126 (2016) 277–281 MR Zbl

[Kawakubo 1969] **K Kawakubo**, *On the inertia groups of homology tori*, J. Math. Soc. Japan 21 (1969) 37–47 MR Zbl

[Keating 2014] **A M Keating**, *Dehn twists and free subgroups of symplectic mapping class groups*, J. Topol. 7 (2014) 436–474 MR Zbl

[Kervaire and Milnor 1963] **M A Kervaire**, **J W Milnor**, *Groups of homotopy spheres, I*, Ann. of Math. 77 (1963) 504–537 MR Zbl

[Khovanov and Seidel 2002] **M Khovanov**, **P Seidel**, *Quivers, Floer cohomology, and braid group actions*, J. Amer. Math. Soc. 15 (2002) 203–271 MR Zbl

[Kragh 2013] **T Kragh**, *Parametrized ring-spectra and the nearby Lagrangian conjecture*, Geom. Topol. 17 (2013) 639–731 MR Zbl

[Levine 1985] **J P Levine**, *Lectures on groups of homotopy spheres*, from "Algebraic and geometric topology" (A Ranicki, N Levitt, F Quinn, editors), Lecture Notes in Math. 1126, Springer (1985) 62–95 MR Zbl

[Li and Wu 2012] **T-J Li**, **W Wu**, *Lagrangian spheres, symplectic surfaces and the symplectic mapping class group*, Geom. Topol. 16 (2012) 1121–1169 MR Zbl

[Mak and Wu 2018] **C Y Mak**, **W Wu**, *Dehn twist exact sequences through Lagrangian cobordism*, Compos. Math. 154 (2018) 2485–2533 MR Zbl

[Mak and Wu 2019] **C Y Mak**, **W Wu**, *Dehn twists and Lagrangian spherical manifolds*, Selecta Math. 25 (2019) art. id. 68 MR Zbl

[Ma'u et al. 2018] **S Ma'u**, **K Wehrheim**, **C Woodward**, *A∞ functors for Lagrangian correspondences*, Selecta Math. 24 (2018) 1913–2002 MR Zbl

[Maydanskiy and Seidel 2010] **M Maydanskiy**, **P Seidel**, *Lefschetz fibrations and exotic symplectic structures on cotangent bundles of spheres*, J. Topol. 3 (2010) 157–180 MR Zbl

[McDuff and Salamon 2017] **D McDuff**, **D Salamon**, *Introduction to symplectic topology*, 3rd edition, Oxford Univ. Press (2017) MR Zbl

[Milnor 1965] **J W Milnor**, *Topology from the differentiable viewpoint*, Univ. Press of Virginia, Charlottesville, VA (1965) MR Zbl

[Moser 1965] **J Moser**, *On the volume elements on a manifold*, Trans. Amer. Math. Soc. 120 (1965) 286–294 MR Zbl

[Novikov 1965] **S P Novikov**, *Topological invariance of rational classes of Pontrjagin*, Dokl. Akad. Nauk SSSR 163 (1965) 298–300 MR Zbl In Russian; translated in Sov. Math., Dokl. 6 (1965) 921–923

[Perutz 2007] **T Perutz**, *Lagrangian matching invariants for fibred four-manifolds, I*, Geom. Topol. 11 (2007) 759–828 MR Zbl

[Perutz 2008]   **T Perutz**, *A symplectic Gysin sequence*, preprint (2008) arXiv 0807.1863

[Poźniak 1994]   **M Poźniak**, *Floer homology*, *Novikov rings and clean intersections*, PhD thesis, University of Warwick (1994)

[Ritter 2014]   **A F Ritter**, *Floer theory for negative line bundles via Gromov–Witten invariants*, Adv. Math. 262 (2014) 1035–1106 MR Zbl

[Seidel 1998]   **P Seidel**, *Symplectic automorphisms of T^*S^2*, preprint (1998) arXiv math/9803084

[Seidel 1999]   **P Seidel**, *Lagrangian two-spheres can be symplectically knotted*, J. Differential Geom. 52 (1999) 145–171 MR Zbl

[Seidel 2000]   **P Seidel**, *Graded Lagrangian submanifolds*, Bull. Soc. Math. France 128 (2000) 103–149 MR Zbl

[Seidel 2003]   **P Seidel**, *A long exact sequence for symplectic Floer cohomology*, Topology 42 (2003) 1003–1063 MR Zbl

[Seidel 2008a]   **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

[Seidel 2008b]   **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

[Seidel and Smith 2005]   **P Seidel**, **I Smith**, *The symplectic topology of Ramanujam's surface*, Comment. Math. Helv. 80 (2005) 859–881 MR Zbl

[Seidel and Thomas 2001]   **P Seidel**, **R Thomas**, *Braid group actions on derived categories of coherent sheaves*, Duke Math. J. 108 (2001) 37–108 MR Zbl

[Smith 2001]   **I Smith**, *Geometric monodromy and the hyperbolic disc*, Q. J. Math. 52 (2001) 217–228 MR Zbl

[Tonkonog 2015]   **D Tonkonog**, *Commuting symplectomorphisms and Dehn twists in divisors*, Geom. Topol. 19 (2015) 3345–3403 MR Zbl

[Wehrheim and Woodward 2010a]   **K Wehrheim**, **C T Woodward**, *Functoriality for Lagrangian correspondences in Floer theory*, Quantum Topol. 1 (2010) 129–170 MR Zbl

[Wehrheim and Woodward 2010b]   **K Wehrheim**, **C T Woodward**, *Quilted Floer cohomology*, Geom. Topol. 14 (2010) 833–902 MR Zbl

[Wehrheim and Woodward 2012]   **K Wehrheim**, **C T Woodward**, *Floer cohomology and geometric composition of Lagrangian correspondences*, Adv. Math. 230 (2012) 177–228 MR Zbl

[Wehrheim and Woodward 2016]   **K Wehrheim**, **C T Woodward**, *Exact triangle for fibered Dehn twists*, Res. Math. Sci. 3 (2016) art. id. 17 MR Zbl

[Wu 2014]   **W Wu**, *Exact Lagrangians in A_n–surface singularities*, Math. Ann. 359 (2014) 153–168 MR Zbl

Centre for Mathematical Sciences, University of Cambridge
Cambridge, United Kingdom

bct@adhoc.ch

On keen weakly reducible bridge spheres

PUTTIPONG PONGTANAPAISAN
DANIEL RODMAN

A bridge sphere is said to be keen weakly reducible if it admits a unique pair of disjoint compressing disks on opposite sides. In particular, such a bridge sphere is weakly reducible, not perturbed, and not topologically minimal in the sense of David Bachman. In terms of Jennifer Schultens' width complex, a link in bridge position with respect to a keen weakly reducible bridge sphere is distance one away from a local minimum. We give infinitely many examples of keen weakly reducible bridge spheres for links in b bridge position for $b \geq 4$.

1 Introduction

Suppose that we have a decomposition of the 3–sphere $S^3 = V_+ \cup_\Sigma V_-$ where V_+ and V_- are 3–balls and Σ is a 2–sphere. A link $L \subset S^3$ intersecting Σ transversely is said to be in *bridge position* with respect to Σ if $L \cap V_+ = \alpha_+$ and $L \cap V_- = \alpha_-$, where α_+ and α_- are b–strand trivial tangles. The punctured sphere $\Sigma_L = \Sigma \backslash L$ is called a b–*bridge sphere*. To each bridge sphere, we can assign a disk complex $\mathscr{D}(\Sigma_L)$, which is a simplicial complex whose vertices are isotopy classes of compressing disks in $S^3 \backslash L$ for Σ_L and whose k simplices are spanned by $k+1$ vertices with pairwise disjoint representatives.

We say that Σ_L is *topologically minimal* if one of the following holds:

(1) $\mathscr{D}(\Sigma_L) = \varnothing$.

(2) There exists $i \in \mathbb{N} \cup \{0\}$ such that the ith homotopy group of $\mathscr{D}(\Sigma_L)$ is nontrivial.

The *topological index* of Σ_L is defined to be 0 if $\mathscr{D}(\Sigma_L) = \varnothing$, or the smallest i such that $\pi_{i-1}(\mathscr{D}(\Sigma_L))$ is nontrivial if $\mathscr{D}(\Sigma_L) \neq \varnothing$. The notion of topological minimality was introduced by David Bachman [2010] as a generalization of useful concepts such as incompressibility and strong irreducibility of surfaces in a 3–manifold. It turns out that topologically minimal surfaces possess desirable properties. For instance, in an irreducible 3–manifold, a topologically minimal surface can be isotoped to intersect an incompressible surface in such a way that any intersection loop is essential in both surfaces. Furthermore, the concept of topological minimality gave rise to examples of 3–manifolds containing arbitrarily many nonminimal genus, unstabilized Heegaard surfaces that are weakly reducible [Bachman 2013]. Moriah [2007] dubbed these examples "the nemesis of Heegaard splittings" as they are difficult to find.

Conjecturally, there is a special and mysterious relationship between topologically minimal surfaces and geometrically minimal surfaces, which are surfaces whose mean curvature is identically zero. Every geometrically minimal surface has a Morse index, which roughly speaking counts the maximal number of directions the surface can be deformed so as to decrease its area. Freedman, Hass and Scott [Freedman et al. 1983] showed that every surface of topological index zero is isotopic to a geometrically minimal surface of Morse index zero. By works of Pitts and Rubinstein [1987] and of Ketover, Liokumovich, and Song [Ketover et al. 2019], a Heegaard surface of topological index one is isotopic to a geometrically minimal surface of Morse index at most one. Campisi and Torres [2020] showed that the genus two Heegaard surface of the 3–sphere has topological index three. By Urbano [1990], this Heegaard surface must have Morse index at least six. Thus, it is not true in general that a surface of topological index k is isotopic to a surface of Morse index at most k, but the precise connection is not well understood.

One can ask the interesting question of which surfaces are topologically minimal. Several authors have given examples of topologically minimal Heegaard surfaces [Bachman and Johnson 2010; Campisi and Rathbun 2018; Campisi and Torres 2020; Lee 2015] and bridge surfaces [Lee 2016; Pongtanapaisan and Rodman 2021; Rodman 2018]. Heegaard surfaces that are not topologically minimal have also been studied by several authors who constructed *keen weakly reducible* Heegaard surfaces. That is, each of these surfaces possesses a unique *weak reducing pair*, a pair of compressing disks on opposite sides of the surface whose boundaries are disjoint. By a result of McCullough [1991], the disk complex of the boundary of a handlebody is contractible. Thus having a unique pair of weak reducing disks on distinct sides of a Heegaard splitting means that in the disk complex, there is a unique edge connecting the two contractible subcomplexes corresponding the two handlebodies, resulting in a contractible disk complex for the Heegaard surface. The examples of keen weakly reducible Heegaard surfaces in the literature with simple descriptions include the canonical Heegaard surface of a surface bundle whose monodromy has sufficiently high translation distance by Johnson [2012], some Heegaard surfaces arising from self-amalgamations by E and Lei [2014], and certain unstabilized genus three Heegaard surfaces in irreducible and orientable 3–manifolds by Kim [2016]. More complicated constructions of keen weakly reducible Heegaard surfaces of genus $g \geq 3$ can also be found in [E 2017; Liang et al. 2018].

The goal of this paper is to provide infinitely many examples of nontopologically minimal bridge spheres, which are lacking in the literature, by verifying that the canonical bridge sphere for certain links in plat position is keen weakly reducible. Such links are obtained by "amalgamating" two types of links whose canonical bridge spheres are topologically minimal. Keen weakly reducible bridge spheres also belong to a family of surfaces with finitely many pairs of disjoint compressing disks [E and Zhang 2023], which is interesting in its own right.

Theorem 1.1 *There exist infinitely many links with keen weakly reducible bridge spheres.*

This paper is organized as follows. In Section 2, we discuss properties of a keen weakly reducible bridge sphere related to perturbations of bridge spheres, thin position of links, and essential surfaces in the link

exterior. In Section 3, we define the notion of a plat position for a link, consider a particular family of links in plat position, and describe useful positions of curves on a punctured sphere with respect to a train track. In Section 4, we characterize the behaviors of curves that bound disks above or below the bridge sphere. In Section 5, we use a criterion presented in [Cho 2008] to show that keen weakly reducible bridge spheres are not topologically minimal.

Acknowledgements

2 Consequences of being keen weakly reducible

In this section, we discuss some consequences of putting a link in bridge position with respect to a keen weakly reducible bridge sphere. We remark that a priori keen weakly reducible bridge spheres are not necessarily canonical bridge spheres for links in plat position.

2.1 Unperturbed bridge spheres

Let L be a link in bridge position with respect to Σ. Then $L \cap V_+ = \alpha_+$ is a collection of disjoint embedded arcs with the property that there exists an isotopy (rel $\partial\alpha^+$) taking α_+ into Σ. For each arc α_+^i of α_+, the trace of such an isotopy is a disk called a *bridge disk* D_+^i. From each bridge disk D_+^i we can obtain a compressing disk dD_+^i called the *frontier* of D_+^i using the construction $dD_+^i = (\partial \overline{N}(D_+^i)) \cap V_+$. Analogous definitions can be made for $L \cap V_- = \alpha_-$.

We say that a bridge sphere Σ_L is *perturbed* if there exist two bridge disks $D_+^1 \subset V_+$ and $D_-^1 \subset V_-$ such that $D_+^1 \cap D_-^1$ is a single point contained in L. It is an interesting problem to search for unperturbed bridge spheres for a link up to isotopy since a perturbed bridge can always be obtained from a bridge sphere that is not perturbed by an isotopy which introduces a maximal point and a minimal point as shown in Figure 1. In some cases, the only destabilized bridge sphere is the one that realizes the bridge number [Otal 1985; Ozawa 2011; Zupan 2011]. Another common way to show that a bridge sphere Σ_L for a nontrivial link L is unperturbed is to show that there is no weak reducing pair for Σ_L. Being keen weakly reducible implies the following.

Proposition 2.1 *If Σ_L is keen weakly reducible, then Σ_L is unperturbed.*

It is well known that a perturbed bridge sphere has a weak reducing pair; we prove that result here, and show that such a bridge sphere in fact has at least two weak reducing pairs.
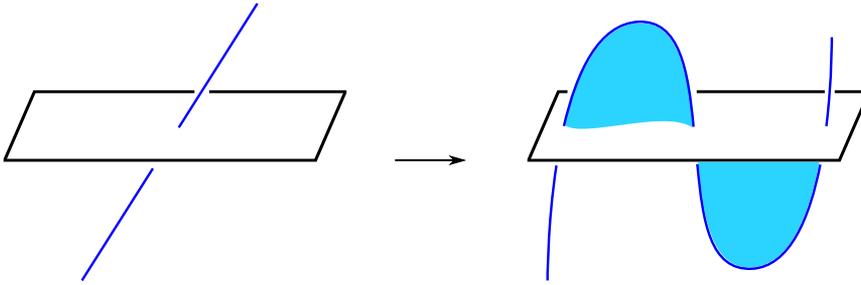
Figure 1: Introducing a canceling pair of critical points.

Proof If Σ_L is a perturbed bridge sphere for a link L in 2–bridge position, then L must be the unknot and there exists a unique compressing disk D above and a unique compressing disk E below. Furthermore, $D \cap E \neq \varnothing$, which implies that Σ_L does not admit a weak reducing pair, and therefore Σ_L cannot be keen weakly reducible. To complete the proof, we consider perturbed bridge spheres for links in b–bridge position, where $b \geq 3$.

Suppose that Σ_L is perturbed. By definition, there exist bridge disks $D_+^1 \subset V_+$ and $D_-^1 \subset V_-$ such that $D_+^1 \cap D_-^1 = \{p\} \in L$. Let $\mathscr{A}_+$ be a set of b disjoint bridge disks for Σ, each corresponding to one of the components of α_+, and suppose further that $D_+^1 \in \mathscr{A}_+$. Let $\mathscr{A}_-$ be a similarly defined set of bridge disks below Σ with $D_-^1 \in \mathscr{A}_-$. (We are able to define $\mathscr{A}_+$ and $\mathscr{A}_-$ *after* D_+^1 and D_-^1 by [Scharlemann 2005, Lemma 3.2].) The elements of $\mathscr{A}_-$ may or may not intersect the interior of the arc $D_+^1 \cap \Sigma$. Below, we describe how if they do, we can replace them with another set of b disjoint bridge disks below Σ, each of which is disjoint from the interior of $D_+^1 \cap \Sigma$.

Suppose that the elements of $\mathscr{A}_-$ intersect the interior of $D_+^1 \cap \Sigma$. Consider a point q of intersection closest to p. Let $D_-' \in \mathscr{A}_-$ denote the bridge disk containing q. We perform a surgery on D_-' as depicted in Figure 2, resulting in a new disk D_-''. Notice that D_-'' is disjoint from the other elements of $\mathscr{A}_-$, and D_-' and D_-'' both correspond to the same bridge arc. In slight abuse of notation, we will replace D_-' with D_-'' in the collection $\mathscr{A}_-$. After this replacement, $\mathscr{A}_-$ remains a collection of pairwise disjoint bridge disks for the bridge arcs below Σ. The difference is that now, the elements of $\mathscr{A}_-$ intersect the interior of $D_+^1 \cap \Sigma$ in one fewer point.
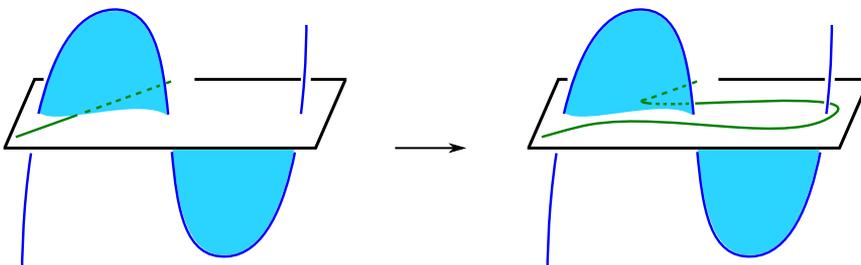
Figure 2: Bridge disks below Σ_K can be isotoped to intersect D_+^1 in two points in K.

We can repeatedly perform such surgeries until the bridge disks of $\mathscr{A}_-$ are all disjoint from the interior of $D^1_+ \cap \Sigma$. It follows that the bridge disks of $\mathscr{A}_-$ intersect D^1_+ only in the two points of $D^1_+ \cap L \cap \Sigma$, each of which intersects a bridge disk of $\mathscr{A}_-$. Since L is in a b bridge position with $b > 2$, these two bridge disks must be distinct. In addition to these two, there must be at least one more bridge disk $D^2_- \in \mathscr{A}_-$ since $b \geq 3$, and so D^2_- is disjoint from D^1_+. Therefore, dD^1_+ and dD^2_- comprise a weak reducing pair for Σ_L.

Now consider dD^1_-. We can mimic the trick in the previous paragraph so that a particular set of b pairwise disjoint bridge disks above Σ intersects D^1_- only in the two points of $D^1_- \cap L \cap \Sigma$. Then there is some bridge disk D^2_+ disjoint from D^1_-, which means that dD^1_- and dD^2_+ comprise another weak reducing pair for Σ_L. Therefore, a perturbed bridge sphere never admits a unique weak reducing pair and can never be keen weakly reducible. $\qquad\square$

2.2 Width complex

Suppose that L is a link and $h: S^3 \to \mathbb{R}$ is the standard Morse function. Assume also that $h|_L$ is a Morse function. Suppose that $c_1 < \cdots < c_n$ are critical values of $h|_L$. Consider $h^{-1}(r_i)$, where r_i is a regular value between c_i and c_{i+1}. We say that a level sphere $h^{-1}(r_i)$ is a *thin level* if $|h^{-1}(r_{i-1}) \cap L| > |h^{-1}(r_i) \cap L|$ and $|h^{-1}(r_i) \cap L| < |h^{-1}(r_{i+1}) \cap L|$. On the other hand, a level sphere $h^{-1}(r_i)$ is a *thick level* if $|h^{-1}(r_{i-1}) \cap L| < |h^{-1}(r_i) \cap L|$ and $|h^{-1}(r_i) \cap L| > |h^{-1}(r_{i+1}) \cap L|$. We say that a disk $D \subseteq S^3 \backslash L$ is a *strong upper* (*resp. lower*) *disk* with respect to $h^{-1}(r_i)$ if

(1) $\partial D = \alpha \cup \beta$ where $\alpha \subset L$ contains exactly one maximal (resp. minimal) point and β is an arc in $h^{-1}(r_i)$, and

(2) the interior of D contains no critical point with respect to the height function h.

If there exists a strong upper disk and a strong lower disk intersecting in exactly one point lying in L (see Figure 1, for instance), then there is an isotopy that cancels a maximal point and a minimal point. We call such a move a *type I move*. On the other hand, if there exists a strong upper disk and a strong lower disk that are disjoint, then there is an isotopy that interchanges a maximal point and a minimal point. We call such a move a *type II move*.

Schultens [2009] associated to a knot K a graph called the *width complex of K* to understand the structure of the collection of Morse embeddings of a fixed knot K. Two embeddings k and k' of K are considered to be *equivalent* if their thin and thick levels are isotopic. With this definition of equivalence, each vertex of the width complex is an equivalence class of embeddings of K such that $h|_K$ is a Morse function. An edge connects two vertices representing embeddings k and k' if k differs from k' by one of the following moves: a type I move, the inverse of a type I move, a type II move, or the inverse of a type II move. Schultens proved the following interesting result.

Theorem 2.2 [Schultens 2009] *The width complex of a knot is connected.*

The proof of Theorem 2.2 uses the fact that projections of k and k' to the vertical plane differ by a finite number of Reidemeister moves and planar isotopy. Furthermore, each of these local moves either affects an embedding by a type I or a type II move or does not alter the equivalence class at all. As any two projections of a multicomponent link L are also related by Reidemeister moves and planar isotopy, it follows that the width complex of a multicomponent link is also connected.

A vertex that is particularly interesting is one representing an embedding that admits no type I or type II moves. Such an embedding is said to be in *locally thin position*.

Proposition 2.3 *Suppose that l is an embedding of a link L in bridge position with respect to a keen weakly reducible bridge sphere. In the width complex of L, there is an edge between l and an embedding l' of L in a locally thin position.*

Proof Let $D \subseteq V_+$ and $E \subseteq V_-$ be a weak reducing pair for a keen weakly reducible bridge sphere Σ_L.

Claim *∂D and ∂E each cut out a twice punctured disk from Σ_L.*

Proof of claim Suppose that ∂D cuts Σ_L into two components F_1 and F_2, where each component is a punctured disk containing more than two punctures. The loop ∂E is contained in one of the components, say F_1. There exists at least one bridge disk D_+^1 such that $\partial D_+^1 = \alpha \cup \beta$ where $\alpha \subset L$ and $\beta \subset F_2$. Then, dD_+^1 and E give rise to a weak reducing pair distinct from D and E, which is a contradiction. The same argument also implies ∂E cuts out a twice-punctured disk from Σ_L.   □

Observe that D cuts off a 3–ball containing a unique bridge disk, which is a strong upper disk disjoint from a strong lower disk contained in a 3–ball cut off by E. This pair of disks gives rise to a type II move. After the type II move is performed, there are neither type I nor type II moves left to perform because any pair of strong upper disk and strong lower disk (intersecting in one point of L or mutually disjoint) that emerges after the type II move on D and E will yield a distinct pair of strong upper disk and strong lower disk on Σ_L, and hence Σ_L admits more than one weak reducing pair, which is a contradiction. □

After a type II move is performed along D and E, a thin level emerges. This thin level is incompressible because a compressing disk for this level would imply the existence of another weak reducing pair different from D and E. Thus, we obtain the following corollary.

Corollary 2.4 *A link with a keen weakly reducible bridge sphere contains an essential meridional surface in its exterior.*

3 Setting

In this section, we redevelop and summarize several tools and concepts of Johnson and Moriah [2016]. Specifically, Section 3.1 is a brief summary of Johnson and Moriah's plat links and accompanying tools such as their σ_i and π_y projection maps. Then in Section 3.3, we develop Johnson and Moriah's taos,
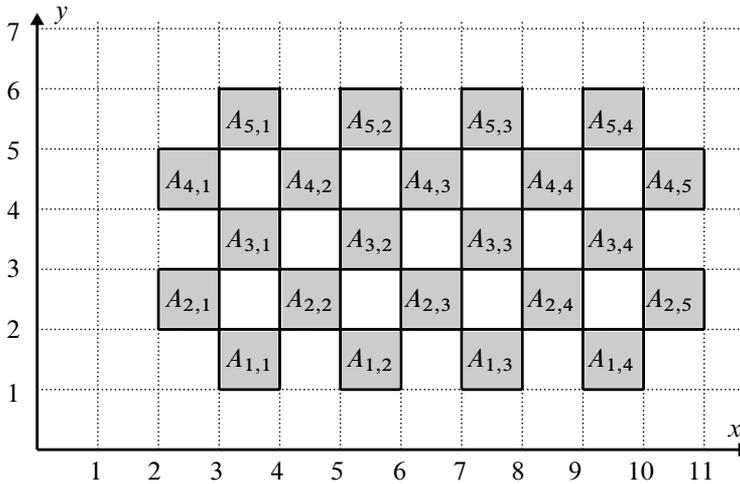
Figure 3: A $(6, 5)$–plat structure.

eyelets, and train tracks and adapt them slightly to our situation. Finally in Section 3.4 we define the concepts of carried and almost carried arcs, loops, and graphs in a manner very similar to that of Johnson and Moriah, differing only in some minor ways that suit our purposes.

3.1 Plat positions

Consider the standard Morse function $h\colon S^3 \to \mathbb{R}$ with exactly one maximum, $+\infty$, and one minimum, $-\infty$. Let $\alpha \subset S^3$ be a strictly increasing arc such that $\partial\alpha = \{\pm\infty\}$. We identify $S^3\backslash\alpha$ with $\mathbb{R}^3$ with Cartesian coordinates (x, y, z) in such a way that the xz–plane lies in $h^{-1}(0)$, and more generally, for each $t \in \mathbb{R}$, the plane $y = t$ lies in $h^{-1}(t)$. We orient our perspective so that the x–axis is horizontal, the y–axis is vertical, and the z–axis points towards the reader. (This allows us to use terms like "up", "down", "left", and "right".) We denote $h^{-1}(t)$ by P_t.

For each $y \in \mathbb{R}$, and $k \in \mathbb{Z}$, let $c_{y,k}$ be the circle of radius $\frac{1}{2}$ in P_y, centered at $x = k + \frac{1}{2}$, $z = 0$. The *plat tube* $A_{i,j}$ is defined to be the annulus

$$A_{i,j} = \begin{cases} \bigcup_{y\in[i,i+1]} c_{y,2j} & \text{if } i \text{ is even,} \\ \bigcup_{y\in[i,i+1]} c_{y,2j+1} & \text{if } i \text{ is odd.} \end{cases}$$

For $n, m \in \{2, 3, 4, \ldots\}$, the (n, m)–*plat structure* is the union of the plat tubes $A_{i,j}$ where i ranges from 1 to $n-1$ and j either ranges from 1 to m or 1 to $m-1$ depending upon whether i is even or odd, respectively.

For $n, m \in \{2, 3, 4, \ldots\}$, an (n, m)–*plat braid* is a union of $2m$ pairwise disjoint arcs in $\mathbb{R}^3$ whose projections to the y–axis are monotonic, satisfying the following properties:

(1) One endpoint of each arc lies in P_1 and the other endpoint lies in P_n.
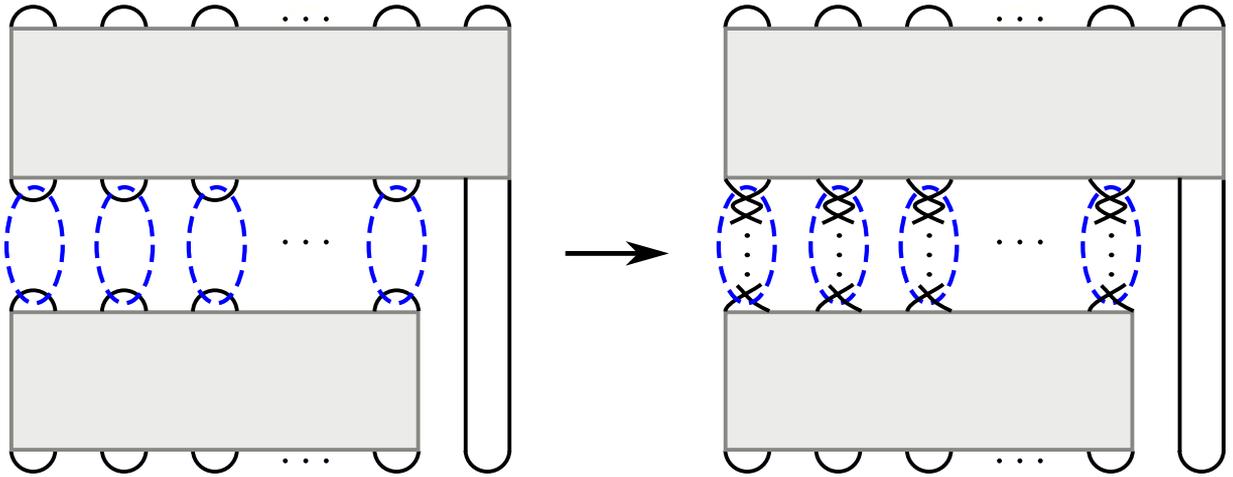
Figure 4: The upper gray box contains D_a, a 4–twisted $(2m-4, m)$–plat braid, where $m \geq 4$. The lower gray box contains D_b a 4–twisted $(n, m-1)$–plat braid.

(2) Each arc can be cut into subarcs, each of which is contained either in the (n, m)–plat structure or in one of the vertical lines $x = 2$ or $x = 2m + 1$ in the xy–plane.

(3) The intersection of the braid with each plat tube consists of a pair of arcs which intersect the plane $z = 0$ in a minimal number of components and whose endpoints lie in $z = 0$.

Observe that the plane $z = 0$ cuts $A_{i,j}$ into two disks. Here we define the *twist number* $a_{i,j}$. If the braid intersects $A_{i,j}$ in vertical arcs, then we define $a_{i,j} = 0$. Otherwise, the disk with nonnegative z–coordinates contains some number of arcs of the plat braid whose projection to the plane $z = 0$ is a set of parallel line segments. We define $|a_{i,j}|$ to be this number of parallel arcs. The sign of $a_{i,j}$ is defined to be the sign of the slope $(\Delta y / \Delta x)$ of the line segments. The integer $a_{i,j}$ is called the *twist number* for $A_{i,j}$.

For our purposes, we will only consider (n, m)–plat braids with n even. In this case, we can obtain a link from an (n, m)–plat braid by first connecting the point $(2j, 1, 0)$ to $(2j + 1, 1, 0)$ for each $1 \leq j \leq m$ with the unique (up to isotopy) arc in the portion of the plane $z = 0$ which lies below the line $y = 1$. Similarly, for each $1 \leq j \leq m$, we also connect the point $(2j, n, 0)$ to the point $(2j + 1, n, 0)$ with the unique arc in the portion of plane $z = 0$ above the line $y = n$. These $2m$ arcs can be isotoped in the plane $z = 0$ (with respect to their endpoints) so that each is injective when projected to the x–axis and each contains either a single maximum or minimum point (with respect to h), with the result that the set of $2m$ arcs is pairwise disjoint. The embedding of a link constructed as the union of the plat braid and these $2m$ arcs in this way is said to be an (n, m)–*plat position of a link*. If a link has an (n, m)–plat position, it is called an (n, m)–*plat link*. A plat link is called k–*twisted* if $|a_{i,j}| \geq k$ for every twist number $a_{i,j}$.

Throughout the rest of the paper, when discussing plat links, $\alpha_+^1 \ldots \alpha_+^m$ will refer to the bridge arcs above P_n, labeled from left to right. Likewise, let $\alpha_-^1 \ldots \alpha_-^m$ be the bridge arcs below P_1, labeled from left to
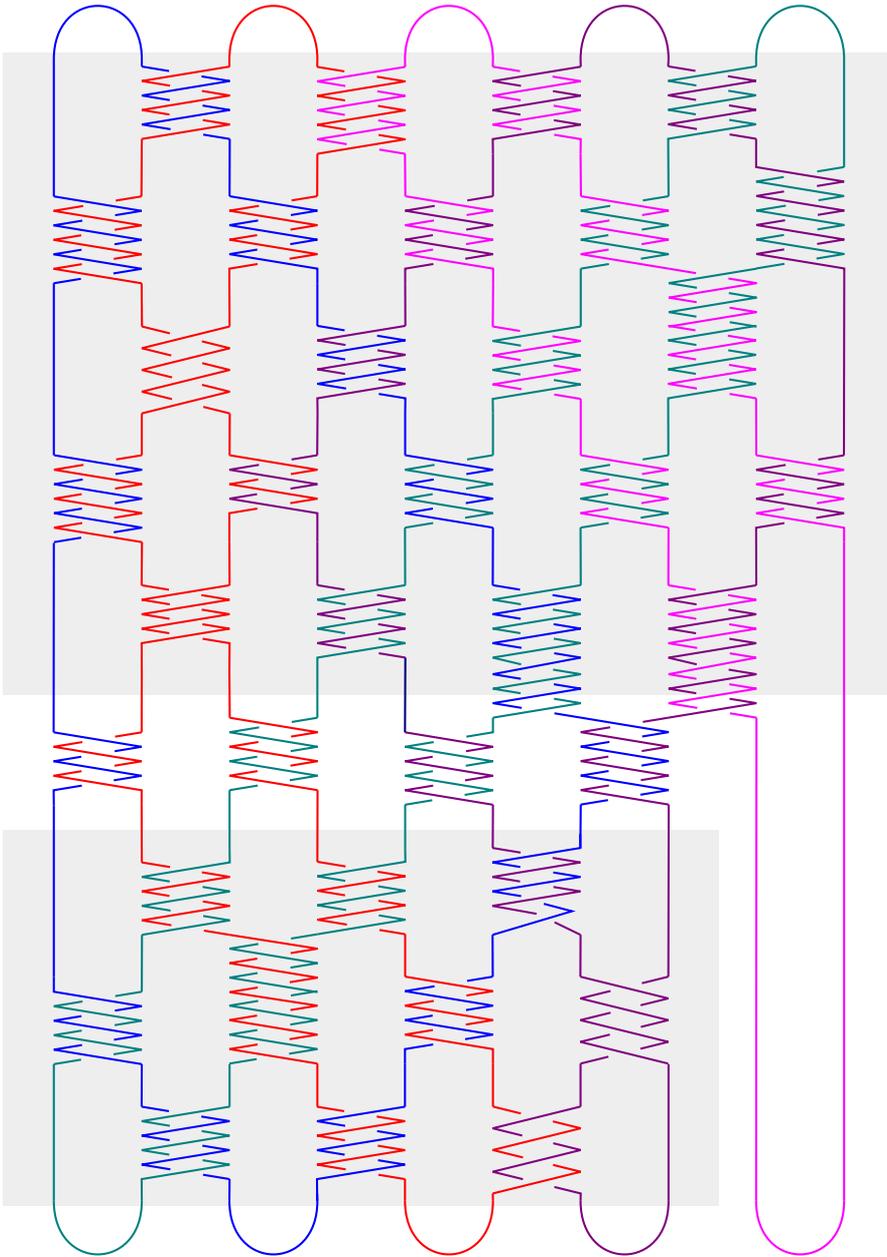
Figure 5: A (10,5)-plat link $L \in \mathscr{L}$.

right. Let L^i be the link component that contains α^i_+. (Of course, since an m–bridge link may have fewer than m components, it may be that the component containing α^i_+ also contains α^j_+ for some $j \neq i$, and so $L^i = L^j$.)

There are two types of projection maps that we will often refer to. The first type of projection map is the Euclidean projection map $\sigma_i \colon \mathbb{R}^3 \to \mathbb{R}^3$ defined by $\sigma_i(x, y, z) = (x, i, z)$. The second type of

projection map is the map $\pi_y \colon \mathbb{R} \times [1, n] \times \mathbb{R} \to P_y$, which sends each component of the plat braid to the corresponding point $(j, y, 0)$, and extends to a homeomorphism from $P_{y'}$ to P_y for each $y' \in [1, n]$. (In a slight abuse of notation, we will refer to this homeomorphism as π_y.)

3.2 The family of links we consider

Let D_a be a 4–twisted (n_a, m)–plat position of a link L_a such that $m \geq 4$ and $n_a = 2m - 4$, and let D_b be a 4–twisted $(n_b, m-1)$–plat position of a link L_b. Position D_a above D_b as shown on the left of Figure 4. Let D_{ab} be an $(n_a + n_b, m)$–plat position obtained from $D_a \sqcup D_b$ by replacing each of the 0–tangles in the dashed ovals with vertical half-twists as shown on the right of Figure 4. (Note: the subscripts a and b are used for "above" and "below".) We define $\mathscr{L}$ to be the family of links constructed in this fashion which have the following additional properties.

(1) The rightmost twist regions of every row alternate in sign from row to row. That is, the sign of the rightmost nonzero twist region of each row is opposite to the sign of the rightmost nonzero twist regions of any adjacent rows.

(2) The sign of $a_{n-1,2}$, the twist number for the second twist region in the top row of D_a, is even, and every other twist region that involves L^3 has an odd twist number. (This forces L^3 to be an unknot component containing the bridge arcs α_+^3 and α_-^m.)

(3) The signs of the twist numbers for the twist regions involving L^m are chosen so that L^m contains the lower left bridge arc α_-^1.

(4) The rest of the twist numbers for D_{ab} are chosen so that D_{ab} is an m–component link and so that the bridges α_+^m and α_-^1 are contained in the same link component, namely L^m. (It follows that for each $i, j \in \{1, \ldots, m\}$, with $i \neq j$, L^i is a distinct link component from L^j.)

(5) Excluding the pair $\{L^1, L^3\}$, every pair of link components comprises a two-bridge nonsplit sublink. (Note: The sublink $L^1 \cup L^3$ will always be split no matter what set of twist numbers is chosen.)

Below in Proposition 3.1, we will show that $\mathscr{L}$ is a nonempty set. First, observe that $\mathscr{L}$ is a family of links in (n, m)–plat position for $m \geq 4$ and $n = n_a + n_b$ with the following conditions on the twist numbers:

(1) For $i > n_b$, $|a_{i,j}| \geq 4$ for all possible values of j.

(2) If i is odd, $1 \leq i \leq n_b$, and $1 \leq j \leq m - 2$, (resp. $j = m - 1$), then $|a_{i,j}| \geq 4$ (resp. $a_{i,j} = 0$).

(3) If i is even, $1 \leq i \leq n_b$, and $1 \leq j \leq m - 1$, (resp. $j = m$), then $|a_{i,j}| \geq 4$ (resp. $a_{i,j} = 0$).

(4) If $a_{i,*}$ denotes the rightmost nonzero twist number in row i, then $a_{i,*} \cdot a_{i-1,*} < 0$ and $a_{i,*} \cdot a_{i+1,*} < 0$. In other words, the signs of the rightmost nonzero twist numbers alternate from row to row.

Figure 5 shows an example of what $L \in \mathscr{L}$ may look like. In this case, $n = 10$, and $m = 5$. It follows from the definition of the family $\mathscr{L}$ that for any $L \in \mathscr{L}$, L^3 is the component containing the lower right bridge arc α_-^m.
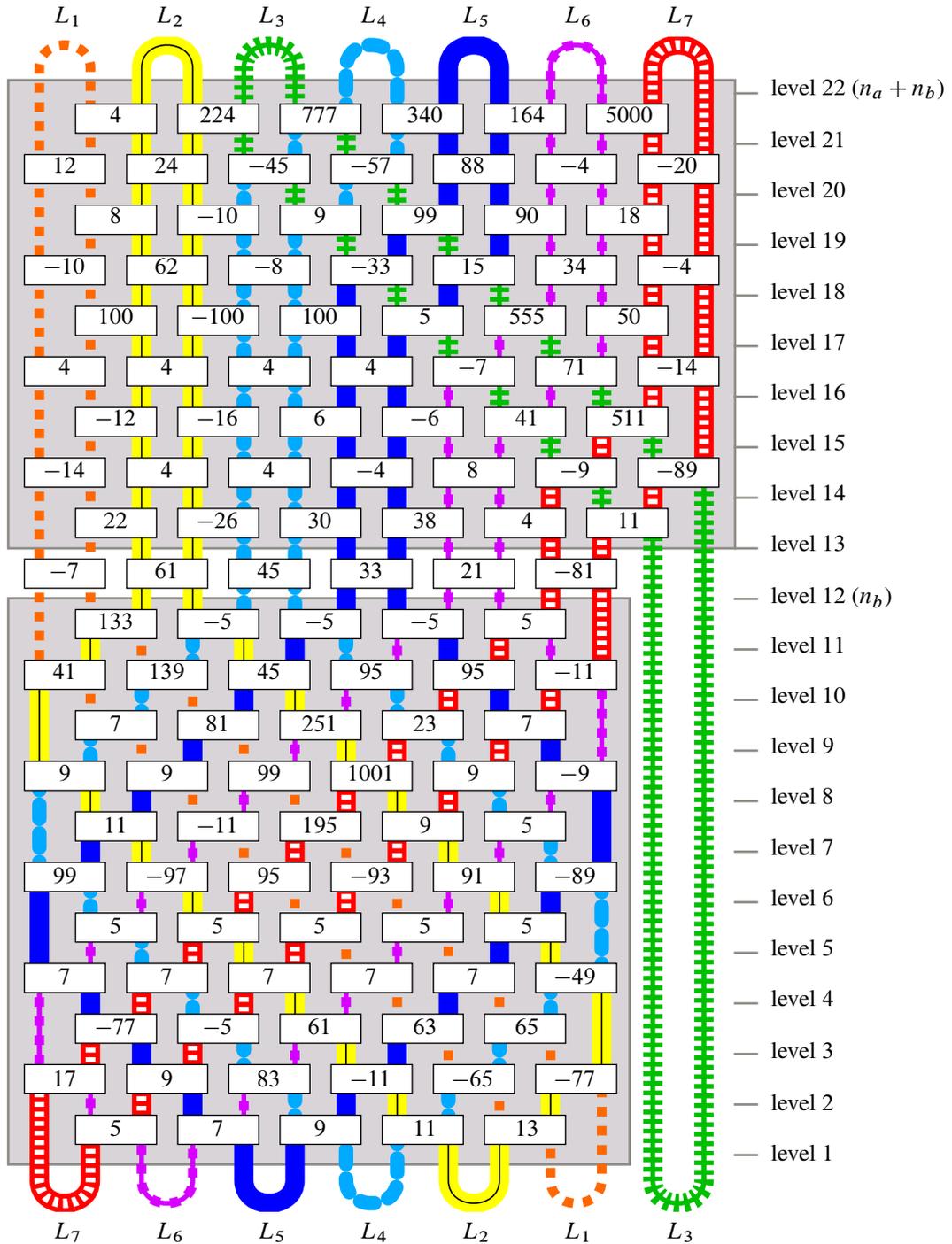
Figure 6: This figure illustrates Proposition 3.1, showing an example a 7–bridge link in $\mathscr{L}$. The seven different colors and line styles represent seven different link components. Each rectangle represents a twist region, and the integer inside each rectangle is the twist number indicating the number and sign of half twists present in that twist region.

Proposition 3.1 *For each integer $m \geq 4$, the family $\mathscr{L}$ contains infinitely many links of bridge number m.*

Proof   It is not difficult to satisfy properties (1) and (2). We need to show that properties (3), (4), and (5) can be satisfied. We will do this by constructing an infinite family of examples for each bridge number $m \geq 4$.

Fix $m \geq 4$. Let D_a and D_b be a $(2m-4, m)$–plat link and an $(2m-2, m-1)$–plat link. We will show that the right choices of the parities and magnitudes of the twist numbers of D_a and D_b will allow us to fulfill the conditions given above.

First, after choosing odd numbers for the particular twist numbers in D_a prescribed by property (2), choose all even numbers for the rest of the twist numbers in D_a. Then choose any integers of magnitude at least four (regardless of parity) for the row of twist regions between D_a and D_b. The twist number choices we have made so far guarantee that at level n_b, the punctures of P_{n_b} occur in pairs corresponding to the link components in this order, from left to right: $L^1, L^2, L^4, L^5, \ldots, L^m, L^3$. That is, they are arranged in numerical order from left to right except that the punctures of L^3 appear at the end of the line.

Then for every twist region below P_{n_b} (ie the twist numbers corresponding to D_b), we choose all odd twist numbers. This guarantees that the punctures of P_1 occur in pairs corresponding from left to right to the link components $L^m, L^{m-1}, \ldots, L^5, L^4, L^2, L^1, L^3$. That is, they area arranged in reverse numerical order, except that again, the punctures of L^3 are at the end of the line. Thus L is an m–component link whose lower left bridge arc is contained in L^m, satisfying conditions (3) and (4).

Since every twist number in D_b is odd, it follows that for each pair $\{L^i, L^j\}$ of distinct link components from the set $\{L^1, L^2, L^4, \ldots, L^m\}$ (the set of all link components excluding L^3), there are exactly four twist regions in D_b which involve both L^i and L^j. The choices of twist numbers in D_a guarantees that there is one twist region containing arcs of both L^2 and L^3, and there are exactly four twist regions containing arcs of both L^3 and L^j for each $j \geq 4$. Now let $\{L^i, L^j\}$ be any pair of link components except for the pair $\{L^1, L^3\}$. To satisfy condition (5), simply choose twist numbers such that the linking number of $L^i \cup L^j$ is nonzero. For example here is one way to do so. There will be some positive number N of twist regions that involve strands from both L^i and L^j. For these twist regions, choose twist numbers $t_1, t_2, \ldots, t_N$ such that $|t_1| > \sum_{k=2}^{N} |t_k|$. $\square$

Proposition 3.2 *Each link in the family $\mathscr{L}$ is nonsplit.*

Proof   Let $L \in \mathscr{L}$, and assume S is a splitting sphere for L.

Case 1   *The link components L^1 and L^2 are both on the same side of S.*

In this case, let L^j be a link component on the other side of S. Then by condition (5) of the definition of $\mathscr{L}$, $L^2 \cup L^j$ is a nonsplit link which is split by S, a contradiction.

Case 2   *The link components L^1 and L^2 are not both on the same side of S.*

Then S is a splitting sphere for the nonsplit link $L^1 \cup L^2$, another contradiction. $\square$

To say that a given compressing disk C is a *cap* is to say that there exists some bridge disk D such that $C = dD$. If α is the bridge arc corresponding to D, then we say that C is a cap *for* α. It follows that ∂C cuts the bridge sphere into two components, one of which is a twice-punctured disk, where the two punctures are the intersection points of the bridge sphere with α.

Proposition 3.3 *Let D and E be compressing disks above and below P, respectively. If $\{D, E\}$ is a weak reducing pair for P, then D is a cap for α_+^1, and E is a cap for α_-^m.*

Proof The loop $\partial D \subset P$ partitions the link components of L into two nonempty sets, A and A' (based on which side of ∂D the punctures of each link component lie). Let A be the set containing L^1. The loop ∂E also partitions the link components into two nonempty sets, B and B'. Let B be the set containing L^3. If A contains L^i for any $i \neq 1$, then $\{D, E\}$ is a weak reducing pair for the sublink $L^i \cup L^3$, a nonsplit 2–bridge link, a contradiction. Similarly, if B contains L^j for any $j \neq 3$, then $\{D, E\}$ is a weak reducing pair for the sublink $L^1 \cup L^j$, a nonsplit 2–bridge link, another contradiction. Therefore D is a cap for α_+^1, the bridge arc above P contained in L^1, and E is a cap for α_-^m, the bridge arc below P contained in L^3. $\quad\square$

The rest of the paper will be devoted to proving that each $L \in \mathscr{L}$ admits a keen weakly reducible bridge sphere. The reason Proposition 3.3 does not immediately imply this is because for any given bridge arc, there are infinitely many distinct caps for that bridge arc, provided there are at least three bridges on each side of the bridge sphere, which is the case for all of the links in $\mathscr{L}$.

3.3 Plat train tracks

Speaking generally, let Σ_L denote a bridge sphere, and let I denote a closed unit interval. A *train track* τ is a compact subsurface of Σ_L whose interior is fibered by open intervals and the fibration extends to a fibration of τ by closed intervals except for at finitely many intervals called *singular fibers*. Let α be a singular fiber, and denote its closed neighborhood in τ by $\overline{N}(\alpha)$. Then there is a homeomorphism $f : \overline{N}(\alpha) \to (I \times I) \backslash \left(\left(\frac{1}{4}, \frac{3}{4}\right) \times \left(\frac{1}{2}, 1\right]\right)$ such that $f(\alpha) = I \times \left\{\frac{1}{2}\right\}$. We will refer to the inverse image of $\left(I \times \left\{\frac{1}{2}\right\}\right) \backslash \left(\left(I \times \left[0, \frac{1}{4}\right]\right) \sqcup \left(I \times \left[\frac{3}{4}, 1\right]\right)\right)$ under f as a *switch* of τ; see Figure 7.
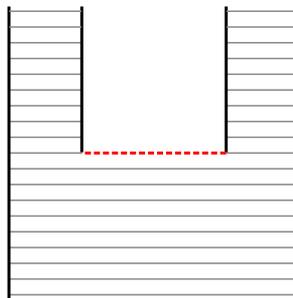
Figure 7: A train track at a singular fiber. The closed red line segment is a switch.
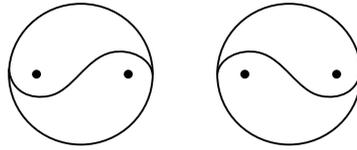
Figure 8: At left, a left-handed tao diagram. At right, a right-handed tao diagram.

In this paper, we will assign a train track τ_i to each bridge sphere P_i for $i = 1, 2, \ldots, n-1$. (There is no need for a train track at the top level P_n.) To this end, we will construct a certain trivalent graph, called a train graph, on each bridge sphere based on the parity of i and the twist numbers $a_{i,j}$ for the row. The train track will then be constructed from the train graph in a natural way.

We define a *train graph* to be a connected trivalent graph with the property that the three edges incident to each vertex are tangent to each other at the vertex, and not all three edges emanate from the vertex in the same direction. (See the left side of Figure 14.) Below, we will construct a specific train graph T_i embedded in P_i for each $i = 1, 2, \ldots, n-1$, and these train graphs will have the property that $P_i \setminus T_i$ consists of $2m$ once-punctured disks and one (nonpunctured) disk. We will informally express this by saying that each puncture is "surrounded by" T_i.

To construct each train graph, there are various cases to consider. Recall from Section 3.2 that $\mathcal{L}$ is a family of links in (n, m)–plat position for $n = n_a + n_b$. If i is odd and $n-1 \geq i \geq n_b + 1$ (resp. $i < n_b + 1$), we define $\ell_{i,j}$ to be the circle in P_i centered at $\left(2j + \frac{3}{2}, i, 0\right)$ with radius $\frac{3}{4}$ for $j = 1, 2, \ldots, m-1$ (resp. for $j = 1, 2, \ldots, m-2$). If i is even and $n-2 \geq i \geq n_b + 1$ (resp. $i < n_b + 1$), we define $\ell_{i,j}$ to be the circle in P_i centered at $\left(2j + \frac{1}{2}, i, 0\right)$ with radius $\frac{3}{4}$ for $j = 1, 2, \ldots, m$ (resp. $j = 1, 2, \ldots, m-1$).

Now, each $\ell_{i,j}$ cuts out a twice-punctured disk from P_i. We will distinguish two types of arcs that separate the two punctures. If $\ell_{i,j}$ is directly below a positive twist region, then we draw a *right-handed tao arc* separating the two punctures as shown on the right of Figure 8. In the case where $\ell_{i,j}$ is directly below a negative twist region, we instead draw a *left-handed tao arc*. The union of a left-handed tao arc (resp. right-handed tao arc) with $\ell_{i,j}$ will be called a *left-handed tao diagram* (resp. *right-handed tao diagram*). An important aspect of these tao diagrams is that at a tao arc's endpoints, the circle and the tao arc are tangent to each other as pictured.
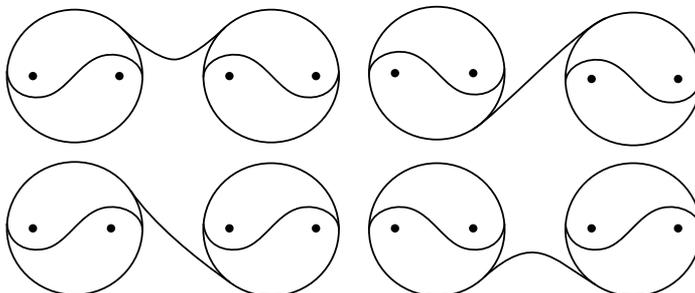
Figure 9: The way we add an edge between two adjacent tao diagrams depends on their handedness.
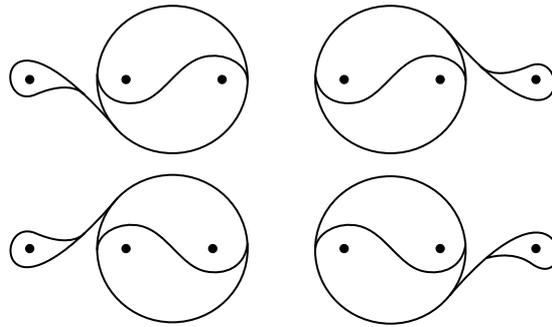
Figure 10: Adding eyelets to "leftover" punctures that are adjacent to a tao diagram.

At this point we have constructed various disconnected tao diagrams in each P_i. We next connect each pair of adjacent tao diagrams with an edge in one of the four ways pictured in Figure 9, depending on the handedness of each tao diagram. If i is even and $i \geq n_b + 1$, the result of this procedure is a train graph which we call T_i.

In all other cases (ie if i is odd and/or $i < n_b + 1$), begin the construction of the train graph as above, combining tao diagrams and connecting edges; however, after doing so, there will be "leftover" punctures that are not surrounded by any tao diagrams. If any such puncture is adjacent to a puncture surrounded by a tao diagram, then we modify our graph according to Figure 10, adding a vertex and two edges to the graph in a way that depends on which side of the tao diagram the puncture is on and the handedness of the tao diagram. The newly added subgraph consists of two edges, one forming a loop around a puncture, and the other connecting the loop to a tao diagram. We refer to such a subgraph as an *eyelet*. If i is odd and $i \geq n_b + 1$, this procedure gives a connected trivalent graph containing two eyelets, surrounding all the punctures. We call this train graph T_i.

For $1 \leq i < n_b + 1$, there are still "leftover" punctures that are not surrounded by a tao diagram or eyelet. Since the sign of the rightmost nonzero twist region of a row is opposite to the sign of the rightmost
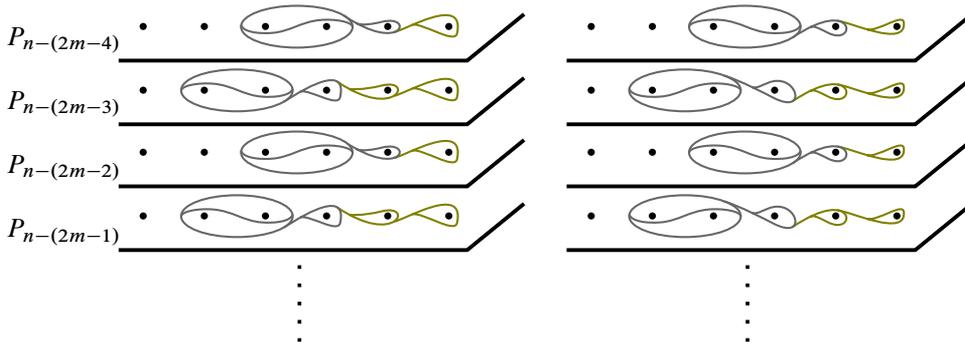
Figure 11: Adding eyelets to "leftover" punctures on P_i for $1 \leq i < n_b + 1$. If the rightmost tao in $P_{n-(2m-4)}$ is left-handed (resp. right-handed), we add eyelets according to the left (resp. right) picture.
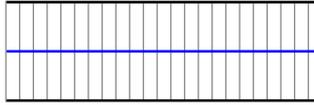
Figure 12: A fibration of $\overline{N}(e)$ by intervals.

nonzero twist region of adjacent rows, there are only two possibilities for the rightmost tao and the eyelet adjacent to it. These are depicted in gray in Figure 11. If $a_{n-1,m-1} < 0$, we add the eyelets according to the left of Figure 11. If $a_{n-1,m-1} > 0$, we add the eyelets according to the right of Figure 11. The newly added eyelets are colored gold. After doing so, we have a train graph that surrounds all the punctures for each P_i for $1 \leq i \leq n - (2m - 4)$, and we call this train graph T_i.

We have constructed a train graph T_i on each bridge sphere P_i. Now we will use each train graph T_i to construct a train track τ_i on each sphere P_i. Let V_i and E_i be the vertex set and the edge set for T_i, respectively. For each vertex $v \in V_i$, let $\overline{N}(v)$ be a closed regular neighborhood of v in P_i.

Let e' denote the connected component of $T_i \setminus \bigcup_{v \in V_i} \overline{N}(v)$ corresponding to the edge e. Let $\overline{N}'(e')$ be a closed regular neighborhood of e' in P_i, and then define $\overline{N}(e) = \overline{N}'(e') \setminus \bigcup_{v \in V_i} \overline{N}(v)$. Notice that $\left(\bigcup_{v \in V_i} \overline{N}(v) \right) \sqcup \left(\bigcup_{e \in E_i} \overline{N}(e) \right)$ is a regular neighborhood of T_i which we call $\overline{N}(T_i)$, and the set $\{ \overline{N}(e), \overline{N}(v) \mid e \in E_i, v \in V_i \}$ is a partition for $\overline{N}(T_i)$.

We fiber each set $\overline{N}(e)$ with interval fibers, each one intersecting e' transversely exactly once as in Figure 12. Then we impose a singular fibration on each $\overline{N}(v)$ containing exactly one singular fiber, as in Figure 13. This makes $\overline{N}(v)$ into a neighborhood of a switch in a train track. The surface $\overline{N}(T_i)$, together with the singular fibration, is a train track which we call τ_i, constructed from the train graph T_i. This construction process is illustrated in Figure 14.

3.4 Carried and almost carried

We want to isotope certain objects in the bridge sphere P_i to a position that behaves nicely with respect to the train track τ_i.

Definition 3.4 For an arc α (not necessarily properly) embedded in P_i, τ_i is said to *almost carry* α if the following are true.
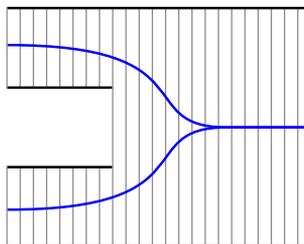
Figure 13: A singular fibration on $\overline{N}(v)$ containing one singular fiber.
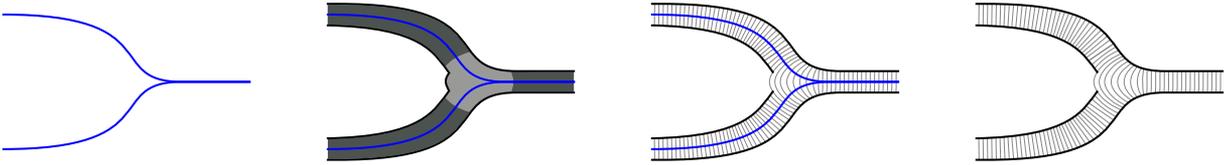
Figure 14: Constructing a train track τ_i from the train graph T_i.

(1) For each point $p \in \alpha$, either $p \notin \tau_i$ or p is a transverse intersection point of α with an interval fiber of τ_i.

(2) No point of α is an endpoint of an interval fiber of τ_i.

(3) No connected component α' of $\alpha \backslash \overset{\circ}{\tau}_i$ is parallel (rel $\partial \alpha'$) into a switch.

(4) No connected component α' of $\alpha \backslash \overset{\circ}{\tau}_i$ is parallel (rel $\partial \alpha'$) into an arc $\alpha'' \subseteq \partial \tau_i$ with the property that α'' is partitioned into three subintervals: the outer two being subintervals of switches and the middle subinterval of α'' being an interval of fiber endpoints of τ_i (see Figure 15).

Remark 3.5 It follows from Definition 3.4 that if τ_i almost carries an arc α, and an endpoint of α lies in $\partial \tau_i$, then that endpoint lies in the interior of a switch.

Remark 3.6 If an arc α satisfies conditions (1), (2), and (3) of Definition 3.4, then each arc of $\alpha \cap (P_i \backslash \overset{\circ}{\tau}_i)$ which is properly embedded in $P_i \backslash \overset{\circ}{\tau}_i$ but which does not satisfy condition (4) can be isotoped into the train track, as illustrated in Figure 15. This results in a position of α which *is* now almost carried.

Definition 3.7 A loop $\ell \subseteq P_i$ is said to be *almost carried* by the train track τ_i if every connected component of $\tau_i \cap \ell$ and every connected component of $(P_i \backslash \overset{\circ}{\tau}_i) \cap \ell$ is an arc which is almost carried by τ_i.

Remark 3.8 An arc or loop in P_i which is completely disjoint from τ_i still satisfies the definition of being almost carried by τ_i.
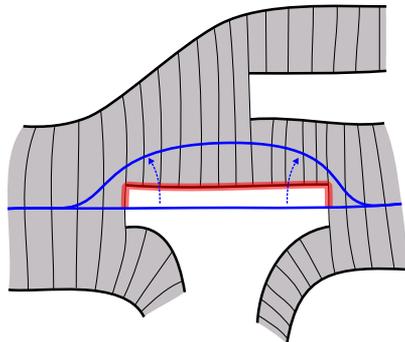
Figure 15: The red shaded arc is α'' from Definition 3.4.
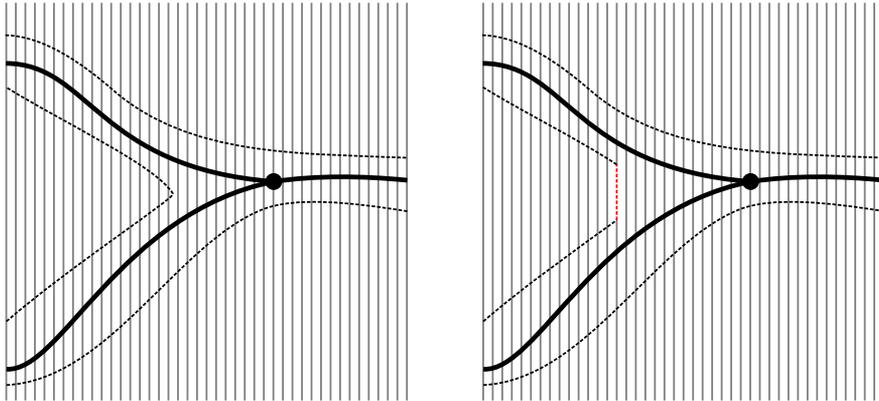
Figure 16: The black graph is T, the vertical gray lines are fibers of σ, and $\overline{N}(T)$, the closed regular neighborhood of T, is outlined with dashed lines. On the right, we see that a slight isotopy of $\overline{N}(T)$ makes it into a train track diagram τ with fibers inherited from σ. That is, every fiber of τ is a subinterval of a fiber of σ.

Definition 3.9 A train graph T is said to be *almost carried* by the train track τ_i if each edge of T is almost carried by τ_i.

As a simple example, for each i, the train graph T_i is almost carried by the train track τ_i.

Definition 3.10 Let T_i' be a subgraph of the train graph T_i (eg a tao), and let $\tau_i' \subseteq \tau_i$ be the sub train track constructed from T_i' following the instruction in Section 3.3. If ℓ is a loop or train graph, then ℓ is said to *cover* T_i' if ℓ is almost carried by τ_i' and ℓ intersects every interval fiber of τ_i'.

Definition 3.11 Let τ and σ be two different train tracks contained in the same bridge sphere P_i. Then σ is said to *almost carry* τ if for each interval fiber I of τ, I is disjoint from σ or I is contained in the interior of some interval fiber of σ.

Proposition 3.12 *Let T be a train graph in the bridge sphere P, corresponding to train track diagram τ, and let σ be another train track diagram in P. If T is almost carried by σ, then τ is almost carried by σ. Furthermore, if p is a point in T which lies in the interval fiber $I \subseteq \tau$, and p also lies in the interval fiber $J \subseteq \sigma$, then $I \subseteq J$.*

Proof Our strategy here is to reexamine the construction of τ and see that it has the desired properties. Let T be a train graph in P, almost carried by σ, and let $\overline{N}(T)$ be a closed regular neighborhood of T.

By definition, every point of T is either disjoint from σ or lies in the interior of a fiber interval of σ. It follows that $\overline{N}(T)$ is disjoint from the fiber endpoints of σ.

Near each vertex v of T which lies in σ, we perform a slight isotopy of $\overline{N}(T)$ (pictured in Figure 16) as follows. We locate an arc λ of $\partial\overline{N}(T)$ located between the two edges of T which emanate from v in the same direction. We isotope $\overline{N}(T)$ so that λ is a subinterval of one of the fiber intervals of σ.

Next we allow the portion of $\overline{N}(T)$ which lies inside σ to inherit a fibering from σ this way: if J is a fiber of σ, then $J \cap \overline{N}(T)$ is a (possibly empty) set of fibers of $\overline{N}(T)$.

After extending this fibration to the rest of $\overline{N}(T)$ which lies outside of σ, $\overline{N}(T)$, endowed with a fibration, is now a train track diagram τ with the desired properties. $\square$

4 How compressing disks meet train tracks

It is desirable to isotope a simple closed curve to intersect a train track in a way over which we can have some control.

For future convenience, we partition the compressing disks into two disjoint sets. Consider α_+^1, the leftmost bridge arc above P_n. A vertical isotopy of α_+^1 into bridge sphere P_n traces out a bridge disk D_+^1. Let $B = dD_+^1$; that is, B is the cap which is the frontier of a regular neighborhood of D_+^1 in V_+. Similarly, consider α_-^m, the rightmost bridge arc below P_1. The bridge arc α_-^m gives rise to a bridge disk D_-^m and a corresponding caps $B' = dD_-^m$. We will refer to these two isotopy classes of caps as *blue* disks. Compressing disks for P_i that are not blue will be referred to as *red* disks.

Proposition 4.1 [Johnson and Moriah 2016, Lemma 8.4] *If D is a compressing disk above P_n, then $\pi_{n-1}(\partial D)$ covers at least one tao of τ_{n-1}, and away from those one or more taos, $\pi_{n-1}(\partial D)$ intersects τ_{n-1} in almost carried arcs or in fiber intervals.*

Definition 4.2 A subgraph T_i' of T_i is called a *mini-graph* of T_i if it has the following properties.

(1) T_i' is a union of taos, connecting arcs, and eyelets of the train graph T_i.

(2) Two adjacent taos of T_i are contained in T_i' if the taos' connecting arc is contained in T_i'.

(3) An eyelet $E \subseteq T_i$ is contained in T_i' only if both the tao T nearest to E in T_i and every other eyelet between E and T are also contained in T_i'.

Definition 4.3 Let T_i' be a mini-graph of T_i. The mini-graph *directly below* T_i' is defined to be the unique mini-graph T_{i-1}' of T_{i-1} with the following properties.

(1) If $T \subseteq T_i'$ is a tao or an eyelet and $\sigma_{i-1}(T)$ intersects a tao T' of T_{i-1}, then $T' \subseteq T_{i-1}'$.

(2) If $T \subseteq T_i'$ is a tao and $\sigma_{i-1}(T)$ intersects two taos of T_{i-1}, say T' and T'', then the connecting arc between T' and T'' is contained in T_{i-1}'.

(3) If $T \subseteq T_i'$ is a tao and $\sigma_{i-1}(T)$ intersects an eyelet E' of T_{i-1}, then $E' \subseteq T_{i-1}'$.

(4) If $T \subseteq T_i'$ is an eyelet and $\sigma_{i-1}(T)$ is an eyelet of T_{i-1}, then $\sigma_{i-1}(T)$ is also an eyelet of T_{i-1}'.

Let $i, i-j \in \{1, 2, \ldots, n-1\}$, with $i > i-j$. Let T_i' and T_{i-j}' be mini-graphs of T_i and T_{i-j}, respectively. We say that T_{i-j}' is *below* T_i' if and only if there exists a sequence of mini-graphs $T_i', T_{i-1}', T_{i-2}', \ldots, T_{i-j}'$ such that for $k = 1, 2, \ldots, j$, the mini-graph T_{i-k-1}' is directly below T_{i-k}. Naturally, we will say that a mini-graph T_i' lies (*directly*) *above* a mini-graph T_{i-j}' if and only if T_{i-1}' lies (directly) below T_i'.
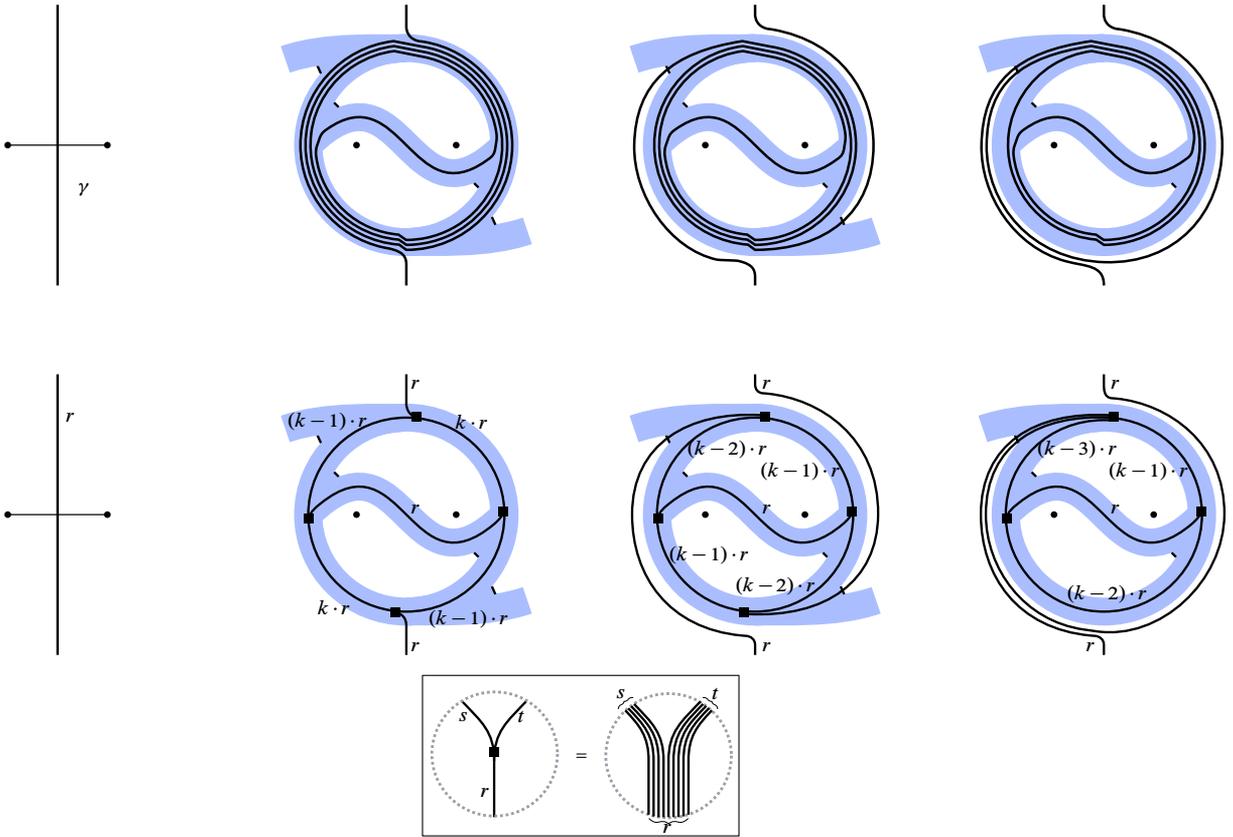
Figure 17: Illustration of Proposition 4.1. In this and following figures, an arc with a label (such as the arc coming out of the bottom of the second picture in the bottom row, marked with an r), represents not just one, but a number of parallel arcs, according to the label. The box at the bottom of the figure illustrates that the small squares placed at various places in this and following figures represents a set of parallel arcs separating into two different sets of parallel arcs. In particular, the square does *not* represent a vertex of a graph.

Observation 4.4 If T_i' and T_i'' are mini-graphs of T_i, then $T_i' \cup T_i''$ is a mini-graph of T_i as well.

Observation 4.5 Suppose that T_i' and T_i'' are mini-graphs of T_i, that T_{i-1}' and T_{i-1}'' are mini-graphs of T_{i-1}, that T_{i-1}' is below T_i', and that T_{i-1}'' is below T_i''. Then $T_{i-1}' \cup T_{i-1}''$ is the mini-graph below $T_i' \cup T_i''$.

Proposition 4.6 *First, if T_{n-1}' is the leftmost tao of T_{n-1}, and if T_1' is the mini-graph of T_1 constructed by excluding from T_1 only the rightmost two eyelets, then T_{n-1}' is above T_1'. Second, if T_{n-1}'' is any other tao of T_{n-1} besides the leftmost tao, then T_{n-1}'' is either above T_1 itself or above some mini-graph T_1'' of T_1 constructed by excluding from T_1 the leftmost eyelet and/or the rightmost eyelet.*

In reading through the following proof, the reader may find it helpful to use the example link in Figure 5 to help locate and visualize the various mini-graphs we discuss.

Proof We will need to define a few more specific mini-graphs. Define T'_{n_b+1} to be the mini-graph constructed from T_{n_b+1} by excluding from it the rightmost eyelet, tao, and connecting arc. Define T'_{n_b} to be the mini-graph of T_{n_b} consisting of all of the taos and connecting arcs of T_{n_b} but excluding both eyelets. Now the reader can observe that by virtue of the dimensions of the link L_a, that is, since $n_a = 2m - 4$, the tao T'_{n-1} is above the mini-graph T'_{n_b+1}. Further, observe that T'_{n_b+1} is directly above T'_{n_b}, which is above T'_1. Therefore, T'_{n-1} is above T'_1.

For the next part of the proof, suppose that T''_{n-1} is a tao of T_{n-1}, but not the leftmost one. Appealing again to the dimensions of L_a, the tao T''_{n-1} is above some mini-graph T''_{n_b+1} of T_{n_b+1} which includes at least all of the taos of T_{n_b+1} but the leftmost one. That is, while T''_{n_b+1} *may* contain the leftmost tao of T_{n_b+1}, T''_{n_b+1} *does* contain all the other taos of T_{n_b+1}. It follows that T''_{n_b+1} is directly above some mini-graph T''_{n_b} of T_{n_b} which contains at least all but the leftmost tao of T_{n_b} and also the first of the two eyelets on the right side of T_{n_b}. Going down another level, T''_{n_b} must be directly above a mini-graph T''_{n_b-1} of T_{n_b-1} which contains all the taos and connecting arcs of T_{n_b-1} as well as the first and second (but not necessarily the third) eyelet on the right.

If $n_b = 2$, then $T_{n_b-1} = T_1$, and we can define $T''_1 = T''_{n_b-1}$, in which case the proof is finished, for the tao T''_{n-1} is above T''_1, which has been shown to have the desired properties.

If $n_b > 2$, then observe that at each level from level $n_b - 1$ down to level 1, T''_{n_b-1} will be above a mini-graph consisting of all of the level's taos and connecting arcs as well as all of the level's eyelets, possibly excluding the leftmost eyelet and/or the rightmost eyelet. Therefore T''_{n_b-1} is above some mini-graph T''_1 with the desired properties, which finishes the proof. □

Definition 4.7 It will be helpful to name a few special types of mini-graphs.

(0) If T is a tao, we will call T a *type* 0 mini-graph.

(1) A *type* 1 mini-graph consists of a final eyelet of T_i and an adjacent tao.

(2) A *type* 2 mini-graph consists of a final eyelet E_2, an eyelet E_1 adjacent to E_2, and a tao adjacent to E_1.

(3) A *type* 3 mini-graph consists of a final eyelet E_3, an eyelet E_2 adjacent to E_3, an eyelet E_1 adjacent to E_2, and a tao adjacent to E_1.

(4) Collectively we will refer to mini-graphs of type 0, 1, 2, or 3 as *typed* mini-graphs.

Observation 4.8 If T'_i is a mini-graph of T_i, then for some positive integer k, T'_i can be decomposed into a union $T'_i = t_1 \cup t_2 \cup \cdots \cup t_k \cup c$, where $t_1, t_2, \ldots, t_k$ are typed mini-graphs and c is a (possibly empty) union of connecting arcs.

Observation 4.9 Suppose T'_i is a mini-graph of T_i, and T'_{i-1} is the mini-graph directly below T'_i. Let T'_i be decomposed into a union $T'_i = t_1 \cup t_2 \cup \cdots \cup t_k \cup c$, where $t_1, t_2, \ldots, t_k$ are typed mini-graphs and c is a union of connecting arcs. For each $j \in \{1, 2, \ldots, k\}$, let u_j be the mini-graph directly below t_j. Then $T'_{i-1} = u_1 \cup u_2 \cup \cdots \cup u_k$.
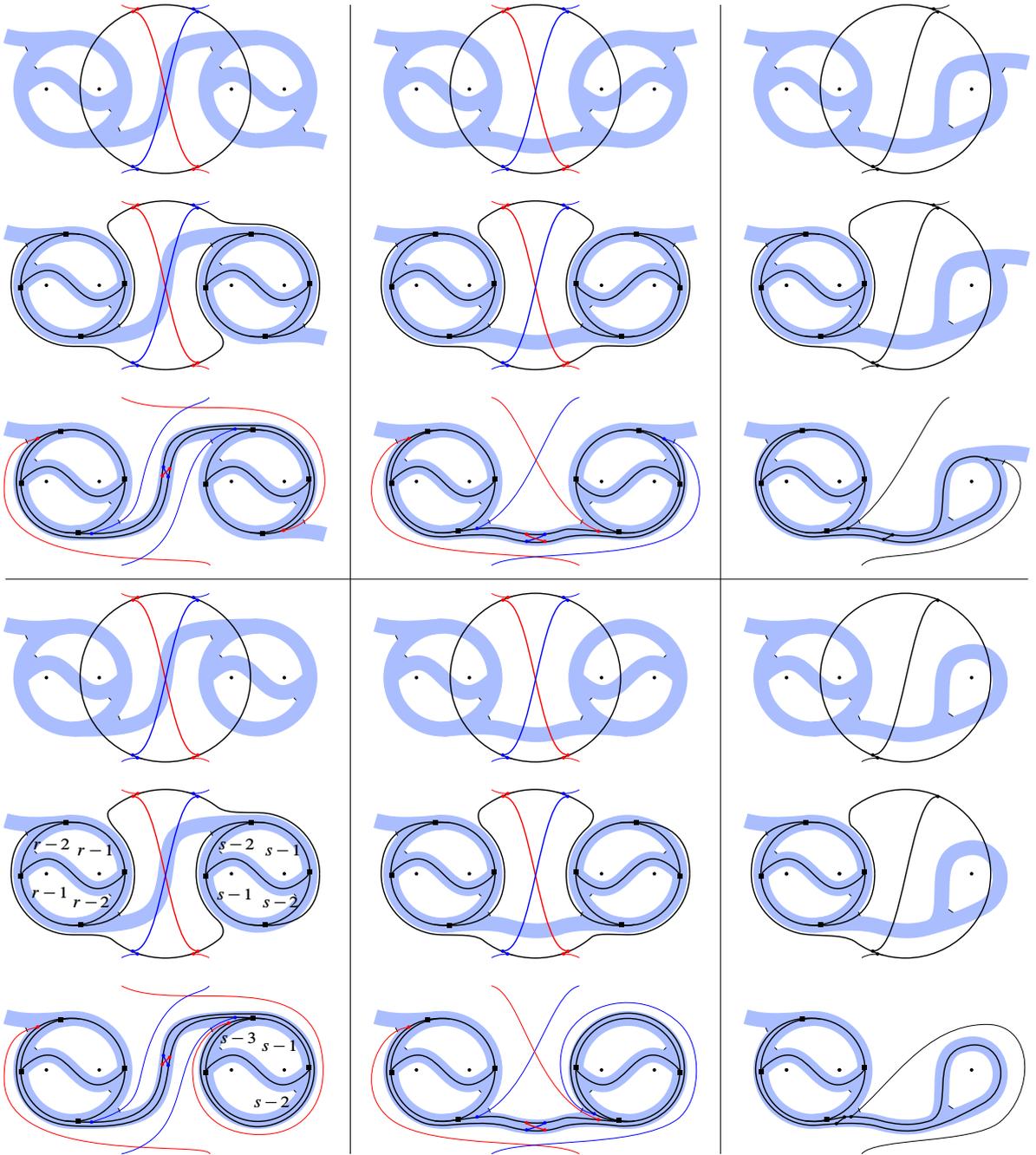
Figure 18: A type 0 mini-graph (ie a tao) covers the mini-graph directly below it. In each picture which includes both red and blue arcs, only either the blue arcs or the red arcs will be present, depending on the handedness of the tao.

Proposition 4.10 For each $i = 2, 3, \ldots, n-1$, if T_i' is a mini-graph of T_i, then $\pi_{i-1}(T_i')$ covers the mini-graph directly below T_i'.
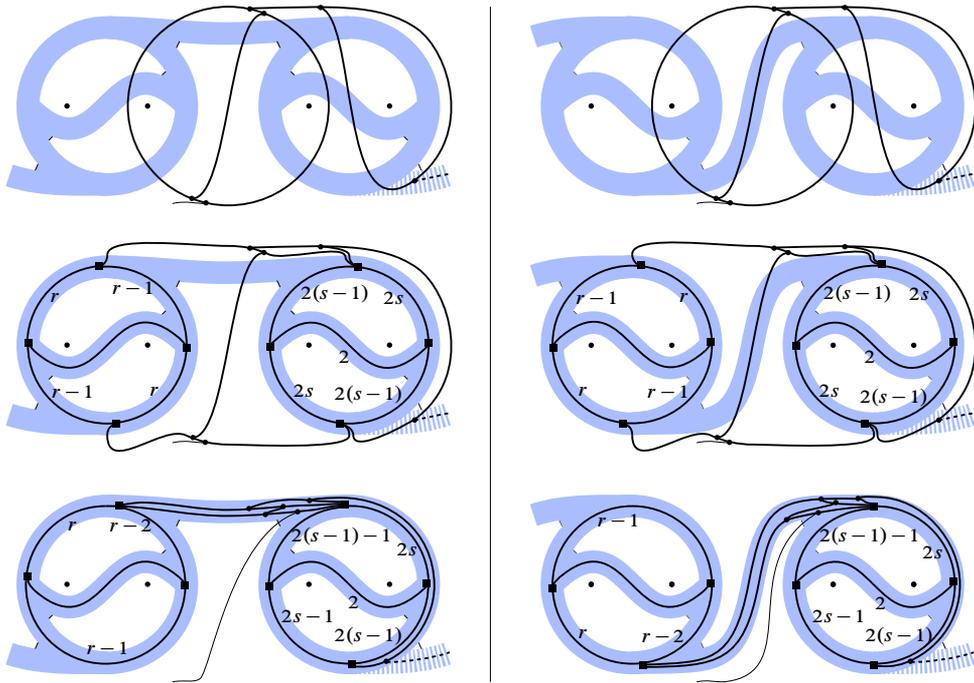
Figure 19: A type 1 mini-graph may lie directly above two taos and their connecting arc. If so, the type 1 mini-graph will cover the two taos and their connecting arc. The dashed lines of the train track diagram and of the train graph are either both present or both absent.
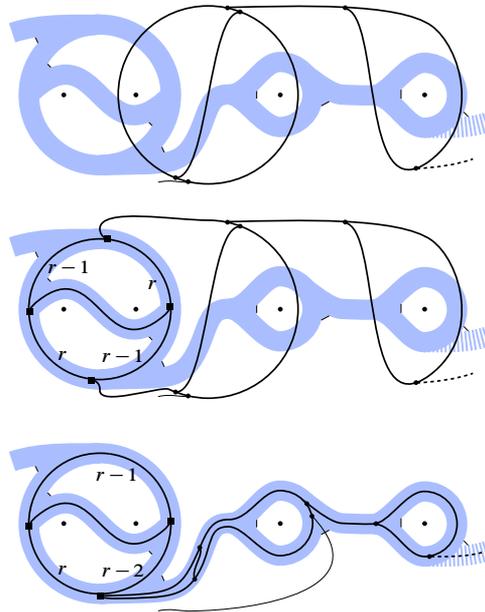
Figure 20: A type 1 mini-graph may lie directly above a type 2 mini-graph, in which case the former will cover the latter.
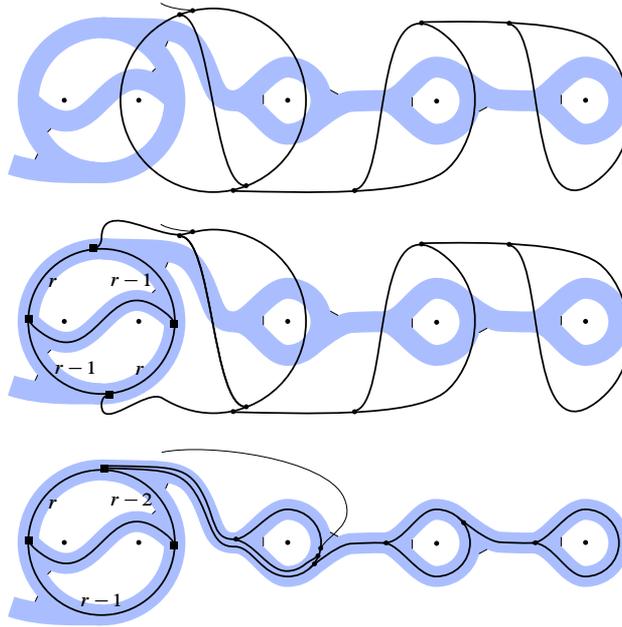
Figure 21: A type 2 mini-graph always lies directly above and covers a type 3 mini-graph.

Proof Fix $2 \le i \le n-1$. Let $T_i' \subseteq T_i$ be a mini-graph, and let T_{i-1}' be the mini-graph directly below T_i'. We will prove this proposition by proving several special cases which will lead us to the general result.

To begin, consider the case in which T_i' is a type 0 mini-graph (a tao). The mini-graph T_{i-1}' may consist of two adjacent taos and their connecting arc, as in the first or second column of pictures in Figure 18, or T_{i-1}' may instead be a type 1 mini-graph, as depicted in the third column of pictures in Figure 18. In any case, the result of the isotopy of the bridge sphere from level i to level $i-1$ is shown from the top row of pictures to the second row, or from the fourth row of pictures to the fifth row.

Observe that an isotopy of $\pi_{i-1}(T_i')$ in P_{i-1} (the result of which is shown in the third and sixth rows of pictures in Figure 18) shows how we may push $\pi_{i-1}(T_i')$ into τ_{i-1} so that $\pi_{i-1}(T_i')$ covers T_{i-1}'.

Next, if T_i' is a type 1 mini-graph, then either T_{i-1}' is a pair of taos (pictured in Figure 19) or T_{i-1}' is a type 2 mini-graph (pictured in Figure 20). Either way, the figures illustrate that $\pi_{i-1}(T_i')$ covers T_{i-1}'.

Now suppose T_i' is a type 2 mini-graph. In this case, T_{i-1}' must be a type 3 mini-graph. Figure 21 depicts this case and shows that $\pi_{i-1}(T_i')$ covers T_{i-1}'.

Finally, suppose T_i' is a type 3 mini-graph. It follows that T_{i-1}' is a type 2 mini-graph, as depicted in Figure 22, which shows that as before, $\pi_{i-1}(T_i')$ covers T_{i-1}'.

Observation 4.11 Notice that in each of the cases above, if c is a connecting arc of T_i which is attached to T_i' at vertex v, then the connected component of $c \cap \tau_{i-1}$ which contains v is almost carried by τ_{i-1}.
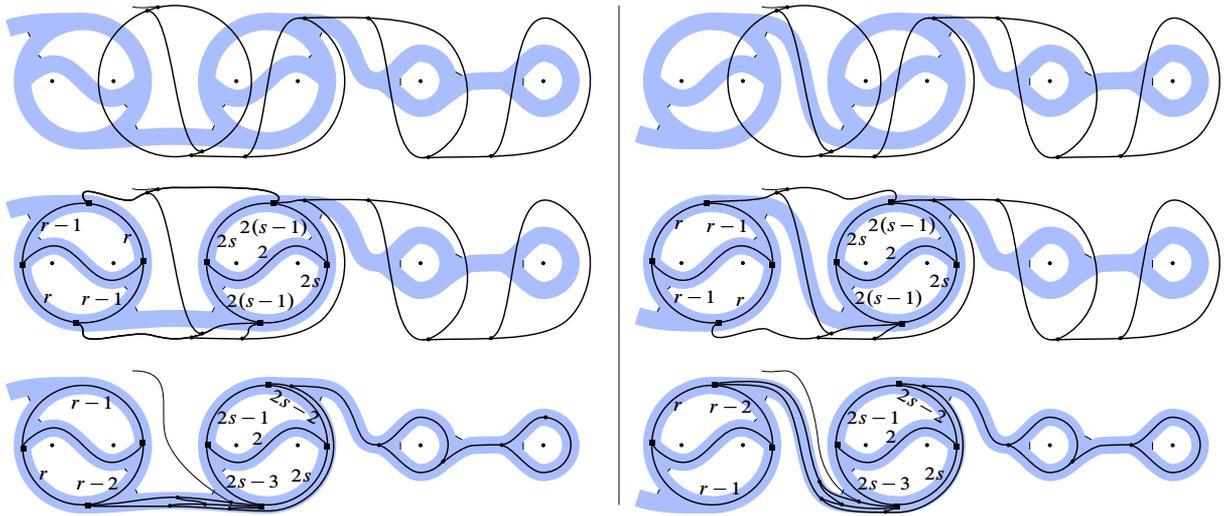
Figure 22: A type 3 mini-graph always lies directly above and covers a type 2 mini-graph (assuming there exists a level below the level of the type 3 minigraph).

Now that we have proven the proposition for cases in which T_i' is a typed mini-graph, we are ready to prove it in the general case where T_i' is an arbitrary mini-graph. According to Observation 4.8, we can view T_i' as a union $T_i' = t_1 \cup t_2 \cup \cdots \cup t_k \cup c$ of typed mini-graphs and connecting arcs. For each $j \in \{1, 2, \ldots, k\}$, define u_j to be the mini-graph of T_{i-1} below t_j. By Observation 4.9, $T_{i-1}' = u_1 \cup u_2 \cup \cdots \cup u_k$. The special cases above demonstrate that for each $j \in \{1, 2, \ldots, k\}$, the mini-graph u_j is covered by t_j, so it follows that $u_1 \cup u_2 \cup \cdots \cup u_k$ is covered by $\pi_{i-1}(t_1 \cup t_2 \cup \cdots \cup t_k)$. Further, if c_0 is one of the connecting arcs of c, then by Observation 4.11, c_0 is also almost carried by τ_{i-1}. Therefore, since $u_1 \cup u_2 \cup \cdots \cup u_k$ is covered by $\pi_{i-1}(t_1 \cup t_2 \cup \cdots \cup t_k)$ and c is almost carried by τ_{i-1}, we can conclude that $u_1 \cup u_2 \cup \cdots \cup u_k$ is covered by $\pi_{i-1}(t_1 \cup t_2 \cup \cdots \cup t_k \cup c)$, or more simply, T_{i-1}' is covered by $\pi_{i-1}(T_i')$. $\qquad \square$

Corollary 4.12 *Let ℓ be a loop which covers a mini-graph $T_i' \subseteq T_i$, and let T_{i-1}' be the mini-graph directly below T_i'. The loop $\pi_{i-1}(\ell)$ covers T_{i-1}'.*

Proof Let J be an interval fiber of τ_{i-1}' that T_{i-1}' intersects. By Proposition 4.10, $\pi_{i-1}(T_i')$ covers T_{i-1}', and so by the definition of covering, J is also intersected by $\pi_{i-1}(T_i')$.

Let p be a point of $(\pi_{i-1}(T_i')) \cap J$, and let I be the interval fiber of $\pi_{i-1}(\tau_i)$ which contains p. By Proposition 3.12, $I \subseteq J$. Further, since ℓ covers T_i', ℓ must by definition intersect I. It follows that since $I \subseteq J$, ℓ intersects J. $\qquad \square$

Corollary 4.13 *Let $i_1 < i_2$, and let $T_{i_1}' \subseteq T_{i_1}$ be the mini-graph below a mini-graph $T_{i_2}' \subseteq T_{i_2}$. If ℓ is a loop which covers $T_{i_2}' \subseteq T_{i_2}$, then $\pi_{i_1}(\ell)$ covers T_{i_1}'.*

Recall the notation of Proposition 4.6. The leftmost blue disk B above P_n is the only disk whose boundary loop covers T'_{n-1} but no other taos. In contrast, the boundary of every red disk above P_n must cover at least one of the other taos. The next corollary then follows from Proposition 4.6 and Corollary 4.13.

Corollary 4.14 *The boundary of the blue disk B above P_n covers T'_1 (the mini-graph defined in Proposition 4.6), and the boundary of every red disk above P covers either T_1 or some mini-graph T''_1 constructed from T_1 by excluding the leftmost and/or the rightmost eyelet of T_1.*

An almost carried loop ℓ that covers enough taos and eyelets is very beneficial in the sense that its presence allows us to predict the behavior of loops which are disjoint from ℓ.

Remark 4.15 The following is [Johnson and Moriah 2016, Lemma 6.5].

Lemma 4.16 *If ℓ is a loop in P_i that covers a mini-graph T''_i of T_i, and if ℓ' is another loop in P_i disjoint from ℓ, then ℓ' can be isotoped to be almost carried by τ''_i, the train track diagram corresponding to T''_i.*

Proof Let $N(\ell)$ be an open regular neighborhood of ℓ disjoint from ℓ'. Since τ''_i is covered by ℓ, every interval fiber of τ''_i intersects $N(\ell)$.

We perform a small isotopy of τ''_i with the following properties:

(1) The image of each interval fiber of τ''_i at each moment of the isotopy is a subinterval of the original interval fiber.

(2) The endpoints of each interval fiber of τ''_i never intersect ℓ throughout the isotopy.

(3) After the isotopy, both endpoints of every interval fiber of τ''_i lie in $N(\ell)$.

The result of this isotopy is illustrated in Figure 23. The point is that each arc of $\partial \tau''_i$ consisting of endpoints of interval fibers gets pushed into $N(\ell)$. Now each component of $\tau''_i \backslash N(\ell)$ is a band in P_i fibered by intervals (each of which is a subinterval of the original interval fibers of τ''_i). The two interval fibers contained in the boundary of a band will be referred to as *exits*. Note that topologically, each of these bands is a closed disk.

Now we isotope ℓ' in $P_i \backslash N(\ell)$ to intersect these bands minimally. Suppose $\ell' \cap \tau''_i = \varnothing$. Then by Remark 3.8, ℓ' is almost carried by τ''_i. If $\ell' \cap (P_i \backslash \mathring{\tau}''_i)$ is empty (that is, ℓ' lies completely in a band), then we can perform an isotopy of ℓ', pushing it out of the band through an exit, contradicting minimality.

Assume then that both $\ell' \cap \tau''_i$ and $\ell' \cap (P_i \backslash \mathring{\tau}''_i)$ are nonempty. Consider a component α of $\ell' \cap (\tau''_i \backslash N(\ell))$. The component α must be an arc properly embedded in a band. Since $\ell' \cap N(\ell)$ is empty, the endpoints of α must lie in exits. Further, the endpoints of α must lie on *different* exits, for otherwise α could be isotoped out of the band, reducing the number of components of $\ell' \cap (\tau''_i \backslash N(\ell))$, again contradicting minimality. Thus α is an arc that travels through a band from one exit to another, so it follows that the
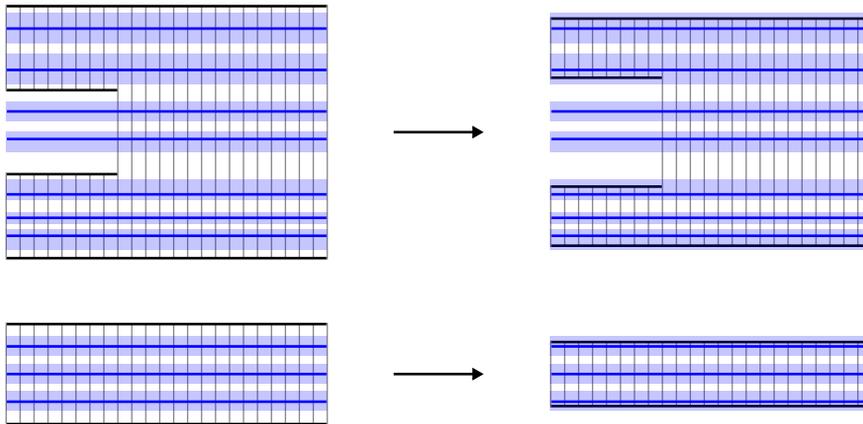
Figure 23: The loop ℓ, shown in blue, covers the train track τ_i''. We perform a small isotopy of τ_i'' which takes each arc of $\partial\tau_i''$ consisting of fiber endpoints into a regular neighborhood $N(\ell)$ of ℓ.

arc α can then be made transverse to each fiber of the band, and thus ℓ' fulfills conditions (1) and (2) of the definition of almost carried. (Note that α also vacuously fulfills condition (3).)

Now let β be a component of $\ell' \cap (P_i \setminus \overset{\circ}{\tau_i}{}'')$. The endpoints of β lie on $\partial\tau_i''$. Since the interval fiber endpoints of τ_i'' all lie in $N(\ell)$ and $\ell' \cap N(\ell)$ is empty, the endpoints of β must lie in exits. Clearly then, β satisfies conditions (1) and (2) of the definition of almost carried. The arc β cannot be parallel in $P_i \setminus N(\ell)$ to a subarc of an exit, for the parallelism would guide an isotopy of ℓ' through a band of $\tau_i'' \setminus N(\ell)$, thereby removing two components of intersection between ℓ' and the bands, once again contradicting minimality. Thus ℓ' fulfills condition (3) of the definition of almost carried.

We have shown that each arc of $\ell \cap \tau_i''$ and each arc of $\ell \cap (P_i \setminus \overset{\circ}{\tau_i}{}'')$ satisfies conditions (1), (2) and (3) of Definition 3.4. Remark 3.6 tells us that each such arc can be isotoped to be almost carried by τ_i''. Therefore ℓ' is by definition almost carried by τ_i''. $\qquad\square$

Henceforth, on P_n, we label the punctures as $p_1, p_2, \ldots, p_{2m}$ in order from left to right. We label the straight arcs connecting the puncture labeled p_{2k-1} to the puncture labeled p_{2k} as β^k. Finally, we label the straight arcs connecting the puncture labeled p_{2k} to the puncture labeled p_{2k+1} as γ^k.

Lemma 4.17 *As above, let T_1' be the mini-graph of T_1 constructed by excluding the rightmost two eyelets, and let τ_1' be the train track diagram corresponding to T_1'. Let T_1'' be a mini-graph of T_1 constructed by possibly excluding the leftmost and/or rightmost eyelet of T_1, and let τ_1'' be the train track diagram corresponding to T_1''. Neither τ_1' nor τ_1'' almost carries the boundary of any red cap for the rightmost bridge arc α_-^m.*

Proof We prove this by contradiction. Let R be a red cap for α_-^m (which implies R is not isotopic to the blue cap $B' = dD_-^m$), and assume ∂R is almost carried by τ_1' or τ_1''. Assume R is in minimal position with respect to the vertical bridge disks below P_1.

We will first establish that R must intersect a lower vertical bridge disk. Let Γ_1 be the unique straight line segment in P_1 which both contains all of the punctures and has p_1 and p_{2m} as endpoints. If ∂R does not intersect Γ_1, then R is trivial, a contradiction. So R must intersect Γ_1. Suppose ∂R passes between the ith and the $(i+1)^{\text{st}}$ vertical bridge disks for $i \in \{1, 2, \ldots, m-3, m-2\}$. (This is equivalent to supposing that ∂R intersects $\sigma_1(\gamma_i)$.) Since $\sigma_1(\gamma_i)$ is surrounded by a tao, and since ∂R is almost carried by the train track τ_1' or τ_1'', the loop ∂R is forced to intersect one of the lower vertical bridge disks. Suppose ∂R passes between the $(m-1)^{\text{st}}$ and m^{th} vertical bridge disk. If ∂R does not also pass between some other pair of bridge disks, then R is the blue disk B', a contradiction. Then ∂R must pass between two of the other bridge disks, and so according to the argument above, ∂R will intersect one of the lower bridge disks. Thus we have established that R is not disjoint from the vertical bridge disks.

Let $\Lambda = R \cap \left(\bigcup_{i=1}^{m} D_-^i \right)$, which is nonempty, as shown above. Since R is in minimal position with respect to the bridge disks, Λ contains no loops of intersection, and so Λ is a collection of arcs. Let $\gamma \subseteq \Lambda$ be an outermost arc in R, cutting off an outermost disk R_{out} from R. Define $q = R_{\text{out}} \cap P_1$. Then q is an arc in P_1 with endpoints on β_i for some i. Let β_* be the subarc of β_i which shares its endpoints with q.

We examine what q can look like (and eventually arrive at a contradiction). First if q never crosses the arc Γ_1, then q would define an isotopy of R through which we could decrease the number of components of Λ, contradicting the fact that R is in minimal position with respect to the lower vertical bridge disks. Since the interior of q is by definition disjoint from the lower vertical bridge disks, q must therefore intersect $\sigma_1(\gamma^k)$ for some k. If q passes between the jth and the $(j+1)^{\text{st}}$ vertical bridge disks for $j \in \{1, 2, \ldots, m-3, m-2\}$, then as above, since q is almost carried and since the point $q \cap \sigma_1(\gamma_k)$ is surrounded by a tao, the train track τ_1' or τ_1'' (whichever is relevant) will force q to intersect two distinct vertical bridge disks, a contradiction. Therefore q must pass between the $(m-1)^{\text{st}}$ and the mth bridge disks.

Suppose that ∂R is almost carried by τ_1''. There are five ways that q, as an almost carried arc, can pass between the $(m-1)^{\text{st}}$ and the mth bridge disks, and they are illustrated in Figure 24. In cases 1 and 2, since q is almost carried, the endpoints of q must lie on both the $(m-1)^{\text{st}}$ and the $(m-2)^{\text{nd}}$ bridge disks, but that contradicts the definition of q, for both endpoints of q must lie on the same vertical bridge disk. Similarly, in cases 4 and 5, the endpoints of q must lie on both the $(m-1)^{\text{st}}$ and the mth bridge disks, which contradicts the definition of q in the same way. Therefore the only case remaining is case 3, in which we see q must enter a switch of τ_1'' and go on to intersect the $(m-1)^{\text{st}}$ bridge disk. By the definition of q, the other endpoint of q must also intersect the $(m-1)^{\text{st}}$ bridge disk on the same side. Thus q has these properties:

- The arc q has both endpoints on the $(m-1)^{\text{st}}$ bridge disk.

- Of the two bands of τ_1'' going through the $(m-1)^{\text{st}}$ bridge disk, an endpoint of q is contained in the rightmost one.
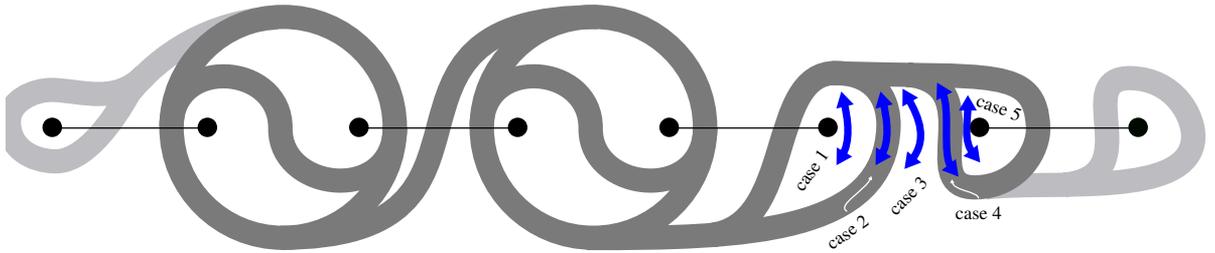
Figure 24: There are only five ways for a loop or arc to pass between the rightmost two bridge arcs while remaining almost carried by τ_1''. The leftmost and rightmost eyelets in this figure are shaded a lighter color to remind the reader that they may or may not be present in τ_1''.

- The arc q is almost carried by τ_1''.
- The interior of q does not intersect any bridge disks below P_1.
- The arc q leaves the $(m-1)^{\text{st}}$ bridge disk in the same direction from both endpoints.

Up to isotopy, there is only one arc that has these properties, and it is depicted in Figure 25. Let a and b be the left and right endpoints, respectively, of q.

The loop $q \cup \beta_*$ cuts P_1 into two punctured disks, one of which contains exactly two punctures: the endpoints of β_m. Call this 2–punctured disk Q (see Figure 25). Observe that ∂R must intersect β_m, or else R would be isotopic to B', contradicting the definition of R. Further, since both endpoints of β_m must be on the same side of ∂R, there must be an even number of points of intersection between β_m and ∂R. It follows that along the interior of β_*, there must be at least four points of intersection with ∂R. Along β_*, let c be the point of $\beta_*^\circ \cap \partial R$ nearest to a, and let d be the point of $\beta_*^\circ \cap \partial R$ nearest to b.

Since R is a cap, R cuts a 2–punctured disk F_R out of P_1 (see Figure 26). Consider the components of $\left(\bigcup_{i=1}^m D_-^i\right) \cap F_R$. There are two components which are arcs that connect a puncture in F_R to ∂R at points we will call x and y. All of the rest of arcs are parallel arcs which separate the two punctures of F_R.

The points x and y cut ∂R into two arcs, one of which must contain both endpoints of q; otherwise q would intersect β_m at point x or at point y, contradicting the definition of q. Now in F_R, a and b are
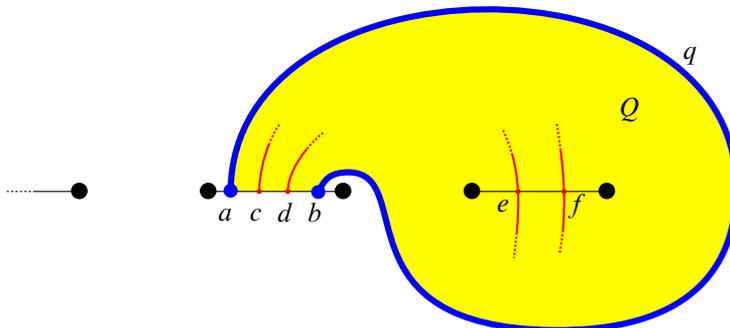
Figure 25: The arc q and the disk Q.
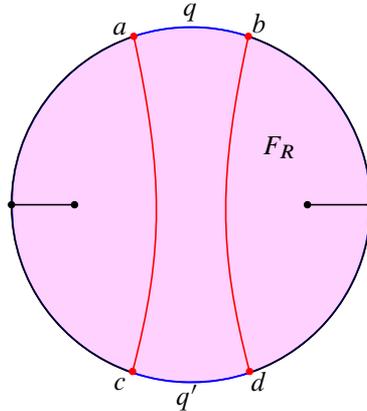
Figure 26: The arc q in relation to the disk F_R. The outer circle is ∂R, and the red vertical arcs are subarcs of β_*.

connected via arcs of ∂R to the points c and d, respectively. These two arcs in F_R with endpoints a, b, c, and d, along with the arc q and another arc of ∂R form a quadrilateral in F_R whose interior is disjoint from $\bigcup_{i=1}^{m} D_{-}^{i}$ (clear from Figure 25). Let the side of this quadrilateral whose endpoints are c and d be called q'.

Now q' is parallel to q. But at this point we could repeat this argument, focusing on q' instead of on q, which would lead us to accept the existence of another parallel arc q'', and we could repeat this infinitely many times, each time obtaining another arc of ∂R with endpoints on β_{m-1}, each of which is nested inside the last one. But this contradicts general position; it cannot be the case that β_{m-1} cuts ∂R into infinitely many subarcs. Therefore we conclude that R cannot be almost carried by τ_1''.

Assume then that ∂R is almost carried by τ_1'. In this case, there are more options for what the arc q may look like, but q still must be an arc with endpoints on a β–arc and with interior disjoint from any β–arcs (see Figure 27). The arc q cannot have its endpoints on β_m, as in Figure 28, because that would force
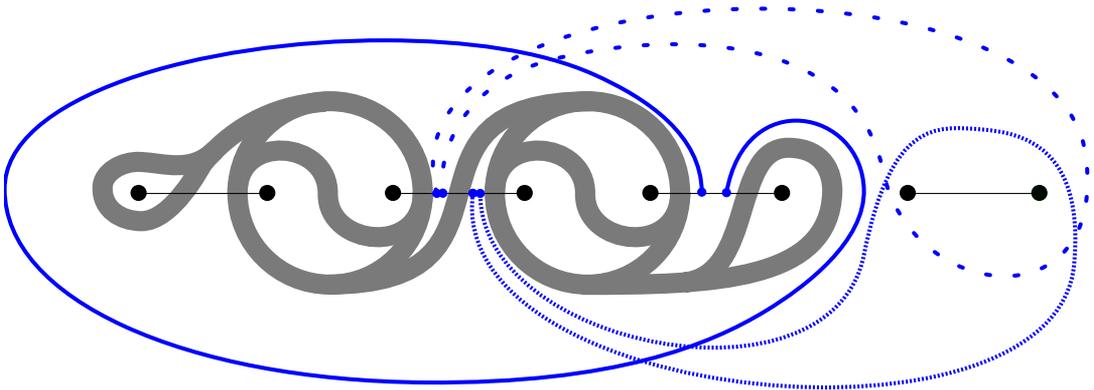
Figure 27: When ∂R is almost carried by τ_1', there are many possibilities for the arc q. Three are illustrated here (each with a different stroke style.)
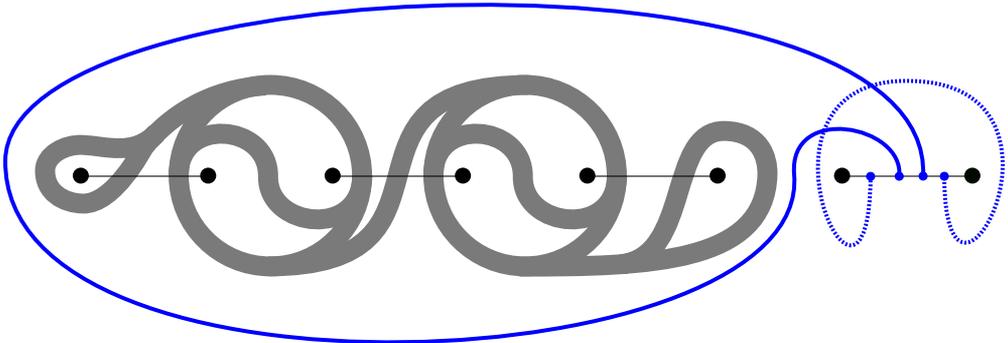
Figure 28: Pictured here are the only nonisotopic possibilities for what q could look like, given that its endpoints lie on β_m. Either way, $\partial R \backslash q$ must contain an arc that cobounds a bigon with β_m, contradicting minimality.

$\partial R \backslash q$ to contain an arc which bounds a bigon with β_m; in other words, ∂R would not be in minimal position with respect to β_m. Then since ∂q lies in β_i for some $i \in \{1, 2, 3, \ldots, m-1\}$, then $q \cup \beta_*$ bounds a disk $Q \subseteq F_R$ which contains β_m. At this point we can apply the same logic as above, leading us to assert the existence of an infinite set $\{q, q', q'', q''', \ldots\}$ of pairwise disjoint arcs of ∂R cut out by β_i, which again contradicts general position. Therefore, we may conclude that ∂R cannot be almost carried by τ_1'' either. $\qquad\square$

Lemma 4.18 *The blue disks B and B' are a weak reducing pair for the bridge sphere.*

Proof Observe that for all points along ∂B, the y–coordinates are all less than 4. (We will speak informally this way even though technically we mean that an isotopy class of ∂B in P_n has the property that all the y–coordinates are less than 4.) Moving down a level, $\pi_{n-1}(\partial B)$ is a loop in P_{n-1} whose y–coordinates are all less than 5. Similarly, $\pi_{n-2}(\partial B)$ is a loop in P_{n-2} whose y–coordinates are all less than 6, and so on. In general, for the levels corresponding to D_a, $\pi_{n-k}(\partial B)$ is a loop in P_{n-k} whose y–coordinates are all less than $4 + k$. Recall that $n = n_a + n_b$, so $n_b + 1 = n - n_a + 1 = n - (n_a - 1)$. Therefore $\pi_{n_b+1}(\partial B) = \pi_{n-(n_a-1)}(\partial B)$ is a loop in $P_{n-(n_a-1)}$ whose y–coordinates are all less than $4 + n_a - 1$. Since the dimensions of D_a were chosen so that $n_a = 2m - 4$, it follows by substitution that the y–coordinates of $\pi_{n_b+1}(\partial B)$ are all less than $4 + (2m-4) - 1 = 2m - 1$. This means that $\pi_{n_b+1}(\partial B)$ is completely to the left of the rightmost two punctures of P_{n_b+1}. Thus $\pi_{n_b+1}(\partial B)$ is disjoint from $\sigma_{n_b+1}(\beta_m)$ (the straight line segment connecting those two punctures).

Consider the π–projections of ∂B at consecutively lower levels. For $1 \leq t \leq n_b$, $\pi_t(\partial B)$ will remain disjoint from $\sigma_t(\beta_m)$ because the isotopy π_t from P_{t+1} to P_t fixes the two rightmost punctures of the bridge sphere pointwise. Observe that $\sigma_1(\beta_m) = \beta_m = D_-^m \cap P_1$. Therefore $\pi_1(\partial B)$ is disjoint from β_m, which implies that $\pi_1(\partial B) \cap \partial B' = \varnothing$, and so $\{B, B\}$ are a weak reducing pair. $\qquad\square$

Observation 4.19 In general, if we compress the cap R for a bridge arc α along a boundary compressing disk Δ, the result will be two disjoint compressing disks whose boundary loops cut the bridge sphere into

two punctured disks and a twice-punctured annulus, the latter of whose punctures correspond to α. This means that neither of the resulting compressing disks are caps for α.

Lemma 4.20 *The blue cap B for α_+^1 intersects all of the red disks below P_1.*

Proof First, by Proposition 3.3, we already know that B intersects all of the red disks below P_1 which are not caps for α_-^m, so it only remains to show that B also intersects all the red caps for α_-^m.

Observe that $\partial B'$ cuts P_1 into a disk F_1 with two punctures and another disk F_2 with $2m - 2$ punctures. By Corollary 4.14, $\pi_1(\partial B)$ covers T_1' (the mini-graph defined in Proposition 4.6 consisting of all of T_1 except the rightmost two eyelets). It follows from Lemma 4.18 that $\pi_1(\partial B)$ is contained in F_2.

Let R be a red cap for α_-^m, and assume by way of contradiction that $\{B, R\}$ is a weak reducing pair. Arrange for R to be in minimal position with respect to the bridge disk D_-^m. An outermost disk Δ on D_-^m cut out by an outermost arc of $D_-^m \cap R$ must be a boundary compressing disk for R, or else D_-^m and R would not be in minimal position. We perform a boundary compression of R along Δ, resulting in a disjoint union $R_1 \sqcup R_2$ of two nonparallel compressing disks for P_1. Since R was disjoint from ∂B, and the boundary compression happened away from ∂B, both R_1 and R_2 are also disjoint from B. Further, by Observation 4.19, neither R_1 nor R_2 are caps for α_-^m, and so we have two weak reducing pairs, $\{B, R_1\}$ and $\{B, R_2\}$ which both contradict Proposition 3.3. $\qquad\square$

Lemma 4.21 *If R_a and R_b are red disks above and below the bridge sphere (respectively), then $\{R_a, R_b\}$ is not a weak reducing pair.*

Proof Assume to the contrary that $\{R_a, R_b\}$ is a weak reducing pair of red disks. By Proposition 3.3, R_a and R_b are caps for α_+^1 and α_-^m, respectively. By Corollary 4.14, the loop $\pi_1(\partial R_a)$ covers either T_1 or some mini-graph T_1'' constructed from T_1 by excluding the leftmost and/or the rightmost eyelet of T_1.

Suppose ∂R_b is disjoint from $\pi_1(\partial R_a)$. Then by Lemma 4.16, ∂R_b is isotopic to an almost carried loop, which contradicts Lemma 4.17. $\qquad\square$

The following lemma is an immediate corollary of [Pongtanapaisan and Rodman 2021, Theorem 5.10] since the upper braid D_a of our link is a (n_a, m) plat link with $n_a = 2m - 4$.

Lemma 4.22 *The cap B' is disjoint from all red disks on the other side of the bridge sphere.*

We have now shown that B and B' are the only weak reducing pair, and so we have proved our main theorem.

Theorem 1.1 *There exist infinitely many links with keen weakly reducible bridge spheres.*

Since a bridge sphere Σ_L of a link L induces a Heegaard surface $\widetilde{\Sigma}_L$ for the 2–fold cover of S^3 branched along L, it is natural to ask whether $\widetilde{\Sigma}_L$ satisfies properties that Σ_L possesses. In our situation, the answer is no.
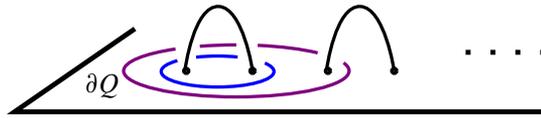
Figure 29: The purple curve bounds a once-punctured disk Q above the bridge sphere and $\partial Q \cap \partial B' = \varnothing$. To visualize Q, imagine a hemisphere shaped disk whose boundary is the purple loop and which is punctured by the second pictured bridge arc.

Proposition 4.23 *Keen weakly reducible bridge spheres in this paper do not lift to keen weakly reducible Heegaard surfaces.*

Proof The blue compressing disk B' is not only disjoint from B, but $\partial B'$ is also disjoint from a curve bounding a once-punctured disk Q above Σ_L (see Figure 29). Such a once-punctured disk lifts to a compressing disk for one of the handlebodies of the Heegaard splitting of the double branched cover. Since B' lifts to a compressing disk in the other handlebody whose boundary is disjoint from lifts of both B and Q, the Heegaard surface $\tilde{\Sigma}_L$ is not keen. $\qquad\square$

5 Nontopologically minimal bridge spheres

One of the main motivations of this article is to search for examples of bridge spheres that are not topologically minimal. The following criterion is needed for our construction of links with nontopologically minimal bridge spheres.

5.1 Cho's criterion

For a link in bridge position, we have that $V_{\pm} \backslash N(\alpha_{\pm})$ is homeomorphic to a handlebody. Therefore, the complex spanned by compressing disks for the bridge sphere for L in $V_{\pm} \backslash N(\alpha_{\pm})$ is a full subcomplex of the disk complex of the handlebody. We recall the following criterion by Cho [2008]:

Theorem 5.1 *If $\mathscr{L}$ is a full subcomplex of the disk complex of the handlebody $\mathscr{K}(V_{\pm} \backslash N(\alpha_{\pm}))$ that satisfies the following condition, then $\mathscr{L}$ is contractible:*

Let D and E be disks representing vertices of $\mathscr{L}$ and suppose that $D \cap E \neq \varnothing$. We assume that D intersects E minimally and transversely. If $\Delta \subset D$ is an outermost subdisk cut off by an outermost arc of $D \cap E$, then at least one of the disks obtained from surgery on E along Δ is also a vertex of $\mathscr{L}$.

Proposition 5.2 *The disk complex of $(V_{\pm}, \alpha_{\pm})$ is contractible*

Proof Suppose that compressing disks D and E in $(V_{\pm}, \alpha_{\pm})$ intersect transversely and minimally. Then the boundary of one of the disks that arises from surgery on E along Δ defined as in Theorem 5.1 must enclose at least two punctures. Otherwise, $D \cap E$ would not be minimal. $\qquad\square$

Using Cho's criterion and Theorem 1.1, we obtain the following corollary.

Corollary 5.3 *There is an infinite family of nontrivial links with bridge spheres that are not topologically minimal.*

Proof Since the bridge sphere P_1 for each link $L \in \mathcal{L}$ contains a unique pair of disjoint compressing disks on opposite sides of P_1, there is exactly one edge connecting the contractible disk complex of (V_+, α_+) to the contractible disk complex of (V_-, α_-) showing that the disk complex of P_1 is contractible.   □

References

[Bachman 2010] **D Bachman**, *Topological index theory for surfaces in 3–manifolds*, Geom. Topol. 14 (2010) 585–609 MR Zbl

[Bachman 2013] **D Bachman**, *Stabilizing and destabilizing Heegaard splittings of sufficiently complicated 3–manifolds*, Math. Ann. 355 (2013) 697–728 MR Zbl

[Bachman and Johnson 2010] **D Bachman**, **J Johnson**, *On the existence of high index topologically minimal surfaces*, Math. Res. Lett. 17 (2010) 389–394 MR Zbl

[Campisi and Rathbun 2018] **M Campisi**, **M Rathbun**, *Hyperbolic manifolds containing high topological index surfaces*, Pacific J. Math. 296 (2018) 305–319 MR Zbl

[Campisi and Torres 2020] **M Campisi**, **L Torres**, *The disk complex and topologically minimal surfaces in the 3–sphere*, J. Knot Theory Ramifications 29 (2020) art. id. 2050092 MR Zbl

[Cho 2008] **S Cho**, *Homeomorphisms of the 3–sphere that preserve a Heegaard splitting of genus two*, Proc. Amer. Math. Soc. 136 (2008) 1113–1123 MR Zbl

[E 2017] **Q E**, *On keen weakly reducible Heegaard splittings*, Topology Appl. 231 (2017) 128–135 MR Zbl

[E and Lei 2014] **Q E**, **F Lei**, *Topologically minimal surfaces versus self-amalgamated Heegaard surfaces*, Sci. China Math. 57 (2014) 2393–2398 MR Zbl

[E and Zhang 2023] **Q E**, **Z Zhang**, *On Heegaard splittings with finitely many pairs of disjoint compression disks*, J. Math. Res. Appl. 43 (2023) 496–504 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

[Johnson 2012] **J Johnson**, *Mapping class groups of Heegaard splittings of surface bundles*, preprint (2012) arXiv 1201.2628

[Johnson and Moriah 2016] **J Johnson**, **Y Moriah**, *Bridge distance and plat projections*, Algebr. Geom. Topol. 16 (2016) 3361–3384 MR Zbl

[Ketover et al. 2019] **D Ketover**, **Y Liokumovich**, **A Song**, *On the existence of minimal Heegaard surfaces*, preprint (2019) arXiv 1911.07161

[Kim 2016] **J Kim**, *A topologically minimal, weakly reducible, unstabilized Heegaard splitting of genus three is critical*, Algebr. Geom. Topol. 16 (2016) 1427–1451 MR Zbl

[Lee 2015] **J H Lee**, *On topologically minimal surfaces of high genus*, Proc. Amer. Math. Soc. 143 (2015) 2725–2730 MR Zbl

[Lee 2016] **J H Lee**, *Bridge spheres for the unknot are topologically minimal*, Pacific J. Math. 282 (2016) 437–443 MR Zbl

[Liang et al. 2018]   **L Liang**, **F Li**, **J Li**, *Reducible handle additions to weakly reducible Heegaard splittings*, Topology Appl. 242 (2018) 33–42   MR Zbl

[McCullough 1991]   **D McCullough**, *Virtually geometrically finite mapping class groups of 3–manifolds*, J. Differential Geom. 33 (1991) 1–65   MR Zbl

[Moriah 2007]   **Y Moriah**, *Heegaard splittings of knot exteriors*, from "Workshop on Heegaard splittings", Geom. Topol. Monogr. 12, Geom. Topol. Publ., Coventry (2007) 191–232   MR Zbl

[Otal 1985]   **J-P Otal**, *Presentations en ponts des nœuds rationnels*, from "Low-dimensional topology", Lond. Math. Soc. Lect. Note Ser. 95, Cambridge Univ. Press (1985) 143–160   MR Zbl

[Ozawa 2011]   **M Ozawa**, *Nonminimal bridge positions of torus knots are stabilized*, Math. Proc. Cambridge Philos. Soc. 151 (2011) 307–317   MR Zbl

[Pitts and Rubinstein 1987]   **J T Pitts**, **J H Rubinstein**, *Applications of minimax to minimal surfaces and the topology of 3–manifolds*, from "Miniconference on geometry and partial differential equations, II", Proc. Centre Math. Anal. Austral. Nat. Univ. 12, Austral. Nat. Univ., Canberra (1987) 137–170   MR Zbl

[Pongtanapaisan and Rodman 2021]   **P Pongtanapaisan**, **D Rodman**, *Critical bridge spheres for links with arbitrarily many bridges*, Rev. Mat. Complut. 34 (2021) 597–614   MR Zbl

[Rodman 2018]   **D Rodman**, *An infinite family of links with critical bridge spheres*, Algebr. Geom. Topol. 18 (2018) 153–186   MR Zbl

[Scharlemann 2005]   **M Scharlemann**, *Thin position in the theory of classical knots*, from "Handbook of knot theory", Elsevier, Amsterdam (2005) 429–459   MR Zbl

[Schultens 2009]   **J Schultens**, *Width complexes for knots and 3–manifolds*, Pacific J. Math. 239 (2009) 135–156   MR Zbl

[Urbano 1990]   **F Urbano**, *Minimal surfaces with low index in the three-dimensional sphere*, Proc. Amer. Math. Soc. 108 (1990) 989–992   MR Zbl

[Zupan 2011]   **A Zupan**, *Properties of knots preserved by cabling*, Comm. Anal. Geom. 19 (2011) 541–562   MR Zbl

Arizona State University
Tempe, AZ, United States

Mathematics Department, Taylor University
Upland, IN, United States

ppongtan@asu.edu,   daniel_rodman@taylor.edu

Upper bounds for the Lagrangian cobordism relation on Legendrian links

JOSHUA M SABLOFF

DAVID SHEA VELA-VICK

C-M MICHAEL WONG

Lagrangian cobordism induces a preorder on the set of Legendrian links in any contact 3–manifold. We show that any finite collection of null-homologous Legendrian links in a contact 3–manifold with a common rotation number has an upper bound with respect to the preorder. In particular, we construct an exact Lagrangian cobordism from each element of the collection to a common Legendrian link. This construction allows us to define a notion of minimal Lagrangian genus between any two null-homologous Legendrian links with a common rotation number.

1 Introduction

The relation $\preceq$ defined by (exact, orientable) Lagrangian cobordism between Legendrian submanifolds in the symplectization of the contact manifold raises a host of surprisingly subtle structural questions. While the Lagrangian cobordism relation is trivially a preorder (ie is reflexive and transitive), it is not symmetric [Baldwin and Sivek 2018; Chantraine 2010; Cornwell et al. 2016]; it is unknown whether the relation is a partial order. Further, not every pair of Legendrians is related by Lagrangian cobordism, with the first obstructions coming from the classical invariants: for links $\Lambda_\pm$ in $\mathbb{R}^3$, if $\Lambda_- \preceq \Lambda_+$ via the Lagrangian $L \subset \mathbb{R} \times \mathbb{R}^3$, then $\mathrm{r}(\Lambda_+) = \mathrm{r}(\Lambda_-)$ and $\mathrm{tb}(\Lambda_+) - \mathrm{tb}(\Lambda_-) = -\chi(L)$ [Chantraine 2010]. A growing toolbox of nonclassical obstructions has been developed to detect this phenomenon; see, just to begin, [Baldwin et al. 2022; Baldwin and Sivek 2018; Ekholm et al. 2016; Golla and Juhász 2019; Pan 2017; Sabloff and Traynor 2013].

If two Legendrians are not related by a Lagrangian cobordism, one may still ask if they have a common upper or lower bound with respect to $\preceq$. Implicit in the work of Boranda, Traynor and Yan [Boranda et al. 2013] is that any finite collection of Legendrian links in the standard contact $\mathbb{R}^3$ with the same rotation number has a lower bound with respect to $\preceq$. In another direction, Lazarev [2020] has shown that any finite collection of formally isotopic Legendrians in a contact $(2n+1)$–manifold with $n \geq 2$ has an upper bound with respect to a moderate generalization of $\preceq$.

The goal of this paper is to find both lower and upper bounds for finite collections of Legendrian links in any contact 3–manifold. On one hand, in contrast to the diagrammatic methods of [Boranda et al. 2013],
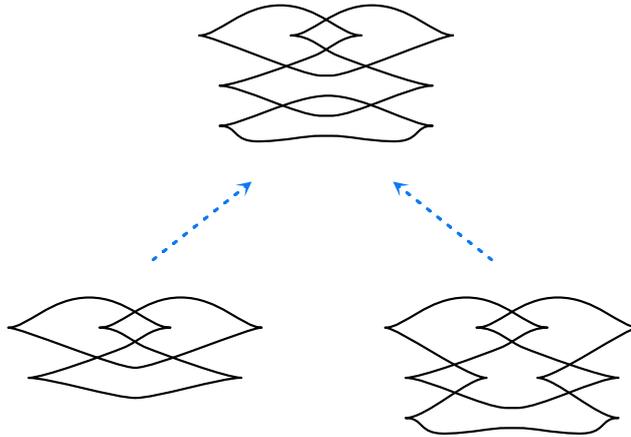
Figure 1: An upper bound for the maximal right-handed trefoil and an $m(5_2)$ knot.

our topological techniques allow us to find lower bounds in any contact 3–manifold, though we also present a refinement of the proof in [Boranda et al. 2013] that better suits our goal of constructing upper bounds. On the other hand, in contrast to Lazarev's use of an h–principle, which restricts his results to higher dimensions, our direct constructions of upper bounds work for Legendrian links in dimension 3.

Theorem 1.1 Let Λ and Λ' be oriented Legendrian links in a contact 3–manifold (Y, ξ), and suppose that there exist Seifert surfaces Σ and Σ' for which $r_{[\Sigma]}(\Lambda) = r_{[\Sigma']}(\Lambda')$. Then there exist oriented Legendrian links $\Lambda_\pm \subset (Y, \xi)$ such that $\Lambda_- \preceq \Lambda \preceq \Lambda_+$ and $\Lambda_- \preceq \Lambda' \preceq \Lambda_+$.

Remark 1.2 For Legendrian links in $\mathbb{R}^3$, the rotation number may be defined without reference to Seifert surfaces, and the hypotheses merely require $r(\Lambda) = r(\Lambda')$.

Remark 1.3 If Λ_- and Λ_+ are connected, then all Lagrangians constructed in the proof of Theorem 1.1 will be connected as well.

Example 1.4 In Figure 1, we display an upper bound for the maximal Legendrian right-handed trefoil and a Legendrian $m(5_2)$ knot. These two Legendrian knots are not related by Lagrangian cobordism. To see why, note that any Lagrangian cobordism between them must be a concordance since they have the same Thurston–Bennequin number, but no such concordance exists even topologically.

Example 1.5 In Figure 2, we display an upper bound for the maximal Legendrian unknot and the maximal Legendrian figure-eight knot. Once again, these two Legendrian knots are not related by Lagrangian cobordism. The fact that the figure-eight has lower Thurston–Bennequin number shows that there cannot be a cobordism from the unknot to the figure-eight; the fact that the figure-eight has two normal rulings shows that there cannot be a cobordism from the figure-eight to the unknot [Cornwell et al. 2016, Theorem 2.7].

In fact, we prove the following strengthened version of Theorem 1.1.
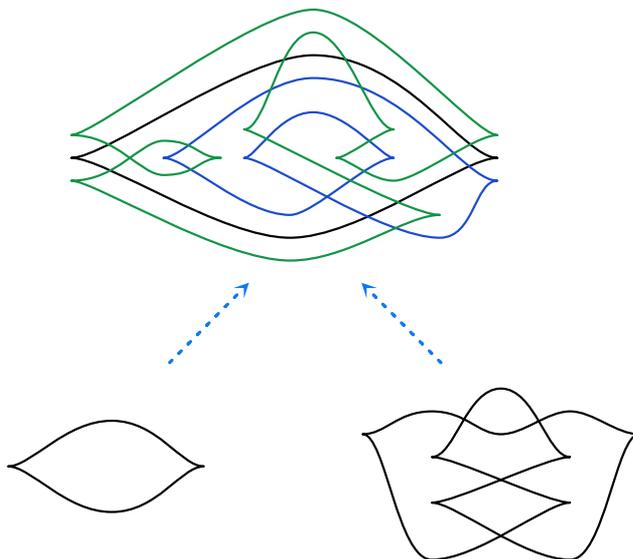
Figure 2: An upper bound for the maximal unknot and the maximal figure-eight knot. The colors in the diagram of the upper bound are only meant to distinguish components of the link to improve readability.

Proposition 1.6 *Under the same hypotheses of* Theorem 1.1, *there exist oriented Legendrian links* $\Lambda_-, \Lambda_+ \subset Y$ *and oriented exact decomposable Lagrangian cobordisms* L *and* L' *from* Λ_- *to* Λ_+, *such that*

- *the Legendrian link* Λ *appears as a collared slice of* L;
- *the Legendrian link* Λ' *appears as a collared slice of* L'; *and*
- L *and* L' *are exact-Lagrangian isotopic.*

Remark 1.7 There are statements analogous to Theorem 1.1 and Proposition 1.6 that hold for unoriented Legendrian links and unoriented (and possibly nonorientable) exact Lagrangian cobordisms, for which there are no requirements on the rotation number.

The main theorem has several interesting consequences. First, we recall that not every Legendrian knot has a Lagrangian filling. The figure-eight knot in Figure 2 is one such example. By transitivity, this implies that not every Legendrian knot lies at the top of a Lagrangian cobordism from a fillable Legendrian. On the other hand, we have the following corollary of the main theorem:

Corollary 1.8 *For any Legendrian link* Λ, *there exists a Legendrian link* Λ_+ *with a Lagrangian filling and a Lagrangian cobordism from* Λ *to* Λ_+.

The proof simply requires us to apply Theorem 1.1 with Λ being the given Legendrian and Λ' being the maximal Legendrian unknot. The upper bound Λ_+ is Lagrangian fillable since there is a cobordism to it from the unknot.

A second consequence of the main theorem is that we are able to define a notion of the minimal genus of a Lagrangian cobordism between *any* two Legendrian links with the same rotation number. Roughly speaking, we define a Lagrangian zigzag-cobordism between Λ and Λ' to be a sequence $\Lambda = \Lambda_0, \Lambda_1, \ldots, \Lambda_n = \Lambda'$ of Legendrian links together with upper (or lower) bounds between each of Λ_i and Λ_{i+1}. The genus of the zigzag-cobordism is the genus of the (smooth) composition of the underlying Lagrangian cobordisms between the Λ_i and their bounds; we may then define $g_L(\Lambda, \Lambda')$ to be the minimal genus of such a Lagrangian zigzag-cobordism. When there is a Lagrangian cobordism from Λ to Λ' and Λ is fillable, $g_L(\Lambda, \Lambda')$ agrees with the relative smooth genus $g_s(\Lambda, \Lambda')$; see Lemma 6.7.

The remainder of the paper is organized as follows. In Section 2, we review key ideas in the definition and construction of Lagrangian cobordisms between Legendrian links. We also define the notion of a Legendrian handle graph, which will form the basis of our later constructions. In Sections 3 and 4, we prove that any two Legendrians in a contact 3–manifold have a lower bound with respect to $\preceq$, and encode the Lagrangian cobordisms involved with Legendrian handle graphs. We present two approaches to this goal: in Section 3, we prove the claim for general contact 3–manifolds using convex surface theory, while in Section 4, we provide a diagrammatic proof in $\mathbb{R}^3$, refining a proof of [Boranda et al. 2013]. We then proceed in Section 5 to prove Proposition 1.6, and hence Theorem 1.1. We end the paper in Section 6 by beginning an exploration of Lagrangian zigzag-cobordisms and their genera, finishing with some open questions.

Acknowledgements

2   A description of Lagrangian cobordisms

In this section, we describe Lagrangian cobordisms, how to construct them, and how to keep track of those constructions.

2.1   Lagrangian cobordisms

We begin with the formal definition of a Lagrangian cobordism between Legendrian links.

Definition 2.1 Let Λ_- and Λ_+ be Legendrian links in the contact manifold (Y, ξ), where $\xi = \ker(\alpha)$ for a contact 1–form α. An (exact, orientable) *Lagrangian cobordism* L from Λ_- to Λ_+ is an exact, orientable, properly embedded Lagrangian submanifold $L \subset (\mathbb{R} \times Y, d(e^t \alpha))$ that satisfies the following:

- there exists $T_+ \in \mathbb{R}$ such that $L \cap ([T_+, \infty) \times Y) = [T_+, \infty) \times \Lambda_+$;

- there exists $T_- < T_+$ such that $L \cap ((-\infty, T_-] \times Y) = (-\infty, T_-] \times \Lambda_-$; and

- the primitive of $(e^t \alpha)|_L$ is constant (rather than locally constant) at each cylindrical end of L.

Note that the last condition enables us to concatenate Lagrangian cobordisms while preserving exactness.

We will use three constructions of Lagrangian cobordisms in this paper, which we will call the *elementary Lagrangian cobordisms*:

- **0–handle** Adding a disjoint, unlinked maximal Legendrian unknot Υ to Λ induces an exact Lagrangian cobordism from Λ to $\Lambda \sqcup \Upsilon$ [Bourgeois et al. 2015; Ekholm et al. 2016].

- **Legendrian isotopy** A Legendrian isotopy from Λ to Λ' induces an exact Lagrangian cobordism from Λ to Λ', though the construction is more complicated than simply taking the trace of the isotopy [Bourgeois et al. 2015; Ekholm et al. 2016; Eliashberg and Gromov 1998].

- **Legendrian ambient surgery** We describe this construction in more detail in Section 2.2, and we will develop a method for keeping track of a set of ambient surgeries in Section 2.3.

2.2 Legendrian ambient surgery

Our next step is to explain Dimitroglou Rizell's [2016] Legendrian ambient surgery construction in the 3–dimensional setting. Similar constructions appear in [Bourgeois et al. 2015; Ekholm et al. 2016], though Dimitroglou Rizell's more flexible language is best suited for our purposes. In dimension 3, Legendrian ambient surgery begins with the data of an oriented Legendrian link $\Lambda \subset (Y, \xi)$ and an embedded Legendrian curve D with endpoints on Λ that is, in a sense to be defined, compatible with the orientation of Λ. The construction then produces a Legendrian Λ_D, contained in an arbitrarily small neighborhood of $\Lambda \cup D$, that is obtained from Λ by ambient surgery along D. Further, the construction produces an exact Lagrangian cobordism from Λ to Λ_D.

More precisely, given $\Lambda \subset (Y, \xi)$ with contact 1–form α, a *surgery disk* is an embedded Legendrian arc $D \subset Y$ such that

(1) $D \cap \Lambda = \partial D$;

(2) the intersection $D \cap \Lambda$ is transverse; and

(3) the vector field $H \subset T_p \Lambda$ defined for all $p \in \partial D$ (up to scaling) by $d\alpha(G, H(p)) > 0$ for all outward-pointing vectors G in $T_p D$ either completely agrees with or completely disagrees with the framing on ∂D induced by the orientation of Λ.
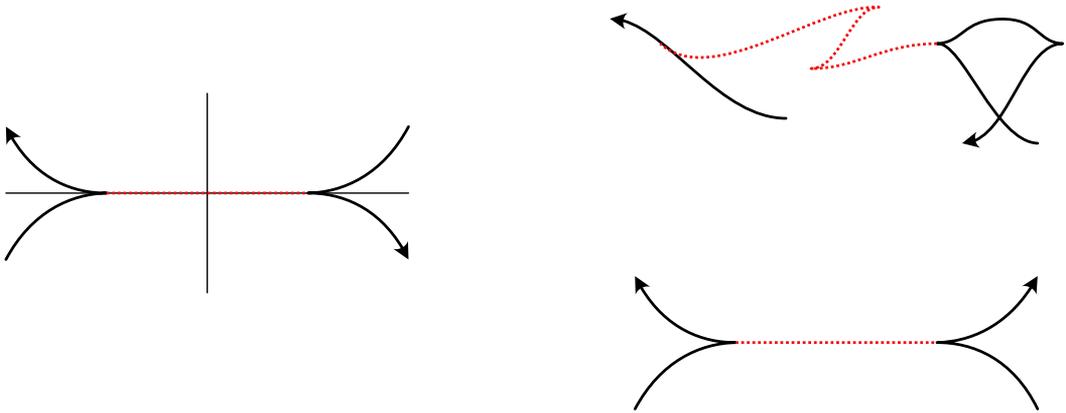
Figure 3: Left: the standard model in $(\mathbb{R}^3, \alpha_0)$ of a surgery disk D_0 with endpoints on a Legendrian Λ_0. Top right: another example of a surgery disk. Bottom right: a disk that fails condition (3).

For an unoriented surgery, we need not specify a framing for ∂D, and the last condition is no longer relevant.

The standard model for such a surgery disk appears in Figure 3, left. In fact, up to an overall orientation reversal on Λ, there is a neighborhood U of D in Y that is contactomorphic to a neighborhood of the standard model for Λ_0 and D_0 [Dimitroglou Rizell 2016, Section 4.4.1]. Working in the standard model, we may replace Λ_0 by the Legendrian arcs Λ_1 as in Figure 4, a process that realizes the ambient surgery on Λ_0 along D_0. Pulling this construction back to the neighborhood of D in Y, we call the resulting link *Legendrian ambient surgery* on Λ along D.

Theorem 2.2 [Dimitroglou Rizell 2016] *Given an oriented Legendrian link Λ and a surgery disk D, let Λ_D be the Legendrian link obtained from Λ by Legendrian ambient surgery along D. Then there exists an exact Lagrangian cobordism from Λ to Λ_D arising from the attachment of a 1–handle to $(-\infty, T] \times \Lambda$.*
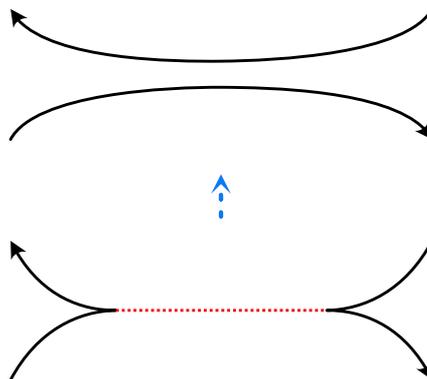
Figure 4: Surgery on the standard model $\Lambda_0 \cup D$ yields a new Legendrian Λ_1.
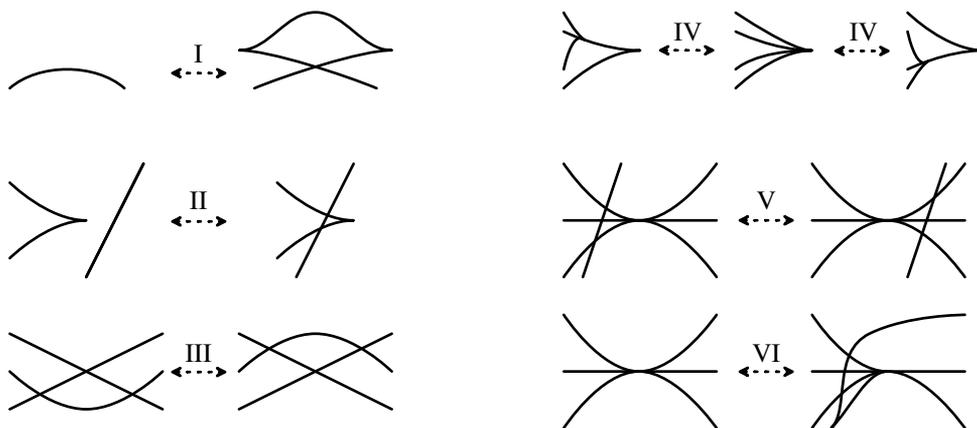
Figure 5: Reidemeister moves for Legendrian graphs in $\mathbb{R}^3$.

Remark 2.3 The construction of Legendrian ambient surgery and the associated Lagrangian cobordism is local. In particular, for a small neighborhood U of D, the surgery construction does not alter $\Lambda \cap (Y \setminus U)$, and the cobordism L outside of $\mathbb{R} \times U$ is cylindrical over $\Lambda \cap (Y \setminus U)$.

2.3 Legendrian handle graphs

In this section, we introduce a structure for keeping track of independent ambient surgeries. We use the notion of a Legendrian graph, following the conventions in [O'Donnol and Pavelescu 2012].

Before we begin, recall from eg [O'Donnol and Pavelescu 2012] that two Legendrian graphs in $(\mathbb{R}^3, \xi_0)$ are Legendrian isotopic if and only if their front diagrams are related by planar isotopy and six Reidemeister moves, as seen in Figure 5.

Definition 2.4 A *Legendrian handle graph* is a pair (G, Λ), where $G \subset (Y, \xi)$ is a trivalent Legendrian graph and $\Lambda \subset (Y, \xi)$ is a Legendrian link (called the *underlying link*), such that

- $\Lambda \subset G$;
- the vertices of G lie on Λ; and
- $G \setminus \Lambda$ is the union of a finite collection of pairwise disjoint Legendrian arcs $\gamma_1, \ldots, \gamma_m$ whose closures satisfy the conditions of surgery disks for Λ.

We also say that G is a *Legendrian handle graph on* Λ. The set of closures of the components of $G \setminus \Lambda$ is denoted by $\mathcal{H}$.

See the bottom of Figure 6 for an example of a Legendrian handle graph whose underlying Legendrian link is a Legendrian Hopf link in $(\mathbb{R}^3, \xi_0)$.

Definition 2.5 Let (G, Λ) be a Legendrian handle graph and let $\mathcal{H}_0$ be a subset of the arcs in $\mathcal{H}$. The *Legendrian ambient surgery* $\mathrm{Surg}(G, \Lambda, \mathcal{H}_0)$ is the Legendrian handle graph (G', Λ') resulting from performing Legendrian ambient surgery along each arc in $\mathcal{H}_0$, as described in Section 2.2.
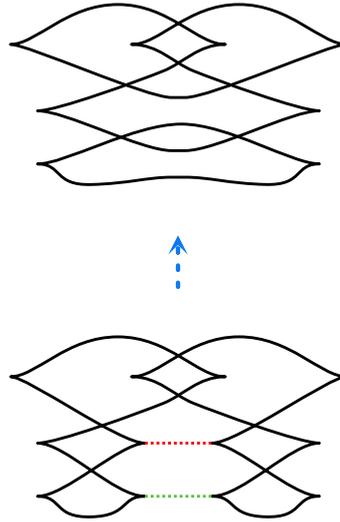
Figure 6: The Legendrian link at the top of the figure is the Legendrian ambient surgery on the Legendrian handle graph (G, Λ) at the bottom.

We will, at times, abuse notation and refer to the underlying Legendrian link Λ' by $\mathrm{Surg}(G, \Lambda, \mathcal{H}_0)$; we will also use $\mathrm{Surg}(G, \Lambda)$ when $\mathcal{H}_0 = \mathcal{H}$. For example, in Figure 6, the Legendrian link at the top is $\mathrm{Surg}(G, \Lambda)$ for the Legendrian handle graph (G, Λ) at the bottom.

By the work of Dimitroglou Rizell [2016] as described in Section 2.2, Legendrian ambient surgery on any given Legendrian arc corresponds to an exact Lagrangian cobordism. This implies that, given an order $\mathfrak{o} = (\gamma_{j_1}, \dots, \gamma_{j_m})$ of the components of $\mathcal{H}_0$, one obtains an exact Lagrangian cobordism $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o})$ from Λ to $\mathrm{Surg}(G, \Lambda, \mathcal{H}_0)$ by performing Legendrian ambient surgery in the order given by $\mathfrak{o}$. The order, in fact, does not matter.

Proposition 2.6 *Suppose (G, Λ) is a Legendrian handle graph, and $\mathfrak{o}_1$ and $\mathfrak{o}_2$ are orders of the components of $\mathcal{H}_0$. The Lagrangian cobordisms $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o}_1)$ and $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o}_2)$ are exact-Lagrangian isotopic.*

Proof It suffices to consider the case where $\mathfrak{o}_1$ and $\mathfrak{o}_2$ differ by an adjacent transposition

$$(\gamma_{j_1}, \gamma_{j_2}) \to (\gamma_{j_2}, \gamma_{j_1}).$$

The cobordism $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o})$ is defined by composing the elementary Lagrangian cobordisms associated to the arcs $\gamma_1, \dots, \gamma_m$. Since there are finitely many of these, by shrinking the neighborhoods of γ_j as in Remark 2.3, we may assume the neighborhoods to be pairwise disjoint. This implies that the elementary Lagrangian cobordisms associated to γ_{j_1} and γ_{j_2} may be constructed simultaneously and shifted past each other along the cylindrical parts of the cobordism. Thus, the parameter given by the relative heights of these two cobordisms gives an exact Lagrangian isotopy between $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o}_1)$ and $L(G, \Lambda, \mathcal{H}_0, \mathfrak{o}_2)$.   □
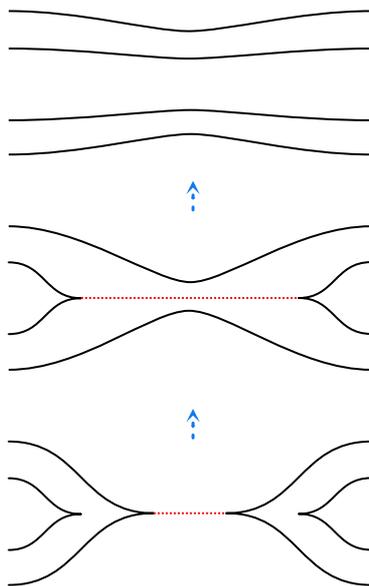
Figure 7: The surgery joining the two inner cusps cannot be performed until after the surgery joining the two outer cusps.

Proposition 2.6 allows us to associate an *isotopy class* $L(G, \Lambda, \mathcal{H}_0)$ of exact Lagrangian cobordisms to a Legendrian handle graph (G, Λ) and $\mathcal{H}_0 \subset \mathcal{H}$.

Remark 2.7 It would be extremely surprising if every decomposable cobordism can be described using a Legendrian handle graph. As shown in Figure 7, one may need to perform one ambient surgery in order for another to be possible; this would violate Proposition 2.6. We emphasize here the particularity of those decomposable cobordisms that can be described by a Legendrian handle graph, as much of Sections 3 and 4 revolves around ensuring the cobordisms we are building belong to this class.

3 Lower bounds via contact topology

In this section, for a pair of Legendrian links with the same rotation number, we construct a pair of exact Lagrangian cobordisms from a common lower bound, encoded by Legendrian handle graphs with the same underlying link.

Proposition 3.1 *Let Λ and Λ' be oriented Legendrian links in a contact manifold (Y, ξ) and suppose that there exist Seifert surfaces Σ and Σ' for which $\mathrm{r}_{[\Sigma]}(\Lambda) = \mathrm{r}_{[\Sigma']}(\Lambda')$. Then there exists an oriented Legendrian link $\Lambda_- \subset (Y, \xi)$ and Legendrian handle graphs G and G' on Λ_- such that $\mathrm{Surg}(G, \Lambda_-)$ (resp. $\mathrm{Surg}(G', \Lambda_-))$ is Legendrian isotopic to Λ (resp. Λ').*

Our proof of Proposition 3.1 relies on convex surface theory applied to the Seifert surfaces Σ and Σ'. To accomplish this, we require two basic results. The first is a lemma extending the work of Boranda, Traynor,
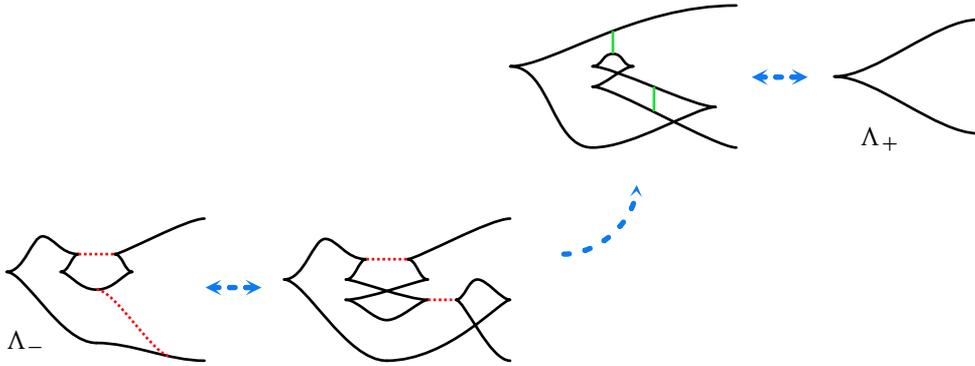
Figure 8: A handle graph giving rise to a Lagrangian cobordism from $S_+ \circ S_-(\Lambda)$ to Λ.

and Yan [Boranda et al. 2013] by placing their result in the context of Legendrian handle graphs. The second translates Dimitroglou Rizell's [2016] Legendrian ambient surgery into a convex surface-theoretic model.

Lemma 3.2 (cf [Boranda et al. 2013, Lemma 3.3]) *Let Λ be an oriented Legendrian link in a contact manifold (Y, ξ), and $S_+ \circ S_-(\Lambda)$ the result of successive negative and positive stabilization on a component of Λ. Then there is a Legendrian handle graph G on $S_+ \circ S_-(\Lambda)$ such that $\mathrm{Surg}(G, S_+ \circ S_-(\Lambda))$ is Legendrian isotopic to Λ.*

Proof The proof is essentially contained in Figure 8, which explicitly identifies a local model for the desired Legendrian handle graph.   □

Our next task is to describe an explicit, convex surface-theoretic local model for Legendrian ambient surgery. In service of this goal, consider the Legendrian graph G depicted in the left part of Figure 9, which will serve as our local model below. The graph G contains three distinguished subsets:

(1) a max-tb unlink Λ of two components, consisting of the two blue arcs and the two black cusps at the two ends;

(2) a dotted red arc D joining the two components of Λ; and

(3) a large, black max-tb unknot Λ'.

Importantly, one can identify Λ' with $\mathrm{Surg}(G, \Lambda)$ in this local model.

The right part of Figure 9 illustrates a convex disk bounded by Λ' and containing the Legendrian graph G as an explicit subset. The key observation is that the above actually yields a convex surface-theoretic local model of the Legendrian ambient surgery operation.

A more general situation is illustrated in Figure 10. On its right, this figure depicts a portion of a convex surface Σ, bounded by a Legendrian link Λ', and containing a Legendrian graph G. The graph G is the union of Λ', the two blue arcs, and the dotted red arc D, and we let Λ be the union of the two blue arcs
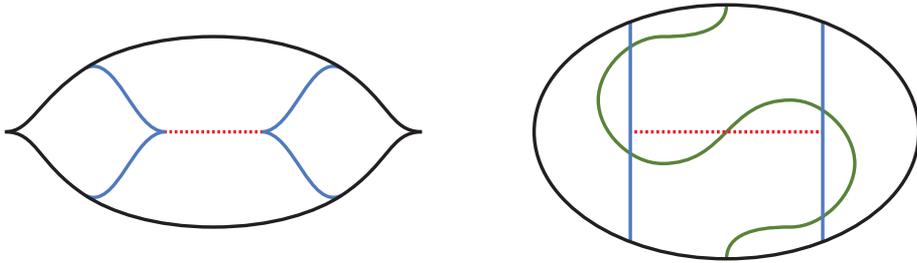
Figure 9: Left: a planar Legendrian graph depicting the Legendrian ambient surgery operation performed on a max-tb unlink. Right: a corresponding convex surface-theoretic interpretation; here, the green arc represents the dividing set.

and Λ', minus the two black boundary arcs between the blue arcs. Then, as in the simpler case above, we have a Legendrian graph G that lies on a convex surface Σ, together with distinguished subsets Λ, D, and Λ', colored blue/black, red, and black respectively. We now claim that this convex surface-theoretic picture corresponds to Legendrian ambient surgery as illustrated on the right of Figure 10.

Lemma 3.3 *Let Σ, G, Λ, Λ', and D be as described in the paragraph above. Then Λ' can be identified with* $\mathrm{Surg}(G, \Lambda, \{D\})$.

Proof This follows immediately from the observation that Legendrian ambient surgery is itself a local operation [Dimitroglou Rizell 2016]. In other words, since Lemma 3.3 is true for a single example — where Λ is a max-tb unlink of two components, D is a trivial arc joining them, and Λ' is a max-tb unknot) — it must be true in general. □

Remark 3.4 While the configuration depicted in Figure 10 provides one possible convex surface-theoretic local model for the Legendrian ambient surgery operation, it is not necessarily unique.

Lemma 3.5 *Let Λ be an oriented, null-homologous Legendrian link in a contact manifold (Y, ξ). Then there exists an oriented Legendrian unknot $\Lambda_U \subset (Y, \xi)$ and a Legendrian handle graph G on Λ_U such that $\mathrm{Surg}(G, \Lambda_U)$ is Legendrian isotopic to Λ.*

Proof Suppose that Σ is a Seifert surface for Λ. Applying Lemma 3.2 to successively double-stabilize each component of Λ if necessary, we obtain a Legendrian handle graph (G_1, Λ_1) and a Seifert surface

Figure 10: A convex surface-theoretic local model for Legendrian ambient surgery. Again, the green arcs represent the dividing set.
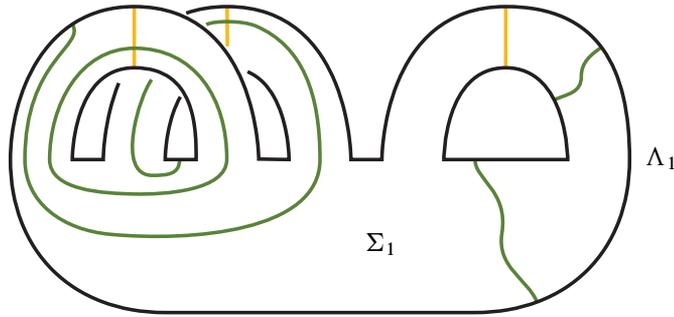
Figure 11: The convex Seifert surface Σ_2, dividing set Γ_Σ, and arc basis, viewed in disk-band form.

Σ_1 for Λ_1 isotopic to Σ, such that the twisting of ξ relative to Σ_1 along each component of $\partial\Sigma_1 = \Lambda_1$ is negative, and $\mathrm{Surg}(G_1, \Lambda_1)$ is Legendrian isotopic to Λ. Below, we will denote this first condition by the shorthand notation $\mathrm{tw}(\xi, \Sigma_1) < 0$, and similarly for other surfaces.

By work of Kanda [1998], since $\mathrm{tw}(\xi, \Sigma_1) < 0$, there is an isotopy of Σ_1 relative to $\partial\Sigma_1 = \Lambda_1$ such that the resulting surface Σ_2 is convex. (While we will not use this, we may assume that this isotopy is a C^0 perturbation near the boundary, followed by a C^∞ perturbation of the interior.) Further, by possibly Legendrian-isotoping the handle arcs of G_1, we obtain a Legendrian handle graph $(G_2, \Lambda_2 = \Lambda_1)$ whose handle arcs $G_2 \setminus \Lambda_2$ intersect Σ_2 transversely in a finite number of points.

To aid the discussion to follow, we picture the convex Seifert surface Σ_2 in disk-band form, meaning that we view it as the union of a 0–handle (disk) and a number g of 1–handles (bands); see Figure 11. Below, we shall fix a particular choice of disk-band decomposition. The cocores $a_1, \dots, a_g$ of the 1–handles form an *arc basis* for Σ_2. (Note that Figure 11 is an abstract diagram of Σ_2; as Σ_2 is embedded in Y, the bands may be "linked".) Since $\mathrm{tw}(\xi, \Sigma_2) < 0$, the dividing set must intersect each component of Λ.

To obtain the desired Legendrian handle graph (G, Λ_U), our strategy is to cut the bands of Σ_2. More precisely, let $\{a_1, \dots, a_g\}$ be an arc basis for Σ_2 consisting of a collection of properly embedded arcs in Σ_2, such that the intersection of each a_i with $G_2 \setminus \Lambda_2$ is empty. Figure 12 depicts a band of Σ_2 and a corresponding basis arc a_i.

Figure 12: A band of the convex Seifert surface Σ_2 and a corresponding basis arc a_i.

Figure 13: A double-stabilization Λ_3 of Λ_2 whose new Seifert surface Σ_3 contains an arc basis disjoint from its dividing set.

Now construct a (not necessarily Legendrian) link $\Lambda_{2.5}$ as follows: Take a parallel push-off of Λ_2 in Σ_2 and, for each basis arc a_i that intersects the dividing set Γ_{Σ_2}, perform a finger move of the pushoff across each of the dividing curves involved and back, as shown in Figure 13. We may choose to perform this finger move from either end of a_i — and it is sufficient to perform it on only one side — and we may choose to turn either left or right on the way back; these choices are immaterial. The goal of the finger move is to obtain a curve that is smoothly isotopic to Λ_2, such that instead of a_i (whose ends are on Λ_2), we may choose a cocore a_i' (with ends on the new curve) that does not intersect the dividing set. Since $a_i \cap (G_2 \setminus \Lambda_2) = \varnothing$ for each i, we may assume that the finger moves avoid all intersection points between $G_2 \setminus \Lambda_2$ and Σ_2.

Recall that the Legendrian realization principle (LeRP) states that, given a convex surface S and a multicurve $C \subset S$ that is transverse to Γ_S, if each component of $S \setminus C$ intersects Γ_S, then S can be isotoped to another convex surface $\phi(S)$ such that $\phi(C)$ is Legendrian (see [Kanda 1998; Honda 2000, Section 3]). Here, since $\mathrm{tw}(\xi, \Sigma_2) < 0$, each component of $\Sigma_2 \setminus \Lambda_{2.5}$ intersects Γ_{Σ_2}, and so we can apply the LeRP to $\Lambda_{2.5} \subset \Sigma_2$ to obtain an isotopy of Σ_2 to a convex surface $\Sigma_{2.5}$ with $\partial \Sigma_{2.5} = \Lambda_2$, such that the image Λ_3 of $\Lambda_{2.5}$ under the isotopy is Legendrian, and $\Gamma_{\Sigma_{2.5}}$ is the image of Γ_{Σ_2}.

A closer look at the proof of [Kanda 1998; Honda 2000, Section 3] reveals that the LeRP in fact applies more generally to graphs $G \subset S$ satisfying the complement condition. We will need this fact towards the end of this proof.

Since the finger moves giving rise to $\Lambda_{2.5}$ — and hence Λ_3 — each involved isotoping across elements of the dividing set an even number of times, we have that the Legendrian link Λ_3 is necessarily an iterated double-stabilization of Λ_2. In fact, Λ_3 is Legendrian isotopic to Λ_2 outside of a tubular neighborhood U_a of the a_i's. We now construct a Legendrian handle graph on Λ_3 as follows: First, extend the Legendrian isotopy between $\Lambda_2 \setminus U_a$ and $\Lambda_3 \setminus U_a$ to a local contact isotopy, and apply the local contact isotopy to the Legendrian handle arcs in $G_2 \setminus \Lambda_2$, obtaining Legendrian handle arcs that are attached to Λ_3. Second, by Lemma 3.2, we may add a collection $\mathcal{H}_3$ of Legendrian handle arcs to this collection to obtain a Legendrian handle graph (G_3, Λ_3) such that $\mathrm{Surg}(G_3, \Lambda_3, \mathcal{H}_3)$ is Legendrian isotopic to (G_2, Λ_2). In particular, this means that $\mathrm{Surg}(G_3, \Lambda_3)$ is Legendrian isotopic to Λ. As before, by possibly Legendrian-isotoping the

Figure 14: The result Λ_4 of iteratively doubly stabilizing Λ_3, whose Seifert surface Σ_4 contains an arc basis, each component of which intersects the dividing set in exactly four points.

handle arcs of G_3, we may assume that the Legendrian handle arcs in (G_3, Λ_3) intersect $\Sigma_{2.5}$ transversely in a finite number of points. (Note that, by Figure 8 in the proof of Lemma 3.2, the Legendrian handle arcs of $\mathcal{H}_3$ can be taken to be contained in an arbitrarily small tubular neighborhood of the a_i's, implying that the complication in Remark 2.7 does not arise, since the Legendrian handle arcs in $\mathcal{H}_3$ are contained in a neighborhood disjoint from $G_{2.5} \setminus \Lambda_3$.)

Let Σ_3 be the closure of the component of $\Sigma_{2.5} \setminus \Lambda_3$ that does not intersect $\partial \Sigma_{2.5}$. Then we have obtained a Legendrian link Λ_3 bounding a convex Seifert surface Σ_3 that contains an arc basis $\{a'_1, \ldots, a'_g\}$ that does not intersect the dividing set Γ_{Σ_3}, and a Legendrian handle graph G_3 on Λ_3 such that $\mathrm{Surg}(G_3, \Lambda_3)$ is Legendrian isotopic to Λ.

We have one final preparatory step before we construct the desired unknot and the accompanying Legendrian handle graph. In this step, we double-stabilize Λ_3 at each point where it intersects the arc basis $\{a'_1, \ldots, a'_g\}$. The result is a Legendrian link Λ_4 bounding a convex Seifert surface Σ_4 whose dividing set differs from that of Σ_3 by a collection of nested pairs of boundary-parallel dividing curves, as shown in Figure 14. We again produce an arc basis $\{a''_1, \ldots, a''_g\}$ which now intersects each of the newly added dividing curves exactly once. As in two paragraphs above, we obtain a Legendrian handle graph G_4 on Λ_4 by applying a local contact isotopy to the Legendrian handle arcs in $G_3 \setminus \Lambda_3$, and then adding the handle arcs required to perform the double-stabilizations.

We are now ready to construct the unknot Λ_U and the desired Legendrian handle graph G on Λ_U. We do this in two steps. First, see Figure 15. Let $G_{4.5}$ be the graph consisting of

(1)   the Legendrian link Λ_4;

(2)   the curves b_i^1 and b_i^2 for $i \in \{1, \ldots, g\}$, which are topologically parallel to the arc basis elements a''_i but have endpoints shifted as in Figure 15; and

(3)   the arcs D_i for $i \in \{1, \ldots, g\}$, each joining b_i^1 to b_i^2 and intersecting the dividing set once.

Since each component of the complement $\Sigma_4 \setminus G_{4.5}$ contains elements of the dividing set, we can apply the LeRP to isotope Σ_4 rel boundary to obtain a convex surface Σ_5 containing $\Lambda_{4.5}$ as a Legendrian graph G.
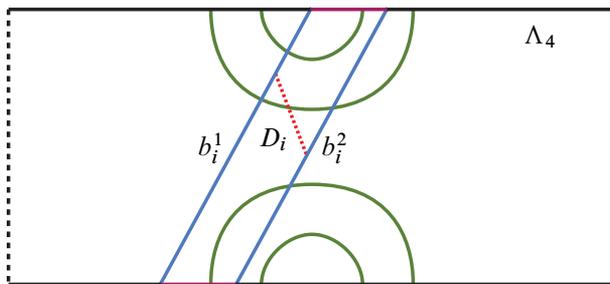
Figure 15: An identification of the convex surface-theoretic local model in Σ_5 for Legendrian ambient surgery.

Finally, we build $\Lambda_U = \Lambda_5$ by taking the segments of Λ_4 in the complement of the short arcs joining the b_i arcs (black in Figure 15) together with the b_i arcs. By construction, this is a Legendrian knot which is topologically trivial.

The key observation is that the Legendrian graph G satisfies the hypotheses of Lemma 3.3 — the "parallelogram" in the center of Figure 15 between the b_i arcs is isotopic to Figure 10. Thus, Legendrian ambient surgery of Λ_U along the collection of arcs $\{D_1, \ldots, D_g\}$ is precisely $\partial \Sigma_4 = \Lambda_4$.

We add to the collection $\{D_1, \ldots, D_g\}$ the handles in G_4 to obtain $G = G_5$. Thus, we obtain a Legendrian handle graph (G, Λ_U) such that, by construction, $\mathrm{Surg}(G, \Lambda_U)$ is Legendrian isotopic to Λ, completing the proof. □

We are now ready to prove the main result of this section.

Proof of Proposition 3.1 According to Lemma 3.5, there are oriented Legendrian unknots Λ_U and Λ'_U and Legendrian handle graphs G and G', such that $\mathrm{Surg}(G, \Lambda_U)$ (resp. $\mathrm{Surg}(G', \Lambda'_U)$) is Legendrian isotopic to Λ (resp. Λ').

Since $\mathrm{r}_{[\Sigma]}(\Lambda) = \mathrm{r}_{[\Sigma']}(\Lambda')$, it follows that $\mathrm{r}(\Lambda_U) = \mathrm{r}(\Lambda'_U)$. This also implies that $\mathrm{tb}(\Lambda_U)$ and $\mathrm{tb}(\Lambda'_U)$ differ by a multiple of 2. Without loss of generality, assume that $\mathrm{tb}(\Lambda_U) \geq \mathrm{tb}(\Lambda'_U)$; then by successively applying Lemma 3.2 to Λ_U if necessary, we obtain a Legendrian handle graph $(\overline{G}, \overline{\Lambda}_U)$ such that $\mathrm{tb}(\overline{\Lambda}_U) = \mathrm{tb}(\Lambda'_U)$ and $\mathrm{Surg}(\overline{G}, \overline{\Lambda}_U)$ is Legendrian isotopic to G. Again, by Figure 8 in the proof of Lemma 3.2, the Legendrian handle arcs of $\overline{G}$ can be taken to be contained in an arbitrarily small neighborhood of a point; thus, we may combine these Legendrian handle arcs with those of G, as in the proof of Lemma 3.5 to obtain a Legendrian handle graph $(\widetilde{G}, \overline{\Lambda}_U)$ such that $\mathrm{Surg}(\widetilde{G}, \overline{\Lambda}_U)$ is Legendrian isotopic to Λ.

Now $\overline{\Lambda}_U$ and Λ'_U are unknots in the contact 3–manifold (Y, ξ) with the same Thurston–Bennequin and rotation numbers. We claim that, possibly after further applying Lemma 3.2 until the Thurston–Bennequin numbers of both unknots are negative, there exists a contact isotopy ϕ_t of (Y, ξ) taking $\overline{\Lambda}_U$ to Λ'_U. If (Y, ξ) is tight, the existence of such an isotopy follows from Eliashberg and Fraser's [2009,
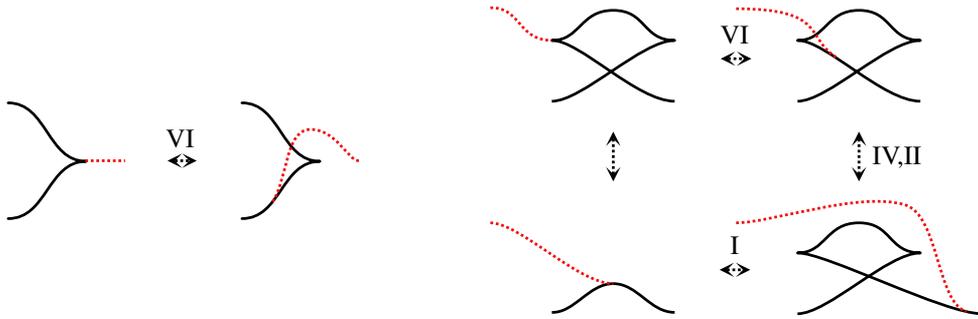
Figure 16: Left: moving a triple point off of a cusp using a Reidemeister VI move from [O'Donnol and Pavelescu 2012]. Right: clearing a cusp of Λ for a Reidemeister I move.

Theorem 1.5] classification of Legendrian unknots in tight contact manifolds. If (Y, ξ) is overtwisted, then our assumption that the Thurston–Bennequin numbers are negative allows us to apply [Eliashberg and Fraser 2009, Proposition 4.12] to find the desired isotopy.

We now apply the isotopy ϕ_t to the Legendrian handle graph $\widetilde{G}$ and perturb the result so that the attached Legendrian handles are disjoint from those of G'. We then obtain a pair of Legendrian handle graphs for the unknot Λ'_U, surgery along which yields Legendrian links isotopic to Λ and Λ' respectively.   □

4 Lower bounds via diagrams

In this section, we reprove Lemma 3.5 for Legendrian links in the standard contact $\mathbb{R}^3$ using diagrammatic techniques rather than convex surface theory. This proof refines that of [Boranda et al. 2013] to produce a handle graph as well as a Lagrangian cobordism from an unknot. While this section is not logically necessary for the proof of Theorem 1.1 given our work in the previous section, the techniques introduced herein are essential for the understanding and practical application of the main ideas of this paper, as justified in Examples 5.1 and 6.3 below.

We begin with a sequence of lemmas that reduce the number of crossings of the front diagram of the Legendrian link in a Legendrian handle graph at the expense of increasing the number of handles. But first, we state a technical general position result.

Lemma 4.1 *For any Legendrian handle graph* (G, Λ)*, there exists a* C^0*–close, isotopic Legendrian handle graph* (G', Λ') *such that all singular points of the front diagram of* G *have distinct* x*–coordinates.*

Proof While this lemma simply expresses general position for the graph G, we note in Figure 16, left, that moving a triple point off of a cusp of Λ is tantamount to using a Reidemeister VI move.   □

First, we remove negative crossings.
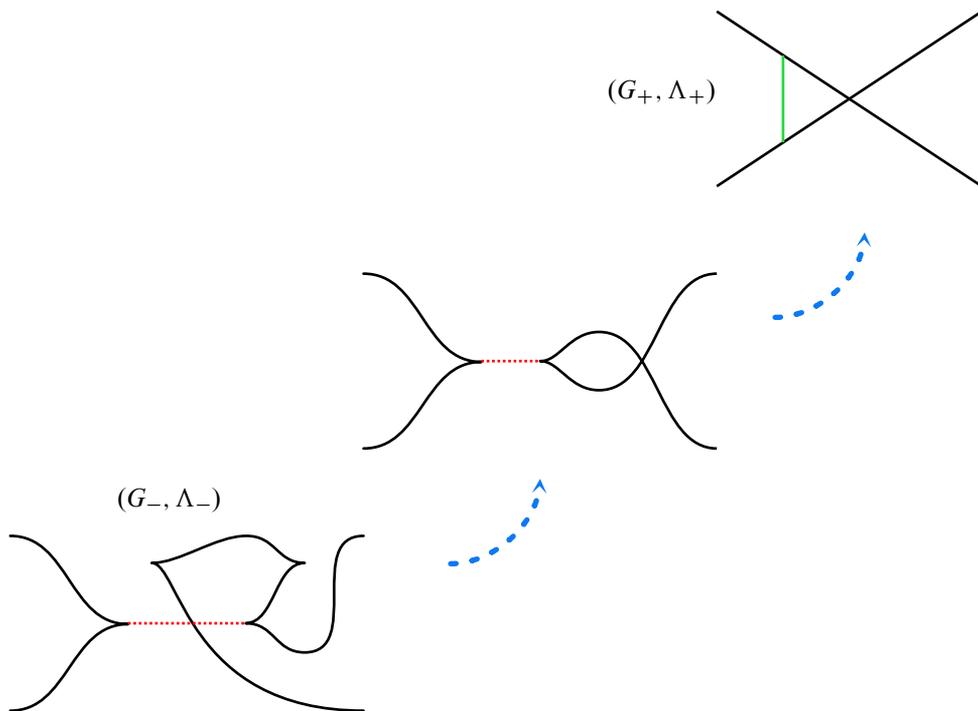
Figure 17: The Legendrian handle graph (G_-, Λ_-) has one fewer negative crossing than (G_+, Λ_+). Red curves represent surgery disks, ie cores of handles, while green curves represent cocores.

Lemma 4.2 *Given a Legendrian link* Λ_+, *whose front diagram has a negative crossing, and a Legendrian handle graph* G_+ *on* Λ_+, *there exists a Legendrian handle graph* (G_-, Λ_-) *and a subset* $\mathscr{H}_0$ *of handles of* G_-, *such that* $\mathrm{Surg}(G_-, \Lambda_-, \mathscr{H}_0)$ *is Legendrian isotopic to* (G_+, Λ_+), *and the front diagram of* Λ_- *has one fewer negative crossing than that of* Λ_+.

After applying Lemma 4.1 to isolate negative crossings, the proof of Lemma 4.2 is contained in Figure 17.

Next, we remove positive crossings.

Lemma 4.3 *Given a Legendrian link* Λ_+, *the leftmost crossing of whose front diagram is positive, and a Legendrian handle graph* G_+ *on* Λ_+, *there exists a Legendrian handle graph* (G_-, Λ_-) *and a subset* $\mathscr{H}_0$ *of handles of* G_-, *such that* $\mathrm{Surg}(G_-, \Lambda_-, \mathscr{H}_0)$ *is Legendrian isotopic to* (G_+, Λ_+) *and the front diagram of* Λ_- *has one fewer positive crossing than that of* Λ_+.

Proof Apply Lemma 4.1 to isolate crossings and cusps of Λ_+ from handles of G_+. Consider the leftmost crossing X_0 of Λ_+. Without loss of generality, we may assume that Λ_+ is oriented from right to left on both strands of X_0. The upper-left strand incident to X_0 must thus next return to X_0; the same is true for the bottom-left strand. Since there are no crossings of Λ_+ to the left of X_0, either the upper left strand must next cross the x–coordinate of X_0 above X_0 or the lower left strand must next cross the
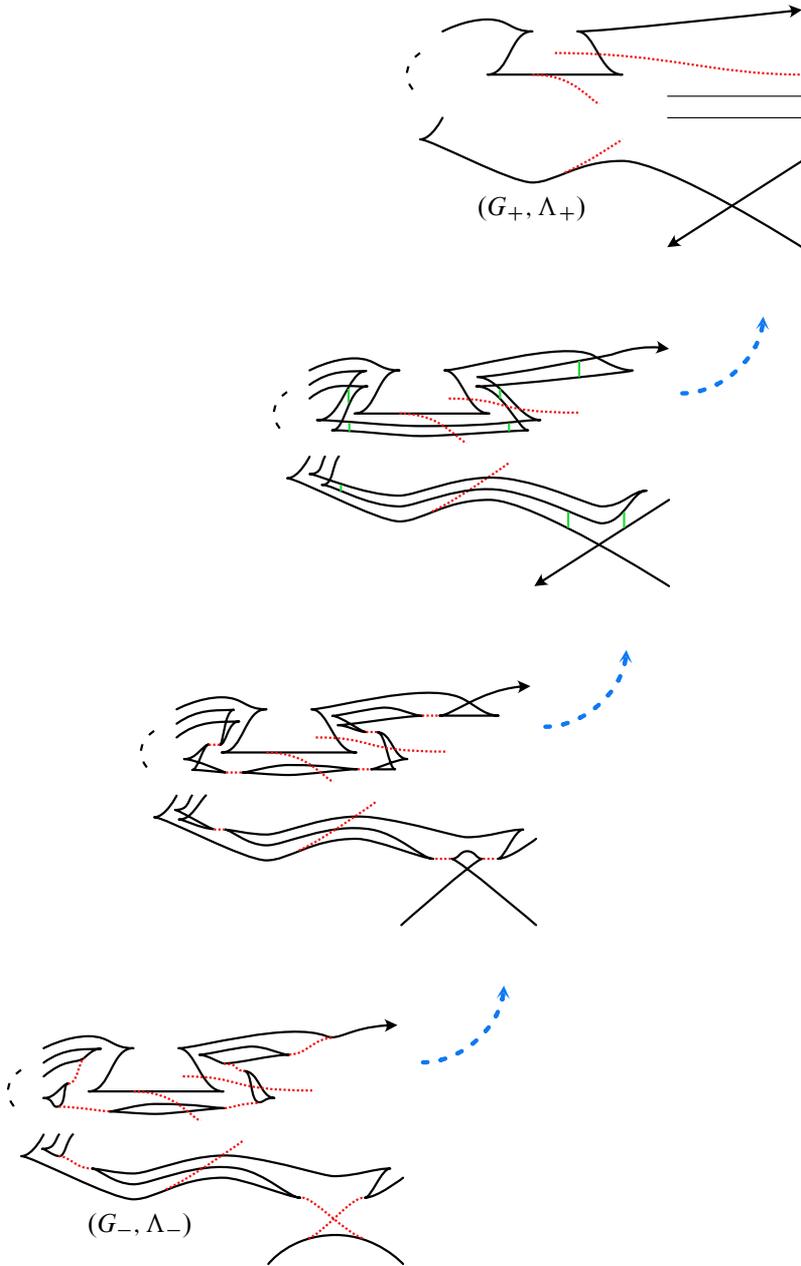
Figure 18: The Legendrian handle graph (G_-, Λ_-) has one fewer positive crossing than (G_+, Λ_+). Red curves represent cores of handles, while green curves represent cocores.

x–coordinate of X_0 below X_0. Without loss of generality, assume that this holds for the upper left strand as in the upper-right portion of Figure 18. Let $-\eta \subset \Lambda_+$ be the compact 1–manifold that starts at X_0 and traverses along the upper-left strand of X_0 until returning to the same x–coordinate, and let η be $-\eta$ with the orientation reversed, so that X_0 is at the end of η.

As in the second diagram down in Figure 18, create a finger of Λ_+ parallel to η using a Reidemeister I move at the initial point of η and next to every cusp of η along with Reidemeister II moves to pass the lead cusp of the finger through handles of G_+ that are incident to η. Move the end of the finger just to the right of X_0. Place a cocore of a handle inside the finger just below each crossing created by a Reidemeister I move. Place two additional cocores from the finger to the original link on either side of the crossing X_0.

Finally, replace the cocores by surgery disks to create a new Legendrian handle graph as in the third row of Figure 18. Isotope the new Legendrian handle graph as at the bottom of Figure 18, using a combination of the move in Figure 16, right, to move the handles away and Reidemeister I moves to remove the crossings. The result is a Legendrian handle graph that has many more surgery disks, but whose underlying Legendrian link has one fewer crossing than before. □

The procedure above may produce a disconnected Legendrian link. We next see how to join these components.

Lemma 4.4 *Let (G_+, Λ_+) be a Legendrian handle graph, where Λ_+ has $n \geq 2$ components, which are mutually disjoint in the front diagram. Suppose that there exists a path γ in the front diagram of G_+ that starts on component $\Lambda'_+ \subset \Lambda_+$, ends on $\Lambda''_+ \subset \Lambda_+$, and does not intersect Λ_+ otherwise. Then there exists a Legendrian handle graph (G_-, Λ_-) and a subset $\mathcal{H}_0$ of handles of G_-, such that $\mathrm{Surg}(G_-, \Lambda_-, \mathcal{H}_0)$ is Legendrian isotopic to (G_+, Λ_+), one component of Λ_- is topologically the connected sum of Λ'_+ and Λ''_+, the other components of Λ_- match the remaining components of Λ_+, and none of the components of Λ_- intersect in the front diagram.*

Proof We may assume that γ intersects Λ'_+ and Λ''_+ away from triple points, crossings, and cusps. Create a finger of Λ'_+ that follows γ, starting with a Reidemeister I move and using Reidemeister II moves to cross handles of G_+ and additional Reidemeister I moves when γ has a vertical tangent; see the middle diagram of Figure 19. Stop the finger just before γ intersects Λ''_+, performing an additional Reidemeister I move if necessary to ensure that the orientations of parallel strands of the finger and Λ''_+ are opposite. Place a cocore of a handle between those two parallel strands. Finally, replace the cocore by a core of a handle to create a new Legendrian handle graph (G_-, Λ_-) as in the bottom-left portion of Figure 19.

That the new component of Λ_- is the connect sum of Λ'_+ and Λ''_+ comes from the facts that the diagrams of Λ'_+ and Λ''_+ are disjoint and that γ is disjoint from the diagram of Λ_+ on its interior. The final two conclusions of the lemma follow immediately from the construction. □

With the tools above in place, we are ready to reprove Lemma 3.5 using the diagrammatic techniques of this section.

Diagrammatic proof of Lemma 3.5 in $(\mathbb{R}^3, \xi_0)$ Given a Legendrian Λ, use Lemma 4.2 repeatedly, and then Lemma 4.3 repeatedly, to obtain a Legendrian handle graph (G_1, Λ_1) such that the front diagram of
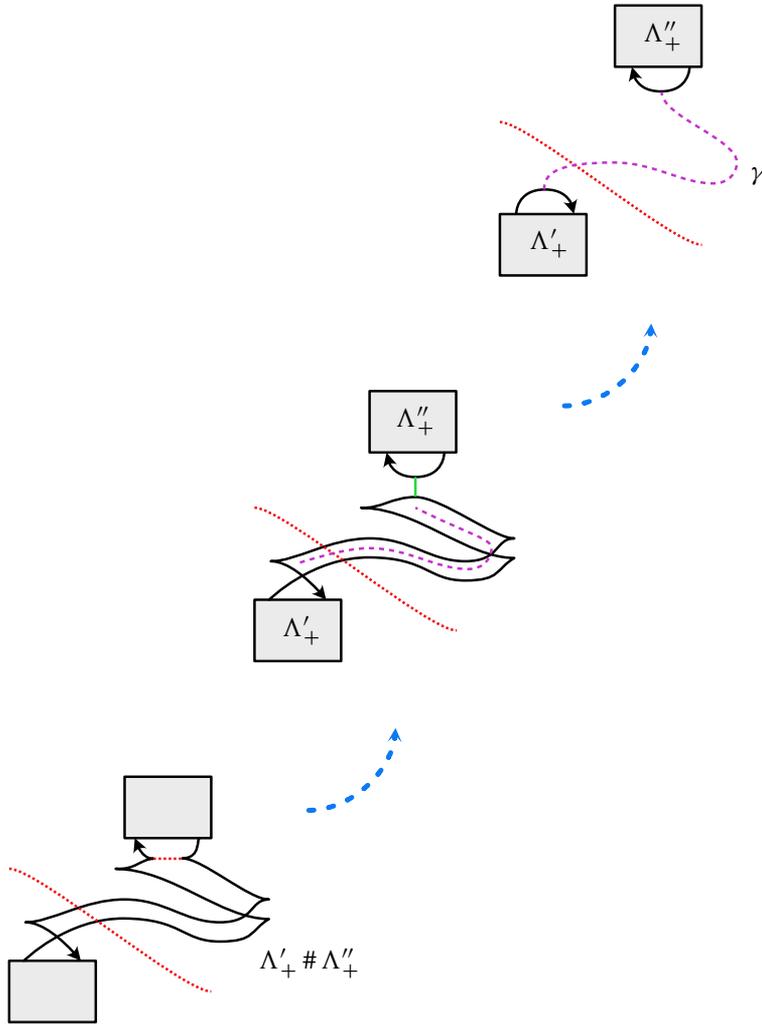
Figure 19: The Legendrian link Λ_- has a component that is topologically the connected sum of two components of Λ_+.

Λ_1 has no crossings, and $\mathrm{Surg}(G_1, \Lambda_1)$ is Legendrian isotopic to Λ. Use Lemma 4.4 to find a Legendrian handle graph (G_2, Λ_2) with a subset $\mathscr{H}_0$ of handles, such that Λ_2 is connected, and $\mathrm{Surg}(G_2, \Lambda_2, \mathscr{H}_0)$ is Legendrian isotopic to $\mathrm{Surg}(G_1, \Lambda_1)$, which implies that $\mathrm{Surg}(G_2, \Lambda_2)$ is Legendrian isotopic to Λ. Finally, note that Λ_2 is a smooth unknot since it is the connected sum of smooth unknots. $\qquad\square$

5 Upper bounds

With constructions of a common lower bound and corresponding handle graphs for Λ and Λ' in hand, we are ready to find an upper bound. The structure of the following proof parallels that of Lazarev [2020] in higher dimensions.

Proof of Proposition 1.6   Given oriented Legendrian links in Λ and Λ' in (Y, ξ), Proposition 3.1 implies that there exist an oriented Legendrian link Λ_- and Legendrian handle graphs (G, Λ_-) and (G', Λ_-) such that $\mathrm{Surg}(G, \Lambda_-)$ (resp. $\mathrm{Surg}(G', \Lambda_-)$) is Legendrian isotopic to Λ (resp. Λ').

We Legendrian isotope the handles $\mathcal{H}'$ of G' to be in general position with respect to the handles $\mathcal{H}$ of G. In particular, we may assume that the Legendrian handle graph (G', Λ_-) has $\mathcal{H}'$ is disjoint from $\mathcal{H}$, with $\mathrm{Surg}(G', \Lambda_-)$ still Legendrian isotopic to Λ'.

Define the Legendrian graph $G_+ = G \cup G'$; it is clear that (G_+, Λ_-) is a Legendrian handle graph. Note that $\mathrm{Surg}(G_+, \Lambda_-, \mathcal{H})$ is Legendrian isotopic to a Legendrian handle graph $(G_{+,1}, \Lambda)$; similarly, $\mathrm{Surg}(G_+, \Lambda_-, \mathcal{H}')$ is Legendrian isotopic to a Legendrian handle graph $(G_{+,2}, \Lambda')$.

Observe that both $\mathrm{Surg}(G_{+,1}, \Lambda)$ and $\mathrm{Surg}(G_{+,2}, \Lambda')$ are Legendrian isotopic to $\mathrm{Surg}(G_+, \Lambda_-)$, which we denote by Λ_+. Let $L: \Lambda_- \to \Lambda_+$ be the concatenation of $L(G, \Lambda_-)$ with $L(G_{+,1}, \Lambda)$; similarly, let $L': \Lambda_- \to \Lambda_+$ be the concatenation of $L(G', \Lambda_-)$ with $L(G_{+,2}, \Lambda')$. Then it is clear that Λ (resp. Λ') appears as a collared slice of L (resp. L'). At the same time, Proposition 2.6 implies that L and L' are exact-Lagrangian isotopic, since they are both obtained from the same Legendrian handle graph (G_+, Λ_-) by Legendrian ambient surgery, only in a different order — in other words, they both belong to the isotopy class $L(G_+, \Lambda_-)$. $\square$

Example 5.1   Figures 20 and 21 display the full process of creating the upper bounds in Figures 1 and 2, respectively.

6   The Lagrangian cobordism genus

In this section, we use the construction of upper and lower bounds for a pair of Legendrian knots to define a new quantity, the relative Lagrangian genus, and a new relation, Lagrangian zigzag-concordance. We explore foundational properties and immediate examples, leaving deeper explorations, as embodied in the list of open questions at the end, for future work. For ease of notation, we work with Legendrian links in the standard contact $\mathbb{R}^3$, though our definitions may easily be adapted to Legendrians in any contact 3–manifold.

6.1   Lagrangian quasicobordism

We begin with a definition that undergirds the two concepts referred to above.

Definition 6.1   A *Lagrangian zigzag-cobordism* between Legendrian knots Λ and Λ' consists of an ordered set of $n + 1$ Legendrian links

$$\boldsymbol{\Lambda} = (\Lambda = \Lambda_0, \Lambda_1, \ldots, \Lambda_n = \Lambda'),$$

another ordered set of n nonempty Legendrian links

$$\boldsymbol{\Lambda}^* = (\Lambda_1^*, \ldots, \Lambda_n^*),$$
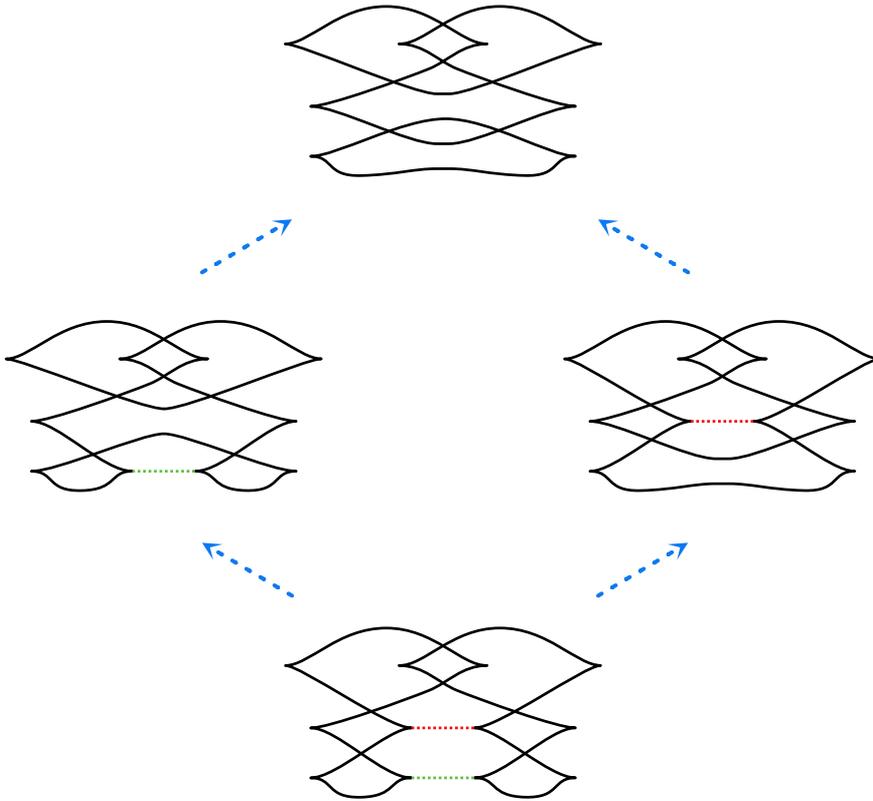
Figure 20: The handle graph at the bottom of the figure is used to create the upper bound of the trefoil and an $m(5_2)$ knot that appeared in Figure 1.

such that Λ_i^* is an upper or lower bound for the pair $(\Lambda_{i-1}, \Lambda_i)$, and connected Lagrangian cobordisms

$$\mathbf{L} = (L_1^<, L_1^>, L_2^<, L_2^>, \ldots, L_n^<, L_n^>)$$

that realize the upper or lower bound constructions.

There are several quantities associated to a Lagrangian zigzag-cobordism.

Definition 6.2 Given a Lagrangian zigzag-cobordism $(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L})$, its *length* is one less than the number of elements in $\mathbf{\Lambda}$, while its *Euler characteristic* $\chi(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L})$ is the sum of the Euler characteristics of the Lagrangians in $\mathbf{L}$ and its *genus* $g(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L})$ defined, as usual, in terms of the Euler characteristic.

Further, we define the *relative Lagrangian genus* $g_L(\Lambda, \Lambda')$ between the Legendrian knots Λ and Λ' as the minimum genus of any Lagrangian zigzag-cobordism between them. Two Legendrian knots Λ and Λ' are *Lagrangian zigzag-concordant* if $g_L(\Lambda, \Lambda') = 0$.

Example 6.3 Let Υ be the maximal Legendrian unknot, and let Λ be a maximal Legendrian representative of $m(6_2)$. Note that both Υ and Λ have Thurston–Bennequin number -1 and that the smooth 4–genus of
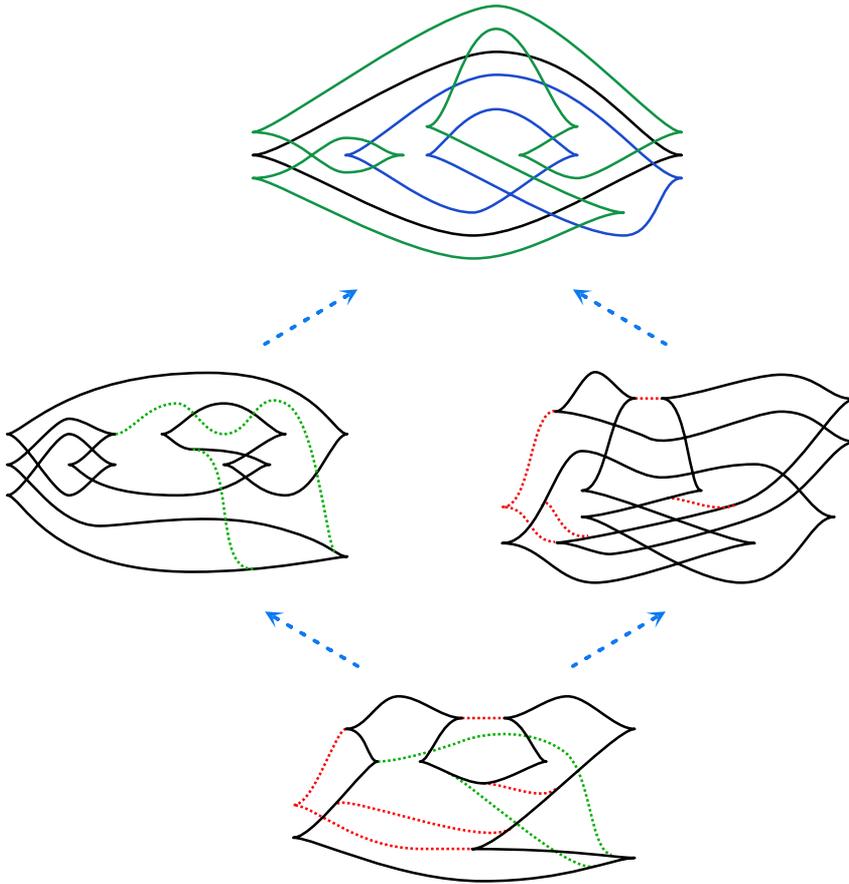
Figure 21: The handle graph at the bottom of the figure is used to create the upper bound of the figure eight knot and the unknot that appeared in Figure 2. Note that the Legendrian knots in the handle graphs in the middle level are isotopic to the unknot (left) and the figure eight (right).

6_2 is equal to 1 [Livingston and Moore 2021]. It follows from the behavior of the Thurston–Bennequin invariant under Lagrangian cobordism that there cannot be a Lagrangian cobordism joining Υ and Λ in either direction. Nevertheless, there is a genus-1 Lagrangian zigzag-cobordism between the two; see Figure 22.

Lagrangian zigzag-cobordism induces an equivalence relation on the set of isotopy classes of Legendrian links. As in the smooth case, this equivalence relation is uninteresting, as shown by the following immediate corollary of Theorem 1.1 or Proposition 3.1, together with Remark 1.3:

Corollary 6.4 *Any two Legendrian knots with the same rotation number are Lagrangian zigzag-cobordant. In fact, the zigzag-cobordism may be chosen to have length* 1.

The corollary shows that the relative Lagrangian genus is defined for any two Legendrian knots of the same rotation number.
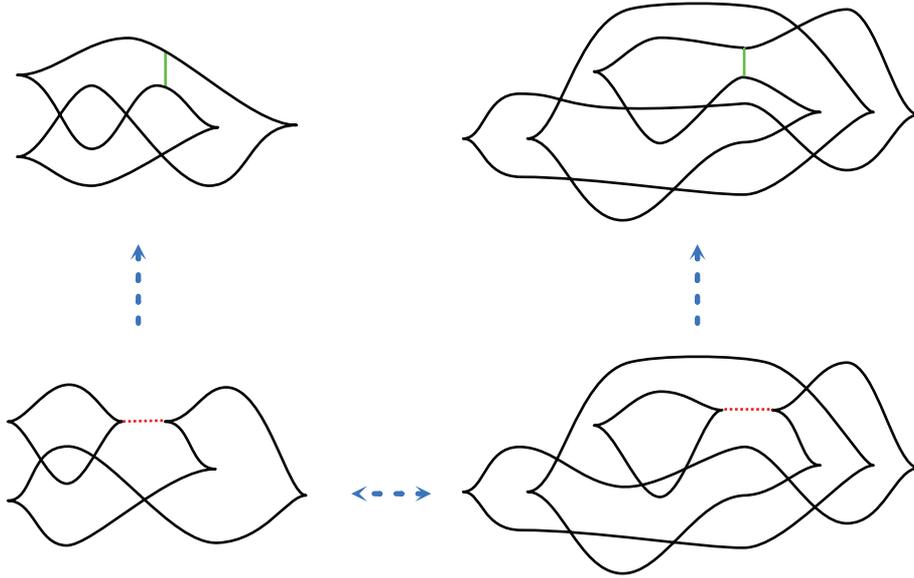
Figure 22: A genus 1 Lagrangian quasicobordism between the maximal unknot Υ and a maximal representative Λ of the mirror of the 6_2 knot. The zigzag-cobordism was produced using the ideas in [Boranda et al. 2013, Section 5], especially Figures 25 and 27.

On the other hand, Lagrangian zigzag-concordance also clearly induces an equivalence relation on the set of isotopy classes of Legendrian knots. The relative Lagrangian genus descends to Lagrangian zigzag-concordance classes. Using [Chantraine 2010] and the connectedness of the Lagrangians, we see that both the rotation number and the Thurston–Bennequin number are invariants of Lagrangian zigzag-concordance, though nonclassical invariants coming from Legendrian contact homology or Heegaard Floer theory will have a more complicated relationship with zigzag-concordance.

6.2 Relation to smooth genus

To connect the relative Lagrangian genus to smooth constructions, note that we may define the smooth cobordism genus between two smooth knots K_1 and K_2 to be the minimum genus of all cobordisms between them; we denote this by $g_s(K_1, K_2)$. Chantraine [2010] proved that Lagrangian *fillings* minimize the smooth 4–ball genus of a Legendrian knot, and so one might ask if this minimization property extends to g_L. We begin with a simple lemma.

Lemma 6.5 *Given Legendrian knots Λ and Λ', we have $g_s(\Lambda, \Lambda') \le g_L(\Lambda, \Lambda')$.*

Proof Let $(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}^*, \mathbf{L})$ be a Lagrangian zigzag-cobordism between Λ and Λ'. Assume for ease of notation that each $\Lambda^* \in \boldsymbol{\Lambda}^*$ is an upper bound. Let $\bar{L}_i^>$ be the smooth cobordism from Λ_i^* to Λ_i obtained from reversing $L_i^>$; note that $\bar{L}_i^>$ is not, in general, a Lagrangian cobordism. Since Euler characteristic is additive under gluing, the smooth cobordism $L_1^< \circ \bar{L}_1^> \circ L_2^< \circ \cdots \circ \bar{L}_n^>$ has genus $g(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}^*, \mathbf{L})$, and hence $g_s(\Lambda, \Lambda') \le g_L(\Lambda, \Lambda')$. □

It is natural to ask under what conditions on Λ_1 and Λ_2 — as Legendrian or as smooth knots — is the inequality in Lemma 6.5 an equality? On one hand, we cannot expect to achieve equality in all cases.

Example 6.6 Let Λ be any Legendrian knot, and let Λ' be a double stabilization of Λ with the same rotation number as Λ. Since Λ and Λ' have the same underlying smooth knot type, we have $g_s(\Lambda, \Lambda') = 0$. On the other hand, let $(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L})$ be a Lagrangian zigzag-cobordism between Λ and Λ'. Note that $\chi(L_i^<), \chi(L_I^>) \leq 0$ for all i, since each of $L_i^<$ and $L_i^>$ is connected and has at least two boundary components. Since $\operatorname{tb}(\Lambda) > \operatorname{tb}(\Lambda')$, some pair Λ_i, Λ_{i+1} in $\mathbf{\Lambda}$ must have different Thurston–Bennequin numbers. In particular, the bound Λ_i^* must have a different Thurston–Bennequin number than at least one of Λ_i or Λ_{i+1}. It follows that $\chi(L_i^<) + \chi(L_i^>) < 0$, and hence that $\chi(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L}) < 0$. Since Λ and Λ' are knots, this implies that $g(\mathbf{\Lambda}, \mathbf{\Lambda}^*, \mathbf{L}) > 0$. In particular, we have $g_L(\Lambda, \Lambda') > 0$ even though $g_s(\Lambda, \Lambda') = 0$.

On the other hand, there is a simple sufficient condition for equality in the lemma above.

Lemma 6.7 *If the Legendrian knot Λ has a Lagrangian filling, and there exists a Lagrangian cobordism from Λ to Λ', then $g_s(\Lambda, \Lambda') = g_L(\Lambda, \Lambda')$.*

Proof We begin by setting notation. Let L_0 be the Lagrangian filling of Λ and let $L_1^<$ be the Lagrangian cobordism from Λ to Λ'. Taking $L_1^>$ to be the trivial cylindrical Lagrangian cobordism from Λ' to itself, and taking $\Lambda_1^* = \Lambda'$, we see that

$$(1) \qquad\qquad g_L(\Lambda, \Lambda') \leq g(L_1^<).$$

Let Σ be the smooth cobordism from Λ to Λ' that minimizes the smooth cobordism genus. We know that $L_0 \circ L_1^<$ is a Lagrangian filling of Λ', and hence that $g(L_0 \circ L_1^<) \leq g(L_0 \circ \Sigma)$. Since Λ is a knot, the genus is additive under composition of cobordisms, and we obtain

$$(2) \qquad\qquad g(L_1^<) \leq g(\Sigma).$$

Combining (1) and (2), we obtain

$$g_L(\Lambda, \Lambda') \leq g(\Sigma) = g_s(\Lambda, \Lambda').$$

The lemma now follows from Lemma 6.5. □

6.3 Open questions

We end with a list of questions about Lagrangian zigzag-cobordism and zigzag-concordance beyond the motivating question above about the relationship between the relative Lagrangian genus and the relative smooth genus.

(1) Building off of Example 6.3, is there an example of a pair Λ and Λ' that are Lagrangian zigzag-concordant but not Lagrangian concordant?

(2) Taking the previous question further, for two Lagrangian zigzag-concordant Legendrians Λ and Λ', what is the minimal length of any Lagrangian zigzag-concordance between them? Are there examples for which this minimal length is arbitrarily high?

(3) Even more generally, define $g_L(\Lambda, \Lambda', n)$ to be the minimal genus of any Lagrangian zigzag-cobordism between Λ and Λ' of length at most n. The sequence $(g_L(\Lambda, \Lambda', n))_{n=1}^{\infty}$ decreases to and stabilizes at $g_L(\Lambda, \Lambda')$. Are there examples for which the number of steps it takes the sequence to stabilize is arbitrarily long?

(4) Can $g_L(\Lambda, \Lambda') - g_s(\Lambda, \Lambda')$ be arbitrarily large when Λ and Λ both have maximal Thurston–Bennequin invariant?

(5) Can the hypotheses of Lemma 6.7 be weakened to Λ having only an augmentation instead of a filling?

(6) Is there a version of this theory for Maslov 0 Lagrangians, which would better allow the use of Legendrian contact homology, especially the tools in [Pan 2017]?

References

[Baldwin and Sivek 2018]   **J A Baldwin**, **S Sivek**, *Invariants of Legendrian and transverse knots in monopole knot homology*, J. Symplectic Geom. 16 (2018) 959–1000 MR Zbl

[Baldwin et al. 2022]   **J A Baldwin**, **T Lidman**, **C-M M Wong**, *Lagrangian cobordisms and Legendrian invariants in knot Floer homology*, Michigan Math. J. 71 (2022) 145–175 MR Zbl

[Boranda et al. 2013]   **B Boranda**, **L Traynor**, **S Yan**, *The surgery unknotting number of Legendrian links*, Involve 6 (2013) 273–299 MR Zbl

[Bourgeois et al. 2015]   **F Bourgeois**, **J M Sabloff**, **L Traynor**, *Lagrangian cobordisms via generating families: construction and geography*, Algebr. Geom. Topol. 15 (2015) 2439–2477 MR Zbl

[Chantraine 2010]   **B Chantraine**, *Lagrangian concordance of Legendrian knots*, Algebr. Geom. Topol. 10 (2010) 63–85 MR Zbl

[Cornwell et al. 2016]   **C Cornwell**, **L Ng**, **S Sivek**, *Obstructions to Lagrangian concordance*, Algebr. Geom. Topol. 16 (2016) 797–824 MR Zbl

[Dimitroglou Rizell 2016]   **G Dimitroglou Rizell**, *Legendrian ambient surgery and Legendrian contact homology*, J. Symplectic Geom. 14 (2016) 811–901 MR Zbl

[Ekholm et al. 2016]   **T Ekholm**, **K Honda**, **T Kálmán**, *Legendrian knots and exact Lagrangian cobordisms*, J. Eur. Math. Soc. 18 (2016) 2627–2689 MR Zbl

[Eliashberg and Fraser 2009]   **Y Eliashberg**, **M Fraser**, *Topologically trivial Legendrian knots*, J. Symplectic Geom. 7 (2009) 77–127 MR Zbl

[Eliashberg and Gromov 1998]   **Y Eliashberg**, **M Gromov**, *Lagrangian intersection theory: finite-dimensional approach*, from "Geometry of differential equations", Amer. Math. Soc. Transl. Ser. 2 186, Amer. Math. Soc., Providence, RI (1998) 27–118 MR Zbl

[Golla and Juhász 2019]   **M Golla**, **A Juhász**, *Functoriality of the EH class and the LOSS invariant under Lagrangian concordances*, Algebr. Geom. Topol. 19 (2019) 3683–3699 MR Zbl

[Honda 2000]   **K Honda**, *On the classification of tight contact structures, I*, Geom. Topol. 4 (2000) 309–368 MR Zbl

[Kanda 1998]   **Y Kanda**, *On the Thurston–Bennequin invariant of Legendrian knots and nonexactness of Bennequin's inequality*, Invent. Math. 133 (1998) 227–242 MR Zbl

[Lazarev 2020]   **O Lazarev**, *Maximal contact and symplectic structures*, J. Topol. 13 (2020) 1058–1083 MR Zbl

[Livingston and Moore 2021]   **C Livingston**, **A H Moore**, *KnotInfo*: *table of knot invariants*, electronic reference (2021) Available at `https://knotinfo.math.indiana.edu/`

[O'Donnol and Pavelescu 2012]   **D O'Donnol**, **E Pavelescu**, *On Legendrian graphs*, Algebr. Geom. Topol. 12 (2012) 1273–1299 MR Zbl

[Pan 2017]   **Y Pan**, *The augmentation category map induced by exact Lagrangian cobordisms*, Algebr. Geom. Topol. 17 (2017) 1813–1870 MR Zbl

[Sabloff and Traynor 2013]   **J M Sabloff**, **L Traynor**, *Obstructions to Lagrangian cobordisms between Legendrians via generating families*, Algebr. Geom. Topol. 13 (2013) 2733–2797 MR Zbl

Department of Mathematics and Statistics, Haverford College
Haverford, PA, United States

Department of Mathematics, Louisiana State University
Baton Rouge, LA, United States

Department of Mathematics and Statistics, University of Ottowa
Ottowa, ON, Canada

jsabloff@haverford.edu,   shea@math.lsu.edu,   mike.wong@uottawa.ca

https://jsabloff.sites.haverford.edu,   https://www.math.lsu.edu/~shea,
https://mysite.science.uottawa.ca/cwong

Interleaving Mayer–Vietoris spectral sequences

ÁLVARO TORRAS-CASAS

ULRICH PENNIG

We discuss the Mayer–Vietoris spectral sequence as an invariant in the context of persistent homology. In particular, we introduce the notion of ε–acyclic carriers and ε–acyclic equivalences between filtered regular CW–complexes and study stability conditions for the associated spectral sequences. We also look at the Mayer–Vietoris blowup complex and the geometric realization, finding stability properties under compatible noise; as a result we prove a version of an approximate nerve theorem. Adapting work by Serre, we find conditions under which ε–interleavings exist between the spectral sequences associated to two different covers.

55N31, 55T99

1 Introduction

One of the benefits of homology as a topological invariant over, for example, the homotopy groups, is its computability via long exact sequences. The classical Mayer–Vietoris exact sequence has been used in countless examples to compute $H_k(X)$ from a decomposition of a space X into two open subsets U and V. When we generalise this concept to open covers $(U_i)_{i \in I}$ consisting of more than just two subsets, the relations between the parts $H_k(U_i)$ become more intricate and are encoded in the Mayer–Vietoris spectral sequence. These sequences first appeared in work of Leray and later Serre, and they proved to be one of the most powerful tools in pure algebraic topology. Applications of spectral sequences in applied algebraic topology, however, is still a young subject.

In [Torras-Casas 2023] it was proven that the persistence Mayer–Vietoris spectral sequence can be used to compute persistent homology. The starting point is a filtered simplicial complex X together with a cover by subcomplexes $\mathcal{U}$. Then, one computes $PH_i(\mathcal{U}_\sigma)$ for all $i \geq 0$ and $\sigma \in N_\mathcal{U}$. Here, notice that $N_\mathcal{U}$ is the nerve of the cover $\mathcal{U}$ whose simplices $\sigma \in N_\mathcal{U}$ are subsets from $\mathcal{U}$; this leads to the notation $\mathcal{U}_\sigma = \bigcap_{U \in \sigma} U$. The Mayer–Vietoris spectral sequence starts from these groups and the morphisms induced by inclusions and converges to $PH_i(X)$. As pointed out in [Yoon and Ghrist 2020], the additional insight gained from the cover $\mathcal{U}$ can be used, for example, for multiscale feature detection. Similar information was also explored much earlier in [Zomorodian and Carlsson 2008] in the form of *localized homology*.

Motivated by these results, we study the spectral sequence $E^*_{p,q}(X, \mathcal{U})$ and answer the following questions:

- Let a pair $(X, \mathcal{U})$ consist of a space, X, and a cover for X, $\mathcal{U}$. The Mayer–Vietoris spectral sequence $E_{p,q}^*(X, \mathcal{U})$ converges to $\mathrm{PH}_*(\Delta^{\mathcal{U}}(X))$. Is $\mathrm{PH}_*(\Delta^{\mathcal{U}}(X))$ stable? Can this result be generalised?

- Suppose that the data in each covering set $\mathcal{U}_\sigma$ for $\sigma \in N_{\mathcal{U}}$ is modified slightly. If the underlying cover $\mathcal{U}$ is ignored, then we would not expect $E_{p,q}^*(X, \mathcal{U})$ to be stable. Are there natural coherence conditions between changes in the sets $\mathcal{U}_\sigma$ that imply stability? If so, what do we mean by stability of spectral sequences?

- Let $\mathcal{U}$ and $\mathcal{V}$ be covers of the same space X. Can we compare $E_{p,q}^*(X, \mathcal{U})$ and $E_{p,q}^*(X, \mathcal{V})$ up to ε–interleavings?

To explain why the first question is important and how it is linked to spectral sequences, we note that $E_{p,q}^*(X, \mathcal{U})$ *converges* to the *target* persistent homology $\mathrm{PH}_*(\Delta^{\mathcal{U}}(X))$ (this is usually denoted by $E_{p,q}^*(X, \mathcal{U}) \Rightarrow \mathrm{PH}_*(\Delta^{\mathcal{U}}(X))$). The blowup complex $\Delta^{\mathcal{U}}(X)$ already appeared in the context of topological data analysis in [Lewis and Morozov 2015] and [Zomorodian and Carlsson 2008]. It is homotopy equivalent to a *homotopy colimit*, and therefore enjoys good properties with respect to local homotopy equivalences. For example, if we assume that $\mathcal{U}_\sigma$ is contractible for all $\sigma \in N_{\mathcal{U}}$, then we can use [Hatcher 2002, Proposition 4G.2] to recover Leray's nerve theorem. That is, there are homotopy equivalences

$$X \simeq \Delta^{\mathcal{U}}(X) \simeq \Delta^{\mathcal{U}}(*) = N(\mathcal{U}),$$

where $*$ denotes the constant *complex of spaces* on $\mathcal{U}$; see [Hatcher 2002, Appendix 4.G]. The fundamental importance of this result in applied topology is underlined by the persistent nerve lemma presented in [Chazal and Oudot 2008]. It is worth mentioning the approximate nerve theorem [Govc and Skraba 2018] and the generalised nerve theorem [Cavanna 2019], which are approximate versions of the Leray theorem within the context of persistence. In particular, in [Govc and Skraba 2018] the spectral sequence $E_{p,q}^*(X, \mathcal{U}) \Rightarrow \mathrm{PH}_*(X)$ is examined, and it is studied how much it differs from another spectral sequence $E_{p,q}^*(*, \mathcal{U}) \Rightarrow \mathrm{PH}_*(N(\mathcal{U}))$, by careful inspection of all pages as well as the extension problem.

Throughout the paper we focus on the category **RCW-cpx** of *regularly filtered regular CW–complexes* as well as the subcategory **FCW-cpx** of *filtered regular CW–complexes*; see Section 2.1. Instead of restricting our attention to a space X together with a cover $\mathcal{U}$, we look at regular diagrams $\mathcal{D}$ in **RCW-cpx** over a simplicial complex K. There is a natural replacement for the Mayer–Vietoris blowup complex in this setting, denoted by $\Delta_K(\mathcal{D})$, as explained in [Hatcher 2002, Appendix 4.G]. This object also appears in the context of semisimplicial spaces, where it is called the *geometric realization* [Ebert and Randal-Williams 2019]; in fact, it has an associated spectral sequence [Ebert and Randal-Williams 2019, Section 1.4]. As we explain in Section 3, there are good reasons why it is worth taking this more general perspective. In particular, we consider the spectral sequence

$$E_{p,q}^2(\mathcal{D}) \Rightarrow \mathrm{PH}_{p+q}(\Delta_K \mathcal{D}).$$

In order to address the first two questions, we introduce the notion of acyclic carriers to define ε–*acyclic equivalences*. Using the acyclic carrier theorem we show the following: Let X and Y be two objects

in **RCW-cpx**. If there exists an ε–acyclic equivalence $F^\varepsilon\colon X \rightrightarrows Y$, then $\mathrm{PH}_*(X)$ is ε–interleaved with $\mathrm{PH}_*(Y)$ (see Corollary 4.7 and Proposition 4.2 for a stronger statement in **FCW-cpx**). These equivalences provide a very flexible notion that works in different contexts as Examples 4.5, 4.6 and 4.8 show.

We address the first question in the following way. Let $\mathscr{D}$ and $\mathscr{L}$ be two diagrams over the same simplicial complex K and assume that for all $\sigma \in K$ there are ε–acyclic equivalences $F^\varepsilon_\sigma\colon \mathscr{D}(\sigma) \rightrightarrows \mathscr{L}(\sigma)$ which satisfy a compatibility condition with respect to composition in the poset category associated to K; see Theorem 5.2 for details. Then, there is an ε–acyclic equivalence $F^\varepsilon\colon \Delta_K(\mathscr{D}) \rightrightarrows \Delta_K(\mathscr{L})$. This result implies stability in the targets of convergence of the spectral sequences. We use this result to show a "strong approximate multinerve theorem" in Corollary 5.3. Later, in Section 6, we introduce (ε, n)–interleavings, which are given by spectral sequence morphisms that start at some page n together with a shift by a persistence parameter $\varepsilon > 0$. Assuming the same conditions as in the geometric realization case, we can obtain a $(\varepsilon, 1)$–interleaving between $E^*_{p,q}(\mathscr{D})$ and $E^*_{p,q}(\mathscr{L})$; see Theorem 5.2. This result appears in Theorem 6.5 and a specialised strong statement for covers of spaces in **FCW-cpx** is given in Proposition 6.4.

As for the third question about the comparison of the spectral sequences associated to two covers $\mathscr{U}$ and $\mathscr{V}$ of a space X, we rely on work of Serre [1955], in which he studied the relation between the Čech cohomology of two different covers; here we adapt this work in the context of cosheaves and cosheaf homology. Take a cosheaf $\mathscr{F}$ of abelian groups on X and assume that there is a refinement $\mathscr{V} \prec \mathscr{U}$. Serre showed that the refinement morphism induced on Čech homology $\rho^{\mathscr{U}\mathscr{V}}\colon \check{\mathscr{H}}_*(\mathscr{V}, \mathscr{F}) \to \check{\mathscr{H}}_*(\mathscr{U}, \mathscr{F})$ is independent of the particular choice of morphism in the cochains. In [Serre 1955] it was also shown that $\rho^{\mathscr{U}\mathscr{V}}$ can be factored through a construction that uses a double complex associated to both covers $C_{p,q}(\mathscr{U}, \mathscr{V}; \mathscr{F})$, see [Serre 1955, Proposition 4, Section 29]. This construction introduces two double complex spectral sequences $^{\mathrm{I}}E^*_{p,q}(\mathscr{U}, \mathscr{V}; \mathscr{F})$ and $^{\mathrm{II}}E^*_{p,q}(\mathscr{U}, \mathscr{V}; \mathscr{F})$, both of which converge to $\check{\mathscr{H}}_*(\mathscr{U} \cap \mathscr{V}; \mathscr{F}) \simeq \check{\mathscr{H}}_*(\mathscr{V}; \mathscr{F})$. Here one might study conditions on $^{\mathrm{II}}E^*_{p,q}(\mathscr{U}, \mathscr{V}; \mathscr{F})$ to find when an inverse of $\rho^{\mathscr{U}\mathscr{V}}$ exists. As an application, Serre [1955, Theorem 1, Section 29] obtained an analogous result to the Leray theorem in the context of sheaves.

We start our analysis of the third question in Section 7. In case $\mathscr{V} \prec \mathscr{U}$ there is a unique morphism induced by the refinement map on the second page

$$\rho^{\mathscr{U}\mathscr{V}}\colon E^*_{p,q}(X, \mathscr{V}) \to E^*_{p,q}(X, \mathscr{U}).$$

On the other hand, Theorem 7.10 tells us under what conditions there exists an ε–shifted morphism $\psi\colon E^*_{p,q}(X, \mathscr{U}) \to E^*_{p,q}(X, \mathscr{V})[\varepsilon]$ such that $\rho^{\mathscr{U}\mathscr{V}}$ and ψ form an $(\varepsilon, 2)$–interleaving between $E^*_{p,q}(X, \mathscr{U})$ and $E^*_{p,q}(X, \mathscr{V})$. Finally, in Proposition 7.12 we give a means of obtaining an $(\varepsilon, 2)$–interleaving between $E^*_{p,q}(X, \mathscr{U})$ and $E^*_{p,q}(X, \mathscr{V})$ through the computation of local Mayer–Vietoris spectral sequences $E^*_{p,q}(\mathscr{U}_\sigma, \mathscr{V}_{|\mathscr{U}_\sigma})$ for all $\sigma \in N_{\mathscr{U}}$. Since the open regions $\mathscr{U}_\sigma$ are assumed to be "small" in comparison to X, this gives a means of using local calculations to deduce the interleaving. As Corollary 7.14 we present the case when $\mathscr{V}$ does not need to refine $\mathscr{U}$.

2 Background

2.1 Regular CW–complexes with filtrations

Recall the definition of CW–complex from [Hatcher 2002, Chapter 0]. In contrast to the usual treatment of CW–complexes, but in line with the structure we are dealing with in TDA, we consider the cell decomposition as part of the data of our CW–complexes. For a CW–complex X, if c is an open cell in X we follow the notation from [Cooke and Finney 1967] and denote this by $c \in X$. We denote by X^n the set of n–dimensional cells from X and we denote by $X^{\leq n}$ the n–skeleton from X. Recall that X has a natural filtration given by its skeleta $X^0 \subseteq X^{\leq 1} \subseteq \cdots \subseteq X^{\leq N} \subseteq \cdots$, and a *cellular morphism* $f : X \to Y$ respects this filtration, in the sense that it restricts to morphisms $f^m : X^{\leq m} \to Y^{\leq m}$ for all $m \geq 0$. We work with regular CW–complexes, which are CW–complexes where the attaching maps are homeomorphisms. It is recommended to consult [Cooke and Finney 1967; Massey 1991] for properties and results related to regular CW–complexes. An intuitive way of understanding incidences of cells in regular complexes is through the barycentric subdivision, as explained in [Ellis 2019, Section 2.1]. Given a pair of cells $a \in X^n$ and $b \in X^{n-1}$, we denote by $[b : a]$ the degree of attaching map $\partial a \to \bar{b}/\partial b$.

Definition 2.1 A cellular morphism $f : X \to Y$ is a *regular morphism* whenever the closure $\overline{f(a)}$ is a subcomplex of Y for all cells $a \in X$. For such a morphism and a pair $a \in X^n$ and $b \in Y^n$, we denote by $[b : f(a)]$ the degree of the morphism f restricted to the open cell a and mapping into the open cell b.

We write **CW-cpx** to denote the category of finite regular CW–complexes and regular morphisms. Denote by **R** the ordered category $(\mathbb{R}, \leq)$ of real numbers. We focus on functors $X : \mathbf{R} \to \mathbf{CW\text{-}cpx}$ which we call *regularly filtered CW–complexes*, and we denote their category by **RCW-cpx**. We say that an object $X \in \mathbf{RCW\text{-}cpx}$ is *tame*, whenever X is constant along a finite number of right open intervals decomposing the poset **R**. For $X \in \mathbf{RCW\text{-}cpx}$, we write X_r for the regular CW–complex $X(r)$ for all $r \in \mathbf{R}$. On the other hand we write $X(r \leq s)$ to denote the morphisms $X_r \to X_s$ for all $r \leq s$ in **R**; we call such morphisms *structure maps*. The reader might find an example of such a regularly filtered complex in the appendix. If the morphisms $X(r \leq s) : X_r \to X_s$ are injections preserving the cellular structure for all $r \leq s$ in **R**, then we call X a *filtered CW–complex*, denoting by **FCW-cpx** the corresponding subcategory of **RCW-cpx**. Notice that objects in **FCW-cpx** can be seen as a pair $(\mathrm{colim}\, X_*, f)$ where $\mathrm{colim}\, X_*$ is a regular CW–complex and $f : \mathrm{colim}\, X_* \to \mathbb{R}$ is a filtration function.

Throughout this text, we work with a fixed field $\mathbb{F}$. Given $X \in \mathbf{RCW\text{-}cpx}$, we define the persistent homology in degree n as the functor $\mathrm{PH}_n(X) : \mathbf{R} \to \mathbf{vect}$ given by computing cellular homology $\mathrm{PH}_n(X)_r = \mathrm{H}_n^{\mathrm{cell}}(X_r; \mathbb{F})$ for all $r \in \mathbf{R}$. As X_r is finite, the vector space $\mathrm{PH}_n(X)_r$ is finite dimensional for all $r \in \mathbf{R}$. If in addition X is tame, $\mathrm{PH}_n(X)$ only changes at a finite number of points $r \in \mathbf{R}$. We call the category of functors $\mathbf{R} \to \mathbf{vect}_\mathbb{F}$ *persistence modules* and denote it by **PMod**. Given $a \in (0, \infty)$ and $X \in \mathbf{RCW\text{-}cpx}$, we write $X[a]$ for the element of **RCW-cpx** such that $X[a]_r = X_{r+a}$ for all $r \in \mathbf{R}$. We use Σ^ε to denote the ε–*shift functor* $\Sigma^\varepsilon : \mathbf{RCW\text{-}cpx} \to \mathrm{Hom}(\mathbf{RCW\text{-}cpx})$ which sends $X \in \mathbf{RCW\text{-}cpx}$ to

$\Sigma^\varepsilon X \colon X \to X[\varepsilon]$, where $\varepsilon \geq 0$. Also, for any morphism of filtered CW–complexes $f \colon A \to B$, one can check that $f[\varepsilon] \circ \Sigma^\varepsilon A = \Sigma^\varepsilon B \circ f$, where we use $f[\varepsilon] \colon A[\varepsilon] \to B[\varepsilon]$. Similarly, there are shift functors for persistence modules $\Sigma^\varepsilon \colon \mathbf{PMod} \to \mathrm{Hom}(\mathbf{PMod})$ for $\varepsilon \geq 0$.

Remark 2.2 Notice that the standard algorithm for the computation of persistent homology cannot be applied to objects in **RCW-cpx**. However, if X is tame and one successfully computes the coefficients for the morphisms $C_*^{\mathrm{cell}}(X_r) \to C_*^{\mathrm{cell}}(X_s)$ for all $r \leq s$ in **R**, then one can use [Torras-Casas 2023, Algorithm 2 image_kernel] to obtain a *barcode basis* for the filtered cellular complex $C_*^{\mathrm{cell}}(X)$. Then we compute homology of the persistence morphisms given by the differentials $d_n \colon C_n^{\mathrm{cell}}(X) \to C_{n-1}^{\mathrm{cell}}(X)$ by the use of image_kernel. See [Torras-Casas 2023] for an explanation.

2.2 Acyclic carriers

Fix a field $\mathbb{F}$. We say that $X \in \mathbf{CW\text{-}cpx}$ is $\mathbb{F}$–*acyclic* if the reduced homology $\widetilde{\mathrm{H}}^*(X; \mathbb{F})$ with $\mathbb{F}$–coefficients vanishes in all dimensions; as the field is understood from the context, we just say that X is *acyclic*. Consider two objects Φ and Γ from **CW-cpx** with their respective pairs of chains and differentials $(C_*^{\mathrm{cell}}(\Phi), \delta^\Phi)$ and $(C_*^{\mathrm{cell}}(\Gamma), \delta^\Gamma)$. Let $\langle \cdot, \cdot \rangle_\Phi$ and $\langle \cdot, \cdot \rangle_\Gamma$ denote the inner products on $C_*^{\mathrm{cell}}(\Phi)$ and $C_*^{\mathrm{cell}}(\Gamma)$, where the cells form an orthonormal basis. We define a relation $\prec$ on Φ by setting $\tau \prec \sigma$ if $\langle \tau, \delta^\Phi(\sigma) \rangle_\Phi \neq 0$ and by taking the transitive closure. We denote by $\preceq$ the partial order generated by $\prec$. Thus, $\tau \prec \sigma$ does not necessarily imply $\dim(\tau) + 1 = \dim(\sigma)$. Also, notice that $\langle \tau, \delta^\Phi(\sigma) \rangle_\Phi = [\tau : \sigma]$; see the cellular boundary formula from [Hatcher 2002, Section 2.2].

Definition 2.3 A *carrier* $F \colon \Phi \rightrightarrows \Gamma$ is a map from the set of cells of Φ to subcomplexes of Γ that is semicontinuous in the sense that for any pair $\tau \prec \sigma$ in Φ, $F(\tau) \subseteq F(\sigma)$. A carrier $F \colon \Phi \rightrightarrows \Gamma$ is called *acyclic*, if for every $\sigma \in \Phi$, $F(\sigma)$ is a nonempty acyclic subcomplex of Γ.

Given a chain map $w_p \colon C_p^{\mathrm{cell}}(\Phi) \to C_{p+r}^{\mathrm{cell}}(\Gamma)$ of degree $r = 0, 1$, we say that it is carried by F if for all cells $\sigma \in \Phi_p$,

$$\{\gamma \in \Gamma_{p+r} \mid \langle w_p(\sigma), \gamma \rangle_\Gamma \neq 0\} \subseteq F(\sigma),$$

where we followed the notation from [Nanda 2012].

The next statement is an application of [Munkres 1984, Theorem 13.4]. In Proposition 4.2 we prove a version of this statement that applies to filtered CW–complexes. Notice that this theorem works for carriers which are $\mathbb{F}$–acyclic and which do not necessarily need to be $\mathbb{Z}$–acyclic; see the proof of Proposition 4.2.

Theorem 2.4 *Let $F \colon \Phi \rightrightarrows \Gamma$ be an acyclic carrier between CW–complexes Φ and Γ. Then*:

- **Existence** *There is a chain map carried by F.*
- **Equivalence** *If F carries two chain maps ϕ and φ, then F carries a chain homotopy between ϕ and φ.*

Given two acyclic carriers $F, G \colon \Phi \rightrightarrows \Gamma$, we write $F \subseteq G$ whenever $F(\sigma) \subseteq G(\sigma)$ for all $\sigma \in \Phi$. Given another pair of acyclic carriers $F' \colon \Phi \rightrightarrows \Gamma$ and $G' \colon \Gamma \rightrightarrows \Psi$, we also define the composition carrier $G' \circ F' \colon \Phi \rightrightarrows \Psi$, where each $\sigma \in \Phi$ is sent to

$$G' \circ F'(\sigma) := \bigcup_{\tau \in F(\sigma)} H(\tau).$$

In particular, notice that if f is carried by F' and g is carried by G', then $g \circ f$ is "carried" by $G' \circ F'$. However, this composition does not need to be acyclic.

Example 2.5 Consider a regular morphism $f \colon \Phi \to \Gamma$. We define the (not necessarily acyclic) carrier $F_f \colon \Phi \rightrightarrows \Gamma$ induced by f that sends $\sigma \in \Phi$ to $\overline{f(\sigma)}$. By continuity of f, for any pair $\tau \prec \sigma$ in Φ, we have that $\overline{f(\tau)} \subseteq \overline{f(\sigma)}$. Also, $\overline{f(\sigma)} \neq \varnothing$ since it must contain at least a point. Given an acyclic carrier $G \colon \Gamma \rightrightarrows \Psi$, we denote by $G(f(\sigma))$ the composition of carriers $G \circ F_f(\sigma)$ for all $\sigma \in \Phi$. This comes up very often in this text and whenever we are looking at the composition $G \circ F_f$ we assume that it is acyclic. Note that F_f is acyclic if f is an embedding of the regular CW–complex Φ as a subcomplex of Γ. The hypothesis that f is regular is key to define the carrier F_f. If we considered a more general continuous morphism $f \colon \Phi \to \Gamma$, a possible strategy would be to use outer approximations [Kaczynski et al. 2004; Nanda 2012]. However, for simplicity, we restrict to regular morphisms in this article.

2.3 Regular diagrams of filtered complexes

First, recall a few gluing constructions that one can perform in algebraic topology. For a brief introduction to these, see [Hatcher 2002, Appendix 4.G]. They are also relevant in Kozlov's approach [2008], where diagrams of spaces over trisps are studied.

Let K be a simplicial complex. We view K as a category whose objects are given by the simplices $\sigma \in K$. For any pair of simplices $\tau, \sigma \in K$ such that $\tau \preceq \sigma$, there is a unique arrow $\tau \to \sigma$ in K. We are particularly interested in K^{op}, the *opposite category* of K whose arrows are given by reversing the arrows of K. The example one should have in mind here is the case where K is the nerve of a cover of a cellular complex. Splitting the input data up by the cover then provides a diagram over the nerve where higher intersections of covering sets are included into smaller degree intersections. We formalise these constructions in the following definition.

Definition 2.6 Let K be a simplicial complex. A functor $\mathscr{D} \colon K^{\mathrm{op}} \to \mathbf{CW\text{-}cpx}$ is called a *regular diagram of CW–complexes* and its category is denoted by $\mathbf{RDiag}(K)$; notice here that, for any pair of simplices $\tau \preceq \sigma$ of K, the morphism $\mathscr{D}(\tau \preceq \sigma) \colon \mathscr{D}(\sigma) \to \mathscr{D}(\tau)$ is regular; we call such morphisms *face maps*. Given a pair of diagrams $\mathscr{D}, \mathscr{L} \in \mathbf{RDiag}(K)$, a morphism of diagrams $\varphi \colon \mathscr{D} \to \mathscr{L}$ is a natural transformation; ie the commutativity relation

$$\mathscr{D}(\tau \preceq \sigma) \circ \varphi(\sigma) = \varphi(\tau) \circ \mathscr{D}(\tau \preceq \sigma)$$

holds for any pair $\tau \preceq \sigma$ of simplices in K.

Example 2.7 Let L be a simplicial complex and suppose that it is covered by a pair of nontrivial subcomplexes L_0 and L_1. Consider a pair of vertices $v, w \in L_0 \cap L_1$ and suppose that both are connected by a pair of paths γ_0 and γ_1 within the respective 1–skeletons of L_0 and L_1. Further, we ask that these paths are simple, in the sense that they do not self intersect. Now, consider a diagram $\mathscr{D} \in \mathbf{RDiag}(\Delta^1)$ given by the closures of the paths on the vertices $\mathscr{D}(0) = \bar{\gamma}_0$ and $\mathscr{D}(1) = \bar{\gamma}_1$, while $\mathscr{D}([0, 1]) = \Delta^1$, the standard one-simplex. We define the face maps of $\mathscr{D}$, for $i = 0, 1$, as the regular morphism mapping $0 \mapsto v$ and $1 \mapsto w$, while $[0, 1]$ is sent to γ_i. On the other hand, we consider a diagram $\mathscr{L} \in \mathbf{RDiag}(\Delta^1)$ which is given by the cover $\{L_0, L_1\}$; that is, we define $\mathscr{L}(0) = L_0$ and $\mathscr{L}(1) = L_1$, while $\mathscr{L}([0, 1]) = L_0 \cap L_1$; also, the face maps of $\mathscr{L}$ are given by inclusions. Then, we might consider a morphism of diagrams $\varphi \colon \mathscr{D} \to \mathscr{L}$ given by inclusions $\mathscr{D}(0) \hookrightarrow \mathscr{L}(0)$ and $\mathscr{D}(1) \hookrightarrow \mathscr{L}(1)$, while $\mathscr{D}([0, 1]) = \Delta^1$ is sent to some path within $L_0 \cap L_1$ so that φ is well defined. In fact, φ can only be well defined whenever $\gamma_0 = \gamma_1$. Later, in Definition 5.1, we introduce (ε, K)–acyclic carriers; in this case, one might be able to consider such a carrier $F^\varepsilon \colon \mathscr{D} \Rightarrow \mathscr{L}$ so that γ_0 and γ_1 are only required to lie within some acyclic complex.

The main object of study in this work are diagrams of filtered CW–complexes. These arise naturally in topological data analysis, for example whenever point clouds come equipped with a cover. We therefore make the following definition:

Definition 2.8 A *regularly filtered regular diagram of CW–complexes* $\mathscr{D}$ over K is a functor

$$\mathscr{D} \colon K^{\mathrm{op}} \to \mathbf{RCW\text{-}cpx};$$

we denote this category by $\mathbf{RRDiag}(K)$. As in $\mathbf{RDiag}(K)$, morphisms in $\mathbf{RRDiag}(K)$ are given by natural transformations. We might consider the subcategory of $\mathbf{RRDiag}(K)$ given by functors

$$\mathscr{D} \colon K^{\mathrm{op}} \to \mathbf{FCW\text{-}cpx},$$

which we call *filtered regular diagrams of CW–complexes*, denoting the corresponding category by $\mathbf{FRDiag}(K)$. If for a diagram $\mathscr{D} \in \mathbf{FRDiag}(K)$ the face maps $\mathscr{D}(\tau \prec \sigma)$ are inclusions respecting the cellular structures for all $\tau \prec \sigma$ from K, then we call $\mathscr{D}$ a *fully filtered diagram of CW–complexes*, whose category we denote by $\mathbf{FFDiag}(K)$.

Example 2.9 Consider a filtered CW–complex X covered by filtered subcomplexes $\mathcal{U}$. We define $X^{\mathcal{U}}$ over the nerve $N_{\mathcal{U}}$ as $X^{\mathcal{U}}(\sigma) = \mathcal{U}_\sigma$ for all $\sigma \in N_{\mathcal{U}}$. This diagram $X^{\mathcal{U}}$ is part of $\mathbf{FFDiag}(N_{\mathcal{U}})$ since all morphisms $X^{\mathcal{U}}(\tau \preceq \sigma)$ are actually embeddings of subcomplexes. On the other hand, we can define the constant diagram $*^{\mathcal{U}}$ as $*^{\mathcal{U}}(\sigma)_r = *$ if $X^{\mathcal{U}}(\sigma)_r \neq \varnothing$ or $*^{\mathcal{U}}(\sigma)_r = \varnothing$ otherwise; for all $\sigma \in N_{\mathcal{U}}$ and all $r \in \mathbf{R}$. We also have that $*^{\mathcal{U}}$ is in $\mathbf{FFDiag}(N_{\mathcal{U}})$. Then, there is an obvious epimorphism of diagrams $X^{\mathcal{U}} \to *^{\mathcal{U}}$. Continuing with the same example, we can also define the complex of spaces $\pi_0^{\mathcal{U}}$ given by $\pi_0^{\mathcal{U}}(\sigma) = \pi_0(U_\sigma)$ for all $\sigma \in N_{\mathcal{U}}$; where for each $r \in \mathbf{R}$, $\pi_0(U_\sigma(r))$ denotes the discrete topological space given by the connected components of $U_\sigma(r)$. Thus, each $\pi_0(U_\sigma)$ is a disjoint union of points that are identified with each other as the filtration value increases and so it cannot be an element in $\mathbf{FCW\text{-}cpx}$,

but rather an element from **RCW-cpx**. Thus, in this case $\pi_0^{\,\mathcal{U}} \in \mathbf{RRDiag}(K)$. For all $r \in \mathbf{R}$, there is an epimorphism $X^{\mathcal{U}}(r) \to \pi_0^{\,\mathcal{U}}(r)$ sending each cell from $X^{\mathcal{U}}(r)$ to its respective connected component from $\pi_0^{\,\mathcal{U}}(r)$; these morphisms are consistent along $\mathbf{R}$. Altogether we have a sequence of epimorphisms $X^{\mathcal{U}} \to \pi_0^{\,\mathcal{U}} \to *^{\mathcal{U}}$.

2.4　Geometric realization

For an abstract simplicial complex K, we denote by $|K|$ its underlying topological space. Given a simplex $\sigma \in K$, we write $|\sigma|$ to denote the number of vertices of σ. We use $\dim(\sigma)$ for the dimension of a simplex σ, that is $\dim(\sigma) = |\sigma| - 1$. We denote by Δ^n the topological space associated to the standard n–simplex. Given a simplex $\sigma \in K$, we use the notation $\Delta^\sigma := \Delta^{\dim(\sigma)}$ for simplicity. Given a pair $\tau \prec \sigma$ in K, we have a corresponding inclusion $\Delta^\tau \hookrightarrow \Delta^\sigma$. As a special case of a CW–complex, we denote by K^n and $K^{\leq n}$ the set of n–cells and the n–skeleton respectively.

Definition 2.10 Let $\mathcal{D} \in \mathbf{RDiag}(K)$. The *geometric realization* $\Delta_K \mathcal{D}$ of $\mathcal{D}$ is the object in **CW-cpx** defined as

$$\Delta_K \mathcal{D} = \bigsqcup_{\sigma \in K} \Delta^\sigma \times \mathcal{D}(\sigma) \big/ \sim$$

where, for any pair $\tau \preceq \sigma$ in K, the relation identifies a pair of points

$$(\Delta^\tau \hookrightarrow \Delta^\sigma)(x) \times y \sim x \times \mathcal{D}(\tau \preceq \sigma)(y)$$

for each pair of points $x \in \Delta^\tau$ and $y \in \mathcal{D}(\sigma)$. This $\Delta_K \mathcal{D}$ has a natural filtration given by

$$F^p \Delta_K \mathcal{D} = \bigcup_{\sigma \in K^{\leq p}} \Delta^\sigma \times \mathcal{D}(\sigma)$$

for all $p \geq 0$. A cell $\tau \times c$ is a face of another cell $\sigma \times a$ if and only if $\tau \preceq \sigma$ and also $c \in \overline{\mathcal{D}(\tau \preceq \sigma)(a)}$. If the underlying simplicial complex K is clear from the context, we write $\Delta \mathcal{D}$ instead of $\Delta_K \mathcal{D}$.

Notice that Definition 2.10 also applies to diagrams $\mathcal{D} \in \mathbf{RRDiag}(K)$. We define $\Delta_K \mathcal{D}$ by setting $(\Delta_K \mathcal{D})_r := \Delta_K(\mathcal{D}_r)$ for all $r \in \mathbf{R}$. Notice that our gluing conditions are consistent in this case as

$$\mathcal{D}(\tau \preceq \sigma) \circ \Sigma^t \mathcal{D}(\sigma)(y) = \Sigma^t \mathcal{D}(\tau) \circ \mathcal{D}(\tau \preceq \sigma)(y)$$

for any pair $\tau \preceq \sigma$ from K and all $t > 0$ and all points $y \in \mathcal{D}(\sigma)$. Altogether we obtain $\Delta_K(\mathcal{D}) \in \mathbf{RCW\text{-}cpx}$. Given a regular morphism $\mathcal{F} \colon \mathcal{D} \to \mathcal{L}$ of diagrams in $\mathbf{RRDiag}(K)$, there is an induced morphism on the geometric realization which we denote by $\Delta \mathcal{F}$. Denote by $*^{\mathcal{D}}$ the diagram given by

$$*^{\mathcal{D}}(\sigma)_r = \begin{cases} * & \text{if } \mathcal{D}(\sigma)_r \neq \varnothing, \\ \varnothing & \text{otherwise.} \end{cases}$$

and note that there is a homotopy equivalence $\Delta(*^{\mathcal{D}})_r \simeq |K_r^{\mathcal{D}}|$, where $K^{\mathcal{D}}$ is the filtered simplicial complex with the same underlying vertex set as K and $\sigma \in K_r^{\mathcal{D}}$ if and only if $\mathcal{D}(\sigma)_r \neq \varnothing$. The projection onto the simplex coordinates gives a *base projection* $p_b \colon \Delta \mathcal{D} \to \Delta(*^{\mathcal{D}}) \simeq |K^{\mathcal{D}}|$.

Example 2.11 Let $\mathscr{D} \in \mathbf{FRDiag}(K)$. We define the *multinerve* of $\mathscr{D}$ as

$$\mathrm{MNerv}(\mathscr{D}) = \Delta(\pi_0(\mathscr{D})).$$

This object was first introduced in [Colin de Verdière et al. 2014] in the case of $\pi_0^{\mathcal{U}}$ for a space X covered by $\mathcal{U}$. In [Colin de Verdière et al. 2014] it was defined as a simplicial poset, a notion that is equivalent to that of a Δ–complex. There are epimorphisms $\Delta\mathscr{D} \to \mathrm{MNerv}(\mathscr{D}) \to \Delta(*^{\mathscr{D}}) \simeq |K|$.

Remark 2.12 Let $\mathscr{D}$ be a diagram of CW–complexes over the simplicial complex K. We can extend $\mathscr{D}$ to a diagram $\mathscr{D}'$ on the barycentric subdivision $\mathrm{Bd}(K)$ by defining $\mathscr{D}'(\tau_0 \prec \cdots \prec \tau_n) = \mathscr{D}(\tau_n)$ on an n–simplex $\tau_0 \prec \tau_1 \prec \cdots \prec \tau_n$ in $\mathrm{Bd}(K)$. A nonidentity morphism in $\mathrm{Bd}(K)$ that has $\tau_0 \prec \tau_1 \prec \cdots \prec \tau_n$ as its codomain must have the same flag with one of the τ_k left out as its domain. The diagram $\mathscr{D}'$ maps such a morphism to the identity in case $k \neq n$ or the morphism $\mathscr{D}(\tau_{n-1} \prec \tau_n)$ in case $k = n$. It is clear from the definition of the homotopy colimit via the simplicial replacement that the geometric realization $\Delta(\mathscr{D}')$ coincides with the definition of $\mathrm{hocolim}\,\mathscr{D}$; see [Dugger 2008, Section 4] and also [Kozlov 2008, Definition 15.8]. Notice that in the category K, each flag is to be interpreted as a sequence of arrows $\tau_0 \leftarrow \tau_1 \leftarrow \cdots \leftarrow \tau_n$. A modified version of the homotopy equivalence $|K| \simeq |\mathrm{Bd}(K)|$ shows that $\Delta(\mathscr{D}) \simeq \Delta(\mathscr{D}')$. Hence, we could have worked with homotopy colimits all throughout, but we chose to work with the geometric realization since it is technically easier to handle and because in some instances it is the Mayer–Vietoris blowup complex, which has already appeared before in TDA [Zomorodian and Carlsson 2008]. An instance of a homotopy colimit in TDA can be found in [Cavanna et al. 2017, Appendix B].

Proposition 2.13 *Let $\mathscr{F}: \mathscr{D} \to \mathscr{L}$ be a morphism of diagrams in $\mathbf{RDiag}(K)$. If $\mathscr{F}(\sigma)$ is a homotopy equivalence for all $\sigma \in K$, then $\Delta\mathscr{F}: \Delta\mathscr{D} \to \Delta\mathscr{L}$ is a homotopy equivalence.*

One way to see this is to view $\Delta\mathscr{D}$ as a homotopy colimit (see Remark 2.12), which is a homotopy invariant functor on diagrams. Also, a proof of this result in the more general context of diagrams of spaces can be found in [Hatcher 2002, Proposition 4G.1].

Example 2.14 Let $X \in \mathbf{CW\text{-}cpx}$ covered by $\mathcal{U}$ and recall the diagram $X^{\mathcal{U}}$ from Example 2.9. In this case $\Delta(X^{\mathcal{U}})$ is the Mayer–Vietoris blowup complex associated to the pair $(X, \mathcal{U})$ and it can be described as a subspace of the product $X \times |N_{\mathcal{U}}|$. This leads to the *fiber projection* $p_f: \Delta(X^{\mathcal{U}}) \to X$ and to the *base projection* $p_b: \Delta(X^{\mathcal{U}}) \to |N_{\mathcal{U}}|$. As shown in [Hatcher 2002, Proposition 4G.2], p_f is a homotopy equivalence $\Delta(X^{\mathcal{U}}) \simeq X$. If each $X^{\mathcal{U}}(\sigma)$ is contractible for all $\sigma \in N_{\mathcal{U}}$, then p_b is also a homotopy equivalence by Proposition 2.13.

An interesting direction of research would be to use Proposition 2.13 to define compatible *collapses*, such as in discrete Morse theory (see [Bauer 2011; Nanda 2012; Sköldberg 2006]) and end up with a diagram of regular CW–complexes. This motivates the study of spectral sequences associated to such diagrams. We see further reasons in Section 3. On the other hand, given the importance of Proposition 2.13, we

would like to adapt it to an approximate version in the context of diagrams in **RRDiag**(K). Instead of studying homotopy equivalences, we consider equivalences induced by acyclic carriers. This is done in Section 5.

2.5 Spectral sequences of bounded filtrations

Let A_* be a graded module with differentials $d_n\colon A_n \to A_{n-1}$ for all $n \geq 1$, and such that $A_m = 0$ for all $m < 0$. Assume that there is a filtration $0 = F^{-1}A_* \subseteq F^0 A_* \subseteq F^1 A_* \subseteq \cdots \subseteq F^N A_* = A_*$ of A_* that is preserved by the differentials d_* in the sense that $d_n(F^p A) \subseteq F^p A$ for all $p \geq 0$. We say that A_* is a *filtered differential graded module* and denote this by the triple (A, d, F). Then there is a spectral sequence

$$E^1_{p,q} = \mathrm{H}_q(F^p A_*/F^{p-1}A_*) \Rightarrow \mathrm{H}_{p+q}(A_*)$$

for all $p, q \geq 0$; see [McCleary 2001, Theorem 2.6]. A morphism of spectral sequences is a sequence of bigraded morphisms $f^r\colon E^r_{p,q} \to \bar{E}^r_{p,q}$ that commute with the spectral sequence differentials, ie $d_r \circ f^r = d_r \circ f^r$ for all $r \geq 0$. Apart from that, these morphisms satisfy $f^{r+1} = H(f^r)$ for all $r \geq 0$.

Suppose that $(\bar{A}_*, \bar{d}, \bar{F})$ is another filtered differential graded module together with its corresponding spectral sequence $\bar{E}^r_{p,q}$. Consider a morphism $f\colon A_* \to B_*$ that commutes with the differential $f \circ d = \bar{d} \circ f$ and also preserves filtrations $f(F^p A_*) \subseteq \bar{F}^p(\bar{A}_*)$ for all $p \geq 0$. This induces a morphism of spectral sequences

$$E^r_{p,q} \to \bar{E}^r_{p,q}$$

by [McCleary 2001, Theorem 3.5]. We denote by **SpSq** the category of spectral sequences, and we denote by **PSpSq** the category of functors $F\colon \mathbf{R} \to \mathbf{SpSq}$.

3 Spectral sequences for geometric realizations

Recall the persistent Mayer–Vietoris spectral sequence [Torras-Casas 2023] associated to a pair $(X, \mathcal{U})$ of a space with a cover:

(1) $$E^1_{p,q}(X, \mathcal{U}) = \bigoplus_{\sigma \in N^p_{\mathcal{U}}} \mathrm{PH}_q(X^{\mathcal{U}}(\sigma)) \Rightarrow \mathrm{PH}_{p+q}(\Delta X^{\mathcal{U}}) \simeq \mathrm{PH}_{p+q}(X).$$

For the details about this spectral sequence in the persistent case we refer the reader to [Torras-Casas 2023]. There are some limitations to the applicability of this spectral sequence to Vietoris–Rips complexes that were already pointed out in [Yoon and Ghrist 2020]: if we choose a cover of a point cloud $\mathbb{X}$ and then deduce a cover $\mathcal{U}$ of the associated Vietoris–Rips complex $\mathrm{VR}_*(\mathbb{X})$ by subcomplexes, then we can only recover $\mathrm{PH}_k(\mathrm{VR}(\mathbb{X}))$ from $\mathrm{PH}_k(\Delta \mathrm{VR}_*(\mathbb{X})^{\mathcal{U}})$ for filtration parameters below an upper bound R determined by the overlaps of the covering sets. In this section we present a regular diagram of CW–complexes that avoids this upper limit problem completely; see Example 3.6.

Before we solve our problem, we need to introduce some chain complexes. We come back to the case of filtrations later, but for now we focus on regular diagrams instead. Given a diagram $\mathcal{D}$ in **RDiag**(K), we denote by $\mathcal{D}(\tau \preceq \sigma)_*$ the induced morphism of cellular chain complexes $C_*^{\text{cell}}(\mathcal{D}(\sigma)) \to C_*^{\text{cell}}(\mathcal{D}(\tau))$. The cellular chain complex $C_*^{\text{cell}}(\Delta\mathcal{D}, \delta^\Delta)$ associated to $\Delta\mathcal{D}$ is defined as follows: For all $m \geq 0$ we have that $C_m^{\text{cell}}(\Delta\mathcal{D})$ is a vector space generated by cells $\sigma \times c$ with $\dim(\sigma) = p$ and $c \in \mathcal{D}(\sigma)_q$ so that $p + q = m$. On such a cell $\sigma \times c$ the differential δ^Δ is given by

$$\sum_{\sigma_i \prec \sigma} (-1)^i \left(\sum_{a \in \mathcal{D}(\sigma_i \preceq \sigma)(c)} [a : \mathcal{D}(\sigma_i \preceq \sigma)(c)] \sigma_i \times a \right) + (-1)^{\dim(\sigma)} \sum_{b \in \bar{c} \setminus c} [b : c] \sigma \times b$$

where the first sum runs over the faces σ_i of σ. As shown in the proof of Lemma 3.1, the map δ^Δ is indeed a differential. In addition, notice that the filtration of $\Delta_K(\mathcal{D})$ carries over to $C_*^{\text{cell}}(\Delta_K\mathcal{D})$ by taking $F^p C_*(\Delta_K\mathcal{D}) := C_*(F^p \Delta_K\mathcal{D})$ for all $p \geq 0$.

Now, consider the double complex $(C_{p,q}(\mathcal{D}), d^V, d^H)$ given by

$$C_{p,q}(\mathcal{D}) = \bigoplus_{\sigma \in K^p} C_q^{\text{cell}}(\mathcal{D}(\sigma))$$

for all $p, q \geq 0$. The vertical differential is defined by the direct sum of chain differentials

$$d_{p,q}^V = (-1)^p \bigoplus_{\sigma \in K^p} d_q^\sigma$$

where d_*^σ denotes the differential from $C_*^{\text{cell}}(\mathcal{D}(\sigma))$ for all $\sigma \in K^p$; of course $d^V \circ d^V = 0$. The horizontal differential is given by the Čech differential $d_{p,q}^H$ which is defined for a cell $a \in \mathcal{D}(\sigma)$ as

$$\sum_{\sigma_i \prec \sigma} (-1)^i \mathcal{D}(\sigma_i \prec \sigma)_*(a),$$

where $\mathcal{D}(\sigma_i \prec \sigma)_*$ denotes the induced chain morphism $C_*^{\text{cell}}(\mathcal{D}(\sigma)) \to C_*^{\text{cell}}(\mathcal{D}(\sigma_i))$ for all faces σ_i from σ. Also, $d^H \circ d^H = 0$ by functoriality of $C_*^{\text{cell}}(\cdot)$ and the fact that $\mathcal{D}(\rho \prec \tau)_* \mathcal{D}(\tau \prec \sigma)_* = \mathcal{D}(\rho \prec \sigma)_*$ for any three simplices $\rho \prec \tau \prec \sigma$. Note that for each pair of indices $i < j$, the face map $\mathcal{D}(\sigma_{ij} \preceq \sigma)_*$ appears twice with respective coefficients $(-1)^i (-1)^j$ and $(-1)^i (-1)^{j-1}$; which have opposite sign and cancel out. On the other hand, anticommutativity $d^V \circ d^H = -d^H \circ d^V$ follows since $\mathcal{D}(\tau \prec \sigma)_*$ is a chain morphism for all $\tau \prec \sigma$ from K.

Now, we consider the double complex spectral sequence from [McCleary 2001, Section 2.4]. Given $\mathcal{D}$ in **RDiag**(K) there is a spectral sequence

$$E_{p,q}^1(\mathcal{D}) = \bigoplus_{\sigma \in K^p} H_q(\mathcal{D}(\sigma)) \Rightarrow H_{p+q}(S_*^{\text{Tot}}(\mathcal{D}))$$

where $S^{\text{Tot}}(\mathcal{D})$ is the *total complex* defined as

$$S_n^{\text{Tot}}(\mathcal{D}) = \bigoplus_{p+q=n} C_{p,q}(\mathcal{D})$$

together with a differential $d^{\mathrm{Tot}} = d^V + d^H$. Also, recall that the total complex has a filtration induced by the vertical filtration on $C_{p,q}(\mathcal{D})$ given by

$$F^m S_*^{\mathrm{Tot}}(\mathcal{D}) = \bigoplus_{\substack{p+q=n \\ p \leq m}} C_{p,q}(\mathcal{D})$$

for all integers $m \geq 0$; see [Torras-Casas 2023] for an explanation. Next, we relate this total complex to the geometric realization from Definition 2.10.

Lemma 3.1 *There is an isomorphism $C_*^{\mathrm{cell}}(\Delta\mathcal{D}, \delta^\Delta) \simeq S_*^{\mathrm{Tot}}(\mathcal{D})$ which preserves filtration. That is, $F^p C_*^{\mathrm{cell}}(\Delta\mathcal{D}, \delta^\Delta) \simeq F^p S_*^{\mathrm{Tot}}(\mathcal{D})$ for all $p \geq 0$.*

Proof First we define a chain morphism $\psi \colon C_m^{\mathrm{cell}}(\Delta\mathcal{D}) \to S_m^{\mathrm{Tot}}(\mathcal{D})$ generated by the assignment: a cell $\sigma \times c \in (\Delta\mathcal{D})_m$ with $\sigma \in K^p$ and $c \in \mathcal{D}(\sigma)^q$ for integers $p+q = m$, is sent to $\psi(\sigma \times c) = (c)_\sigma \in S_m^{\mathrm{Tot}}(\mathcal{D})$; where by $(c)_\sigma$ we refer to the vector from $S_m^{\mathrm{Tot}}(\mathcal{D})$ which is zero in all components except at the component indexed by σ, where it is equal to c. On the other hand, ψ is a chain morphism since we have the equality

$$\begin{aligned}
\psi(\delta^\Delta(\sigma \times c)) &= \sum_{\sigma_i \prec \sigma} (-1)^i \left(\sum_{a \in \mathcal{D}(\sigma_i \leq \sigma)(c)} ([a : \mathcal{D}(\sigma_i \leq \sigma)(c)]a)_{\sigma_i} \right) + (-1)^{\dim(\sigma)} \sum_{b \in \bar{c} \setminus c} ([b : c]b)_\sigma \\
&= \sum_{\sigma_i \prec \sigma} (-1)^i (\mathcal{D}(\sigma_i \leq \sigma)_*(c))_{\sigma_i} + (-1)^{\dim(\sigma)} (d_q^\sigma(c))_\sigma \\
&= (d^H + d^V)((c)_\sigma) \\
&= d^{\mathrm{Tot}}((c)_\sigma).
\end{aligned}$$

One can see that ψ is injective, and admits an inverse $\psi^{-1} \colon S_m^{\mathrm{Tot}}(\mathcal{D}) \to C_m^{\mathrm{cell}}(\Delta\mathcal{D})$ that sends $(\sigma)_c$ to $\sigma \times c$. Notice that by definition ψ sends a chain in $F^p C_n^{\mathrm{cell}}(\Delta\mathcal{D})$ to a chain in $F^p S_n^{\mathrm{Tot}}(\mathcal{D})$ for all $p \geq 0$ and in particular it preserves filtration. □

Remark 3.2 Continuing with Remark 2.12, as both $\Delta_{\mathrm{Bd}(K)}\mathcal{D}'$ and hocolim$(\mathcal{D})$ refer to the same space, we could have considered the homotopy colimit spectral sequence

$$E_{p,q}^1(\mathrm{Bd}(K), \mathcal{D}') = \bigoplus_{\sigma \in \mathrm{Bd}(K)^p} \mathrm{H}_q(\mathcal{D}'(\sigma)) \Rightarrow \mathrm{H}_{p+q}(\mathrm{hocolim}\,\mathcal{D}).$$

Let us construct a diagram of spaces whose geometric realization is homeomorphic to $|K|$ for any finite simplicial complex K. We start by taking a finite partition $\mathcal{P}$ of the vertex set $V(K)$ and denote by $K(U)$ the maximal subcomplex of K with vertices in $U \in \mathcal{P}$. We denote by $\Delta^{\mathcal{P}}$ the standard simplex with vertices in $\mathcal{P}$. For a simplex $\tau \in K$, we define $\mathcal{P}(\tau) \in \Delta^{\mathcal{P}}$ to be the simplex consisting of all partitioning sets $U \in \mathcal{P}$ such that $\tau \cap U \neq \varnothing$. In particular if $U \in \mathcal{P}(\tau)$, then it determines a standard simplex $\tau(U) \in K(U)$ of dimension $|\tau \cap U| - 1 \geq 0$ whose vertices are precisely those from $\tau \cap U$, so that there is an inclusion $\Delta^{\tau(U)} \hookrightarrow |K(U)|$. For a vertex $v \in K$, we denote by $\mathcal{P}(v)$ the partitioning set from $\mathcal{P}$ which contains v.

We define the $(K, \mathcal{P})$–*join* diagram $\mathcal{J}_{\mathcal{P}}^K : (\Delta^{\mathcal{P}})^{\mathrm{op}} \to$ **FCW-cpx** for all $\sigma \subseteq \mathcal{P}$ by assigning the subspace formed by the union of products of images

$$\mathcal{J}_{\mathcal{P}}^K(\sigma) = \bigcup_{\substack{\rho \in K \\ \mathcal{P}(\rho) = \sigma}} \prod_{U \in \sigma} \mathrm{Im}(\Delta^{\rho(U)} \hookrightarrow |K(U)|)$$

for all $\sigma \in \Delta^{\mathcal{P}}$; by definition, notice that $\mathcal{J}_{\mathcal{P}}^K(\sigma) \subseteq \prod_{U \in \sigma} |K(U)|$. Additionally, $\mathcal{J}_{\mathcal{P}}^K(U) = |K(U)|$ for all $U \in \mathcal{P}$. However, $\mathcal{J}_{\mathcal{P}}^K(\sigma)$ could even be empty for $\sigma \in \Delta^{\mathcal{P}}$ with $\dim(\sigma) > 0$. For any pair $\tau \preceq \sigma$ in $\Delta^{\mathcal{P}}$, we consider the projection $\pi_{\tau \preceq \sigma} : \prod_{U \in \sigma} |K(U)| \to \prod_{U \in \tau} |K(U)|$ that forgets all product components which are indexed by vertices of σ that are not vertices of τ. We claim that $\pi_{\tau \preceq \sigma}$ restricts to a well-defined face map $\mathcal{J}_{\mathcal{P}}^K(\tau \preceq \sigma) : \mathcal{J}_{\mathcal{P}}^K(\sigma) \to \mathcal{J}_{\mathcal{P}}^K(\tau)$. In order to show this, we consider an arbitrary simplex $\rho \in K$ such that $\mathcal{P}(\rho) = \sigma$. Next, we consider the face $\lambda(\tau) \preceq \rho$ which is spanned by the vertices from $\rho \cap U$ for all $U \in \tau$, so that $\mathcal{P}(\lambda(\tau)) = \tau$ and also $\lambda(\tau)(U) = \rho(U)$ for all $U \in \mathcal{P}$. Then, we obtain the equality

$$\pi_{\tau \preceq \sigma}\left(\prod_{U \in \sigma} \mathrm{Im}(\Delta^{\rho(U)} \hookrightarrow |K(U)|) \right) = \prod_{U \in \tau} \mathrm{Im}(\Delta^{\lambda(\tau)(U)} \hookrightarrow |K(U)|),$$

so the face maps are well defined, as claimed.

Lemma 3.3 *Let K be a simplicial complex together with a partition $\mathcal{P}$ of its vertex set $V(K)$. There is a CW–complex homeomorphism $\Delta(\mathcal{J}_{\mathcal{P}}^K) \simeq |K|$.*

Proof Consider the continuous map $f : \Delta(\mathcal{J}_{\mathcal{P}}^K) \to |K|$ defined by mapping a point

$$\left(\sum_{U \in \mathcal{P}(\tau)} y_U U, \left(\sum_{v \in U} x_v v \right)_{U \in \mathcal{P}(\tau)} \right) \in \Delta^{\mathcal{P}(\tau)} \times \prod_{U \in \mathcal{P}(\tau)} \Delta^{\tau(U)} \Big/ \sim$$

to $\sum_{v \in \tau} y_{\mathcal{P}(v)} x_v v$ in Δ^τ for all $\tau \in K$, where we have values $0 \le y_U \le 1$ and $0 \le x_v \le 1$ for all $U \in \mathcal{P}(\tau)$ and all $v \in U$, and such that $\sum_{U \in \mathcal{P}(\tau)} y_U = 1$ and $\sum_{v \in U} x_v = 1$ for all $U \in \mathcal{P}$. On the other hand, let $\sum_{v \in \tau} x_v v \in \Delta^\tau$ be a point such that $0 \le x_v \le 1$ for all $v \in \Delta^\tau$ and such that $\sum_{v \in \tau} x_v = 1$. Then we can define the inverse continuous map

$$f^{-1}\left(\sum_{v \in \tau} x_v v \right) = \left(\sum_{U \in \mathcal{P}(\tau)} \left(\sum_{v \in U} x_v \right) U, \left(\psi_U\left(\sum_{v \in \tau} x_v v \right) \right)_{U \in \mathcal{P}(\tau)} \right),$$

where we consider a map $\psi_U : \Delta^\tau \to \Delta^{\tau(U)}$ given by

$$\psi_U\left(\sum_{v \in \tau} x_v v \right) = \begin{cases} \sum_{v \in \tau(U)} \left(\dfrac{x_v}{\sum_{v \in \tau(U)} x_v} \right) v & \text{if } \sum_{v \in \tau(U)} x_v \neq 0, \\ * \in \Delta^{\tau(U)} & \text{otherwise, where } * \text{ denotes any point (see below).} \end{cases}$$

By the equivalence relation used to define $\Delta(\mathcal{J}_{\mathcal{P}}^K)$, the product factor $\Delta^{\tau(U)}$ is collapsed to a single point for the subset of points whose U–coordinate in $\Delta^{\mathcal{P}(\tau)}$ vanishes. If $\sum_{v \in \tau(U)} x_v = 0$, then $x_v = 0$ for all $v \in \tau(U)$ and the U–coordinate of the point $\sum_{v \in \tau} x_v v$ in $\Delta^{\mathcal{P}(\tau)}$ is 0. It is straightforward to check that f and f^{-1} are well defined and consistent along K. $\qquad \square$
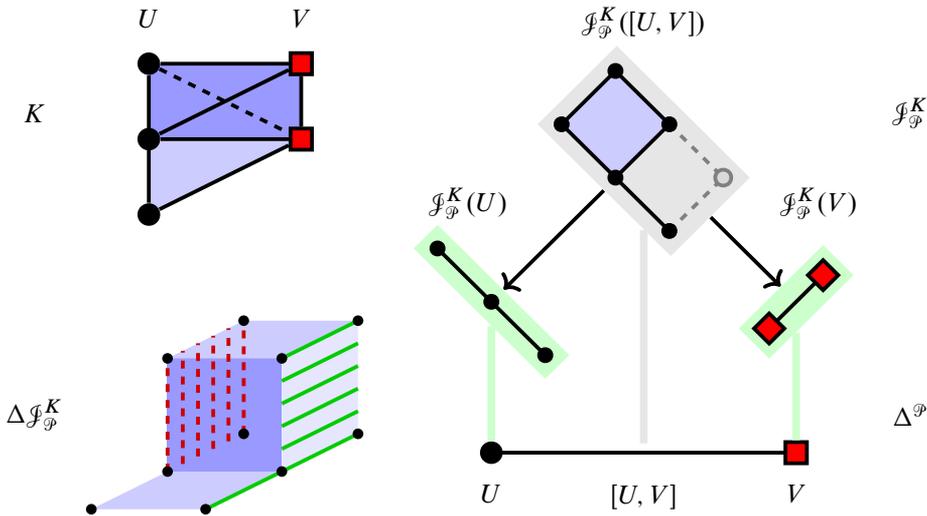
Figure 1: Depiction of K, $\mathcal{J}_{\mathcal{P}}^K$ and $\Delta\mathcal{J}_{\mathcal{P}}^K$ from Example 3.4. Over the edge $[U, V]$, we consider $\mathcal{J}_{\mathcal{P}}^K([U, V]) \subset |K(U)| \times |K(V)|$, where we add dashed lines to illustrate the embedding into the product. On the bottom left we depict $\Delta\mathcal{J}_{\mathcal{P}}^K$, where each red dashed line and each green line is collapsed to a single point.

Example 3.4 Consider the simplicial complex K depicted in the top left part of Figure 1, formed by gluing a 2–simplex to a 4–simplex along an edge. We consider a partition of the vertex set of K into the two subsets $\mathcal{P} = \{U, V\}$, where points in U are indicated by black circles and points in V are indicated by red squares. In the top right of Figure 1, we depict the standard 1–simplex $\Delta^{\mathcal{P}}$ together with the diagram $J_{\mathcal{P}}^K$ over it. In particular, notice that $J_{\mathcal{P}}^K([U, V])$ is a subset of the product $|K(U)| \times |K(V)|$ and that the morphisms $J_{\mathcal{P}}^K([U, V]) \to J_{\mathcal{P}}^K(V) = |K(V)|$ and $J_{\mathcal{P}}^K([U, V]) \to J_{\mathcal{P}}^K(U) = |K(U)|$ are both projections. In addition, notice that $\mathcal{J}_{\mathcal{P}}^K([U, V])$ has five vertices corresponding to the five different edges connecting vertices from U to V, five edges corresponding to five 2–simplices containing vertices in both U and V and a single 2–cell corresponding to the unique 4–simplex in K. Finally, the bottom left of Figure 1 shows the geometric realization $\Delta J_{\mathcal{P}}^K$.

Observe that $\mathcal{J}_{\mathcal{P}}^K$ is a diagram of *prodsimplicial* complexes as in [Kozlov 2008, Definition 2.43], which are in particular regular CW–complexes. By the observations above we can therefore consider the associated double complex spectral sequence

$$E_{p,q}^1(\mathcal{J}_{\mathcal{P}}^K) = \bigoplus_{\sigma \in \Delta^{\mathcal{P}}} \mathrm{H}_q(\mathcal{J}_{\mathcal{P}}^K(\sigma)) \Rightarrow \mathrm{H}_{p+q}(\Delta\mathcal{J}_{\mathcal{P}}^K) \simeq \mathrm{H}_{p+q}(K).$$

Next, we show that the "size" of K is the same as the "size" of the diagram $\mathcal{J}_{\mathcal{P}}^K$. For this, recall that each simplex $\sigma \in K$ corresponds to a unique simplex $\mathcal{P}(\sigma) \in \Delta^{\mathcal{P}}$. This is different to the case of a cover $\mathcal{U}$ for K, where a simplex in K might correspond to several simplices from the nerve $N_{\mathcal{U}}$. Here, we write $\#L$ for the number of cells in a complex L.

Proposition 3.5
$$\#K = \sum_{\sigma \in \Delta^{\mathscr{P}}} \#\mathcal{J}_{\mathscr{P}}^K(\sigma).$$

Proof Consider an assignment $\psi \colon K \to \bigsqcup_{\sigma \in \Delta^{\mathscr{P}}} \mathcal{J}_{\mathscr{P}}^K(\sigma)$ given by sending $\rho \in K$ to $(\rho(U))_{U \in \mathscr{P}(\rho)}$ in $\mathcal{J}_{\mathscr{P}}^K(\mathscr{P}(\rho))$, where $(\rho(U))_{U \in \mathscr{P}(\rho)} \in \prod_{U \in \mathscr{P}(\rho)} |K(U)|$. By the definition of $\mathcal{J}_{\mathscr{P}}^K$, ψ is well defined and surjective. Also, ψ is injective as the vertex set from $\rho \in K$ is uniquely determined by the simplices $\rho(U)$ for all $U \in \mathscr{P}(\rho)$. $\qquad\square$

Now, let us consider a filtered simplicial complex $K_* \in$ **FCW-cpx** such that its vertex set $V(K_*)$ is fixed throughout all values of **R**. Let $\mathscr{P}$ be a partition of $V(K_*)$. We define the filtered regular diagram $\mathcal{J}_{\mathscr{P}}^K \in$ **FRDiag**$(\mathscr{P})$ by sending $r \in$ **R** to $\mathcal{J}_{\mathscr{P}}^{K_r}$. These diagrams inherit the shift morphisms ΣK_* from K_* in the following way: Let $\sigma \in \Delta^{\mathscr{P}}$ and notice that we have restrictions $\Sigma^{s-r} K_{|U} \colon |K_r(U)| \to |K_s(U)|$ for all $U \in \sigma$, so that we have induced morphisms

$$\prod_{U \in \sigma} \Sigma^{s-r} K_{|U} \colon \mathcal{J}_{\mathscr{P}}^{K_r}(\sigma) \to \mathcal{J}_{\mathscr{P}}^{K_s}(\sigma)$$

for all $\sigma \in \Delta^{\mathscr{P}}$. In turn, these induce a shift morphism on $\Delta \mathcal{J}_{\mathscr{P}}^K$ which respect filtrations, so that we have a commutative diagram

$$\begin{array}{ccccc}
E_{p,q}^*(\mathcal{J}_{\mathscr{P}}^{K_r}) & \Longrightarrow & \Delta \mathcal{J}_{\mathscr{P}}^{K_r} & \xrightarrow{\simeq} & K_r \\
\downarrow & & \downarrow & & \downarrow \\
E_{p,q}^*(\mathcal{J}_{\mathscr{P}}^{K_s}) & \Longrightarrow & \Delta \mathcal{J}_{\mathscr{P}}^{K_s} & \xrightarrow{\simeq} & K_s
\end{array}$$

and thus $\mathrm{PH}_*(\Delta \mathcal{J}_{\mathscr{P}}^K) \simeq \mathrm{PH}_*(K_*)$. For each simplex $\sigma \in \Delta^{\mathscr{P}}$ one can see $\mathcal{J}_{\mathscr{P}}^K(\sigma)$ as a filtered simplicial complex, so that

$$E_{p,q}^1(\mathcal{J}_{\mathscr{P}}^K) = \bigoplus_{\sigma \in (\Delta^{\mathscr{P}})^p} \mathrm{PH}_q(\mathcal{J}_{\mathscr{P}}^K(\sigma)) \Rightarrow \mathrm{PH}_{p+q}(K).$$

Example 3.6 Consider a point cloud $\mathbb{X}$, a partition $\mathscr{P}$ and consider its Vietoris Rips complex $\mathrm{VR}_*(\mathbb{X})$ in **FCW-cpx**. In this case we have a fixed partition of the vertex set of $\mathrm{VR}_*(\mathbb{X})$, which allows us to consider the spectral sequence

$$E_{p,q}^1(\mathcal{J}_{\mathscr{P}}^{\mathrm{VR}_*(\mathbb{X})}) = \bigoplus_{\sigma \in \Delta^{\mathscr{P}}} \mathrm{PH}_q(\mathcal{J}_{\mathscr{P}}^{\mathrm{VR}_*(\mathbb{X})}(\sigma)) \Rightarrow \mathrm{PH}_{p+q}(\mathrm{VR}_*(\mathbb{X})).$$

This is very convenient as it avoids the main difficulties with the Mayer–Vietoris blowup complex associated to a cover. Namely, one recovers $\mathrm{PH}_*(K)$ completely without any bounds depending on the cover overlaps. In addition, notice that $\Delta \mathcal{J}_{\mathscr{P}}^{\mathrm{VR}_*(\mathbb{X})}$ has the same number of cells as $\mathrm{VR}_*(\mathbb{X})$, contrary to the Mayer–Vietoris blowup complex, whose number of cells is much larger, as shown in Proposition 3.5.

The $(K, \mathscr{P})$–join diagram is related to [Robinson 2020, Example 4]. There the motivation behind the filtrations is given by a consistency radius and a filtration based on the differences between local measurements. The same example appears (without a filtration) as one of the opening examples in [Hatcher 2002, Appendix 4.G].

4 ε–acyclic carriers

The following definition encodes our notion of "noise".

Definition 4.1 Let $X, Y \in \mathbf{RCW}\text{-}\mathbf{cpx}$. An ε–acyclic carrier $F_*^\varepsilon \colon X_* \rightrightarrows Y[\varepsilon]_*$ is a family of acyclic carriers $F_a^\varepsilon \colon X_a \rightrightarrows Y_{a+\varepsilon}$ for all $a \in \mathbf{R}$ such that

$$Y(a + \varepsilon \leq b + \varepsilon) F_a^\varepsilon(c) \subseteq F_b^\varepsilon(X(a \leq b)(c))$$

for all cells c of X_a and $a, b \in \mathbf{R}$ with $a \leq b$.

The proposition below is an adaptation of [Munkres 1984, Theorem 13.4] or [Cooke and Finney 1967, Lemma 2.4] to the context of tame filtered CW–complexes.

Proposition 4.2 *Let $X_*, Y_* \in \mathbf{FCW}\text{-}\mathbf{cpx}$ be tame, and assume that there exists an ε–acyclic carrier*

$$F_*^\varepsilon \colon X_* \rightrightarrows Y[\varepsilon]_*.$$

Then there exist chain morphisms $f_a^\varepsilon \colon C_*(X_a) \to C_*(Y_{a+\varepsilon})$ carried by F_a^ε for all $a \in \mathbf{R}$, so that $Y(a + \varepsilon \leq b + \varepsilon) \circ f_a^\varepsilon = f_b^\varepsilon \circ X(a \leq b)$. Furthermore, given another such sequence of morphisms $g_a^\varepsilon \colon C_*(X_a) \to C_*(Y_{a+\varepsilon})$, there exist chain homotopy equivalences $H_a^\varepsilon \colon g_a^\varepsilon \simeq f_a^\varepsilon$ which are carried by F_a^ε for all $a \in \mathbf{R}$.

Proof Let $b \in \mathbf{R}$ and assume that f_a^ε has already been defined for all values $a < b$, where we allow for $b = -\infty$. We first define f_b^ε on all cells which are in the image of $X(a < b)$ for any $a < b$ using the definition

$$f_b^\varepsilon \circ X(a < b) = Y(a + \varepsilon < b + \varepsilon) \circ f_a^\varepsilon.$$

Notice that the assumption that $X_a \subseteq X_b$ is crucial for this to work. By hypotheses, given a cell $c \in \mathrm{Im}(X(a < b))$, its image $f_b^\varepsilon(c)$ is then contained in

$$Y(a + \varepsilon < b + \varepsilon) F_a^\varepsilon(\tilde{c}) \subseteq F_b^\varepsilon(X(a < b)(\tilde{c})),$$

where $\tilde{c} \in X_a$ is such that $c = X(a < b)(\tilde{c})$. Hence, f_b^ε satisfies the carrier condition. Next we define f_b^ε on the remaining cells in

$$\tilde{X}_b = X_b \setminus \left(\bigcup_{a < b} X(a < b) \right).$$

We proceed to prove this by induction. First, choose a 0–cell $f_b^\varepsilon(v) \in F_b^\varepsilon(v)$ for each remaining 0–cell $v \in \tilde{X}_b$, and notice that $d_* f_b^\varepsilon(v) = 0 = f_b^\varepsilon(d_* v)$, where we use d_* for the chain complex differentials. Next, by induction, assume that for a fixed $p \geq 0$, the p–cells $s \in X_b$ have image $f_b^\varepsilon(s)$ carried by $F_b^\varepsilon(s)$ and such that $d_* \circ f_b^\varepsilon(s) = f_b^\varepsilon \circ d_*(s)$. We would like to extend f_b^ε to the $(p+1)$–cells. By semicontinuity, given such a cell $c \in X_b$, its boundary $d_* c$ is contained in $F_b^\varepsilon(c)$. On the other hand, notice that by linearity and the induction hypotheses $d_* f_b^\varepsilon(d_* c) = f_b^\varepsilon(d_* d_* c) = 0$; thus $f_b^\varepsilon(d_* c)$ is a

cycle in $C_*(F_b^\varepsilon(c))$. By acyclicity, there exists $h \in C_*(F_b^\varepsilon(c))$ such that $d_* h = f_b^\varepsilon(d_* c)$ and thus we can define $f_b^\varepsilon(c) = h$. Altogether, we have defined a chain morphism f_b^ε which is carried by F_b^ε.

Since X_* is tame, there exist a finite sequence of values $a_1 < a_2 < \cdots < a_N$ such that $X_s = X_{a_i}$ for all $s \in (a_{i-1}, a_i)$ where we define $a_0 = -\infty$ and $a_{N+1} = \infty$. We apply the construction of f_b^ε for all values b ranging over a_i from $i = 1$ up to $i = N$. This determines the chain morphism $f_*^\varepsilon \colon C_*(X_*) \to C_*(Y[\varepsilon]_*)$, where we set $f_s^\varepsilon = f_{a_i}^\varepsilon$ for all $s \in (a_{i-1}, a_i]$ where $i = 1, 2, \dots, N$ and also $f_t^\varepsilon = f_{a_N}^\varepsilon$ for all $t > a_N$.

Now, assume that g_b^ε is also carried by F_b^ε for all $b \in \mathbb{R}$. Following [May 1999, Section 12.3], we define the chain complex $\mathcal{I}$ given by $\mathcal{I}_0 = \langle [0], [1] \rangle$ and $\mathcal{I}_1 = \langle [0, 1] \rangle$ and $\mathcal{I}_k = 0$ for $k \geq 2$. This is the cellular chain complex of the unit interval I decomposed into two 0–cells and one 1–cell. A chain homotopy $h_b^\varepsilon \colon f_b^\varepsilon \simeq g_b^\varepsilon$ corresponds to a chain map $h_b^\varepsilon \colon C_*^{\text{cell}}(X_b) \otimes \mathcal{I} \to C_*^{\text{cell}}(Y_b)$ such that $h_b^\varepsilon(x, [0]) = f_b^\varepsilon(x)$ and $h_b^\varepsilon(x, [1]) = g_b^\varepsilon(x)$ for all $x \in X_b$. Let $H_b^\varepsilon(c, i) = F_b^\varepsilon(c)$ for a cell $(c, i) \in X \times I$. By assumption, $H^\varepsilon \colon X \times I \Rightarrow Y$ is an ε–acyclic carrier. Note that $C_*^{\text{cell}}(X_b) \otimes \mathcal{I} \cong C_*^{\text{cell}}(X_b \times I)$. Replicating the first part of the proof we can now extend any map $h_b^\varepsilon \colon C_*^{\text{cell}}(X_b) \otimes \mathcal{I}_0 \to C_*^{\text{cell}}(Y_b)$ with the above properties to all cells of $X \times I$. $\qquad \square$

Definition 4.3 Let $X_*, Y_* \in \mathbf{RCW\text{-}cpx}$. A *shift* carrier is an ε–acyclic carrier $I_X^\varepsilon \colon X_* \Rightarrow X_{*+\varepsilon}$ carrying the standard shift $\Sigma^\varepsilon X_*$. Let two ε–acyclic carriers

$$F^\varepsilon \colon X_* \Rightarrow Y_{*+\varepsilon}, \quad G^\varepsilon \colon Y_* \Rightarrow X_{*+\varepsilon},$$

together with shift carriers $I_X^{2\varepsilon}$ and $I_Y^{2\varepsilon}$. We say that X_* and Y_* are *ε–acyclic equivalent* whenever we have inclusions $G^\varepsilon \circ F^\varepsilon \subseteq I_X^{2\varepsilon}$ and $F^\varepsilon \circ G^\varepsilon \subseteq I_Y^{2\varepsilon}$.

The motivation for the definition of ε–acyclic equivalences is the following lemma:

Proposition 4.4 Let X_* and Y_* be two tame elements from $\mathbf{FCW\text{-}cpx}$ which are **ε–acyclic equivalent**. Then $\mathrm{PH}(X_*)$ and $\mathrm{PH}(Y_*)$ are ε–interleaved.

Proof By Proposition 4.2 we know that there exist two chain maps $f_*^\varepsilon \colon C_*(X_*) \to C_*(Y_{*+\varepsilon})$ and $g^\varepsilon \colon C_*(Y_*) \to C_*(X_{*+\varepsilon})$ carried by F^ε and G^ε respectively. By hypothesis the compositions $g^\varepsilon \circ f^\varepsilon$ and $f^\varepsilon \circ g^\varepsilon$ are carried by corresponding shift carriers $I_X^{2\varepsilon}$ and $I_Y^{2\varepsilon}$. Thus, using the second part of Proposition 4.2 we obtain chain homotopies $g^\varepsilon \circ f^\varepsilon \simeq \Sigma^{2\varepsilon} C_*(X)$ and $f^\varepsilon \circ g^\varepsilon \simeq \Sigma^{2\varepsilon} C_*(Y)$. Altogether, in homology these compositions are equal to the corresponding shifts, and $\mathrm{PH}_*(X_*)$ and $\mathrm{PH}_*(Y_*)$ are ε–interleaved. $\qquad \square$

Example 4.5 Consider two finite metric spaces $\mathbb{X}$ and $\mathbb{Y}$. Let $d_H(\mathbb{X}, \mathbb{Y})$ be their Hausdorff distance and set $\varepsilon = 2d_H(\mathbb{X}, \mathbb{Y})$. Given a subcomplex $K \subseteq \mathrm{VR}(\mathbb{X})$, we denote its vertex set by $\mathbb{X}(K) \subseteq \mathbb{X}$. Likewise for a simplex $\sigma \in \mathrm{VR}(\mathbb{X})$, we write $\mathbb{X}(\sigma) \subseteq \mathbb{X}$ for the vertices spanning σ. Define a carrier $F^\varepsilon \colon \mathrm{VR}(\mathbb{X}) \Rightarrow \mathrm{VR}(\mathbb{Y})$ by mapping a simplex $\sigma \in \mathrm{VR}(\mathbb{X})_a$ to

$$F^\varepsilon(\sigma) = \big| \sup\{K \subseteq \mathrm{VR}(\mathbb{Y})_{a+\varepsilon} \mid d_H(\mathbb{X}(\sigma), \mathbb{Y}(K)) \leq \varepsilon/2\} \big|.$$

This is clearly semicontinuous. If $v_0, \ldots, v_n$ are vertices in $F^\varepsilon(\sigma)$, then by definition $\{v_0, \ldots, v_n\}$ is an n–simplex in $F^\varepsilon(\sigma)$. Therefore we have $F^\varepsilon(\sigma) \simeq \Delta^N$ for some $N \in \mathbb{Z}_{\geq 0}$, which is acyclic. In particular, F^ε is an ε–acyclic carrier. Interchanging the roles of $\mathbb{X}$ and $\mathbb{Y}$ we also obtain an ε–acyclic carrier $G^\varepsilon \colon \mathrm{VR}(\mathbb{Y}) \rightrightarrows \mathrm{VR}(\mathbb{X})$. Similarly, we define for a simplex $\sigma \in \mathrm{VR}(\mathbb{X})_a$ the shift carrier

$$I_{\mathbb{X}}^{2\varepsilon}(\sigma) = \big| \sup\{K \subseteq \mathrm{VR}(\mathbb{X})_{a+2\varepsilon} \mid d_{\mathrm{H}}(\mathbb{X}(\sigma), \mathbb{X}(K)) \leq \varepsilon\} \big|.$$

Analogously one defines $I_{\mathbb{Y}}^{2\varepsilon}$. Since $G^\varepsilon \circ F^\varepsilon \subseteq I_{\mathbb{X}}^{2\varepsilon}$ and $F^\varepsilon \circ G^\varepsilon \subseteq I_{\mathbb{Y}}^{2\varepsilon}$, Proposition 4.4 implies that $\mathrm{PH}_*(\mathrm{VR}(\mathbb{X}))$ and $\mathrm{PH}_*(\mathrm{VR}(\mathbb{Y}))$ are ε–interleaved. This is similar to the proof using *correspondences*; see [Oudot 2015, Proposition 7.8, Section 7.3].

Example 4.6 Consider $\mathbb{R}^N$ together with a 1–Lipschitz function $f \colon \mathbb{R}^N \to \mathbb{R}$ with constant $\varepsilon > 0$. On the other hand, consider the lattices $\mathbb{Z}^N$ and $r\mathbb{Z}^N + l$ for a pair of vectors $r, l \in \mathbb{R}^N$ such that the coordinates of r satisfy $0 < r_i \leq 1$ for all $1 \leq i \leq N$. Then we take their corresponding cubical complexes $\mathscr{C}(\mathbb{Z}^N)$ and $\mathscr{C}(r\mathbb{Z}^N + l)$ thought as embedded in $\mathbb{R}^N$. The function f induces a natural filtration for these cubical complexes: a vertex $v \in \mathscr{C}(\mathbb{Z}^N)$ is contained in $\mathscr{C}(\mathbb{Z}^N)_{f(v)}$, while a cell $a \in \mathscr{C}(\mathbb{Z}^N)$ appears at the maximum filtration value on its vertices. There is an ε–acyclic carrier $F^\varepsilon \colon \mathscr{C}(\mathbb{Z}^N) \rightrightarrows \mathscr{C}(r\mathbb{Z}^N + l)$ sending each cell $a \in \mathscr{C}(\mathbb{Z}^N)$ to the smallest subcomplex $F^\varepsilon(a)$ containing all $b \in \mathscr{C}(r\mathbb{Z}^N + l)$ such that $\bar{b} \cap a \neq \varnothing$. In an analogous way the inverse acyclic carrier can be defined, and the compositions $F^\varepsilon \circ G^\varepsilon$ and $G^\varepsilon \circ F^\varepsilon$ define the shift carriers. Thus, using Proposition 4.4, one shows that $\mathrm{PH}_*(\mathscr{C}(\mathbb{Z}^N))$ and $\mathrm{PH}_*(\mathscr{C}(r\mathbb{Z}^N + l))$ are ε–interleaved.

An important assumption of Proposition 4.2 is that we are dealing with tame filtered CW–complexes. However, what if we considered a pair of elements $X_*, Y_* \in \mathbf{RCW\text{-}cpx}$ instead? In this context, we notice that given an ε–acyclic carrier $F^\varepsilon \colon X_* \to Y_*[\varepsilon]$, it is not necessarily true that the compositions

$$Y(a + \varepsilon \leq b + \varepsilon)F_a^\varepsilon(c) \quad \text{and} \quad F_b^\varepsilon(X(a \leq b)(c))$$

are still acyclic for all pairs $a \leq b$ from $\mathbf{R}$. Thus, whenever we talk about ε–acyclic carriers $F^\varepsilon \colon X_* \to Y_*[\varepsilon]$ in this context we assume that $F_b^\varepsilon(X(a \leq b)(c))$ is acyclic for all pairs $a, b \in \mathbf{R}$ with $a \leq b$ and all cells $c \in X(a)$.

Corollary 4.7 *Let $X_*, Y_* \in \mathbf{RCW\text{-}cpx}$ be a pair of elements such that both are ε–acyclic equivalent in the above sense. Then $d_I(\mathrm{PH}_*(X_*), \mathrm{PH}_*(Y_*)) \leq \varepsilon$.*

Proof For each persistence value $a \in \mathbf{R}$, we use Theorem 2.4 twice to obtain a pair of chain morphisms $f_a \colon C_a^{\mathrm{cell}}(X) \to C_{a+\varepsilon}^{\mathrm{cell}}(Y)$ and $g_{a+\varepsilon} \colon C_{a+\varepsilon}^{\mathrm{cell}}(Y) \to C_{a+2\varepsilon}^{\mathrm{cell}}(X)$. In a similar way we obtain a pair of chain homotopies $g_{a+\varepsilon} \circ f_a \simeq (\Sigma^{2\varepsilon} C_*^{\mathrm{cell}}(X))_a$ and $f_{a+\varepsilon} \circ g_a \simeq (\Sigma^{2\varepsilon} C_*^{\mathrm{cell}}(Y))_a$ so that we have equalities between the induced homology morphisms $[g_{a+\varepsilon}] \circ [f_a] = [(\Sigma^{2\varepsilon} C_*^{\mathrm{cell}}(X))_a]$ and $[f_{a+\varepsilon}] \circ [g_a] = [(\Sigma^{2\varepsilon} C_*^{\mathrm{cell}}(Y))_a]$ for all $a \in \mathbf{R}$. Now, for a pair of values $a \leq b$ from $\mathbf{R}$, it is not necessarily true that

$$Y(a + \varepsilon \leq b + \varepsilon) \circ f_a = f_b \circ X(a \leq b).$$

However, since $Y(a + \varepsilon \leq b + \varepsilon) \circ f_a$ and $f_b \circ X(a \leq b)$ are both included in $F_b^\varepsilon(X(a \leq b)(c))$ by hypotheses, then by applying Theorem 2.4 again there is a chain homotopy equivalence

$$Y(a + \varepsilon \leq b + \varepsilon) \circ f_a \simeq f_b \circ X(a \leq b),$$

which implies

$$[Y(a + \varepsilon \leq b + \varepsilon)] \circ [f_a] = [f_b] \circ [X(a \leq b)],$$

and we have defined a persistence morphism $[f_*] \colon \mathrm{PH}_*(X_*) \to \mathrm{PH}_*(Y_*[\varepsilon])$. Similarly, we can also put together the g_a for all $a \in \mathbf{R}$ so that we obtain a morphism $[g_*] \colon \mathrm{PH}_*(Y_*) \to \mathrm{PH}_*(X_*[\varepsilon])$. This leads to the claimed ε–interleaving. $\qquad\square$

Example 4.8   In the appendix, we describe a filtered CW–complex X, a regularly filtered CW–complex Y, together with a pair of 0–acyclic carriers (ie $\varepsilon = 0$) $F \colon Y \rightrightarrows X$ and $G \colon X \rightrightarrows Y$ which, together with the compositions $G \circ F$ and $G \circ F$ as shift carriers, define a 0–acyclic equivalence between Y and X. Therefore, by Corollary 4.7 we obtain isomorphisms $\mathrm{PH}_n(X) \cong \mathrm{PH}_n(Y)$ for all $n \geq 0$. In this case, notice that Y is much smaller than X; thus it is worth considering the regularly filtered complex Y in place of X. Next, we briefly describe how one could use ε–equivalences. In this case, one could have considered a filtered complex $\widetilde{X}$ which is equal to X_* outside the intervals $(i - \varepsilon, i + \varepsilon)$ for values $i = 1, 2, 3, 4$ and for some $\varepsilon < 1/2$. Notice that in this case one should be able to obtain an ε–acyclic equivalence between $\widetilde{X}$ and Y, so that by Corollary 4.7 $\mathrm{PH}_n(\widetilde{X})$ and $\mathrm{PH}_n(Y)$ are ε–interleaved for all $n \geq 0$.

Remark 4.9   Notice that our notion of acyclicity is different from that in [Cavanna 2019] and [Govc and Skraba 2018]. In [Govc and Skraba 2018] a filtered complex K_* is called ε–acyclic whenever the induced homology maps $\mathrm{H}_*(K_r) \to \mathrm{H}_*(K_{r+\varepsilon})$ vanish for all $r \in \mathbb{R}$. In this case, one can still (trivially) define acyclic carriers between $*$ and K_*. The problem arises when defining the shift carrier $I_K^{A\varepsilon}$ for some constant $A > 0$, which does not exist in general. One can however, adapt the proof of Proposition 4.2 so that there is a chain morphism $\psi^{\varepsilon(\dim(K_r)+1)} \colon C_*(K_r) \to C_*(K_{r+\varepsilon(\dim(K_r)+1)})$; and that this coincides up to chain homotopy with the composition through $C_*(*)$. One does this by following the same proof as in Proposition 4.2, but increasing the filtration value by ε each time we assume that some cycle lies in an acyclic carrier. Thus, if we have $\dim(K) = \sup_{r \in \mathbb{R}}(\dim(K_r)) < \infty$, then one could say that there is an $\varepsilon(\dim(K)+1)$–*approximate* chain homotopy equivalence between $C(*)$ and $C(K_*)$.

5   Interleaving geometric realizations

Next, we focus on acyclic carrier equivalences between a pair of diagrams $\mathscr{D}, \mathscr{L} \in \mathbf{RRDiag}(K)$. We start by taking ε–acyclic carriers $F_\sigma^\varepsilon \colon \mathscr{D}(\sigma) \rightrightarrows \mathscr{L}(\sigma)$ for all $\sigma \in K$ which have to be compatible in the sense that for any pair $\tau \preceq \sigma$ and any cell $c \in \mathscr{D}(\sigma)$, there is an inclusion

$$(2) \qquad \mathscr{L}(\tau \preceq \sigma)(F_\sigma^\varepsilon(c)) \subseteq F_\tau^\varepsilon(\mathscr{D}(\tau \preceq \sigma)(c))$$

and we assume in addition that $F_\tau^\varepsilon(\mathcal{D}(\tau \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c))$ is acyclic for all $r \geq 0$. This compatibility leads to "local" diagrams of spaces. That is, given a pair of values $a \in \mathbf{R}$ and $r \geq 0$ and a cell $c \in \mathcal{D}(\sigma)_a$, we consider an object $F_{\sigma \times c}^{r,\varepsilon} \in \mathbf{RDiag}(\Delta^\sigma)$. It is given by the space $F_{\sigma \times c}^{r,\varepsilon}(\tau) = F_\tau^\varepsilon(\mathcal{D}(\tau \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c))$ for all faces $\tau \preceq \sigma$. For any sequence $\rho \preceq \tau \preceq \sigma$ in K, there are morphisms in $F_{\sigma \times c}^{r,\varepsilon}$ given by restricting morphisms from $\mathcal{L}$:

$$
\begin{array}{ccc}
\tau \longrightarrow F_{\sigma \times c}^{r,\varepsilon}(\tau) & = = & F_\tau^\varepsilon(\mathcal{D}(\tau \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c)) \\
\preceq \Big\uparrow \qquad \Big\downarrow & & \Big\downarrow \mathcal{L}(\rho \preceq \tau) \\
\rho \longrightarrow F_{\sigma \times c}^{r,\varepsilon}(\rho) & = = & F_\rho^\varepsilon(\mathcal{D}(\rho \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c))
\end{array}
$$

Using condition (2) repeatedly on the cells from $L = \mathcal{D}(\tau \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c)$, we see that we have an inclusion

$$
\mathcal{L}(\rho \preceq \tau)(F_\tau^\varepsilon(L)) \subseteq F_\sigma^\varepsilon(\mathcal{D}(\rho \preceq \tau)(L)).
$$

Thus the diagram $F_{\sigma \times c}^{r,\varepsilon}$ is indeed well defined, and we may consider the geometric realization $\Delta F_{\sigma \times c}^{r,\varepsilon}$. By hypothesis, each $F_{\sigma \times c}^{r,\varepsilon}(\tau)$ is acyclic for all $\tau \preceq \sigma$, so the first page of the spectral sequence $E_{p,q}^*(F_{\sigma \times c}^{r,\varepsilon}) \Rightarrow \mathrm{H}_{p+q}(\Delta F_{\sigma \times c}^{r,\varepsilon})$ is equal to

$$
E_{p,q}^1(F_{\sigma \times c}^{r,\varepsilon}) = \bigoplus_{\tau \in (\Delta^\sigma)^p} \mathrm{H}_q(F_{\sigma \times c}^{r,\varepsilon}(\tau)) = \begin{cases} \bigoplus_{\tau \in (\Delta^\sigma)^p} \mathbb{F} & \text{if } q = 0, \\ 0 & \text{otherwise.} \end{cases}
$$

In fact, computing the homology with respect to the horizontal differentials on the first page corresponds to computing the homology of Δ^σ. Thus, $E_{p,q}^2(F_{\sigma \times c}^{r,\varepsilon})$ is zero everywhere except at $p = q = 0$ where it is equal to $\mathbb{F}$. Thus, the spectral sequence collapses on the second page, and $\Delta F_{\sigma \times c}^{r,\varepsilon}$ is acyclic. We use the notation $F_{\sigma \times c}^\varepsilon = F_{\sigma \times c}^{0,\varepsilon}$.

Definition 5.1 Let $\mathcal{D}$ and $\mathcal{L}$ be two diagrams in $\mathbf{RRDiag}(K)$. Suppose that there are ε–acyclic carriers $F_\sigma^\varepsilon: \mathcal{D}(\sigma) \rightrightarrows \mathcal{L}(\sigma)$ for all $\sigma \in K$, and that

$$
\mathcal{L}(\tau \preceq \sigma)(F_\sigma^\varepsilon(c)) \subseteq F_\sigma^\varepsilon(\mathcal{D}(\tau \preceq \sigma)(c))
$$

for all $c \in \mathcal{D}(\sigma)$ and in addition $F_\tau^\varepsilon(\mathcal{D}(\tau \preceq \sigma)\Sigma^r\mathcal{D}(\sigma)(c))$ is acyclic for all $r \geq 0$. Then we say that the set $\{F_\sigma^\varepsilon\}_{\sigma \in K}$ is an (ε, K)–acyclic carrier between $\mathcal{D}$ and $\mathcal{L}$. We denote this by $F^\varepsilon: \mathcal{D} \rightrightarrows \mathcal{L}$.

Theorem 5.2 *Let $\mathcal{D}$ and $\mathcal{L}$ be two diagrams in $\mathbf{RRDiag}(K)$. Suppose that there are (ε, K)–acyclic carriers $F^\varepsilon: \mathcal{D} \rightrightarrows \mathcal{L}$ and $G^\varepsilon: \mathcal{L} \rightrightarrows \mathcal{D}$, together with a pair of shift (ε, K)–acyclic carriers $I_{\mathcal{D}}^{2\varepsilon}: \mathcal{D} \rightrightarrows \mathcal{D}$ and $I_{\mathcal{L}}^{2\varepsilon}: \mathcal{L} \rightrightarrows \mathcal{L}$, and such that these restrict to acyclic equivalences*

$$
G_\tau^\varepsilon \circ F_\tau^\varepsilon \subseteq (I_{\mathcal{D}}^{2\varepsilon})_\tau \quad \text{and} \quad F_\tau^\varepsilon \circ G_\tau^\varepsilon \subseteq (I_{\mathcal{L}}^{2\varepsilon})_\tau
$$

for each simplex $\tau \in K$. Then there is an ε–acyclic equivalence $F^\varepsilon: \Delta\mathcal{D} \rightrightarrows \Delta\mathcal{L}$ which preserves filtration. That is, there are ε–acyclic equivalences $F^p F^\varepsilon: F^p\Delta\mathcal{D} \rightrightarrows F^p\Delta\mathcal{L}$ for all $p \geq 0$.

Proof Let $\sigma \times c \in \Delta\mathcal{D}$ be a cell, where c is an m–cell in $\mathcal{D}(\sigma)$. Define an acyclic carrier $F^\varepsilon \colon \Delta\mathcal{D} \rightrightarrows \Delta\mathcal{L}$ by sending $\sigma \times c$ to the acyclic carrier $\Delta F^\varepsilon_{\sigma \times c}$, which is a subcomplex of $\Delta\mathcal{L}$. Let us first check semicontinuity. For any pair of cells $\tau \times a \preceq \sigma \times c$ in $\Delta\mathcal{D}$, the cell a is contained in the subcomplex $\overline{\mathcal{D}(\tau \preceq \sigma)(c)}$, and by continuity of $\mathcal{D}(\rho \preceq \tau)$ we have that $\mathcal{D}(\rho \preceq \tau)(a) \subseteq \overline{\mathcal{D}(\rho \preceq \sigma)(c)}$. Thus there are inclusions

$$F^\varepsilon_\rho(\mathcal{D}(\rho \preceq \tau)(a)) \subseteq F^\varepsilon_\rho(\overline{\mathcal{D}(\rho \preceq \sigma)(c)}) = F^\varepsilon_\rho(\mathcal{D}(\rho \preceq \sigma)(c))$$

for all $\rho \preceq \tau$. More concisely, $F^\varepsilon_{\tau \times a}(\rho) \subseteq F^\varepsilon_{\sigma \times c}(\rho)$ for all $\rho \preceq \tau$. As a consequence $\Delta F^\varepsilon_{\tau \times a} \subseteq \Delta F^\varepsilon_{\sigma \times c}$ and semicontinuity holds.

Next, notice that $F^\varepsilon(\Sigma^r \Delta\mathcal{D}(\sigma \times c)) = F^\varepsilon(\sigma \times \Sigma^r \mathcal{D}(\sigma)(c)) = \Delta F^{r,\varepsilon}_{\sigma \times c}$ which is an acyclic carrier. In order for F^ε to be an ε–acyclic carrier, it remains to show the inclusion $\Sigma^r \Delta\mathcal{L} \circ F^\varepsilon \subseteq F^\varepsilon \circ \Sigma^r \Delta\mathcal{D}$ for all $r \geq 0$. For this, take $\sigma \times c \in \Delta\mathcal{D}$ and see that

$$\Sigma^r \Delta\mathcal{L} \circ F^\varepsilon(\sigma \times c) = \Sigma^r \Delta\mathcal{L}\left(\bigcup_{\tau \preceq \sigma} \tau \times F^\varepsilon_\tau(\mathcal{D}(\tau \preceq \sigma)(c)) \right)$$

$$= \bigcup_{\tau \preceq \sigma} \tau \times \Sigma^r \mathcal{L}(\tau)\left(F^\varepsilon_\tau(\mathcal{D}(\tau \preceq \sigma)(c)) \right)$$

$$\subseteq \bigcup_{\tau \preceq \sigma} \tau \times F^\varepsilon_\tau(\Sigma^r \mathcal{D}(\tau)\mathcal{D}(\tau \preceq \sigma)(c))$$

$$= \bigcup_{\tau \preceq \sigma} \tau \times F^\varepsilon_\tau(\mathcal{D}(\tau \preceq \sigma)\Sigma^r \mathcal{D}(\sigma)(c))$$

$$= F^\varepsilon(\sigma \times \Sigma^r \mathcal{D}(\sigma)(c)) = F^\varepsilon \circ \Sigma^r \Delta\mathcal{D}(\sigma \times c).$$

Similarly, one can define an ε–acyclic carrier $G^\varepsilon \colon \Delta\mathcal{L} \rightrightarrows \Delta\mathcal{D}$ sending $\sigma \times c \in \Delta\mathcal{L}$ to $\Delta G^\varepsilon_{\sigma \times c}$. In addition, we define respective shift ε–acyclic carriers $I^{2\varepsilon}_\mathcal{D} \colon \Delta\mathcal{D} \rightrightarrows \Delta\mathcal{D}$ and $I^{2\varepsilon}_\mathcal{L} \colon \Delta\mathcal{L} \rightrightarrows \Delta\mathcal{L}$ sending, respectively, $\sigma \times c \in \Delta\mathcal{D}$ to $\Delta(I^{2\varepsilon}_\mathcal{D})_{\sigma \times c}$ and $\tau \times a \in \Delta\mathcal{L}$ to $\Delta(I^{2\varepsilon}_\mathcal{L})_{\tau \times a}$. Then we have

$$G^\varepsilon \circ F^\varepsilon(\sigma \times c) = G^\varepsilon(\Delta F^\varepsilon_{\sigma \times c})$$

$$= G^\varepsilon\left(\bigcup_{\tau \preceq \sigma} \tau \times F^\varepsilon_\tau(\mathcal{D}(\tau \preceq \sigma)(c)) \right)$$

$$= \bigcup_{\rho \preceq \tau \preceq \sigma} \rho \times G^\varepsilon_\rho\left(\mathcal{L}(\rho \preceq \tau) F^\varepsilon_\tau(\mathcal{D}(\tau \preceq \sigma)(c)) \right)$$

$$\subseteq \bigcup_{\rho \preceq \sigma} \rho \times G^\varepsilon_\rho F^\varepsilon_\rho(\mathcal{D}(\rho \preceq \sigma)(c)) \subseteq \Delta(I^{2\varepsilon}_\mathcal{D})_{\sigma \times c} = I^{2\varepsilon}_\mathcal{D}(\sigma \times c),$$

where we have used the commutativity condition and equivalence of F^ε_ρ and G^ε_ρ. Consequently, $G^\varepsilon \circ F^\varepsilon \subseteq I^{2\varepsilon}_\mathcal{D}$; the other inclusion $F^\varepsilon \circ G^\varepsilon \subseteq I^{2\varepsilon}_\mathcal{L}$ follows by symmetry. Altogether, we have obtained an ε–equivalence $F^\varepsilon \colon \Delta\mathcal{D} \rightrightarrows \Delta\mathcal{L}$. Finally, notice that for all $p \geq 0$ and for each cell $\sigma \times c \in F^p \Delta\mathcal{D}$, its carrier $\Delta F^\varepsilon_{\sigma \times c}$ is contained in $F^p \Delta\mathcal{D}$ and so it preserves filtration. The same follows for the other acyclic carriers. $\qquad\square$

Let $X \in \mathbf{FCW\text{-}cpx}$ together with a cover $\mathcal{U}$. Recall the definitions of the diagrams $X^{\mathcal{U}}$ and $\pi_0^{\mathcal{U}}$ over $N_{\mathcal{U}}$ from Example 2.9. Let $d_I\big(\mathrm{PH}_*(X^{\mathcal{U}}(\sigma)), \mathrm{PH}_*(\pi_0^{\mathcal{U}}(\sigma))\big) \leq \varepsilon$ for all $\sigma \in N_{\mathcal{U}}$. This example has been of interest before; see for example [Govc and Skraba 2018] or [Cavanna 2019]. As mentioned in Remark 4.9, our notion of ε–acyclicity is much stronger than that from [Govc and Skraba 2018]. This is why we obtain a result closer to the *persistence nerve theorem* from [Chazal and Oudot 2008] than to the *approximate nerve theorem* from [Govc and Skraba 2018].

Given a diagram $\mathcal{D} \in \mathbf{FRDiag}(K)$, recall the diagram $\pi_0\mathcal{D}$ from Example 2.11. We may define an (ε, K)–acyclic carrier $\pi_0^{\varepsilon}\mathcal{D} \colon \mathcal{D} \Rightarrow \pi_0\mathcal{D}$ where we send cells to their corresponding connected component classes. The compatibility condition $\pi_0(\mathcal{D}(\tau \preceq \sigma))(\pi_0^{\varepsilon}\mathcal{D}(\mathcal{D}(\sigma))) \subseteq \pi_0^{\varepsilon}\mathcal{D}(\mathcal{D}(\tau))$ also follows for any pair of simplices $\tau \preceq \sigma$ from K.

Corollary 5.3 (strong approximate multinerve theorem) *Consider a diagram $\mathcal{D}$ in $\mathbf{FRDiag}(K)$. Assume that there is a (ε, K)–acyclic carrier $F^{\varepsilon} \colon \pi_0\mathcal{D} \Rightarrow \mathcal{D}$ such that the composition $F_{\sigma}^{\varepsilon} \circ \pi_0^{\varepsilon}\mathcal{D}_{\sigma}$ carries the shift morphism $\Sigma^{2\varepsilon}\mathcal{D}_{\sigma}$ for all $\sigma \in K$. Then, there is an ε–acyclic equivalence $F^{\varepsilon} \colon \mathrm{MNerv}(\mathcal{D}) \Rightarrow \Delta\mathcal{D}$. Consequently,*

$$d_I(\mathrm{PH}_*(\mathrm{MNerv}(\mathcal{D})), \mathrm{PH}_*(\Delta\mathcal{D})) \leq \varepsilon.$$

Proof The shift $(2\varepsilon, K)$–carrier $I_{\pi_0\mathcal{D}}^{2\varepsilon}$ sends points to points, while the other $I_{\mathcal{D}}^{2\varepsilon}$ is defined as the composition $F_{\sigma}^{\varepsilon} \circ \pi_0^{\varepsilon}\mathcal{D}_{\sigma}$, which is a $(2\varepsilon, K)$–acyclic carrier by hypotheses. Thus, by Theorem 5.2 there exists an ε–acyclic equivalence $F^{\varepsilon} \colon \mathrm{MNerv}(\mathcal{D}) \Rightarrow \Delta\mathcal{D}$. $\square$

Example 5.4 Consider a filtered simplicial complex L_* together with a partition of its vertex set $\mathcal{P}$. Assume that the $(L_*, \mathcal{P})$–join diagram $\mathcal{J}_{\mathcal{P}}^{L_*}$ is such that there exists a (ε, K)–acyclic carrier $F^{\varepsilon} \colon \pi_0\mathcal{J}_{\mathcal{P}}^{L_*} \Rightarrow \mathcal{J}_{\mathcal{P}}^{L_*}$ such that $F_{\sigma}^{\varepsilon} \circ \pi_0^{\varepsilon}\mathcal{J}_{\mathcal{P}}^{L_*}(\sigma)$ is a carrier for $\Sigma^{2\varepsilon}\mathcal{J}_{\mathcal{P}}^{L_*}(\sigma)$ for all $\sigma \in \Delta^P$. Then, by Corollary 5.3, there is an ε–acyclic equivalence $\Delta\pi_0(\mathcal{J}_{\mathcal{P}}^{L_*}) \Rightarrow \Delta\mathcal{J}_{\mathcal{P}}^{L_*}$ such that

$$d_I\big(\mathrm{PH}_*(\mathrm{MNerv}(\mathcal{J}_{\mathcal{P}}^{L_*})), \mathrm{PH}_*(L_*)\big) \leq \varepsilon.$$

Acyclic carriers have been used in [Kaczynski et al. 2004] and in [Nanda 2012] for approximating continuous morphisms by means of simplicial maps. Here we have used the same tools to obtain an approximate homotopy colimit theorem. The acyclic carrier theorem is an instance of the more general acyclic model theorem; see [Eilenberg and MacLane 1953, Section 2]. An interesting future research direction would be to see how that general result can bring new insights into applied topology.

6 Interleaving spectral sequences

Definition 6.1 Let $\mathcal{A}$ and $\mathcal{B}$ be from $\mathbf{SpSq}$. A n–spectral sequence morphism $f \colon \mathcal{A} \to \mathcal{B}$ is a spectral sequence morphism $f \colon \mathcal{A} \to \mathcal{B}$ which is defined from page n.

Definition 6.2 Given two objects $\mathscr{A}$ and $\mathscr{B}$ in **PSpSq**. We say that $\mathscr{A}$ and $\mathscr{B}$ are (ε, n)–interleaved whenever there exist two n–morphisms $\psi \colon \mathscr{A} \to \mathscr{B}[\varepsilon]$ and $\varphi \colon \mathscr{B} \to \mathscr{A}[\varepsilon]$ such that the diagram

(3)

commutes for all pages $r \geq n$. This interleaving defines a pseudometric in **PSpSq**,

$$d_I^n(\mathscr{A}, \mathscr{B}) := \inf\{\varepsilon \mid \mathscr{A} \text{ and } \mathscr{B} \text{ are } (\varepsilon, n)\text{–interleaved}\}.$$

Proposition 6.3 *Suppose that $\mathscr{A}$ and $\mathscr{B}$ are (ε, n)–interleaved. Then these are (ε, m)–interleaved for all $m \geq n$. In particular, we have that*

$$d_I^m(\mathscr{A}, \mathscr{B}) \leq d_I^n(\mathscr{A}, \mathscr{B})$$

for any pair of integers $m \geq n$.

Proof This follows directly from the definitions. $\square$

We start now by considering Mayer–Vietoris spectral sequences. Under some conditions which are a special case of Theorem 5.2, one can obtain one-page stability. In fact, this stability is due to morphisms directly defined on the underlying double complexes, which is a very strong property.

Proposition 6.4 *Let X and Y be two tame elements in **FCW**-**cpx** together with a pair of respective finite covers $\mathcal{U}$ and $\mathcal{V}$ by subcomplexes such that $K = N_{\mathcal{U}} = N_{\mathcal{V}}$. Suppose that there are (ε, K)–acyclic carriers $F^\varepsilon \colon X^{\mathcal{U}} \rightrightarrows Y^{\mathcal{V}}$ and $G^\varepsilon \colon Y^{\mathcal{V}} \rightrightarrows X^{\mathcal{U}}$, together with a pair of shift (ε, K)–acyclic carriers $I_{X^{\mathcal{U}}}^{2\varepsilon} \colon X^{\mathcal{U}} \rightrightarrows X^{\mathcal{U}}$ and $I_{Y^{\mathcal{V}}}^{2\varepsilon} \colon Y^{\mathcal{V}} \rightrightarrows Y^{\mathcal{V}}$, and such that these restrict to acyclic equivalences*

$$G_\tau^\varepsilon \circ F_\tau^\varepsilon \subseteq (I_{X^{\mathcal{U}}}^{2\varepsilon})_\tau \quad \text{and} \quad F_\tau^\varepsilon \circ G_\tau^\varepsilon \subseteq (I_{Y^{\mathcal{V}}}^{2\varepsilon})_\tau$$

for each simplex $\tau \in K$. Then there is a pair of double complex morphisms

$$\phi^\varepsilon \colon C_{*,*}(X, \mathcal{U}) \to C_{*,*}(Y, \mathcal{V})[\varepsilon] \quad \text{and} \quad \psi^\varepsilon \colon C_{*,*}(Y, \mathcal{V}) \to C_{*,*}(X, \mathcal{U})[\varepsilon]$$

inducing a first page interleaving between $E_{*,*}^*(X, \mathcal{U})$ and $E_{*,*}^*(Y, \mathcal{V})$.

Proof Unpacking the definitions, this means we have to give chain homomorphisms

$$(\phi_\sigma^\varepsilon)_r \colon C_*(X^{\mathcal{U}}(\sigma)_r) \to C_*(Y^{\mathcal{V}}(\sigma)_{r+\varepsilon}), \quad (\psi_\sigma^\varepsilon)_r \colon C_*(Y^{\mathcal{V}}(\sigma)_r) \to C_*(X^{\mathcal{U}}(\sigma)_{r+\varepsilon})$$

that are natural in $\sigma \in K$ and in $r \in \mathbf{R}$. Since K is a poset category, these can be constructed inductively as follows: As in Proposition 4.2 we may define ϕ_σ^ε on all simplices $\sigma \in K$ of dimension $\dim(\sigma) = \dim(K)$.

Note that $(\phi_\sigma^\varepsilon)_r$ is carried by $(F_\sigma^\varepsilon)_r$ for all $r \in \mathbf{R}$. Assume by (reverse) induction that ϕ_τ^ε are defined and carried by F_τ^ε for all $\tau \in K$ with $n \leq \dim(\tau) \leq \dim(K)$ in such a way that for all cofaces $\tau \preceq \sigma$ the naturality condition $\phi_\tau^\varepsilon \circ X^{\mathcal{U}}(\tau \prec \sigma) = Y^{\mathcal{V}}(\tau \prec \sigma)[\varepsilon] \circ \phi_\sigma^\varepsilon$ holds. Now let $\tau \in K$ have dimension $\dim(\tau) = n - 1 \geq 0$. The naturality condition on the simplices fixes ϕ_τ^ε on the filtered subcomplex $X^\tau = \bigcup_{\tau \prec \sigma} \operatorname{Im}(X^{\mathcal{U}}(\tau \prec \sigma))$, where the union is taken over all cofaces σ of τ. Here notice that we can assume that ϕ_τ^ε is well defined since the previous choices of ϕ_σ^ε for all cofaces $\tau \prec \sigma$ are consistent due to the fact that for each cell $c \in X^\tau$ there exists a unique maximal simplex $\sigma \in N_{\mathcal{U}}$ such that $c \in X^{\mathcal{U}}(\sigma)$. In addition, notice that by hypotheses $Y^{\mathcal{V}}(\tau \prec \sigma)((F_\sigma^\varepsilon)(c)) \subseteq F_\tau^\varepsilon(X^{\mathcal{U}}(\tau \prec \sigma)(c))$ for all $a \in \mathbf{R}$ and $c \in X^{\mathcal{U}}(\sigma)$, so that our definition of ϕ_τ^ε in X^τ is indeed carried by F_τ^ε. We then proceed as in Proposition 4.2 to define $(\phi_\tau^\varepsilon)_a$ on all simplices in the subset $X^{\mathcal{U}}(\tau)_a \setminus X_a^\tau$ for all $a \in \mathbf{R}$. The resulting chain map $(\phi_\tau^\varepsilon)_a$ is carried by $(F_\tau^\varepsilon)_a$ for all $a \in \mathbf{R}$. Since $X^{\mathcal{U}}$ is tame, we only need finitely many steps to obtain a morphism $\phi_\tau^\varepsilon \colon C_*(X^{\mathcal{U}}(\tau)) \to C_*(Y^{\mathcal{V}}(\tau)[\varepsilon])$ that satisfies the induction hypotheses.

Thus, we obtain double complex morphisms $\phi_{p,q}^\varepsilon \colon C_{p,q}(X, \mathcal{U}) \to C_{p,q}(Y, \mathcal{V})[\varepsilon]$ for all $p, q \geq 0$ by adding up our defined local morphisms

$$\phi_{p,q}^\varepsilon \colon \bigoplus_{\sigma \in K^p} \phi_\sigma^\varepsilon \colon \bigoplus_{\sigma \in K^p} C_q(X^{\mathcal{U}}(\sigma)) \longrightarrow \bigoplus_{\sigma \in K^p} C_q(Y^{\mathcal{V}}(\sigma))[\varepsilon].$$

Notice that the $\phi_{p,q}^\varepsilon$ commute both with horizontal and vertical differentials since we assumed that each ϕ_σ^ε is a chain morphism and these satisfy a naturality condition with respect to K. Thus, this double complex morphism induces a spectral sequence morphism $\phi_{p,q}^\varepsilon \colon E_{p,q}^*(X^{\mathcal{U}}) \to E_{p,q}^*(Y^{\mathcal{V}})[\varepsilon]$. By doing the same construction, we can obtain local chain morphisms $\psi_\sigma^\varepsilon \colon C_*(Y^{\mathcal{V}}(\sigma)) \to C_*(X^{\mathcal{U}}(\sigma))[\varepsilon]$ so that by Proposition 4.2 we have equalities $[\psi_\sigma^\varepsilon] \circ [\phi_\sigma^\varepsilon] = [\Sigma^{2\varepsilon} C_*(X^{\mathcal{U}}(\sigma))]$ and also $[\phi_\sigma^\varepsilon] \circ [\psi_\sigma^\varepsilon] = [\Sigma^{2\varepsilon} C_*(Y^{\mathcal{V}}(\sigma))]$ for all $\sigma \in K$. Then we can construct a double complex morphism $\psi_{p,q}^\varepsilon \colon C_{p,q}(Y, \mathcal{V}) \to C_{p,q}(X, \mathcal{U})[\varepsilon]$ inducing an "inverse" spectral sequence morphism $\psi_{p,q}^\varepsilon \colon E_{p,q}^*(Y, \mathcal{V}) \to E_{p,q}^*(X, \mathcal{U})[\varepsilon]$. These are such that from the first page, $\phi_{*,*}^\varepsilon$ and $\psi_{*,*}^\varepsilon$ form a $(\varepsilon, 1)$–interleaving of spectral sequences. $\qquad\square$

Notice that the proof of Proposition 6.4 relies heavily on the fact that the diagrams we are considering come from a cover. This allows us to define a pair of double complex morphisms that are compatible along the common indexing nerve. However, in Theorem 5.2 we observed that, under some conditions, the geometric realizations of regularly filtered regular diagrams are stable. Does this stability carry over to the associated spectral sequences? The next theorem shows that this is indeed the case.

Theorem 6.5 *Let $\mathcal{D}$ and $\mathcal{L}$ be two diagrams in $\mathbf{RRDiag}(K)$. Suppose that there are (ε, K)–acyclic carriers $F^\varepsilon \colon \mathcal{D} \Rightarrow \mathcal{L}$ and $G^\varepsilon \colon \mathcal{L} \Rightarrow \mathcal{D}$, together with a pair of shift (ε, K)–acyclic carriers $I_\mathcal{D}^{2\varepsilon} \colon \mathcal{D} \Rightarrow \mathcal{D}$ and $I_\mathcal{L}^{2\varepsilon} \colon \mathcal{L} \Rightarrow \mathcal{L}$, and such that these restrict to acyclic equivalences*

$$G_\tau^\varepsilon \circ F_\tau^\varepsilon \subseteq (I_\mathcal{D}^{2\varepsilon})_\tau \quad \text{and} \quad F_\tau^\varepsilon \circ G_\tau^\varepsilon \subseteq (I_\mathcal{L}^{2\varepsilon})_\tau$$

for each simplex $\tau \in K$. Then

$$d_I^1(E(\mathcal{D}, K), E(\mathcal{L}, K)) \leq \varepsilon.$$

Proof Recall from Theorem 5.2 that there is a filtration-preserving ε–acyclic carrier $F^\varepsilon \colon \Delta_K \mathcal{D} \rightrightarrows \Delta_K \mathcal{L}$. Given $r \in \mathbf{R}$, this implies that there is a chain complex morphism $f_r^\varepsilon \colon C_*(\Delta \mathcal{D})_r \to C_*(\Delta \mathcal{L})_{r+\varepsilon}$ carried by F_r^ε and which respects filtrations in the sense that $f_r^\varepsilon(F^p C_*(\Delta \mathcal{D})_r) \subseteq F^p C_*(\Delta \mathcal{L})_{r+\varepsilon}$ for all $p \geq 0$. By Lemma 3.1 this defines a morphism $f_r^\varepsilon \colon S_*^{\mathrm{Tot}}(\mathcal{D})_r \to S_*^{\mathrm{Tot}}(\mathcal{L})_{r+\varepsilon}$ which respects filtrations. Altogether we deduce that f_r^ε determines a morphism of spectral sequences $f_r^\varepsilon \colon E_{p,q}^*(\mathcal{D})_r \to E_{p,q}^*(\mathcal{L})_{r+\varepsilon}$. Similarly as in Corollary 4.7, the commutativity

$$\Sigma^s E_{p,q}^*(\mathcal{L})_{r+\varepsilon} \circ f_r^\varepsilon = f_{r+s}^\varepsilon \circ \Sigma^s E_{p,q}^*(\mathcal{D})_r \tag{4}$$

does not need to hold for all $r \in \mathbf{R}$ and all $s \geq 0$. However, by definition of ε–acyclic carrier, there is an inclusion $\Sigma^s \Delta \mathcal{L} \circ F^\varepsilon \subseteq F^\varepsilon \circ \Sigma^s \Delta \mathcal{D}$ where the superset is acyclic, so $\Sigma^s C_*(\Delta \mathcal{L})_{r+\varepsilon} \circ f_r^\varepsilon$ and $f_{r+s}^\varepsilon \circ \Sigma^s C_*(\Delta \mathcal{D})_r$ are both carried by the filtration preserving acyclic carrier $F^\varepsilon \circ \Sigma^s \Delta \mathcal{D}_r$. This implies that there exist chain homotopies $h_r^\varepsilon \colon C_n(\Delta \mathcal{D})_r \to C_{n+1}(\Delta \mathcal{L})_{r+s+\varepsilon}$ which respect filtrations and satisfy

$$f_{r+s}^\varepsilon \circ \Sigma^s C_*(\Delta \mathcal{D})_r - \Sigma^s C_*(\Delta \mathcal{L})_{r+\varepsilon} \circ f_r^\varepsilon = \delta^\Delta \circ h_r^\varepsilon + h_r^\varepsilon \circ \delta^\Delta.$$

for all $r \in \mathbf{R}$ and all $s \geq 0$. Recall that the zero page terms are given as quotients on successive filtration terms $E_{p,q}^0(\mathcal{D})_r = F^p S_{p+q}^{\mathrm{Tot}}(\mathcal{D})_r / F^{p-1} S_{p+q}^{\mathrm{Tot}}(\mathcal{D})_r$, for all $r \in \mathbf{R}$ and all integers $p, q \geq 0$. Thus, by Lemma 3.1, these chain homotopies carry over to $S_*^{\mathrm{Tot}}(\mathcal{D})_r$ and the commutativity relation from (4) holds from the first page onwards.

Similarly, we can define spectral sequence morphisms $g_r^\varepsilon \colon E_{p,q}^*(\mathcal{L})_r \to E_{p,q}^*(\mathcal{D})_{r+\varepsilon}$ for all $r \in \mathbf{R}$ which commute with the shift morphisms from the first page. Also, by inspecting the shift carriers, we can obtain equalities of 1–spectral sequence morphisms $g_{r+\varepsilon}^\varepsilon \circ f_r^\varepsilon = \Sigma^{2\varepsilon} E_{p,q}^*(\mathcal{D})_r$ and also $f_{r+\varepsilon}^\varepsilon \circ g_r^\varepsilon = \Sigma^{2\varepsilon} E_{p,q}^*(\mathcal{L})_r$ for all $r \in \mathbf{R}$, and the result follows. $\qquad\square$

Example 6.6 Consider a pair of point clouds $\mathbb{X}, \mathbb{Y} \in \mathbb{R}^N$, together with partitions $\mathcal{P}$ and $\mathcal{Q}$ for $\mathbb{X}$ and $\mathbb{Y}$ respectively. Also, assume that there is an isomorphism $\phi \colon \Delta^{\mathcal{P}} \to \Delta^{\mathcal{Q}}$ such that $d_H(\mathbb{X} \cap V, \mathbb{Y} \cap \phi(V)) < \varepsilon$ for all $V \in \mathcal{P}$. As defined in Example 4.5, there are ε–acyclic carrier equivalences

$$F_V^\varepsilon \colon \mathrm{VR}_*(\mathbb{X} \cap V) \rightrightarrows \mathrm{VR}_*(\mathbb{Y} \cap V)$$

for all $V \in \mathcal{U}$. Now suppose that, for some $\eta > 0$, if $\mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{X})}(\sigma)_r \neq \varnothing$ then $\mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{Y})}(\phi(\sigma))_{r+\eta} \neq \varnothing$ for all $\sigma \in \Delta^{\mathcal{P}}$ and all $r \in \mathbf{R}$. For any $\sigma \in \Delta^{\mathcal{P}}$, one can define $(\varepsilon+\eta)$–acyclic carriers

$$\widetilde{F}_\sigma^{(\varepsilon+\eta)} \colon \mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{X})}(\sigma) \rightrightarrows \mathcal{J}_{\mathcal{Q}}^{\mathrm{VR}_*(\mathbb{Y})}(\sigma)$$

by sending a cell $\prod_{V \in \sigma} \tau_V \in \mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{X})}(\sigma)_r$ to $\prod_{V \in \sigma} \Sigma^\eta \mathrm{VR}_*(\mathbb{Y} \cap V)(F_V^\varepsilon(\tau_V)) \in \mathcal{J}_{\mathcal{Q}}^{\mathrm{VR}_*(\mathbb{Y})}(\sigma)_{r+(\varepsilon+\eta)}$ for all $r \in \mathbf{R}$. Similarly, we assume the converse that $\mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{Y})}(\tilde\sigma)_r \neq \varnothing$ implies $\mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{X})}(\phi^{-1}(\tilde\sigma))_{r+\eta} \neq \varnothing$ for all $\tilde\sigma \in \Delta^{\mathcal{Q}}$ and all $r \in \mathbf{R}$. With an analogous definition to that of $\widetilde{F}_\sigma^{(\varepsilon+\eta)}$, we obtain "inverses" for the carriers $\widetilde{F}_\sigma^{(\varepsilon+\eta)}$, so that these become $(\varepsilon+\eta)$–acyclic equivalences. One can check that these are compatible along $\Delta^{\mathcal{P}}$ and $\Delta^{\mathcal{Q}}$, so by Theorem 6.5

$$d_I^1\big(E_{*,*}^*(\mathcal{J}_{\mathcal{P}}^{\mathrm{VR}_*(\mathbb{X})}, \Delta^{\mathcal{P}}), E_{*,*}^*(\mathcal{J}_{\mathcal{Q}}^{\mathrm{VR}_*(\mathbb{Y})}, \Delta^{\mathcal{Q}})\big) \leq \varepsilon + \eta.$$

7 Interleavings with respect to different covers

7.1 Refinement induced interleavings

In the previous sections we considered general diagrams in **FRDiag**(K) for some simplicial complex K. We now focus on the situation where we have a filtered complex X together with a cover $\mathcal{U}$, which provides a diagram $X^{\mathcal{U}} \colon N_{\mathcal{U}} \to$ **FCW-cpx**. The associated spectral sequence is denoted by $E_{*,*}^*(X, \mathcal{U})$, as done at the start of Section 3. We want to measure how $E_{*,*}^*(X, \mathcal{U})$ changes depending on $\mathcal{U}$ and follow ideas from [Serre 1955] to achieve this. First we consider a refinement $\mathcal{V} \prec \mathcal{U}$, which means that for all $V \in \mathcal{V}$, there exists $U \in \mathcal{U}$ such that $V \subseteq U$. In particular, one can choose a morphism $\rho^{\mathcal{U},\mathcal{V}} \colon N_{\mathcal{V}} \to N_{\mathcal{U}}$ such that $\mathcal{V}_\sigma \subseteq \mathcal{U}_{\rho\sigma}$ for all $\sigma \in N_{\mathcal{V}}$. This choice is of course not necessarily unique. We would like to compare the Mayer–Vietoris spectral sequences of both covers. For this, we recall the definition of the Čech chain complex outlined in the introduction of [Torras-Casas 2023], which leads to the following isomorphism on the terms from the 0–page:

$$(5) \qquad E_{p,q}^0(X, \mathcal{U}) = \check{\mathscr{C}}_p(\mathcal{U}; C_q^{\mathrm{cell}}) := \bigoplus_{\sigma \in N_{\mathcal{U}}^p} C_q^{\mathrm{cell}}(\mathcal{U}_\sigma) \simeq \bigoplus_{s \in X^q} f_*^{\sigma(s, \mathcal{U})}(C_p^{\mathrm{cell}}(\Delta^{\sigma(s, \mathcal{U})})).$$

Here, $\sigma(s, \mathcal{U})$ is the simplex of maximal dimension in $N_{\mathcal{U}}$ such that $s \in X^{\mathcal{U}}(\sigma(s, \mathcal{U}))$, and

$$f^{\sigma(s, \mathcal{U})} \colon \Delta^{\sigma(s, \mathcal{U})} \hookrightarrow N_{\mathcal{U}}$$

denotes the inclusion. The isomorphism in (5) is given by sending a generator

$$(a)_\sigma \in \bigoplus_{\sigma \in N_{\mathcal{U}}^p} C_q^{\mathrm{cell}}(\mathcal{U}_\sigma)$$

to its transpose $(\sigma)_a$, for all cells $a \in X$ and all $\sigma \in N_{\mathcal{U}}$.

Returning to a refinement $\mathcal{V} \prec \mathcal{U}$ and a morphism $\rho^{\mathcal{U},\mathcal{V}} \colon N_{\mathcal{V}} \to N_{\mathcal{U}}$, there is an induced double complex morphism $\rho_{p,q}^{\mathcal{U},\mathcal{V}} \colon C_{p,q}(X, \mathcal{V}) \to C_{p,q}(X, \mathcal{U})$ given by

$$\rho_{p,q}^{\mathcal{U},\mathcal{V}}((\sigma)_a) = \begin{cases} (\rho^{\mathcal{U},\mathcal{V}}\sigma)_a & \text{if } \dim(\rho^{\mathcal{U},\mathcal{V}}\sigma) = p, \\ 0 & \text{otherwise,} \end{cases}$$

for all generators $(\sigma)_a \in C_{p,q}(X, \mathcal{V})$ with $\sigma \in N_{\mathcal{V}}^p$ and $a \in X^q$.

Lemma 7.1 $\rho_{*,*}^{\mathcal{U},\mathcal{V}}$ *is a morphism of double complexes. Thus, it induces a morphism of spectral sequences*

$$\rho_{p,q}^{\mathcal{U},\mathcal{V}} \colon E_{p,q}^*(X, \mathcal{V}) \to E_{p,q}^*(X, \mathcal{U})$$

dependent on the choice of $\rho^{\mathcal{U},\mathcal{V}}$.

Proof Let $\delta^{\mathcal{V}}$ and $\delta^{\mathcal{U}}$ denote the respective Čech differentials from $\check{\mathscr{C}}_p(\mathcal{V}; C_q^{\mathrm{cell}})$ and $\check{\mathscr{C}}_p(\mathcal{U}; C_q^{\mathrm{cell}})$. The refinement $\rho^{\mathcal{U},\mathcal{V}} \colon N_{\mathcal{V}} \to N_{\mathcal{U}}$ induces a chain morphism $\rho_*^{\mathcal{U},\mathcal{V}} \colon C_*^{\mathrm{cell}}(N_{\mathcal{V}}) \to C_*^{\mathrm{cell}}(N_{\mathcal{U}})$, so that we have commutativity $\rho_{*,*}^{\mathcal{U},\mathcal{V}} \circ \delta^{\mathcal{V}} = \delta^{\mathcal{U}} \circ \rho_{*,*}^{\mathcal{U},\mathcal{V}}$. This implies that $\rho_{*,*}^{\mathcal{U},\mathcal{V}}$ commutes with the horizontal differential d^H.

For commutativity with d^V, we consider a generating chain $(\sigma)_a \in E^0_{p,q}(X, \mathcal{V})$ with $\sigma \in N^p_{\mathcal{V}}$ and $a \in X^q$. Then, if $\dim(\rho^{\mathcal{U},\mathcal{V}}\sigma) = p$,

$$\rho^{\mathcal{U},\mathcal{V}}_{p,q-1} \circ d^V((\sigma)_a) = \rho^{\mathcal{U},\mathcal{V}}_{p,q-1}\left((-1)^p \sum_{b \preceq \bar{a}} ([b:a]\sigma)_b\right)$$

$$= (-1)^p \sum_{b \preceq \bar{a}} ([b:a]\rho^{\mathcal{U},\mathcal{V}}\sigma)_b = (-1)^p d^{\text{cell}}_q((\rho^{\mathcal{U},\mathcal{V}}\sigma)_a) = d^V \circ \rho^{\mathcal{U},\mathcal{V}}_{p,q}((\sigma)_a)$$

and for $\dim(\rho^{\mathcal{U},\mathcal{V}}\sigma) < p$ commutativity follows since both terms vanish.

A morphism of double complexes gives rise to a morphism of the vertical filtration. By [McCleary 2001, Theorem 3.5] this induces a morphism of spectral sequences $\rho^{\mathcal{U},\mathcal{V}}_{*,*}$. $\qquad\square$

Since $\rho^{\mathcal{U},\mathcal{V}}: N_{\mathcal{V}} \to N_{\mathcal{U}}$ is not unique, the induced morphism $\rho^{\mathcal{U},\mathcal{V}}_{*,*}$ on the 0–page does not need to be unique either. We have, however, the following:

Proposition 7.2 *The 2–morphism obtained by restricting $\rho^{\mathcal{U},\mathcal{V}}_{*,*}$ is independent of the particular choice of refinement map $\rho^{\mathcal{U},\mathcal{V}}: N_{\mathcal{V}} \to N_{\mathcal{U}}$.*

Proof We have to show that $\rho^{\mathcal{U},\mathcal{V}}_{*,*}$ is independent of the particular choice of the refinement morphism. First, define a carrier $R: N_{\mathcal{V}} \rightrightarrows N_{\mathcal{U}}$ by the assignment

$$\sigma \mapsto R(\sigma) = \{\nu \in N_{\mathcal{U}} \mid V_\sigma \subseteq U_\nu\}.$$

The geometric realization of the subcomplex $R(\sigma)$ is homeomorphic to a standard simplex, in particular contractible, so R is acyclic. Note that $\rho^{\mathcal{U},\mathcal{V}}_{*,*}$ is carried by R. Hence, by Theorem 2.4 for any pair of refinement maps $\rho^{\mathcal{U},\mathcal{V}}, \tau^{\mathcal{U},\mathcal{V}}: N_{\mathcal{V}} \to N_{\mathcal{U}}$, there exists a chain homotopy $k_*: C_n(N_{\mathcal{V}}) \to C_{n+1}(N_{\mathcal{U}})$ carried by R such that

$$k_* \delta^{\mathcal{V}} + \delta^{\mathcal{U}} k_* = \tau^{\mathcal{U},\mathcal{V}}_* - \rho^{\mathcal{U},\mathcal{V}}_*$$

for all $n \geq 0$ and where $\tau^{\mathcal{U},\mathcal{V}}_*$ and $\rho^{\mathcal{U},\mathcal{V}}_*$ are induced morphisms of chain complexes $C_*(N_{\mathcal{V}}) \to C_*(N_{\mathcal{U}})$. In particular, using the same notation, this translates into chain homotopies $k_*: E^0_{p,q}(X, \mathcal{V}) \to E^0_{p+1,q}(X, \mathcal{U})$ on the 0–page such that

$$k_* \delta^{\mathcal{V}} + \delta^{\mathcal{U}} k_* = \tau^{\mathcal{U},\mathcal{V}}_{*,*} - \rho^{\mathcal{U},\mathcal{V}}_{*,*}.$$

Thus, $\tau^{\mathcal{U},\mathcal{V}}_{*,*} = \rho^{\mathcal{U},\mathcal{V}}_{*,*}$ from the second page onward. $\qquad\square$

Example 7.3 Consider a filtered cubical complex $\mathcal{C}_*$. At value 0, $\mathcal{C}_*$ is given by the vertices on $\mathbb{R}^2$ at the coordinates $a = (0,0)$, $b = (1,0)$, $c = (2,0)$, $d = (3,0)$, $e = (0,1)$, $f = (1,1)$, $g = (2,1)$ and $h = (3,1)$, together with all edges contained in the boundary of the rectangle $adhe$. Then, at value 1 there appears the edge bf with the face $abfe$. At value 2 the edge gc with the face $fgcb$, and finally at value 3 the face $ghdc$ appears. This is depicted on Figure 2. Then, consider the cover $\mathcal{U}_0$ by three subcomplexes on the squares $A = (a, b, f, e)$, $B = (b, c, g, f)$ and $C = (c, d, h, g)$. Also, we consider the cover $\mathcal{U}_1$ given by A and $C \cup B$, and $\mathcal{U}_2$ given by all $\mathcal{C}_*$. The induced morphisms on second-page terms at different filtration values are either null or the identity, as illustrated on Figure 3.
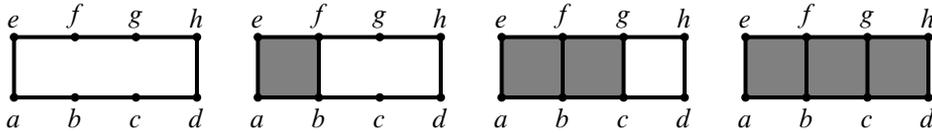
Figure 2: Cubical complex $\mathscr{C}_*$ at values 0, 1, 2 and 3.

A consequence of Proposition 7.2 is that if we have a space X together with covers $\mathcal{U} \prec \mathcal{V} \prec \mathcal{U}$, then by uniqueness the morphism on the second page induced by the consecutive inclusions coincides with the identity. This gives rise to the next result.

Proposition 7.4 *Suppose a pair of covers $\mathcal{U}$ and $\mathcal{V}$ of X are a refinement of one another. Then there is a 2–spectral sequence isomorphism $E^2_{*,*}(X, \mathcal{U}) \simeq E^2(X, \mathcal{V})$.*

This implies that for any cover $\mathcal{U}$ of X, the cover $\mathcal{U} \cup X$ obtained by adding the extra covering element X is such that the second page $E^2_{p,q}(X, \mathcal{U} \cup X)$ has only the first column nonzero.

Lemma 7.5 *Consider a cover $\mathcal{U}$ of a space X, and suppose that $X \in \mathcal{U}$. Then $E^2_{p,q}(X, \mathcal{U}) = 0$ for all $p > 0$.*

Proof This follows from the observation that the cover $\{X\}$ consisting of a single element satisfies $\{X\} \prec \mathcal{U} \prec \{X\}$. Using Proposition 7.4 we therefore obtain isomorphisms $E^2_{p,q}(X, \mathcal{U}) \simeq E^2_{p,q}(X, \{X\})$, and the result follows.   □

Suppose that none of the two covers $\mathcal{V}$ and $\mathcal{U}$ refines the other. One can still compare them using the common refinement $\mathcal{V} \cap \mathcal{U} = \{V \cap U\}_{V \in \mathcal{V}, U \in \mathcal{U}}$ which is a cover of X. Thus, there are two refinement morphisms

$$(6) \qquad E^2_{p,q}(X, \mathcal{U}) \xleftarrow{\rho^{\mathcal{U}, \mathcal{V} \cap \mathcal{U}}_{p,q}} E^2_{p,q}(X, \mathcal{V} \cap \mathcal{U}) \xrightarrow{\rho^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}_{p,q}} E^2_{p,q}(X, \mathcal{V}).$$
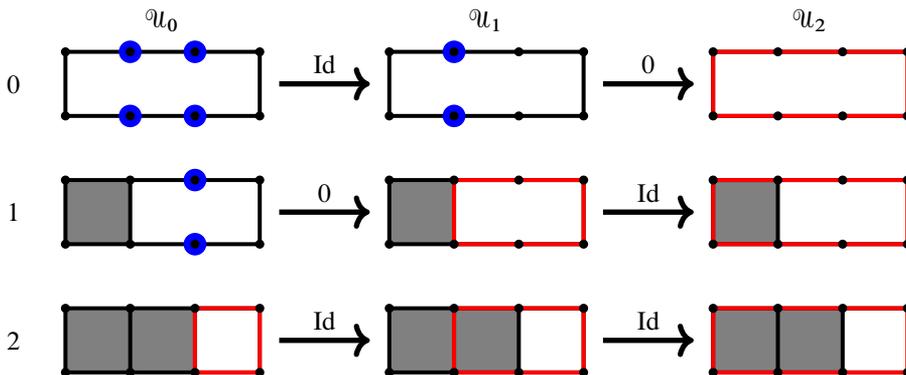
Figure 3: Cubical complex $\mathscr{C}_*$ with covers $\mathcal{U}_0$, $\mathcal{U}_1$ and $\mathcal{U}_2$, and with filtration values 0, 1 and 2. Blue dots represent classes in $E^2_{1,0}(\mathscr{C}, \mathcal{U}_i)$ and red loops represent classes on $E^2_{0,1}(\mathscr{C}, \mathcal{U}_i)$, for $i = 0, 1, 2$.
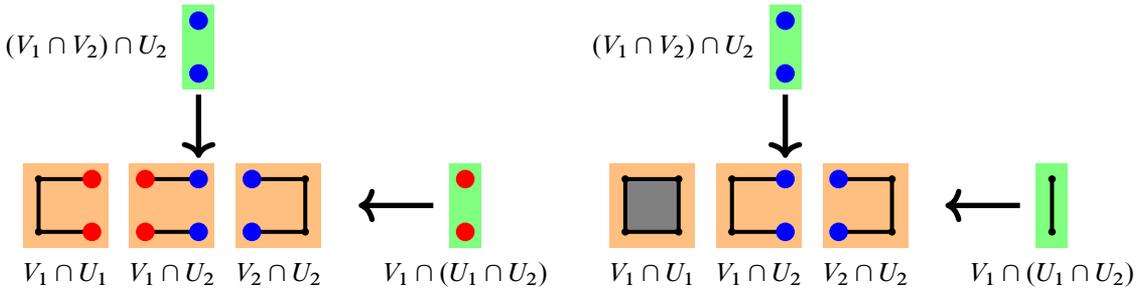
Figure 4: $C_{p,q}(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k)$ at filtration values 0 and 1.

Following [Serre 1955, Section 28] we can now build the double complex $C_{p,q}(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k)$ which, for each $k \geq 0$, is given by

$$
\begin{array}{ccc}
\displaystyle\bigoplus_{\substack{\sigma \in N_{\mathcal{V}}^{p+1} \\ \tau \in N_{\mathcal{U}}^{q}}} \mathrm{PH}_k(\mathcal{V}_\sigma \cap \mathcal{U}_\tau) & \xleftarrow{(-1)^{p+1}\delta^{\mathcal{U}}} & \displaystyle\bigoplus_{\substack{\sigma \in N_{\mathcal{V}}^{p+1} \\ \tau \in N_{\mathcal{U}}^{q+1}}} \mathrm{PH}_k(\mathcal{V}_\sigma \cap \mathcal{U}_\tau) \\
\downarrow{\scriptstyle \delta^{\mathcal{V}}} & & \downarrow{\scriptstyle \delta^{\mathcal{V}}} \\
\displaystyle\bigoplus_{\substack{\sigma \in N_{\mathcal{V}}^{p} \\ \tau \in N_{\mathcal{U}}^{q}}} \mathrm{PH}_k(\mathcal{V}_\sigma \cap \mathcal{U}_\tau) & \xleftarrow{(-1)^{p}\delta^{\mathcal{U}}} & \displaystyle\bigoplus_{\substack{\sigma \in N_{\mathcal{V}}^{p} \\ \tau \in N_{\mathcal{U}}^{q+1}}} \mathrm{PH}_k(\mathcal{V}_\sigma \cap \mathcal{U}_\tau)
\end{array}
$$

for any pair of integers $p, q \geq 0$. From this double complex we can study the two associated spectral sequences

$$
{}^{\mathrm{I}}E_{p,q}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) = \bigoplus_{\sigma \in N_{\mathcal{V}}^p} \check{\mathcal{H}}_q(\mathcal{V}_\sigma \cap \mathcal{U}; \mathrm{PH}_k),
$$

$$
{}^{\mathrm{II}}E_{p,q}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) = \bigoplus_{\tau \in N_{\mathcal{U}}^q} \check{\mathcal{H}}_p(\mathcal{V} \cap \mathcal{U}_\tau; \mathrm{PH}_k),
$$

whose common target of convergence is $\check{\mathcal{H}}_n(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k)$ with $p + q = n$. For details about the spectral sequence associated to a double complex, the reader is recommended to look at [McCleary 2001, Theorem 2.15].

Example 7.6 Consider the cubical complex $\mathcal{C}_*$ from Example 7.3. Set $\mathcal{U} = \mathcal{U}_1$, that is, $\mathcal{U}$ is the cover by the sets $U_1 = A$ and $U_2 = B \cup C$. On the other hand, consider $\mathcal{V}$ to be formed of $V_1 = A \cup B$ and $V_2 = C$. The double complex $C_{p,q}(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k)$ is illustrated on Figure 4 for filtration values 0 and 1, and for $k = 0$. We encourage the reader to work out the refinement morphisms from (6) and see that these are actually projections.

Consider the nerve $N_{\mathcal{V} \cap \mathcal{U}}$ as a subset of the product of nerves $N_{\mathcal{V}} \times N_{\mathcal{U}}$. We have then two projections $\pi^{\mathcal{V}} \colon N_{\mathcal{V} \cap \mathcal{U}} \to N_{\mathcal{V}}$ and $\pi^{\mathcal{U}} \colon N_{\mathcal{V} \cap \mathcal{U}} \to N_{\mathcal{U}}$, both of which induce chain morphisms $\pi_*^{\mathcal{V}} \colon C_*(N_{\mathcal{V} \cap \mathcal{U}}) \to C_*(N_{\mathcal{V}})$

and $\pi_*^{\mathcal{U}} \colon C_*(N_{\mathcal{V} \cap \mathcal{U}}) \to C_*(N_{\mathcal{U}})$. For example, $\pi_*^{\mathcal{V}}$ is given by $\pi_*^{\mathcal{V}}(\sigma \times \tau) = \sigma$ if $\dim(\tau) = 0$ or $\pi_*^{\mathcal{V}}(\sigma \times \tau) = 0$ otherwise, for all $\sigma \in N_{\mathcal{V}}$ and $\tau \in N_{\mathcal{U}}$. These induce a pair of morphisms

$$\bigoplus_{\sigma \in N_{\mathcal{V}}^p} C_k^{\text{cell}}(\mathcal{V}_\sigma) \xleftarrow{\pi_{p,k}^{\mathcal{V}}} \bigoplus_{\substack{\sigma \in N_{\mathcal{V}}^p \\ \tau \in N_{\mathcal{U}}^q}} C_k^{\text{cell}}(\mathcal{V}_\sigma \cap \mathcal{U}_\tau) \xrightarrow{\pi_{q,k}^{\mathcal{U}}} \bigoplus_{\tau \in N_{\mathcal{U}}^q} C_k^{\text{cell}}(\mathcal{U}_\tau),$$

for any pair of integers $p, q \geq 0$. The induced map $\pi_{p,k}^{\mathcal{V}}$ on $C_k(\mathcal{V}_\sigma \cap \mathcal{U}_\tau)$ satisfies

$$\pi_{p,k}^{\mathcal{V}}((\sigma \times \tau)_a) = (\pi_*^{\mathcal{V}}(\sigma \times \tau))_a$$

for all $\sigma \in N_{\mathcal{V}}^p$, $\tau \in N_{\mathcal{U}}$ and all $a \in (\mathcal{V}_\sigma \cap \mathcal{U}_\tau)^k$. The map $\pi_{*,*}^{\mathcal{U}}$ acts similarly. By definition $\pi_{*,*}^{\mathcal{U}}$ and $\pi_{*,*}^{\mathcal{V}}$ both commute with the Čech differentials $\delta^{\mathcal{U}}$ and $\delta^{\mathcal{V}}$ respectively. Let $\sigma \in N_{\mathcal{V}}^p$ and $\tau \in N_{\mathcal{U}}^0$. Then we have

$$
\begin{array}{ccc}
(\sigma \times \tau)_a & \xrightarrow{\quad \pi_{*,*}^{\mathcal{V}} \quad} & (\sigma)_a \\
\downarrow{\scriptstyle d_n} & & \downarrow{\scriptstyle d_n} \\
\sum_{b \in \bar{a}} ([b:a]\sigma \times \tau)_b & \xrightarrow{\quad \pi_{*,*}^{\mathcal{V}} \quad} & \sum_{b \in \bar{a}} ([b:a]\sigma)_b
\end{array}
$$

for all cells $a \in (\mathcal{V}_\sigma \cap \mathcal{U}_\tau)^k$. This implies that $\pi_{*,*}^{\mathcal{V}}$ commutes with d_n and the same holds for $\pi_{*,*}^{\mathcal{U}}$. We obtain a morphism $\pi_{p,k}^{\mathcal{V}} \colon \check{\mathscr{C}}_p(\mathcal{V} \cap \mathcal{U}; C_k^{\text{cell}}) \to \check{\mathscr{C}}_p(\mathcal{V}; C_k^{\text{cell}})$ commuting with d_* and $\delta^{\mathcal{V} \cap \mathcal{U}}$ and $\delta^{\mathcal{V}}$. This induces $\kappa_{p,k}^{\mathcal{V}} \colon \check{\mathscr{C}}_p(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \to \check{\mathscr{C}}_p(\mathcal{V}; \mathrm{PH}_k)$ and, in turn, this induces

$$\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}} \colon \check{\mathscr{H}}_p(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \to \check{\mathscr{H}}_p(\mathcal{V}; \mathrm{PH}_k).$$

There is a very natural way of understanding how much $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}$ fails to be an isomorphism. To start, notice that $\kappa_{p,k}^{\mathcal{V}}$ is equal to the composition

$$\check{\mathscr{C}}_p(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \twoheadrightarrow {}^{\mathrm{I}}E_{p,0}^0(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \xrightarrow{{}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}} \check{\mathscr{C}}_p(\mathcal{V}; \mathrm{PH}_k),$$

where the first morphism forgets the summands with $\tau \notin N_{\mathcal{U}}^0$; the second morphism is the restriction of $\kappa_{p,k}^{\mathcal{V}}$ to the remaining terms. Next, we take for each simplex $\sigma \in N_{\mathcal{V}}^p$, the Mayer–Vietoris spectral sequence for $\mathcal{V}_\sigma$ covered by $\mathcal{V}_\sigma \cap \mathcal{U}$

$$M_{q,k}^2(\mathcal{V}_\sigma \cap \mathcal{U}) \Rightarrow \mathrm{PH}_{q+k}(\mathcal{V}_\sigma),$$

where we changed the notation from $E_{q,k}^2(\mathcal{V}_\sigma, \mathcal{V}_\sigma \cap \mathcal{U})$ to $M_{q,k}^2(\mathcal{V}_\sigma \cap \mathcal{U})$ as it helps distinguishing this spectral sequence from ${}^{\mathrm{I}}E_{p,q}^*$. Then, we write more compactly

$$ {}^{\mathrm{I}}E_{p,q}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) = \bigoplus_{\sigma \in N_{\mathcal{V}}^p} M_{q,k}^2(\mathcal{V}_\sigma \cap \mathcal{U}).$$

Taking ${}^{\mathrm{I}}E_{p,0}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k)$ as a chain complex, ${}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}$ induces a chain morphism

$$ {}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}} \colon {}^{\mathrm{I}}E_{p,0}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \to \check{\mathscr{C}}_p(\mathcal{V}; \mathrm{PH}_k)$$

for all $p \geq 0$. In particular, the restriction of ${}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}$ to the summand $M_{0,k}^2(\mathcal{V}_\sigma \cap \mathcal{U})$ equals the composition

$$M_{0,k}^2(\mathcal{V}_\sigma \cap \mathcal{U}) \twoheadrightarrow M_{0,k}^\infty(\mathcal{V}_\sigma \cap \mathcal{U}) \hookrightarrow \mathrm{PH}_k(\mathcal{V}_\sigma).$$

Notice that PH_0 is a cosheaf, and in this case $M_{0,0}^2(\mathcal{V}_\sigma \cap \mathcal{U}) = \mathrm{PH}_0(\mathcal{V}_\sigma)$ for all $\sigma \in N_{\mathcal{V}}^p$. This implies that ${}^{\mathrm{I}}\pi_{p,0}^{\mathcal{V}}$ is an isomorphism for all $p \geq 0$. By the same argument, there is another chain morphism for all $q \geq 0$,

$$ {}^{\mathrm{II}}\pi_{q,k}^{\mathcal{U}} : {}^{\mathrm{II}}E_{0,q}^1(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \to \check{\mathcal{C}}_q(\mathcal{U}; \mathrm{PH}_k).$$

Going back to the morphism $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}$, it is given by the composition

$$\check{\mathcal{H}}_p(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \twoheadrightarrow {}^{\mathrm{I}}E_{p,0}^\infty(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \hookrightarrow {}^{\mathrm{I}}E_{p,0}^2(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k) \xrightarrow{{}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}} \check{\mathcal{H}}_p(\mathcal{V}; \mathrm{PH}_k).$$

Using Lemma 7.5, if $\mathcal{V} \prec \mathcal{U}$ then $M_{q,k}^2(\mathcal{V}_\sigma \cap \mathcal{U}) = 0$ for all $q > 0$ and ${}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}$ becomes an isomorphism. In addition, ${}^{\mathrm{I}}E_{p,q}^1 = 0$ for all $q > 0$ and the first two arrows in the above factorisation of $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}$ are isomorphisms. Altogether, the inverse $(\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}})^{-1}$ is well defined, and by composition we define morphisms $\theta_{p,k}^{\mathcal{U}, \mathcal{V}} = \theta_{p,k}^{\mathcal{U}, \mathcal{V} \cap \mathcal{U}} \circ (\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}})^{-1}$. Here notice that $\theta_{p,k}^{\mathcal{U}, \mathcal{V} \cap \mathcal{U}}$ is defined in an analogous way to $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}$, but it factors through ${}^{\mathrm{II}}\pi_{q,k}^{\mathcal{U}}$ instead of ${}^{\mathrm{I}}\pi_{p,k}^{\mathcal{V}}$. The following proposition should also follow from applying an appropriate version of the universal coefficient theorem to [Serre 1955, Proposition 4.4]. Instead, we prove the dual statement of this proposition by means of acyclic carriers.

Proposition 7.7 *Suppose that $\mathcal{V} \prec \mathcal{U}$, and let $\rho^{\mathcal{U}, \mathcal{V}}$ denote a refinement map. The morphism*

$$\theta_{p,k}^{\mathcal{U}, \mathcal{V}} : E_{p,k}^2(X, \mathcal{V}) \to E_{p,k}^2(X, \mathcal{U})$$

coincides with the standard morphism induced by $\rho^{\mathcal{U}, \mathcal{V}}$.

Proof Since $\mathcal{V} \prec \mathcal{U}$, the morphism $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}} : \check{\mathcal{H}}_p(\mathcal{V} \cap \mathcal{U}, \mathrm{PH}_k) \to \check{\mathcal{H}}_p(\mathcal{V}, \mathrm{PH}_k)$ is an isomorphism. Now consider the diagram

$$
\begin{array}{ccc}
\check{\mathcal{H}}_p(\mathcal{V}; \mathrm{PH}_k) & \xrightarrow{\rho_{p,k}^{\mathcal{U}, \mathcal{V}}} & \check{\mathcal{H}}_p(\mathcal{U}; \mathrm{PH}_k) \\
\uparrow {\scriptstyle \simeq} & & \uparrow {\scriptstyle {}^{\mathrm{II}}\pi_{p,k}^{\mathcal{U}}} \\
\check{\mathcal{H}}_p(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \twoheadrightarrow & {}^{\mathrm{II}}E_{0,p}^\infty(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \hookrightarrow & {}^{\mathrm{II}}E_{0,p}^2(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k)
\end{array}
$$

To check that it commutes we study triangles of acyclic carriers

where R is defined in Proposition 7.2. The carrier F is given for every $\sigma \in N_{\mathcal{V}}$ by $F(\sigma) = \Delta^\sigma \times |R(\sigma)|$. In fact, F defines an acyclic equivalence by considering the inverse carrier $P_{\mathcal{V}} : N_{\mathcal{V} \cap \mathcal{U}} \rightrightarrows N_{\mathcal{V}}$ sending $\sigma \times \tau$ to Δ^σ. In this case the shift carrier $I_{\mathcal{V}} : N_{\mathcal{V}} \rightrightarrows N_{\mathcal{V}}$ is given by the assignment $\sigma \mapsto \Delta^\sigma$, and

$I_{\mathcal{V} \cap \mathcal{U}} \colon N_{\mathcal{V} \cap \mathcal{U}} \rightrightarrows N_{\mathcal{V} \cap \mathcal{U}}$ is given by $\sigma \times \tau \mapsto \Delta^\sigma \times \Delta^{\tau \cup \tau'}$, where $\tau' \in N_{\mathcal{U}}$ is such that $|R(\sigma)| = \Delta^{\tau'} \subseteq N_{\mathcal{U}}$. Here, we need to show that $\Delta^\sigma \times \Delta^{\tau \cup \tau'}$ is a subcomplex of $N_{\mathcal{V} \cap \mathcal{U}}$. First notice that, by hypotheses, $\mathcal{V}_\sigma \cap \mathcal{U}_\tau \neq \varnothing$ and, by definition of $R(\sigma)$, we have $\mathcal{V}_\sigma \subseteq \mathcal{U}_{\tau'}$. Consequently $\mathcal{V}_\sigma \cap (\mathcal{U}_\tau \cap \mathcal{U}_{\tau'}) \neq \varnothing$, which accounts for $\Delta^\sigma \times \Delta^{\tau \cup \tau'}$ being a subcomplex of $N_{\mathcal{V} \cap \mathcal{U}}$.

Since F is acyclic, there exists $\nu_* \colon C_*(N_{\mathcal{V}}) \to C_*(N_{\mathcal{V} \cap \mathcal{U}})$ carried by F and inducing a chain morphism $f_* \colon \check{\mathscr{C}}_p(\mathcal{V}, C_k^{\mathrm{cell}}) \to \check{\mathscr{C}}_p(\mathcal{V} \cap \mathcal{U}, C_k^{\mathrm{cell}})$ by the assignment $(\sigma)_s \mapsto (\nu_*(\sigma))_s$ for all cells $s \in X$ and all $\sigma \in N_{\mathcal{V}}$. On the other hand, recall that $\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}}$ is induced by $\pi_{p,k}^{\mathcal{V}}$, which is given as an assignment

$$(\sigma \times \tau)_s \to (\pi_*^{\mathcal{V}}(\sigma \times \tau))_s.$$

As $\pi_*^{\mathcal{V}}$ is carried by $P_{\mathcal{V}}$ and, as noted earlier, F defines an acyclic equivalence, it follows that $\pi_*^{\mathcal{V}} \circ \nu_*$ is the identity in $C_*(N_{\mathcal{V}})$ up to boundary. Thus, $\pi_{p,k}^{\mathcal{V}} \circ f_*$ is the identity in $\check{\mathscr{C}}_p(\mathcal{V}, C_k^{\mathrm{cell}})$ up to the Čech boundary $\check{\delta}_{\mathcal{V}}$. This implies that $f_* = (\theta_{p,k}^{\mathcal{V}, \mathcal{V} \cap \mathcal{U}})^{-1}$ as morphisms $\check{\mathscr{H}}_p(\mathcal{V}, \mathrm{PH}_k) \to \check{\mathscr{H}}_p(\mathcal{V} \cap \mathcal{U}, \mathrm{PH}_k)$. Consequently, $\theta_{p,k}^{\mathcal{U}, \mathcal{V}}$ is induced by the assignment $(\sigma)_s \mapsto (\pi_*^{\mathcal{U}} \circ \nu_*(\sigma))_s$ for all $\sigma \in N_{\mathcal{V}}$ and all $s \in X$, where $\pi_*^{\mathcal{U}} \circ \nu_*$ is carried by $P_{\mathcal{U}} F = R$. Altogether, as $\rho^{\mathcal{U}, \mathcal{V}}$ is carried by R, we obtain the equality $\theta_{p,k}^{\mathcal{U}, \mathcal{V}} = \rho_{p,k}^{\mathcal{U}, \mathcal{V}}$ as morphisms $\check{\mathscr{H}}_p(\mathcal{V}, \mathrm{PH}) \to \check{\mathscr{H}}_p(\mathcal{U}, \mathrm{PH})$. $\qquad \square$

Still assuming that $\mathcal{V} \prec \mathcal{U}$, we now look for conditions for the existence of an inverse of $\theta_{p,k}^{\mathcal{U}, \mathcal{V}}$,

$$\varphi_{p,k}^{\mathcal{V}, \mathcal{U}} \colon E_{p,k}^2(X, \mathcal{U}) \to E_{p,k}^2(X, \mathcal{V}).$$

Proposition 7.8 *Suppose that $\mathcal{V} \prec \mathcal{U}$. If $M_{p,k}^2(\mathcal{V} \cap \mathcal{U}_\tau) = 0$ for all $p > 0$, $k \geq 0$ and all $\tau \in N_{\mathcal{U}}^q$, then the maps $\theta_{*,*}^{\mathcal{U}, \mathcal{V}}$ induce a 2–isomorphism of spectral sequences*

$$E_{*,*}^{\geq 2}(X, \mathcal{U}) \simeq E_{*,*}^{\geq 2}(X, \mathcal{V}).$$

Proof By Propositions 7.2 and 7.7 we can choose a refinement map $\rho^{\mathcal{U}, \mathcal{V}} \colon N_{\mathcal{V}} \to N_{\mathcal{U}}$ giving a morphism of spectral sequences

$$\rho_{*,*}^{\mathcal{U}, \mathcal{V}} \colon E_{*,*}^{\geq 2}(X, \mathcal{V}) \to E_{*,*}^{\geq 2}(X, \mathcal{U})$$

that coincides with $\theta_{*,*}^{\mathcal{U}, \mathcal{V}}$. Our assumption about $M_{p,k}^2$ implies $^{\mathrm{II}}E_{p,q}^2(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) = 0$ for all $p > 0$, which in turn, gives

(7) $$\mathrm{Ker}\big(\check{\mathscr{H}}_q(\mathcal{V} \cap \mathcal{U}; \mathrm{PH}_k) \twoheadrightarrow {}^{\mathrm{II}}E_{0,q}^\infty(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k)\big) = 0$$

and

(8) $$\mathrm{Coker}\big({}^{\mathrm{II}}E_{0,q}^\infty(\mathcal{V}, \mathcal{U}; \mathrm{PH}_k) \hookrightarrow {}^{\mathrm{II}}E_{0,q}^2(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k)\big) = 0.$$

Now note that $^{\mathrm{II}}\pi_{q,k}^{\mathcal{U}}$ yields an isomorphism $^{\mathrm{II}}E_{0,q}^2(\mathcal{V}, \mathcal{U}, \mathrm{PH}_k) \simeq \check{\mathscr{H}}_q(\mathcal{U}, \mathrm{PH}_k)$. This shows that $\theta_{q,k}^{\mathcal{U}, \mathcal{V}}$ is a composition of isomorphisms; thus the statement follows. $\qquad \square$

We now relax the conditions in Proposition 7.8 and use the relations of *left-interleaving* and *right-interleaving* of persistence modules (denoted by $\sim_L^\varepsilon$ and $\sim_R^\varepsilon$, respectively) to achieve this (see [Govc and Skraba 2018, Section 4]). We have to adapt [Govc and Skraba 2018, Proposition 4.14].

Lemma 7.9 *Suppose that we have persistence modules A, B and C, and a parameter $\varepsilon \geq 0$ such that $A \sim_R^\varepsilon B$ and $B \sim_L^\varepsilon C$. Denote by Φ the morphism $\Phi \colon A \to C$ given by the composition $A \twoheadrightarrow B \hookrightarrow C$. Then there exists $\Psi \colon C \to A[2\varepsilon]$ such that Φ and Ψ define a 2ε–interleaving $A \sim^{2\varepsilon} C$.*

Proof By hypothesis, we have a sequence

$$\mathscr{E}_1 \to A \xrightarrow{f} B \xhookrightarrow{g} C \to \mathscr{E}_2$$

which is exact in A and C and where $\mathscr{E}_1 \sim^\varepsilon 0$ and $\mathscr{E}_2 \sim^\varepsilon 0$. Then, let $v \in C$ and notice that $\Sigma^\varepsilon C(v) \in \mathrm{Im}(g)$. Thus, there exists a unique vector $w \in B$ such that $g(w) = \Sigma^\varepsilon C(v)$. On the other hand, there exists $z \in A$, not necessarily unique, such that $f(z) = w$. This defines a unique element $\Sigma^\varepsilon A(z) \in A$. To see this, suppose that another $z' \in A$ is such that $f(z') = w$. Then $f(z - z') = 0$ and $z - z' \in \mathrm{Ker}(f)$, which implies $0 = \Sigma^\varepsilon A(z - z') = \Sigma^\varepsilon A(z) - \Sigma^\varepsilon A(z')$, and then $\Sigma^\varepsilon A(z) = \Sigma^\varepsilon A(z')$. Altogether, we set $\Psi = \Sigma^\varepsilon A \circ \Phi^{-1} \circ \Sigma^\varepsilon C$, which is well defined. $\qquad\square$

Recall that for $\mathscr{V} \prec \mathscr{U}$ we have that $\check{\mathscr{H}}_q(\mathscr{V}; \mathrm{PH}_k) \simeq \check{\mathscr{H}}_q(\mathscr{V} \cap \mathscr{U}; \mathrm{PH}_k)$ for all $k \geq 0$ and $q \geq 0$. There is a natural way to relax (7) and (8) to the persistent case. We assume that for $\varepsilon \geq 0$, there are right and left interleavings

$$(9) \qquad \check{\mathscr{H}}_q(\mathscr{V} \cap \mathscr{U}; \mathrm{PH}_k) \sim_R^\varepsilon {}^{\mathrm{II}}E_{0,q}^\infty(\mathscr{V}, \mathscr{U}; \mathrm{PH}_k) \sim_L^\varepsilon {}^{\mathrm{II}}E_{0,q}^2(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k).$$

If we define $\Phi_{q,k} \colon \check{\mathscr{H}}_q(\mathscr{V} \cap \mathscr{U}; \mathrm{PH}_k) \to {}^{\mathrm{II}}E_{0,q}^2(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k)$ to be the composition of the associated persistence morphisms as in Lemma 7.9, then there exists

$$\Psi_{q,k} \colon {}^{\mathrm{II}}E_{0,q}^2(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k) \to \check{\mathscr{H}}_q(\mathscr{V} \cap \mathscr{U}; \mathrm{PH}_k)[2\varepsilon]$$

such that $\Phi_{q,k}$ and $\Psi_{q,k}$ define a 2ε–interleaving. We repeat this argument for the local Mayer–Vietoris spectral sequences. Assume that for some $\nu \geq 0$ there are interleavings

$$(10) \qquad {}^{\mathrm{II}}E_{0,q}^1(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k) \sim_R^\nu \bigoplus_{\tau \in N_{\mathscr{U}}^q} M_{k,0}^\infty(\mathscr{V} \cap \mathscr{U}_\tau) \sim_L^\nu \bigoplus_{\tau \in N_{\mathscr{U}}^q} \mathrm{PH}_k(\mathscr{U}_\tau).$$

Let $\Pi_{q,k} \colon {}^{\mathrm{II}}E_{0,q}^1(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k) \to \bigoplus_{\tau \in N_{\mathscr{U}}^q} \mathrm{PH}_k(\mathscr{U}_\tau)$ be the composition of the associated morphisms. By Lemma 7.9 there exists $\Xi_{q,k}$ such that $\Pi_{q,k}$ and $\Xi_{q,k}$ define a 2ν–interleaving. By slight abuse of notation we continue to denote the induced 2ν–interleaving between ${}^{\mathrm{II}}E_{0,q}^2(\mathscr{V}, \mathscr{U}, \mathrm{PH}_k)$ and $\check{\mathscr{H}}_q(\mathscr{U}; \mathrm{PH}_*)$ by $\Pi_{q,k}$ and $\Xi_{q,k}$. Altogether we have that

$$\theta_{q,k}^{\mathscr{U},\mathscr{V}} = \Pi_{q,k} \circ \Phi_{q,k} \circ (\theta_{q,k}^{\mathscr{V},\mathscr{V}\cap\mathscr{U}})^{-1}$$

and in this situation there is an "inverse" $\psi_{q,k}^{\mathscr{V},\mathscr{U}} = \theta_{q,k}^{\mathscr{V},\mathscr{V}\cap\mathscr{U}} \circ \Psi_{q,k} \circ \Xi_{q,k}$, which increases the persistence values by $2(\varepsilon + \nu)$.

Theorem 7.10 *Suppose that $\mathscr{V} \prec \mathscr{U}$ and for $\varepsilon \geq 0$ and $\nu \geq 0$ the interleavings in (9) and (10) hold. Then*

$$\psi_{p,q}^{\mathscr{V},\mathscr{U}} \colon E_{p,q}^*(X, \mathscr{U}) \to E_{p,q}^*(X, \mathscr{V})[2(\varepsilon + \nu)]$$
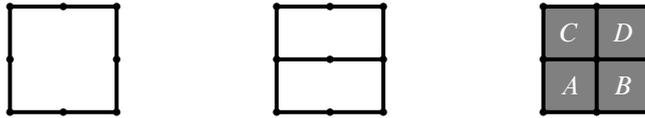
Figure 5: Cubical complex $\mathscr{C}_*$ at values 0, 1 and $1 + \varepsilon$.

is a 2–morphism of spectral sequences such that $\theta_{p,q}^{\mathscr{U},\mathscr{V}}$ and $\psi_{p,q}^{\mathscr{V},\mathscr{U}}$ define a second page $2(\varepsilon+v)$–interleaving between $E_{p,q}^*(X,\mathscr{U})$ and $E_{p,q}^*(X,\mathscr{V})$.

Proof The only thing that remains to be proved is that $\psi_{p,q}^{\mathscr{V},\mathscr{U}}$ commutes with the spectral sequence differentials d_n for all $n \geq 2$. Since these differentials commute with the shift morphisms $\Sigma^{2(\varepsilon+v)}$, this follows from considering the diagram

$$
\begin{array}{ccc}
E_{p,q}^n(X,\mathscr{U}) & \xrightarrow{\quad d_n \quad} & E_{p-n,q+n-1}^n(X,\mathscr{U}) \\
\end{array}
$$

in which the two trapeziums and the two triangles commute. $\qquad\square$

Example 7.11 Consider a cubical complex $\mathscr{C}_*$ as shown in Figure 5, together with the covers

$$
\mathscr{V} = \{\overline{A}, \overline{B}, \overline{C}, \overline{D}\} \quad \text{and} \quad \mathscr{U} = \{\overline{A \cup B}, \overline{C \cup D}\};
$$

see Figure 5 for the cells A, B, C and D. In this case, we have

$$
\check{\mathscr{H}}_1(\mathscr{V}; \mathrm{PH}_0) \simeq \check{\mathscr{H}}_1(\mathscr{V} \cap \mathscr{U}; \mathrm{PH}_0) \simeq \mathrm{I}(0, 1+\varepsilon) \oplus \mathrm{I}(1, 1+\varepsilon) \sim^\varepsilon \mathrm{I}(0, 1) \simeq {}^{\mathrm{II}}E_{0,1}^2(\mathscr{V}, \mathscr{U}, \mathrm{PH}_0)
$$

and also

$$
{}^{\mathrm{II}}E_{0,0}^1(\mathscr{V}, \mathscr{U}, \mathrm{PH}_1) \simeq 0 \sim^\varepsilon \mathrm{I}(1, 1+\varepsilon) \oplus \mathrm{I}(1, 1+\varepsilon) \simeq \bigoplus_{\dim(\tau)=0} \mathrm{PH}_1(\mathscr{U}_\tau).
$$

These interleavings are shown in Figure 6. Theorem 7.10 implies that there is a 4ε–interleaving between $E_{p,q}^*(X,\mathscr{U})$ and $E_{p,q}^*(X,\mathscr{V})$. Notice that in this example, the nontrivial interleaved terms are in different positions of the spectral sequences. Therefore we can improve the upper bound to 2ε. We use this observation later in Proposition 7.12.
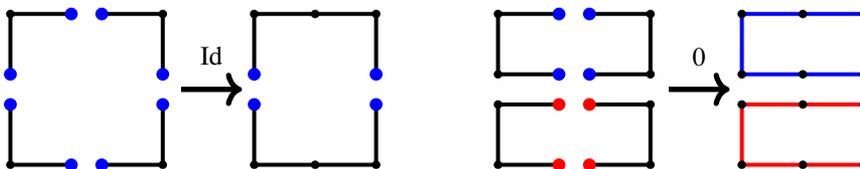
Figure 6: Morphisms $\theta_{1,0}^{\mathscr{U},\mathscr{V}}$ along $[0, 1)$ and along $[1, 1+\varepsilon)$.

7.2 Interpolating covers and spectral sequence interleavings

Consider $X \in \mathbf{FCW\text{-}cpx}$, together with a pair of covers $\mathcal{W}$ and $\mathcal{U}$ such that $\mathcal{W} \prec \mathcal{U}$. Motivated by the interleaving constructed in Theorem 7.10 we take a closer look at the following finite sequence of covers interpolating between $\mathcal{W}$ and a cover that both refines and is refined by $\mathcal{U}$. Let the strict r^{th} intersections of $\mathcal{U}$ be the family of sets $\mathcal{U}^r = \{\mathcal{U}_\tau\}_{\tau \in N_{\mathcal{U}}^r}$ for all $r \geq 0$. We define the $(r, \mathcal{W}, \mathcal{U})$–*interpolation* as the covering set $\mathcal{W}^r = \mathcal{W} \cup \mathcal{U}^r$. In particular, note that the $(0, \mathcal{W}, \mathcal{U})$–interpolation has the property that $\mathcal{W}^0 \prec \mathcal{U} \prec \mathcal{W}^0$, and consequently $E_{p,q}^2(X, \mathcal{U}) \simeq E_{p,q}^2(X, \mathcal{W}^0)$. In addition if $\mathcal{U}$ is a finite cover, then we have $\mathcal{U}^N = \varnothing$ for $N \geq 0$ sufficiently large and consequently $\mathcal{W}^N = \mathcal{W}$.

Proposition 7.12 (local checks) *Let $\mathcal{W} \prec \mathcal{U}$ be a pair of covers for X, where $\mathcal{U}$ is finite. Let $N \geq 0$ be such that $\mathcal{U}^N = \varnothing$. For every $0 \leq r \leq N$, we assume that there exist $\varepsilon_r \geq 0$ and $\nu_r \geq 0$ such that for all $\tau \in N_{\mathcal{U}}^r$,*

$$E_{0,q}^2(\mathcal{U}_\tau, \mathcal{W}_{|\mathcal{U}_\tau}^{r+1}) \sim_R^{\nu_r} E_{0,q}^\infty(\mathcal{U}_\tau, \mathcal{W}_{|\mathcal{U}_\tau}^{r+1}) \sim_L^{\nu_r} \mathrm{PH}_q(\mathcal{U}_\tau),$$

and also

$$d_I(E_{p,q}^2(\mathcal{U}_\tau, \mathcal{W}_{|\mathcal{U}_\tau}^{r+1}), 0) \leq \varepsilon_r$$

for all $p > 0$ and $q \geq 0$. Then we have that

$$d_I^2\big(E_{p,q}^*(X, \mathcal{W}^r), E_{p,q}^*(X, \mathcal{W}^{r+1})\big) \leq 2 \max(\varepsilon_r, \nu_r).$$

Therefore, by using the triangle inequality, we obtain

$$d_I^2\big(E_{p,q}^*(X, \mathcal{U}), E_{p,q}^*(X, \mathcal{W})\big) \leq \sum_{r=0}^{N} 2 \max(\varepsilon_r, \nu_r).$$

Proof We need to consider the spectral sequence $^{\mathrm{II}}E_{p,q}^2(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k)$. Note that, by the construction of $\mathcal{W}^r$, for each $\tau \in N_{\mathcal{W}^r}$ with $\dim(\tau) > 0$ the set $\mathcal{W}_\tau^r$ is contained in one of the open sets from $\mathcal{W}^{r+1}$. By Lemma 7.5 this implies that $^{\mathrm{II}}E_{p,q}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) = 0$ for all $p > 0$, $q > 0$ and $k \geq 0$. Moreover, we have that $^{\mathrm{II}}E_{0,q}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) = \bigoplus_{\tau \in N_{\mathcal{W}^r}^q} \mathrm{PH}_k(\mathcal{W}_\tau^r)$ for all $q > 0$ and $k \geq 0$. The resulting spectral sequence is shown in Figure 7.

As a consequence of these observations condition (10) holds for these indices with $\nu = 0$. In addition, $^{\mathrm{II}}E_{0,q}^2(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) = E_{q,k}^2(X, \mathcal{W}^r)$ holds for all $q \geq 2$ and $k \geq 0$ (see Figures 7 and 8). In particular,

$$^{\mathrm{II}}E_{2,0}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) \qquad\qquad 0 \qquad\qquad\qquad 0 \qquad\qquad\qquad \ddots$$

$$^{\mathrm{II}}E_{1,0}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) \qquad\qquad 0 \qquad\qquad\qquad 0 \qquad\qquad\qquad 0$$

$$^{\mathrm{II}}E_{0,0}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) \xleftarrow{\ d_1\ } \bigoplus_{\tau \in N_{\mathcal{W}^r}^1} \mathrm{PH}_k(\mathcal{W}_\tau^r) \longleftarrow \bigoplus_{\tau \in N_{\mathcal{W}^r}^2} \mathrm{PH}_k(\mathcal{W}_\tau^r) \longleftarrow \bigoplus_{\tau \in N_{\mathcal{W}^r}^3} \mathrm{PH}_k(\mathcal{W}_\tau^r)$$

Figure 7: First page of $^{\mathrm{II}}E_{p,q}^*(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k)$.
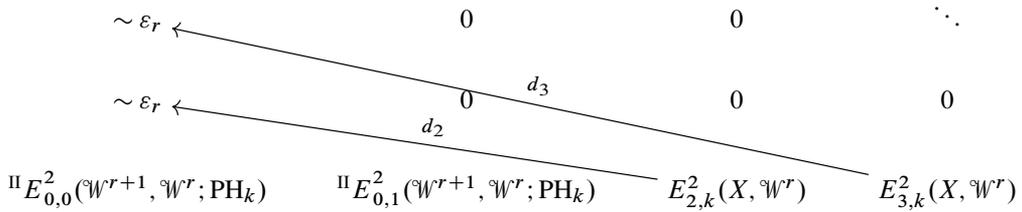
Figure 8: Second page of $^{II}E_{p,q}^*(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k)$ together with higher differentials.

there is only one possible nontrivial differential for each entry in the bottom row as indicated in Figure 8. Note that our hypothesis $d_I(E_{p,q}^2(\mathcal{U}_\tau, \mathcal{W}_{|\mathcal{U}_\tau}^{r+1}), 0) \leq \varepsilon_r$ applies to the entries in the first column with $p > 0$ and gives left and right interleavings of the form

$$\check{\mathcal{H}}_q(\mathcal{W}^{r+1} \cap \mathcal{W}^r; \mathrm{PH}_k) \sim_R^{\varepsilon_r} \; ^{II}E_{0,q}^\infty(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) \sim_L^{\varepsilon_r} \; ^{II}E_{0,q}^2(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k)$$

for all $q > 0$ and $k \geq 0$. Hence, condition (9) holds with value ε_r.

Let us look now at the case $q = 0$. Here we have $\check{\mathcal{H}}_0(\mathcal{W}^{r+1} \cap \mathcal{W}^r; \mathrm{PH}_k) = \; ^{II}E_{0,0}^2(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k)$ and consequently (9) holds with value $\varepsilon = 0$. Next, by hypothesis, for all $k \geq 0$ we have right and left interleavings

$$M_{0,k}^2(\mathcal{U}_\tau \cap \mathcal{W}^{r+1}) \sim_R^{\nu_r} M_{0,k}^\infty(\mathcal{U}_\tau \cap \mathcal{W}^{r+1}) \sim_L^{\nu_r} \mathrm{PH}_k(\mathcal{U}_\tau),$$

for all $\tau \in N_{\mathcal{U}}^r$. Thus by taking the direct sum of these interleavings we obtain

$$^{II}E_{0,0}^1(\mathcal{W}^{r+1}, \mathcal{W}^r; \mathrm{PH}_k) \sim_R^{\nu_r} \bigoplus_{\tau \in N_{\mathcal{W}^r}^0} M_{0,k}^\infty(\mathcal{W}_\tau^r \cap \mathcal{W}^{r+1}) \sim_L^{\nu_r} E_{0,k}^1(X, \mathcal{W}^r).$$

and condition (10) also holds for $q = 0$. The result now follows from Theorem 7.10.

Notice that we can slightly improve the statement of Theorem 7.10 here: for each term in the bottom row of the spectral sequence in this particular example only one of the two conditions (9) and (10) is nontrivial, and the proof of Theorem 7.10 carries over with $2\max(\varepsilon_r, \nu_r)$ replacing $2(\varepsilon_r + \nu_r)$.   □

Remark 7.13 Notice that for reasonable cases the parameters ν_r are bounded above by $K\varepsilon_r$ for some constant $K > 0$ by a result from [Govc and Skraba 2018]. Nevertheless, we would like to keep ν_r and ε_r separated here, since we hope to compute it from $M_{p,k}^*(\mathcal{U}_\tau, \mathcal{W}_{|\mathcal{U}_\tau}^{r+1})$ for $\tau \in N_{\mathcal{U}}^r$ hereby get more accurate estimates. Intuitively, asking for ε_r and ν_r to be small is equivalent to asking for cycle representatives in covers from $\mathcal{W}^r$ to be approximately contained in covering sets from $\mathcal{W}^{r+1}$.

Finally, we would like to compare two separate covers $\mathcal{U}$ and $\mathcal{V}$ and have an estimate for the interleaving distance between the associated spectral sequences. The main idea of Proposition 7.12 is to translate this comparison problem into a few local checks that can be run in parallel. We formalize this in the following corollary.

Corollary 7.14 (stability of covers) *Consider two pairs $(X, \mathcal{U})$ and $(X, \mathcal{V})$, where X is a space and $\mathcal{U}$ and $\mathcal{V}$ are covers. Let $\mathcal{W} = \mathcal{U} \cap \mathcal{V}$ and denote by $\mathcal{W}_{\mathcal{U}}^r$ and $\mathcal{W}_{\mathcal{V}}^r$ the respective $(r, \mathcal{W}, \mathcal{U})$ and $(r, \mathcal{W}, \mathcal{V})$*

interpolations. For every $0 \le r \le N$, we assume that there exist $\varepsilon_r, \varepsilon'_r \ge 0$ and $\nu_r, \nu'_r \ge 0$ such that for all $\tau \in N^r_{\mathcal{U}}$ and $\sigma \in N^r_{\mathcal{V}}$

$$E^2_{0,q}(\mathcal{U}_\tau, \mathcal{W}^{r+1}_{\mathcal{U}}) \sim^{\nu_r}_R E^\infty_{0,q}(\mathcal{U}_\tau, \mathcal{W}^{r+1}_{\mathcal{U}}) \sim^{\nu_r}_L \mathrm{PH}_q(\mathcal{U}_\tau),$$

$$E^2_{0,q}(\mathcal{V}_\sigma, \mathcal{W}^{r+1}_{\mathcal{V}}) \sim^{\nu'_r}_R E^\infty_{0,q}(\mathcal{V}_\sigma, \mathcal{W}^{r+1}_{\mathcal{V}}) \sim^{\nu'_r}_L \mathrm{PH}_q(\mathcal{V}_\sigma),$$

for all $r \ge 0$, and also

$$d_I(E^2_{p,q}(\mathcal{U}_\tau, \mathcal{W}^{r+1}_{\mathcal{U}}), 0) \le \varepsilon_r, \quad d_I(E^2_{p,q}(\mathcal{V}_\sigma, \mathcal{W}^{r+1}_{\mathcal{V}}), 0) \le \varepsilon'_r$$

for all $p > 0$, and $q \ge 0$. Then we have that

$$d^2_I\big(E^*_{p,q}(X, \mathcal{U}), E^*_{p,q}(X, \mathcal{V})\big) \le R(\mathcal{U}, \mathcal{V})$$

where $R(\mathcal{U}, \mathcal{V}) = \max\big(\sum_{r=0}^N 2\max(\varepsilon_r, \nu_r), \sum_{r=0}^N 2\max(\varepsilon'_r, \nu'_r)\big)$.

Proof By Lemma 7.1 there are double complex morphisms given by the refinement maps

$$\check{\mathfrak{C}}_p(\mathcal{U}, C^{\mathrm{cell}}_q) \xleftarrow{\rho^{\mathcal{U},\mathcal{W}}_{p,q}} \check{\mathfrak{C}}_p(\mathcal{W}, C^{\mathrm{cell}}_q) \xrightarrow{\rho^{\mathcal{V},\mathcal{W}}_{p,q}} \check{\mathfrak{C}}_p(\mathcal{V}, C^{\mathrm{cell}}_q).$$

In turn, these induce 2–morphisms of spectral sequences

$$E^2_{p,q}(X, \mathcal{U}) \xleftarrow{\rho^{\mathcal{U},\mathcal{W}}_{p,q}} E^2_{p,q}(X, \mathcal{W}) \xrightarrow{\rho^{\mathcal{V},\mathcal{W}}_{p,q}} E^2_{p,q}(X, \mathcal{V}).$$

Let $\psi^{\mathcal{U},\mathcal{W}}_{p,q}$ and $\psi^{\mathcal{V},\mathcal{W}}_{p,q}$ be the "inverses" of $\rho^{\mathcal{U},\mathcal{W}}_{p,q}$ and $\rho^{\mathcal{V},\mathcal{W}}_{p,q}$, respectively, witnessing the interleavings of the two spectral sequences (see Theorem 7.10 and Proposition 7.12). The result follows from considering the commutative diagram

where all arrows are 2–morphisms of spectral sequences. $\qquad\square$

8 Outlook

We expect spectral sequences associated to the geometric realizations of diagrams of CW–complexes to have a natural use in the distributed computation of persistent homology. The first future research direction is to develop further examples and use cases that benefit from the theory developed in this article.

The ε–acyclic carriers and equivalences which we introduced here in the context of persistent homology are of course based on acyclic carriers, which are similar to the ones used for example in [Björner 2003, Theorem 6] to prove a generalisation of the nerve theorem. A possible future research direction might

be to ask for conditions on the acyclic carriers with the goal of obtaining similar results as those from [Björner 2003] within the category of regularly filtered diagrams.

The bounds obtained in Section 7 for the interleavings between the second pages of two spectral sequences can certainly be improved; one possible direction is to explore similar examples as those in [Govc and Skraba 2018, Section 9] where the authors found sharp bounds.

In general, we think that spectral sequences deserve a more prominent role in applied algebraic topology and hope that the tools we developed here will motivate further study.

Appendix   Example of acyclic equivalence in RCW-cpx

Consider a filtered regular CW–complex X which is constant along $\mathbf{R}$, except at values 1, 2, 3 and 4, where it changes; see Figure 9. In order to describe X, we use the notation $(CD)_1$ for the edge between C and D, $(FGIJ)_2$ for a two cell whose vertices are F, G, I and J, and so on. By regularity of X, and since we do not define multiple edges between the same pair of vertices, X is determined by

$X_1 = \{A, B, C, D, E, F, H\} \cup \{(AH)_1, (BC)_1, (CD)_1, (EF)_1\}$,

$X_2 = X_1 \cup \{G\} \cup \{(AB)_1, (DE)_1, (FG)_1, (GH)_1\}$,

$X_3 = X_2 \cup \{I, J\} \cup \{(BI)_1, (CJ)_1, (FJ)_1, (GI)_1, (IJ)_1\} \cup \{(FGIJ)_2\}$,

$X_4 = X_3 \cup \{K\} \cup \{(AK)_1, (CK)_1, (EK)_1, (GK)_1\} \cup \{(ABCK)_2, (CDEK)_2, (EFGK)_2, (AKGH)_2\}$,

where $X_0 = \varnothing$; this is shown in Figure 9, which illustrates X. Of course, as X is a filtered complex, the structure maps of X are given by inclusions $X_s \hookrightarrow X_t$ for all $s < t$ from $\mathbf{R}$. Next, we describe the regularly filtered CW–complex Y, which is constant along $\mathbf{R}$, except at values 1, 2, 3 and 4, where it changes; this is also depicted in Figure 9. We define Y_* by

$$Y_1 = \{\alpha, \beta, \gamma\},$$
$$Y_2 = Y_1 \cup \{(\alpha\beta)_1, (\alpha\gamma)_1, (\beta\gamma)_1\},$$
$$Y_3 = (Y_2 \setminus \{(\alpha\gamma)_1\}) \cup \{\delta, \tau\} \cup \{(\gamma\tau)_1, (\tau\delta)_1, (\alpha\delta)_1, (\beta\delta)_1, (\beta\tau)_1\},$$
$$Y_4 = Y_3 \setminus \{\alpha, (\alpha\beta)_1, (\alpha\delta)_1\},$$

and $Y_0 = \varnothing$.

The structure maps of Y are defined as follows, where we use the overline notation $\bar{*}$ to denote the closure of some cell:

- $Y(1 \leq 2)$ is an inclusion,
- $Y(2 \leq 3)$ restricts to an inclusion in the subcomplex $\overline{(\alpha\beta)}_1 \cup \overline{(\beta\gamma)}_1$, while $\overline{(\alpha\gamma)}_1$ is sent to $\overline{(\alpha\delta)}_1 \cup \overline{(\delta\tau)}_1 \cup \overline{(\tau\gamma)}_1$.
- $Y(3 \leq 4)$ restricts to the identity in $Y_3 \setminus \{(\alpha\beta)_1, \alpha, (\alpha\delta)_1\}$ while it maps the vertex α to γ, the edge $(\alpha\beta)_1$ to $(\beta\gamma)_1$ and the edge $(\alpha\delta)_1$ to $\{(\gamma\tau)_1, \tau, (\tau\delta)_1\}$.
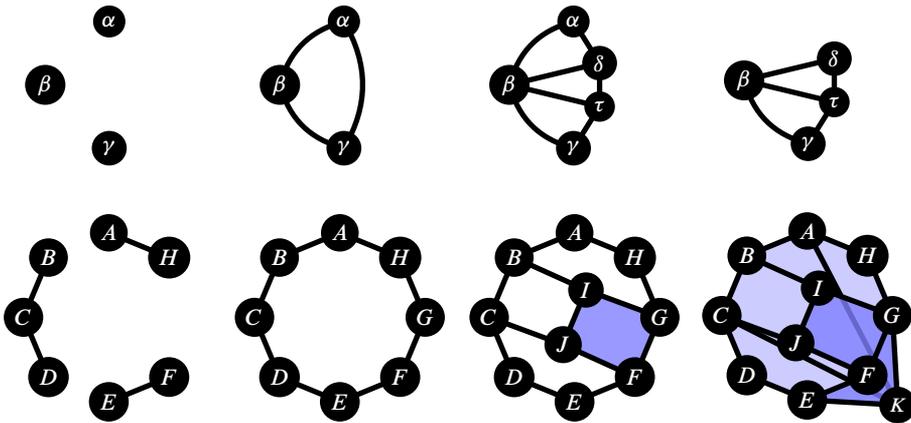
Figure 9: The spaces Y_i are shown at the top and X_i are at the bottom for values $i = 1, 2, 3, 4$. In filtration value 4, a cone with vertex in K is attached along the octahedron at the boundary of X_3; notice that we used 2–cells which are not 2–simplices.

One might check that Y is well defined according to Section 2.1. Next, we proceed to define an acyclic carrier $F : Y \rightrightarrows X$, which we depict in Figure 10, as follows:

- $F_1(\alpha) = \overline{(AH)}_1$, $F_1(\beta) = \overline{(BC)}_1 \cup \overline{(CD)}_1$, $F_1(\gamma) = \overline{(EF)}_1$,
- $F_2((\alpha\beta)_1) = F_1(\alpha) \cup F_1(\beta) \cup \{(AB)_1\}$, $F_2((\alpha\gamma)_1) = F_1(\alpha) \cup F_1(\gamma) \cup \{(HG)_1, G, (FG)_1\}$, $F_2((\beta\gamma)_1) = F_1(\beta) \cup F_1(\gamma) \cup \{(DE)_1\}$,
- $F_3(\delta) = G$, $F_3(\tau) = F$, $F_3((\alpha\delta)_1) = \overline{(AH)}_1 \cup \overline{(HG)}_1$, $F_3((\delta\tau)_1) = \overline{(IJFG)}_2$, $F_3((\gamma\tau)_1) = \overline{(EF)}_1$, $F_3((\beta\delta)_1) = \overline{(BC)}_1 \cup \overline{(CD)}_1 \cup \overline{(BI)}_1 \cup \overline{(IG)}_1$, $F_3((\beta\tau)_1) = \overline{(BC)}_1 \cup \overline{(CD)}_1 \cup \overline{(CJ)}_1 \cup \overline{(JF)}_1$,
- $F_4(\gamma) = F_4((\beta\gamma)_1) = F_4((\gamma\tau)_1) = \overline{\mathrm{St}(K)}$.

If we did not define a carrier, this is because we assume it is continued from an earlier definition. On the other hand, we define the carrier $G : X \rightrightarrows Y$ as follows:

- $G_1(A) = G_1(H) = G_1((AH)_1) = \alpha$, $G_1(E) = G_1(F) = G_1((EF)_1) = \gamma$, $G_1(B) = G_1(C) = G_1(D) = G_1((BC)_1) = G_1((CD)_1) = \beta$,
- $G_2((AB)_1) = \overline{(\alpha\beta)}_1$, $G_2((DE)_1) = \overline{(\beta\gamma)}_1$, $G_2((HG)_1) = G_2(G) = G_2((GF)_1) = \overline{(\alpha\gamma)}_1$,
- Define $A_3 = \{I, J, G, (IJ)_1, (GI)_1, (FJ)_1, (HG)_1, (GF)_1, (FGIJ)_2\}$; then for all $\sigma \in A_3$, we have $G_3(\sigma) = \overline{(\alpha\delta)}_1 \cup \overline{(\delta\tau)}_1 \cup \overline{(\tau\gamma)}_1$, $G_3((BI)_1) = \overline{(\beta\delta)}_1$, $G_3((CJ)_1) = \overline{(\beta\tau)}_1$,
- for all $\sigma \in X_4 \setminus \{(BI)_1, (CJ)_1\}$, $G_4(\sigma) = \overline{(\beta\gamma)}_1 \cup \overline{(\gamma\tau)}_1 \cup \overline{(\tau\delta)}_1$.

We define the shift carriers on X and Y by composition, that is, $I_X^0 = G \circ F$ and $I_Y^0 = F \circ G$, which in this particular case lead to well-defined acyclic carriers as one can check; to illustrate this, we write a couple of compositions:

$$G_3 \circ F_3((\beta\tau)_1) = \overline{(\alpha\delta)}_1 \cup \overline{(\delta\tau)}_1 \cup \overline{(\tau\gamma)}_1 \cup \overline{(\beta\tau)}_1,$$
$$F_3 \circ G_3((IJ)_1) = \overline{(AH)}_1 \cup \overline{(HG)}_1 \cup \overline{(IJFG)}_2 \cup \overline{(EF)}_1.$$
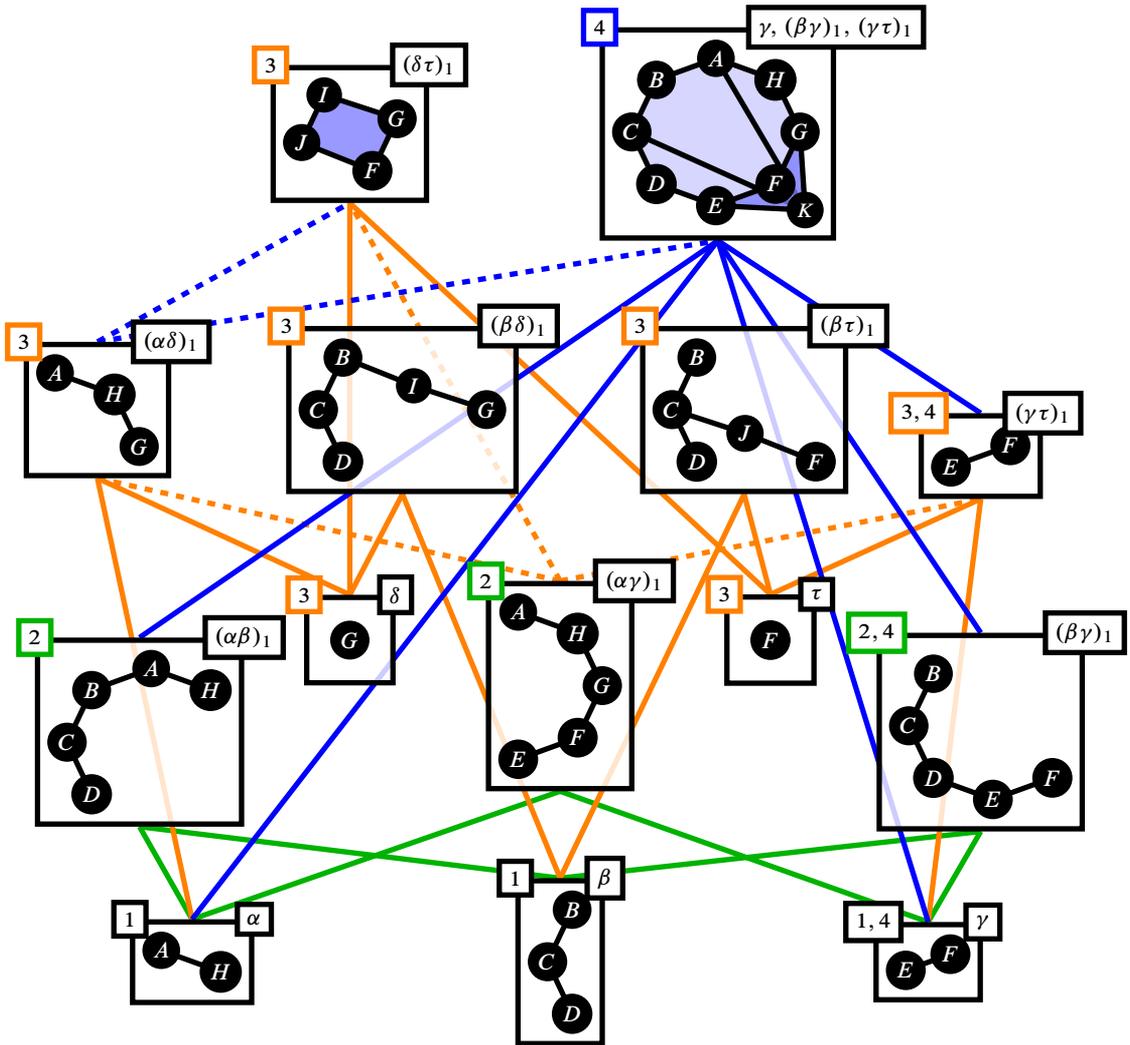
Figure 10: We depict the acyclic carriers from F. For each acyclic carrier we include its initial filtration value within a square on the top left while we write the cell(s) it corresponds to within a square on the top right; sometimes we write a pair of numbers a, b to indicate that the carrier applies for the filtration values in $[a, b)$ and that a new carrier is defined at b. Solid lines connecting the middle top of a box to the middle bottom of another box indicate that the containment relation must hold, where the carrier in the lower box needs to be embedded into the carrier on the upper box. We use dashed lines for containment relations involving a union of carriers, eg $F_3((\alpha\delta)_1) \subseteq F_4((\gamma\tau)_1) \cup F_4((\delta\tau)_1)$.

One can check that the conditions from Definition 4.3 are satisfied and so by Corollary 4.7 we obtain isomorphisms $\mathrm{PH}_*(X) \cong \mathrm{PH}_*(Y)$.

References

[Bauer 2011]   **U Bauer**, *Persistence in discrete Morse theory*, PhD thesis, Universität Göttingen (2011) Available at http://dx.doi.org/10.53846/goediss-2536

[Björner 2003]   **A Björner**, *Nerves, fibers and homotopy groups*, J. Combin. Theory Ser. A 102 (2003) 88–93 MR Zbl

[Cavanna 2019]   **N J Cavanna**, *Methods in homology inference*, doctoral dissertations, University of Connecticut (2019) Available at https://opencommons.uconn.edu/dissertations/2118

[Cavanna et al. 2017]   **N J Cavanna**, **K P Gardner**, **D R Sheehy**, *When and why the topological coverage criterion works*, from "Proceedings of the Twenty-Eighth Annual ACM–SIAM Symposium on Discrete Algorithms" (P N Klein, editor), SIAM, Philadelphia, PA (2017) 2679–2690 MR Zbl

[Chazal and Oudot 2008]   **F Chazal**, **S Y Oudot**, *Towards persistence-based reconstruction in Euclidean spaces*, from "Computational geometry", ACM, New York (2008) 232–241 MR Zbl

[Colin de Verdière et al. 2014]   **É Colin de Verdière**, **G Ginot**, **X Goaoc**, *Helly numbers of acyclic families*, Adv. Math. 253 (2014) 163–193 MR Zbl

[Cooke and Finney 1967]   **G E Cooke**, **R L Finney**, *Homology of cell complexes*, Princeton Univ. Press (1967) MR Zbl

[Dugger 2008]   **D Dugger**, *A primer on homotopy colimits*, book project (2008) Available at https://pages.uoregon.edu/ddugger/hocolim.pdf

[Ebert and Randal-Williams 2019]   **J Ebert**, **O Randal-Williams**, *Semisimplicial spaces*, Algebr. Geom. Topol. 19 (2019) 2099–2150 MR Zbl

[Eilenberg and MacLane 1953]   **S Eilenberg**, **S MacLane**, *Acyclic models*, Amer. J. Math. 75 (1953) 189–199 MR Zbl

[Ellis 2019]   **G Ellis**, *An invitation to computational homotopy*, Oxford Univ. Press (2019) MR Zbl

[Govc and Skraba 2018]   **D Govc**, **P Skraba**, *An approximate nerve theorem*, Found. Comput. Math. 18 (2018) 1245–1297 MR Zbl

[Hatcher 2002]   **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002)

[Kaczynski et al. 2004]   **T Kaczynski**, **K Mischaikow**, **M Mrozek**, *Computational homology*, Appl. Math. Sci. 157, Springer (2004) MR Zbl

[Kozlov 2008]   **D Kozlov**, *Combinatorial algebraic topology*, Algor. Comput. Math. 21, Springer (2008) MR Zbl

[Lewis and Morozov 2015]   **R Lewis**, **D Morozov**, *Parallel computation of persistent homology using the blowup complex*, from "Proceedings of the 27th ACM Symposium on Parallelism in Algorithms and Architectures", ACM, New York (2015) 323–331 Zbl

[Massey 1991]   **W S Massey**, *A basic course in algebraic topology*, Grad. Texts in Math. 127, Springer (1991) MR Zbl

[May 1999]   **J P May**, *A concise course in algebraic topology*, Univ. Chicago Press (1999) MR Zbl

[McCleary 2001]   **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Stud. Adv. Math. 58, Cambridge Univ. Press (2001) MR Zbl

[Munkres 1984]   **J R Munkres**, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA (1984) MR Zbl

[Nanda 2012]   **V Nanda**, *Discrete Morse theory for filtrations*, PhD thesis, Rutgers, The State University of New Jersey (2012) Available at https://www.proquest.com/docview/1312508982

[Oudot 2015]   **S Y Oudot**, *Persistence theory: from quiver representations to data analysis*, Math. Surv. Monogr. 209, Amer. Math. Soc., Providence, RI (2015) MR Zbl

[Robinson 2020]   **M Robinson**, *Assignments to sheaves of pseudometric spaces*, Compositionality 2 (2020) 25 MR Zbl

[Serre 1955]   **J-P Serre**, *Faisceaux algébriques cohérents*, Ann. of Math. 61 (1955) 197–278 MR Zbl

[Sköldberg 2006]   **E Sköldberg**, *Morse theory from an algebraic viewpoint*, Trans. Amer. Math. Soc. 358 (2006) 115–129 MR Zbl

[Torras-Casas 2023]   **Á Torras-Casas**, *Distributing persistent homology via spectral sequences*, Discrete Comput. Geom. 70 (2023) 580–619 MR Zbl

[Yoon and Ghrist 2020]   **I H R Yoon**, **R Ghrist**, *Persistence by parts*: *multiscale feature detection via distributed persistent homology*, preprint (2020) arXiv 2001.01623

[Zomorodian and Carlsson 2008]   **A Zomorodian**, **G Carlsson**, *Localized homology*, Comput. Geom. 41 (2008) 126–148 MR Zbl

School of Mathematics, Cardiff University
Cardiff, United Kingdom

School of Mathematics, Cardiff University
Cardiff, United Kingdom

TorrasCasasA@cardiff.ac.uk,   pennigu@cardiff.ac.uk

Slope norm and an algorithm to compute the crosscap number

WILLIAM JACO
JOACHIM HYAM RUBINSTEIN
JONATHAN SPREER
STEPHAN TILLMANN

We give three algorithms to determine the crosscap number of a knot in the 3–sphere using 0–efficient triangulations and normal surface theory. Our algorithms are shown to be correct for a larger class of complements of knots in closed 3–manifolds. The crosscap number is closely related to the minimum over all spanning slopes of a more general invariant, the slope norm. For any irreducible 3–manifold M with incompressible boundary a torus, we give an algorithm that, for every slope on the boundary that represents the trivial class in $H_1(M; \mathbb{Z}_2)$, determines the maximal Euler characteristic of any properly embedded surface having a boundary curve of this slope. We complement our theoretical work with an implementation of our algorithms, and compute the crosscap number of knots for which previous methods would have been inconclusive. In particular, we determine 196 previously unknown crosscap numbers in the census of all knots with up to 12 crossings.

1 Introduction

Let K be a knot in the 3–sphere with knot exterior M. The *crosscap number* $c(K)$ of K denotes the smallest genus of a nonorientable surface $S \subset M$ such that $\partial S = K$. It is a classical knot invariant that is defined for all knots in the 3–sphere.

Our main contributions are algorithms to compute the crosscap number of a knot in the 3–sphere. Efforts to compute the crosscap number of a knot have been at the centre of various other research projects using a variety of techniques. Among these is a formula for the crosscap number of torus knots by Teragaito [2004], an algorithm for alternating knots developed by Adams and Kindred [2013], and upper and lower bounds for the general case via the Jones polynomial by Kalfagianni and Lee [2016]. Recent work by Ito and Takimura [2018; 2020a; 2020b] establishes various further bounds. The KnotInfo database [Livingston and Moore 2021], and in particular their page on crosscap numbers, gives a detailed overview and results for specific knots.

Our work completes an approach put forward by Burton and Ozlen [2012]. Our starting point is Jaco and Sedgwick's generalisation [2003] of a celebrated result by Hatcher [1982]: they showed that in any orientable irreducible 3–manifold with incompressible boundary a torus, there are only finitely many

boundary slopes of *geometrically* incompressible and ∂–incompressible surfaces. This paper rests on a technical observation in [Jaco and Sedgwick 2003] that concerns fundamental surfaces (stated here as Proposition 12). We refer the reader to [Burton and Ozlen 2012; Jaco and Rubinstein 2003; Jaco and Sedgwick 2003; Matveev 2007] for the definitions and basic properties of normal surface theory in (singular) triangulations, [Burton and Ozlen 2012; Jaco and Rubinstein 2003] for basic facts concerning 0–efficient triangulations used herein, and [Burton and Tillmann 2018; Tillmann 2008; Tollefson 1998] for basic properties of working with quadrilateral coordinates only.

In Sections 2–4, we first develop our algorithms using standard coordinates for normal surfaces under varying hypotheses on the triangulations. We then extend the theory to work in quadrilateral coordinates in Section 5, and report on our computational results within this framework in Section 6. Throughout this paper, a *fundamental surface* is a normal surface whose normal coordinates are fundamental in standard (triangle–quadrilateral) normal surface space, and a Q–*fundamental surface* is a connected normal surface whose normal Q–coordinates are fundamental in quadrilateral space.

Slope norm   The dual tree to the Farey tessellation of the hyperbolic plane is used to organise the set of all boundary slopes of properly embedded surfaces with a single boundary curve. This allows us to give an algorithm that, for an irreducible 3–manifold with boundary a torus and a slope on the boundary that represents the trivial class in $H_1(M; \mathbb{Z}_2)$, determines the maximal Euler characteristic of any properly embedded connected surface having connected boundary of this slope (Theorem 15). We call the negative of this number the *norm of the slope*, and the minimum over all these norms the *slope norm* of M. This norm is used in forthcoming work to apply the complexity bounds given in [Jaco et al. 2020b; 2020a; 2009] to infinite families of Dehn fillings. The existence of such an algorithm goes back to Schubert [1961]; see also Matveev [2007, Theorems 4.1.10 and 4.1.11]. Our new contribution is that we do not need to adapt a triangulation to the slope. Similar results to our slope norm algorithm were obtained independently by Howie [2021], and used for different applications. Our work here on slope norm feeds into the proof of the main theorem. We later show that there is an algorithm to determine the slope norm of M and the set of all minimising slopes in quadrilateral space (Corollary 22).

Crosscap number   We next give two algorithms to determine the crosscap number of a knot in the 3–sphere (Theorems 1 and 3). Both use standard coordinates for normal surfaces, but they make different assumptions on the underlying triangulation. Our algorithms are shown to be correct for a larger class of complements of knots in closed 3–manifolds N that represent the trivial class in $H_1(N; \mathbb{Z}_2)$, including all that have a complete hyperbolic structure of finite volume.

Burton and Ozlen [2012] introduce triangulations that contain no normal 2–spheres and have an edge in the boundary that represents the meridian. These triangulations are called *efficient suitable*, and they guarantee the existence of fundamental spanning surfaces of maximal Euler characteristic. Efficient suitable triangulations can be constructed from any input triangulations, and we outline the algorithm in Section 3.1. Burton and Ozlen describe a procedure (Algorithm 3), where on input a knot in the 3–sphere the output

is either one integer (the crosscap number) or a pair of consecutive integers (one of which is the crosscap number). Theorem 1 shows that in the latter case, the crosscap number is the larger integer. The apparent difficulty of determining the crosscap number algorithmically lies in the case where every maximal Euler characteristic fundamental spanning surface is orientable. This is solved in Lemma 19. The following result thus improves and generalises [loc. cit., Algorithm 3], and the result is given in Algorithm 18.

Theorem 1 *Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathcal{T}$ be an efficient suitable triangulation of M. Then $c(K) = \min(A, B)$, where*

$$A = \min\{1 - \chi(S) \mid S \text{ is a nonorientable fundamental spanning surface for } K\},$$
$$B = \min\{2 - \chi(S) \mid S \text{ is an orientable fundamental spanning surface for } K\},$$

and we let $\min \varnothing = \infty$.

Theorem 1 has the following consequence when a knot has no orientable spanning surface:

Corollary 2 *Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$ and $[K] \neq 0 \in H_1(N; \mathbb{Z})$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathcal{T}$ be an efficient suitable triangulation of M. Then*

$$c(K) = \min\{1 - \chi(S) \mid S \text{ is a fundamental spanning surface for } K\}.$$

For arbitrary 0–efficient triangulations (which do not contain properly embedded nonvertex linking normal spheres or discs), we give a more general algorithm (Theorem 3) that uses the slope norm algorithm. The basic idea is that Burton and Ozlen's suitable triangulations ensure that minimal spanning surfaces can be found amongst the normal surfaces even though in general they may be ∂–compressible. In an arbitrary triangulation, there may be no nonorientable normal spanning surfaces of maximal Euler characteristic, but our slope norm algorithm keeps track of optimal boundary compression sequences. To do this we define what we call the *even integral subtree distance* as the length $d(\partial S, \mathcal{F}_e)$ of the shortest path from a given boundary slope ∂S to the subtree $\mathcal{F}_e$ corresponding to the slopes of spanning surfaces; see Section 3.3 for details.

Theorem 3 *Let M be the exterior of a knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathcal{T}$ be a 0–efficient triangulation of M and suppose that the coordinates for a meridian for K on the induced triangulation $\mathcal{T}_\partial$ of ∂M are given. Then $c(K) = \min(A, B, Z)$, where*

$A = \min\{1 - \chi(S) \mid S \text{ is a nonorientable fundamental spanning surface for } K\},$

$B = \min\{2 - \chi(S) \mid S \text{ is an orientable fundamental spanning surface for } K\},$

$Z = \min\{1 - \chi(S) + d(\partial S, \mathcal{F}_e)$

$\qquad\qquad\qquad \mid S \text{ is a fundamental nonspanning surface with connected essential boundary}\},$

and we let $\min \varnothing = \infty$.

Since $d(\partial S, \mathcal{F}_e) = 0$ for a spanning surface S, the above theorem could have been stated using two terms rather than three, but we wanted to keep the notation in line with Theorem 1. Again, when there is no orientable spanning surface, this specialises to:

Corollary 4 *Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$ and $[K] \neq 0 \in H_1(N; \mathbb{Z})$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathcal{T}$ be a 0–efficient triangulation of M and suppose that the coordinates for a meridian for K on the induced triangulation $\mathcal{T}_{\partial}$ of ∂M are given. Then*

$$c(K) = \min\{1 - \chi(S) + d(\partial S, \mathcal{F}_e) \mid S \text{ is a fundamental surface with connected essential boundary}\}.$$

Knot genus As an interlude, we show in Section 4 that in our framework one can give a short proof of Schubert's classical result that the genus of a knot is realised by one of the fundamental surfaces. Schubert [1961] originally proved this in the context of normal surfaces with respect to handle decompositions.

Theorem 5 *Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z})$. Suppose that M is irreducible and let $\mathcal{T}$ be a 0–efficient triangulation of M. Then an orientable spanning surface of maximal Euler characteristic is amongst the fundamental surfaces.*

Quadrilateral space All results up to this point were stated in the context of normal surface theory with standard coordinates. For computations, it is of advantage to be able to work with quadrilateral coordinates only, as this makes otherwise impossible calculations feasible. In Section 5, we give some extensions of the previous results in this context; see for instance [Burton and Tillmann 2018] for similar results for closed normal surfaces. For definitions and basic properties of working with quadrilateral coordinates only, we refer to [Burton and Tillmann 2018; Tillmann 2008; Tollefson 1998]. The following is the main result of this paper; whilst the previous results were given for either efficient suitable triangulations or for 0–efficient triangulations, we now need to combine these properties:

Theorem 6 *Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$. Suppose M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathcal{T}$ be a 0–efficient suitable triangulation of M. Then $c(K) = \min(A', B')$, where*

$$A' = \min\{1 - \chi(S) \mid S \text{ is a nonorientable } Q\text{–fundamental spanning surface for } K\},$$
$$B' = \min\{2 - \chi(S) \mid S \text{ is an orientable } Q\text{–fundamental spanning surface for } K\}.$$

Computations We use our methods to determine the crosscap numbers of 196 knots with up to 12 crossings for which the crosscap number was previously not known. As a result, crosscap numbers of all knots up to and including ten crossings are now known. Our algorithms give a *theoretical* method to determine the crosscap numbers of knots with a unified method, where previously different techniques were needed. From a *practical* viewpoint, using our algorithm in standard coordinates allows us to handle triangulations of up to about 26 tetrahedra. Making use of our results in quadrilateral space allows us to push this limit to about 30 tetrahedra. See Section 6 for our computational results.

2   The norm of an even slope

Throughout this section let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Suppose $\mathcal{T}$ is a (singular or semisimplicial) triangulation of M with the property that the induced triangulation $\mathcal{T}_\partial$ of ∂M has exactly one vertex. For instance, a 0–efficient triangulation has this property [Jaco and Rubinstein 2003]. In the following, we choose the single vertex of $\mathcal{T}$ as the basepoint for the fundamental group of M and omit it in the notation.

We first show that every connected essential curve $c \subset \partial M$ with $[c] = 0 \in H_1(M, \mathbb{Z}_2)$ bounds a properly embedded connected surface $S \subset M$ with $\partial S = c$ (Corollary 10). We then use a result by Jaco and Sedgwick (Proposition 12) to conclude that if $\mathcal{T}$ is 0–efficient and S is of maximal Euler characteristic amongst all connected surfaces with this slope and ∂–incompressible, then a surface of equal boundary slope and Euler characteristic must be represented by a fundamental surface in $\mathcal{T}$. This then leads to Algorithm 16, which computes the smallest norm of a given boundary slope on ∂M with respect to a given framing.

Lemma 7   *Let $\rho: \pi_1(M) \to \mathbb{Z}_2$ be a homomorphism with the property that $\rho(\mathfrak{m}) = 1$ for some primitive peripheral element $\mathfrak{m} \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$. Then for every primitive peripheral element $\gamma \in \ker(\rho)$, there is a properly embedded surface S in M with ∂S a single boundary curve that satisfies $[\partial S] = \gamma^{\pm 1}$ as free homotopy classes of unoriented loops. In particular, $[\gamma] = 0 \in H_1(M, \mathbb{Z}_2)$.*

The proof of the lemma introduces the way we will use the Farey graph in our later algorithms.

Proof   Note that there is $\mathfrak{l} \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$ with $\rho(\mathfrak{l}) = 0$ and $\langle \mathfrak{m}, \mathfrak{l} \rangle = \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$. The primitive peripheral elements in $\ker \rho$ are precisely $\mathfrak{m}^{2k} \mathfrak{l}^q$, where $k \in \mathbb{Z}$ and $\gcd(2k, q) = 1$. Since the boundary of M is incompressible and has abelian fundamental group, we have

(2-1) $$\mathrm{im}(\pi_1(\partial M) \to \pi_1(M)) \cong \pi_1(\partial M) \cong H_1(\partial M, \mathbb{Z}).$$

We therefore identify all groups and freely switch between additive and multiplicative notation for peripheral elements. Since we are only interested in unoriented isotopy classes of primitive elements, we always choose $\mathfrak{m}^p \mathfrak{l}^q$ with $q \geq 0$.

We first show that there is a surface for some peripheral element in the kernel. Our 0–efficient triangulation $\mathcal{T}$ of M has exactly one vertex. Hence every edge is a loop and represents an element of $\pi_1(M)$. This maps to either 0 or 1 under ρ. Place a normal corner on an edge if and only if the corresponding element
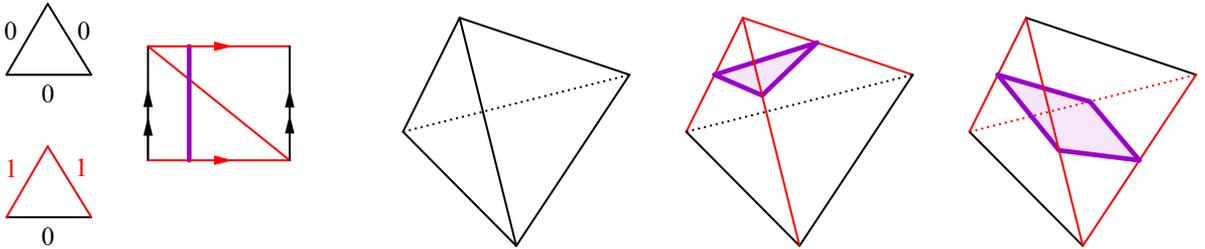
Figure 1: Labelling of edges and canonical normal curves and surfaces.

maps to 1. As observed in [Jaco et al. 2009], this results in a normal surface in M having at most a single triangle or a single quadrilateral in each tetrahedron, as shown in Figure 1. Since $\rho(\mathfrak{m}) = 1$, this normal surface has a single boundary curve $[\partial S] = \mathfrak{m}^{2k'}\mathfrak{l}^{q'}$ for some $q' \geq 0$ and $\gcd(2k', q') = 1$. This single boundary curve meets each boundary triangle in a single normal arc.

We now show how all other boundary curves $\mathfrak{m}^{2k}\mathfrak{l}^q$, where $q > 0$, $k \in \mathbb{Z}$ and $\gcd(2k, q) = 1$, can be obtained by adding saddles to S. To this end, we use the *layering procedure* (see [loc. cit.] and Figure 3) in conjunction with the Farey tessellation as an organising principle for the set of isotopy classes of triangulations of the torus with a marked point. This is different from the L–graph used in [loc. cit.].

We treat the single vertex in the induced triangulation $\mathcal{T}_\partial$ of ∂M as a marked point, and give the torus ∂M a Euclidean structure with the property that the marked point lifts to the integer lattice via a universal covering map $\mathbb{R}^2 \to \partial M$. Moreover, up to the action of the Deck group, we may assume that $\mathfrak{m}$ and $\mathfrak{l}$ lift to horizontal and vertical lines, respectively. The map

$$\mathfrak{m}^p\mathfrak{l}^q \mapsto \frac{p}{q}$$

gives a bijection between the set of isotopy classes of primitive curves and $\mathbb{Q} \cup \{\infty\}$. Now the isotopy classes of triangulations with a single vertex at the marked point correspond to the orbit of the triple $\left(\frac{1}{0}, \frac{0}{1}, \frac{-1}{1}\right)$ under the action of $\mathrm{SL}(2, \mathbb{Z})$ by Möbius transformations. Identifying $\mathbb{R} \cup \{\infty\}$ with the boundary of the Poincaré disc model of the hyperbolic plane and each triple with an ideal triangle gives the well-known Farey tessellation.

Each triple $(p_0/q_0, p_1/q_1, p_2/q_2)$ contains precisely one fraction with even numerator, say p_0/q_0. Note that the corresponding primitive element is in the kernel of ρ, whilst the other two are mapped to 1. The normal curve resulting from our above procedure of assigning 0 or 1 to each edge, when applied to the corresponding marked triangulation of the boundary, results in a normal curve of slope $\mathfrak{m}^{p_0}\mathfrak{l}^{q_0}$. We call p_0/q_0 the *even slope* of the triple.

The dual 1–skeleton to the Farey tessellation is an infinite trivalent tree. Any two triangulations that correspond to triangles sharing an edge in the tessellation are related by an *edge flip*. If one of the triangulations corresponds to $\mathcal{T}_\partial$, then, as an operation on the triangulation $\mathcal{T}$ of M, one can *layer* a tetrahedron σ on $\mathcal{T}$ along the edge that is being flipped; see Figure 3. This results in a new triangulation
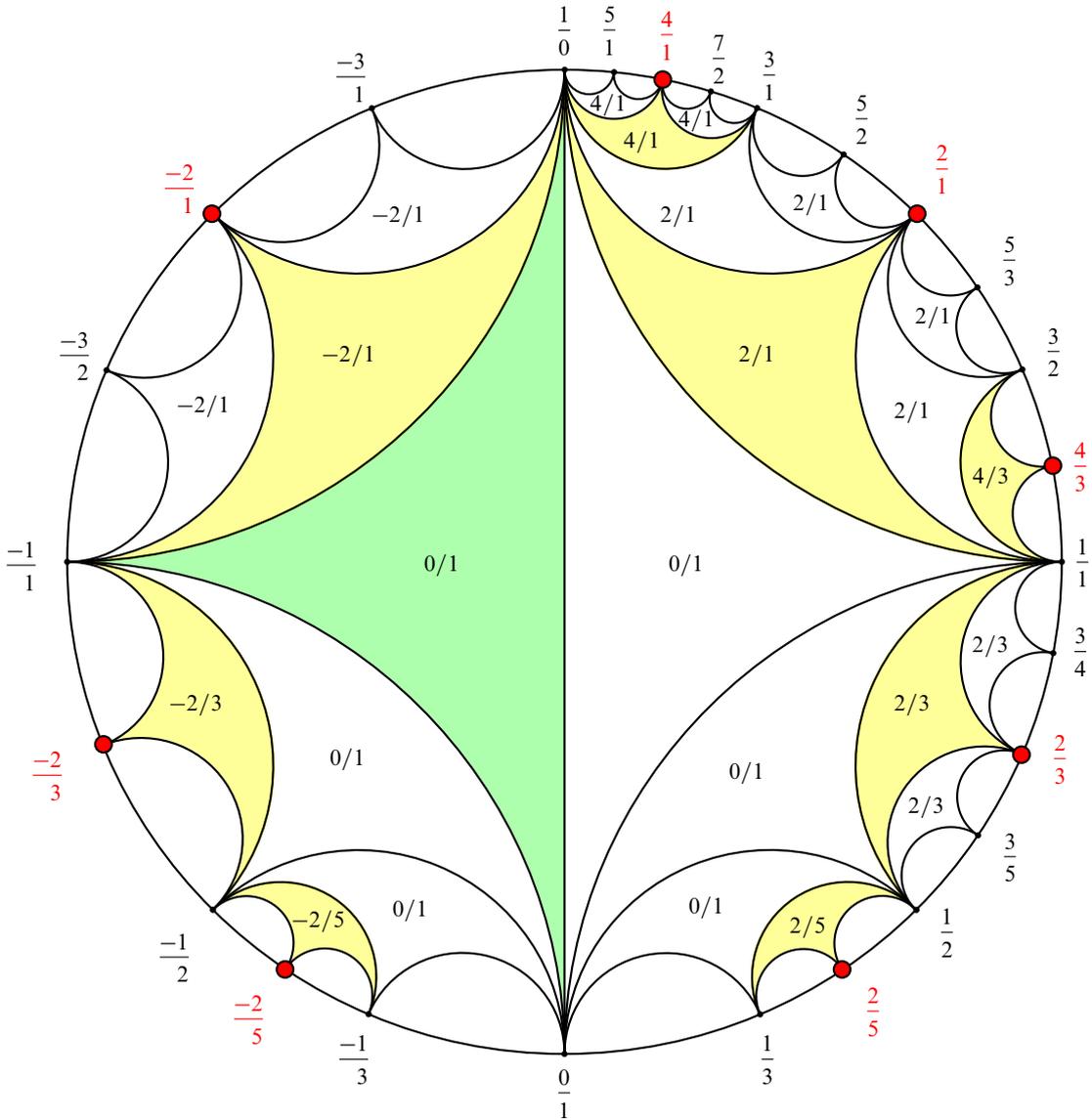
Figure 2: The Farey tessellation. Each ideal triangle corresponds to an isotopy class of 1–vertex triangulations of the torus. The ideal vertices are labelled with the slopes of the edges, and each ideal triangle is labelled with its unique even slope. The base triangle is marked in green, and the canonical triangles for the even slopes in yellow. Adjacent triangles differ by an edge flip.

$\mathscr{T}' = \mathscr{T} \cup \sigma$ of M with the property that the isotopy class of the triangulation of the boundary has changed. Since the trivalent tree is connected, one observes that all isotopy classes of triangulations of ∂M can be realised as the induced triangulations of the boundary of M. In particular, every even slope is an edge in some triangulation of the boundary of M. It remains to relate this information to triangulations of M and boundary curves of properly embedded surfaces.
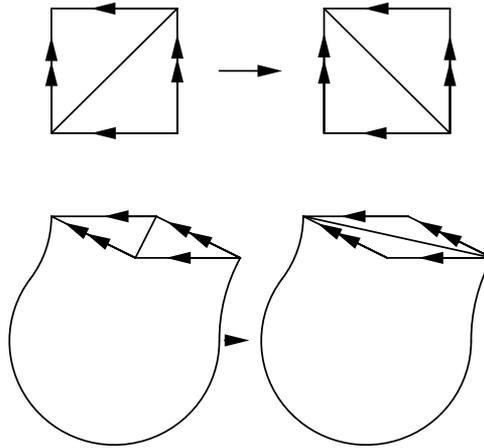
Figure 3: Layering on a boundary edge adds a tetrahedron to the triangulation of M.

Each triangulation of ∂M allows three distinct flips, corresponding to the three edges of the ideal triangle in the Farey tessellation. Layering a tetrahedron on the triangulation has the effect of changing the normal surface S to a normal surface S' by adding a quadrilateral. If the corresponding ideal edge in the Farey tessellation has endpoint at the even slope, then the layering adds a pinched annulus to S and maintains its boundary slope. If the ideal edge does not have an endpoint at the even slope, then a punctured and pinched Möbius strip (which we call a *saddle*) is added to S. Hence its Euler characteristic is lowered by 1, and its slope changes to the even slope of the adjacent triangle in the Farey tessellation. Topologically, the relationship between the surfaces is that for two of the three layerings, S is obtained from S' by deleting a pinched annulus, whilst for the last, S is obtained from S' by performing a boundary compression. The three possibilities are shown in Figure 4.

This completes the proof of the lemma, since starting with S, for every even slope, this constructs a properly embedded surface with boundary that slope. This construction and our observations about the Farey tessellation are key to the algorithm given in Theorem 15.   □

The inclusion map induces a homomorphism $H_1(\partial M, \mathbb{Z}_2) \to H_1(M, \mathbb{Z}_2)$, which we precompose with the natural map $H_1(\partial M, \mathbb{Z}) \to H_1(\partial M, \mathbb{Z}_2)$ to obtain $\varphi\colon H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2)$. The next lemma shows that the homomorphism in the hypothesis of Lemma 7 exists and is unique:

Lemma 8 *We have $H = \operatorname{im}(\varphi\colon H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2)) \cong \mathbb{Z}_2$.*

It follows from the lemma that we may choose a basis $(\mathfrak{m}_2, \mathfrak{l}_2)$ of $H_1(\partial M, \mathbb{Z})$ with the property that $\varphi(\mathfrak{m}_2) \neq 0 = \varphi(\mathfrak{l}_2)$. We say that the basis $(\mathfrak{m}_2, \mathfrak{l}_2)$ is a 2–*torsion framing* of ∂M and call $\mathfrak{m}_2$ a 2–*meridian* and $\mathfrak{l}_2$ a 2–*longitude*.

First proof **(geometric topology)** First assume that $H = \operatorname{im}(\varphi\colon H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2)) = \{0\}$. Let $(\mathfrak{m}, \mathfrak{l})$ be any basis of $H_1(\partial M, \mathbb{Z})$. Choose a sufficiently fine simplicial triangulation of M so that
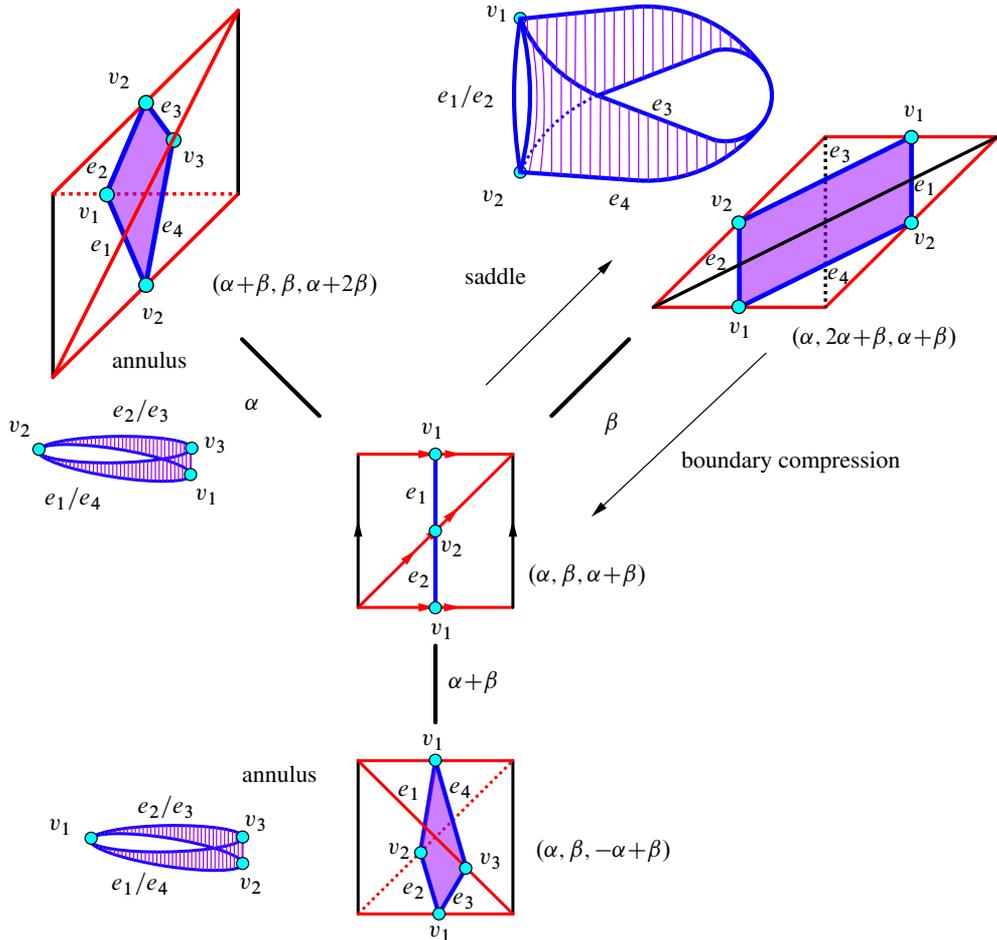
Figure 4: The three possible layerings on the boundary torus. Two add a pinched annulus to the surface and do not change the boundary slope. The last adds a saddle and changes the boundary slope according to the labelling of the associated ideal triangles in the Farey graph.

we may choose a simple closed curve L in the 1–skeleton in ∂M that is isotopic to $\mathfrak{l}$. Since $\varphi(\mathfrak{l}) = 0$, there is a 2–chain C in the 2–skeleton with $\partial C = L$. Since we are working with $\mathbb{Z}_2$ coefficients, C is an assignment of 0 or 1 to each 2–simplex in the triangulation. We add a product collar to ∂M and add the annulus $L \times [0, 1]$ to C. We let M and C denote the resulting manifold and chain again.

Each 1–simplex in the interior of M meets C in an even number of 2–simplices. Hence, away from the vertices, we can resolve the 2–simplices in pairs to obtain a properly embedded but possibly singular surface S' in M with the property that its singularities are contained in the set of interior vertices of the triangulation of M. Now a small regular neighbourhood N of the union of all vertices meets S' in a union of circles. Hence, we replace $S \cap N$ with a union of discs, giving a properly embedded surface S in M with $\partial S = L$.

Since $\mathfrak{m}$ meets $\mathfrak{l}$ in a single point, the intersection pairing implies that $\varphi(\mathfrak{m}_\infty) \neq 0 \in H_1(M, \mathbb{Z}_2)$. But this contradicts our hypothesis that $H = \{0\}$.

Hence $H \neq \{0\}$. It now follows from [Lemma 7](#) that the rank of H cannot be two, so it must be one. $\square$

Second proof (algebraic topology) First assume that $H = \mathrm{im}(\varphi : H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2)) = \{0\}$. Glue a solid torus to M, resulting in a closed 3–manifold N. Since N is closed, the Euler characteristic of N is zero and we have $b_1(N, \mathbb{Z}_2) = b_2(N, \mathbb{Z}_2)$. Consider the following part of the Mayer–Vietoris long exact sequence in homology with $\mathbb{Z}_2$ coefficients:

$$\cdots \to H_1(\partial M, \mathbb{Z}_2) \to H_1(M, \mathbb{Z}_2) \oplus H_1(S^1 \times D^2, \mathbb{Z}_2) \to H_1(N, \mathbb{Z}_2) \to \cdots.$$

Since $H = \{0\}$, we have $H_1(N, \mathbb{Z}_2) \cong H_1(M, \mathbb{Z}_2)$. Now $H = \{0\}$ also implies that there is a relative $\mathbb{Z}_2$–chain in M that attaches to the meridian disc of the solid torus. The intersection pairing with the core curve of the solid torus implies that the rank of $H_2(N, \mathbb{Z}_2)$ is one larger than the rank of $H_2(M, \mathbb{Z}_2)$. In particular, $b_1(M, \mathbb{Z}_2) = b_2(M, \mathbb{Z}_2) + 1$.

Now consider the long exact sequence for the pair $(M, \partial M)$ with $\mathbb{Z}_2$–coefficients. We obtain

$$0 \to H_2(M, \mathbb{Z}_2) \to H_2(M, \partial M, \mathbb{Z}_2) \to H_1(\partial M, \mathbb{Z}_2) \to 0.$$

Using Poincaré–Lefschetz duality and the universal coefficient theorem, we have

$$H_1(M, \mathbb{Z}_2) \cong H^1(M, \mathbb{Z}_2) \cong H_2(M, \partial M, \mathbb{Z}_2) \cong H_2(M, \mathbb{Z}_2) \oplus H_1(\partial M, \mathbb{Z}_2)$$

This gives $b_1(M, \mathbb{Z}_2) = b_2(M, \mathbb{Z}_2) + 2$, contradicting the calculation in the first paragraph.

Hence $H \neq \{0\}$. It now follows from [Lemma 7](#) that the rank of H is one. $\square$

Remark 9 The standard half-lives half-dies argument [Hatcher 2023, Lemma 3.5] implies for homology with rational coefficients that one may choose a basis $\langle \mathfrak{m}_\infty, \mathfrak{l}_\infty \rangle = H_1(\partial M, \mathbb{Z})$ with the property that $\mathfrak{m}_\infty$ maps to an element of infinite order whilst $\mathfrak{l}_\infty$ maps to an element of finite order under the inclusion map to $H_1(M, \mathbb{Z})$. In particular, $\mathfrak{l}_\infty$ is uniquely determined up to sign, whilst $\mathfrak{m}_\infty$ is only well defined up to sign and a power of $\mathfrak{l}_\infty$. We call $\mathfrak{l}_\infty$ *the homological longitude* and $\mathfrak{m}_\infty$ *a homological meridian*.

It is not necessarily the case that one may choose $(\mathfrak{m}_2, \mathfrak{l}_2) = (\mathfrak{m}_\infty, \mathfrak{l}_\infty)$. To make this statement less mysterious, we give a third proof of the lemma that does not appeal to a contradiction:

Third proof (geometric topology) Suppose the order of $\mathfrak{l}_\infty$ is m in $H_1(M, \mathbb{Z})$. The significance of the order is that $\mathfrak{m}_\infty$ maps to an element of the form $a^m h \in H_1(M, \mathbb{Z})$, where a generates a free $\mathbb{Z}$ summand and h is a torsion element. A geometric interpretation of this algebraic relationship arises from a construction due to Stallings [1961] that produces a properly embedded connected oriented surface S in M with $[\partial S] = \mathfrak{l}_\infty^{\pm m}$ dual to the action of $\pi_1(M)$ on $\mathbb{R}$ associated with a homomorphism $\pi_1(M) \to H_1(M, \mathbb{Z}) \to \mathbb{Z}$ with $a \mapsto 1$. Moreover, S has exactly m boundary components, which implies that all have the same induced orientation, and the meridian has algebraic intersection number $\pm m$ with the surface.

Note that if one connects any two adjacent boundary components with a boundary parallel annulus, then one obtains a (nonorientable) surface. In particular, if m is odd, then one may connect (possibly zero) pairs of boundary components with annuli to obtain a properly embedded surface S in M with a single boundary curve $[\partial S] = \mathfrak{l}_\infty^{\pm 1}$. In particular, $\mathfrak{m}_\infty$ maps to a generator of the image and $\mathfrak{l}_\infty$ is contained in the kernel of $\varphi \colon H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2)$. Hence we may choose $\mathfrak{m}_2 = \mathfrak{m}_\infty$ and $\mathfrak{l}_2 = \mathfrak{l}_\infty$.

If m is even, the same construction of connecting boundary components in pairs results in a closed nonorientable surface in M. Since m is the order of $\mathfrak{l}_\infty$ in $H_1(M, \mathbb{Z})$, there are two cases, depending on whether $\varphi(\mathfrak{l}_\infty)$ maps to zero or not.

First assume that $m = 2$ and $\varphi(\mathfrak{l}_\infty)$ is the generator of a $\mathbb{Z}_2$–summand. Then there is a homomorphism $\rho \colon \pi_1(M) \to \mathbb{Z}_2$ with $\rho(\mathfrak{l}_\infty) = 1$. So then $\rho(\gamma) = 0$, where either $\gamma = \mathfrak{m}_\infty$ or $\gamma = \mathfrak{m}_\infty \mathfrak{l}_\infty$. Now $[\gamma] = 0 \in H_1(M, \mathbb{Z}_2)$ according to Lemma 7. It follows that we may choose $\mathfrak{m}_2 = \mathfrak{l}_\infty$ and $\mathfrak{l}_2 = \gamma$.

The remaining case is that $\varphi(\mathfrak{l}_\infty) = 0$. In this case, the construction from the first proof of the surface S with $[\mathfrak{l}_\infty] = [\partial S] = 0 \in H_1(M, \mathbb{Z}_2)$ can be applied, and we let $\mathfrak{l}_2 = \mathfrak{l}_\infty$. Since $\mathfrak{m}_\infty$ meets $\mathfrak{l}_\infty$ in a single point, $\varphi(\mathfrak{m}_\infty) \neq 0 \in H_1(M, \mathbb{Z}_2)$. So we let $\mathfrak{m}_2 = \mathfrak{m}_\infty$. $\qquad\square$

We use the following terminology and notation for *unoriented isotopy classes* of nontrivial simple closed loops on the boundary torus. Let $\alpha \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$. Recalling the identification (2-1), consider $[\alpha] \in H_1(\partial M, \mathbb{Z})$. If $[\alpha] = \mathfrak{m}_2^p \mathfrak{l}_2^q$ is a nontrivial primitive class in $H_1(\partial M, \mathbb{Z})$ with $q \geq 0$, then we call α a *slope*. We may therefore identify a slope with an unoriented isotopy class of a nontrivial simple closed loop on the torus. Conversely, each such unoriented isotopy class arises from a unique slope. A slope α is an *even slope* if α maps to zero in $H_1(M, \mathbb{Z}_2)$. We remark that this is consistent with the terminology concerning even slopes in the Farey construction in Lemma 7, and that the notion of an even slope is independent of the chosen 2–torsion framing.

Given a surface S with connected boundary, we give ∂S the unique orientation that makes $[\partial S] \in \pi_1(M)$ a slope. Now Lemmas 7 and 8 imply:

Corollary 10 *Let $\alpha \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$ be a slope. There is a properly embedded surface S in M with $[\partial S] = \alpha$ if and only if α is an even slope.*

The *norm* of an even slope α is defined as

$$\|\alpha\| = \min\{-\chi(S) \mid S \text{ is a properly embedded surface in } M \text{ with } [\partial S] = \alpha\}.$$

Since ∂M is incompressible, $\|\alpha\| \geq 0$ for all even slopes. We say that S is *taut* for α if S is connected, $[\partial S] = \alpha$ and $\|\alpha\| = -\chi(S)$.

Remark 11 We emphasise here that the slope norm is defined on isotopy classes of simple closed connected curves on the boundary of M, and that it gives the maximal Euler characteristic of a surface with connected boundary of this slope. A related approach, which we do not take, would be to also consider surfaces with multiple boundary components.

The *slope norm* of M is defined as

$$\|M\| = \min\{\|\alpha\| \mid \alpha \text{ is an even slope}\}.$$

Each even slope α satisfying $\|\alpha\| = \|M\|$ is called a *minimising slope* for M.

We will use key results of Jaco and Sedgwick [2003, Proposition 3.7 and Corollary 3.8]. We require a few definitions in order to state them. We refer readers unfamiliar with the concepts in normal surface theory that are mentioned here to either of [Jaco and Rubinstein 2003; Jaco and Sedgwick 2003].

The boundary curves of a normal surface S form a collection of normal curves on ∂M. It is shown in [Jaco and Sedgwick 2003] that two normal curves are normally isotopic with respect to $\mathscr{T}_\partial$ if and only if they are isotopic on ∂M, and that a normal curve is trivial if and only if it is vertex linking. It follows that we can identify a slope with the normal isotopy class of a nontrivial normal curve on the torus. For a collection C of pairwise disjoint normal curves on the torus containing at least one nontrivial component, the slope of C is the isotopy class of a nontrivial component. Two slopes are *complementary* if their Haken sum is a collection of trivial curves.

Jaco and Sedgwick [2003, Proposition 3.7] show that if two normal surfaces are compatible and meet ∂M in nontrivial slopes, then these slopes are either equal or complementary. This allows the possibility that some or all boundary curves of a normal surface are trivial. The projective solution space of normal surface theory is denoted by $\mathscr{P}(\mathscr{T})$. Given a normal surface S, its *carrier* is the unique minimal face $\mathscr{C}(S) \subset \mathscr{P}(\mathscr{T})$ that contains the projectivised normal coordinates of S.

Proposition 12 [Jaco and Sedgwick 2003, Corollary 3.8] *Let M be an orientable compact connected 3–manifold with ∂M a single torus. Suppose $\mathscr{T}$ is a triangulation of M that restricts to a one-vertex triangulation of ∂M. Suppose S is a normal surface and $\partial S \neq 0$. Assume also that ∂S contains at least one essential curve. There are at most two slopes (complementary ones) for all surfaces in the carrier $\mathscr{C}(S) \subset \mathscr{P}(\mathscr{T})$.*

Lemma 13 *Let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Let $\mathscr{T}$ be a 0–efficient triangulation of M and let $\alpha \in \mathrm{im}(\pi_1(\partial M) \to \pi_1(M))$ be an even slope. If there is an incompressible and ∂–incompressible surface S in M with $[\partial S] = \alpha$ and $\chi(S) = -\|\alpha\|$, then there is a fundamental surface F of $\mathscr{T}$ with $[\partial F] = \alpha$ and $\chi(F) = -\|\alpha\|$.*

Proof Let S be an incompressible and ∂–incompressible surface S with $[\partial S] = \alpha$ and $\chi(S) = -\|\alpha\|$. Since M is irreducible and has incompressible boundary, we can isotope S to be a normal surface. If S is not fundamental, then S is a sum of fundamental surfaces. Now S has boundary a single curve. It follows from Proposition 12 that there is only a single summand, F, with nonempty boundary and all other summands are closed normal surfaces. To see this, note that otherwise S would have disconnected boundary or a trivial curve in the boundary, whence $[\partial F] = [\partial S] = \alpha$. Since $\chi(S) = -\|\alpha\|$, we have $\chi(F) \leq \chi(S)$. Since Euler characteristic is additive under Haken sums and the triangulation is 0–efficient, this forces $\chi(F) = \chi(S)$. $\qquad\square$

Corollary 14 *Let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Let $\mathcal{T}$ be a 0–efficient triangulation of M. Then α is a minimising slope for M if and only if there is a fundamental surface F of $\mathcal{T}$ with $[\partial F] = \alpha$ and $\chi(F) = -\|M\|$.*

Proof Suppose α is a minimising slope for M. Then there is a surface S in M with $[\partial S] = \alpha$ and $\chi(S) = -\|M\|$. Since $\|M\|$ is minimal, there is no surface in M with a single boundary component and larger Euler characteristic than S. Hence S is incompressible and ∂–incompressible. The result therefore follows from Lemma 13. $\square$

The above corollary gives an algorithm to compute $\|M\|$ and the set of all minimising slopes. This can be improved to compute the norm of every even slope:

Theorem 15 *Let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Suppose $\mathcal{T}$ is a 0–efficient triangulation of M, and $(\mathfrak{m}_2, \mathfrak{l}_2)$ a 2–torsion framing of ∂M.*

There is an algorithm that, upon input $\mathcal{T}$ with a 2–torsion framing and an even slope, computes the norm of this slope.

Proof Let δ be an even slope and S be a taut surface for δ. Hence S has a single boundary curve of slope δ, is incompressible, and satisfies $\|\delta\| = -\chi(S)$. If S is not ∂–incompressible, we can perform boundary compressions on S until we have an incompressible and ∂–incompressible surface. Denote the resulting sequence of surfaces $S = S_0, S_1, \ldots, S_n$ with boundary slopes $\delta = \delta_0, \delta_1, \ldots, \delta_n$, where S_n is an incompressible and ∂–incompressible surface, and S_i is obtained from S_{i-1} by a single boundary compression. In particular, $\chi(S_i) = \chi(S_{i-1}) + 1$.

Let δ_i be the slope of S_i. Since $S = S_0$ is a taut surface for δ_0, it follows inductively that S_i is a taut surface for δ_i, since otherwise reversing the process of boundary compressions by adding saddles to a taut surface would result in a surface of higher Euler characteristic. It now follows from Lemma 13 that S_n is isotopic to a fundamental surface with respect to the 0–efficient triangulation $\mathcal{T}$.

A priori, there are infinitely many possibilities for the slope δ_1 that result from a boundary compression on S_0, as can be seen from the Farey tree described in Figure 2. However, the facts that we have a sequence of boundary compressions terminating at a fundamental surface (and hence at one of only a finite list of slopes) and that the set of boundary slopes can be organised via the trivalent tree that is the dual 1–skeleton of the Farey tessellation, makes the problem of determining $\chi(S)$ finite.

In order to compute the norm of the given even slope δ, we first compute the (finite) set of all fundamental surfaces. This is equivalent to enumerating a Hilbert basis on the projective solution space. Amongst these, we select the surfaces with connected boundary. Denote these surfaces by F_i and their slopes by α_i. We remark that F_i may not be taut for α_i. For all the surfaces in the list that have the same slope, we only keep one of maximal Euler characteristic.

Consider the Farey tessellation associated with the framing $(\mathfrak{m}_2, \mathfrak{l}_2)$. Recall that the dual 1–skeleton is an infinite trivalent tree. We use this to define distances between ideal triangles (equivalently, isotopy classes

of 1–vertex triangulations). For every slope, there are infinitely many ideal triangles with a vertex at this slope. We canonically (but arbitrarily) choose, for each slope α_i of a fundamental normal surface F_i, the ideal triangle $\tau(\alpha_i) = (\alpha_i, \beta_i, \gamma_i)$ that is at shortest distance from the base triangle $\left(\frac{1}{0}, \frac{0}{1}, \frac{-1}{1}\right)$. This is characterised by the relationship $\alpha_i = \beta_i \oplus \gamma_i$ in Farey addition (whereby numerators and denominators are simply added). Note that the set of all ideal triangles with α_i as an ideal vertex corresponds to an infinite path in the dual tree; see Figure 9 for a number of these infinite paths in the case where α_i is an integral even slope.

The effect of a boundary compression on the slope of a surface is exactly one step in the dual tree between triangles with distinct even slopes. The reverse step is the addition of a saddle. Hence the sequence $S = S_0, S_1, \dots, S_n$ described above corresponds to a path without backtracking between $\tau(\delta_0)$ and $\tau(\delta_n)$ in the dual tree. The difference $\chi(S_n) - \chi(S_0)$ is the number of edges in this path that have endpoints in triangles of different even slopes. This difference is well defined since the dual 1–skeleton of the Farey tessellation is a tree.

Given even slopes α and β, let $d(\alpha, \beta)$ be the number of edges in the shortest path between $\tau(\alpha)$ and $\tau(\beta)$ that have endpoints in triangles of different even slopes. Note that this is independent of the choice of $\tau(\alpha)$ and $\tau(\beta)$ as triangles with those even slopes. It follows that

$$\|\delta\| = -\chi(S) = \min_{F_i}\{-\chi(F_i) + d(\alpha_i, \delta)\}. \qquad\qquad \square$$

The proof leads to the following algorithm:

Algorithm 16 Compute the norm of a given even slope.

    Input:
    (1) $\mathcal{T}$, a 0–efficient triangulation of M
    (2) $(\mathfrak{m}_2, \mathfrak{l}_2)$, a 2–torsion framing of ∂M
    (3) p/q, an even boundary slope
    compute fundamental surfaces of $\mathcal{T}$
    for S fundamental surface with connected boundary **do**
        compute boundary slope of S with respect to $(\mathfrak{m}_2, \mathfrak{l}_2)$
        if boundary slope already in Farey tessellation data structure **then**
            update norm $= \min(\text{norm}, -\chi(S))$
        else
            insert boundary slope and norm into Farey tessellation data structure
        end if
    end for
    insert p/q into boundary slope into Farey tessellation data structure
    return minimum of (distance + norm) of p/q to boundary slopes in the Farey tessellation data structure

Remark 17 (Farey tessellation data structure) In the proof, we have chosen to associate a canonical triangle in the Farey tessellation with an even slope. In practice, one may use any triangle with the slope, thus saving some computations. If the norms of many slopes are to be computed, then it would be worthwhile to first set up a data structure containing the norm for each slope of a fundamental surface.

It is important to note that we do not have to deal with the infinite object that is the Farey tessellation, but with a finite tree that is modelled on the dual 1–skeleton of the Farey tessellation. The nodes of our data structure are those of the dual 1–skeleton corresponding to even boundary slopes realised as boundary slopes of fundamental surfaces. Arcs are established between nodes along edges in the dual 1–skeleton of the Farey tessellation. This is an efficient procedure thanks to the Euclidean algorithm. It may add auxiliary nodes that are common to paths coming from different nodes. The arcs are assigned weights equal to the number of edges in the dual 1–skeleton of the Farey tessellation that have endpoints in triangles of different even slopes. With this setup, providing a query boundary slope amounts to inserting this extra slope into the data structure (exactly as before), and computing weighted path lengths without backtracking to all other nodes.

3 Crosscap number of knots

We now restrict our view to knot exteriors with the following special property. Suppose N is a closed orientable 3–manifold, and $K \subset N$ a knot with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$.

Let $\nu(K)$ be an open regular neighbourhood of K. We assume that the exterior $M = N \setminus \nu(K)$ is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. For instance, this is the case for any knot in $N = \mathbb{S}^3$, and it is the case for any hyperbolic knot in N, ie when M has a complete hyperbolic structure of finite volume.

The knot K is *trivial* in N if there is a properly embedded disc in M that has nontrivial slope on ∂M. In particular, K is nontrivial if and only if M has incompressible boundary.

The *crosscap number* $c(K)$ of a nontrivial knot $K \subset N$ is defined by

$$c(K) = \min\{1 - \chi(S) \mid S \text{ is a nonorientable spanning surface for } K\}.$$

The crosscap number of a trivial knot is defined to be zero. The subtle difference between crosscap number and slope norm is that the crosscap number is not simply obtained by computing the minimal norm over all slopes of spanning surfaces, since it also takes into account the orientability class of a taut surface.

3.1 Geometric framings and spanning surfaces

The closure of $\nu(K)$ is a solid torus, and the curve $\mathfrak{m}_g$ on ∂M bounding the meridian disc for this solid torus is *the geometric meridian* for K. This information allows us to pass between (N, K) and $(M, \mathfrak{m}_g)$.

A *spanning surface* for K in N is an embedded connected surface S in N with $\partial S = K$. If S is a spanning surface for K in N, then $F = S \cap M$ has the property that ∂F has algebraic intersection

number ± 1 with $\mathfrak{m}_g$ on the boundary torus (after choosing orientations for both curves). Conversely, suppose F is a properly embedded surface in M with a single boundary component. If ∂F has algebraic intersection number ± 1 with $\mathfrak{m}_g$, then F extends to a spanning surface of K in N.

The condition that $[K] = 0 \in H_1(N; \mathbb{Z}_2)$ implies that K has a (possibly nonorientable) spanning surface. The intersection pairing with the meridian shows that $\mathfrak{m}_g$ maps to a nontrivial element of $\mathrm{im}(H_1(\partial M, \mathbb{Z}) \to H_1(M, \mathbb{Z}_2))$, and hence is a 2–meridian.

Recall the definition of a homological framing in Remark 9. In the setting here, it is more natural to define a geometric framing that includes the class of the boundary of the meridian disc as a generator.

Geometric framing, I Suppose K has an orientable spanning surface S. Then the boundary of S is the homological longitude of M and, due to the intersection pairing, the geometric meridian is a homological meridian. In particular, K has an orientable spanning surface if and only if $[K] = 0 \in H_1(N; \mathbb{Z})$. By the discussion above, in this case, every nonorientable spanning surface has boundary slope of the form $\mathfrak{m}_g^{2k} \mathfrak{l}_\infty$ (recall that we represent isotopy classes of unoriented curves by choosing a representative with nonpositive longitudinal coordinates). Any orientable spanning surface can be turned into a nonorientable spanning surface by attaching a saddle, and iteratively adding saddles shows that the set of all slopes of spanning surfaces is precisely the set

$$S(K) = \{\mathfrak{m}_\infty^{2k} \mathfrak{l}_\infty \mid k \in \mathbb{Z}\}$$

for any homological meridian $\mathfrak{m}_\infty$. Alternatively, one may define a *geometric longitude* $\mathfrak{l}_g = \mathfrak{l}_\infty$, and hence

$$S(K) = \{\mathfrak{m}_g^{2k} \mathfrak{l}_g \mid k \in \mathbb{Z}\}$$

is the set of all boundary slopes of spanning surfaces. See Figure 9 for the subtree of spanning slopes sitting inside the dual of the Farey tessellation.

Geometric framing, II Now suppose K has no orientable, but a nonorientable spanning surface S. This is the case if and only if $[K] = 0 \in H_1(N; \mathbb{Z}_2)$ and $[K] \neq 0 \in H_1(N; \mathbb{Z})$. The "only if" direction follows from the existence of S and the discussion above, and the "if" direction from the construction given in the proof of Lemma 7. By the above, we have $[\partial S] = \mathfrak{m}_\infty^{2k_0} \mathfrak{l}_\infty^{q_0}$ for some $q_0, k_0 \in \mathbb{Z}$, q_0 and $2k_0$ coprime, and $q_0 > 0$. Since this is isotopic to the core curve of $\nu(K)$, $\mathfrak{m}_g = \mathfrak{m}_\infty^{r_0} \mathfrak{l}_\infty^{s_0}$, where $2k_0 s_0 - q_0 r_0 = \pm 1$. Each slope of a nonorientable surface is zero in $H_1(M; \mathbb{Z}_2)$ and meets the geometric meridian once algebraically. Again, by adding saddles to S, one can obtain all possible slopes that arise this way, and we conclude that the slopes of all spanning surfaces are precisely the set

$$S(K) = \{\mathfrak{m}_g^{2k} (\mathfrak{m}_\infty^{2k_0} \mathfrak{l}_\infty^{q_0}) \mid k \in \mathbb{Z}\}.$$

To obtain a more pleasing description, we define a *geometric longitude* $\mathfrak{l}_g$ as follows. Fix a Euclidean norm on $H_1(\partial M, \mathbb{Z})$ with the property that $\|\mathfrak{l}_\infty\| = 1$ and $\|\mathfrak{m}_g\| = 1$. Then define $\mathfrak{l}_g$ to be a shortest curve in $S(K)$. It then follows that

$$S(K) = \{\mathfrak{m}_g^{2k} \mathfrak{l}_g \mid k \in \mathbb{Z}\}.$$
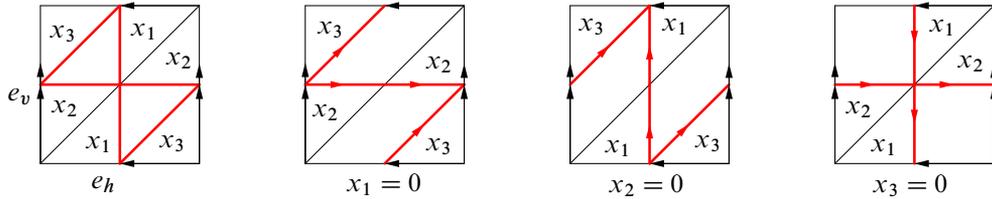
Figure 5: The figure-8 graphs and their orientations for the three types of essential normal curves.

Computing intersection numbers Suppose $\mathscr{T}$ is a triangulation of M with the property that the induced triangulation $\mathscr{T}_{\partial}$ of ∂M is a two-triangle triangulation of the torus. The purpose of this section is to show how to set up the equations to determine whether a given normal surface is a spanning surface.

Given a normal surface S with nonempty boundary, we obtain $\partial S = c$ as a (not necessarily connected) normal curve in $\mathscr{T}_{\partial}$. The normal curve c is connected if and only if the greatest common divisor of its coordinates equals one. Hence S can only be a spanning surface if this is the case.

We assume that the geometric meridian $\mathfrak{m}_g$ is given as a normal curve with respect to $\mathscr{T}_{\partial}$. We now explain how to compute the minimal number of intersections between the isotopy classes of any two essential connected normal curves (such as $\mathfrak{m}_g$ and $\partial S = c$). This makes use of the fact that, on the torus, the geometric and the algebraic intersection numbers of oriented curves coincide.

Represent the triangulation $\mathscr{T}_{\partial}$ as the identification space of a square with a diagonal and label the normal coordinates, as in Figure 5. It was shown in [Jaco and Sedgwick 2003] that the normal coordinates (x_1, x_2, x_3) of one triangle determine the normal coordinates in the second triangle, as indicated in the figure. A connected essential normal curve contains no vertex linking curves, and hence its normal coordinates satisfy $x_i = 0$ for at least one $i \in \{1, 2, 3\}$. Again, we check that $\partial S = c$ satisfies this requirement.

It follows that an essential normal curve can be isotoped to be in the neighbourhood of a figure-8 graph on the torus that is composed of four normal arcs. There are three such graphs, as shown in Figure 5. The normal coordinates can be viewed as weights on the edges of this graph. It was observed in [Bachman et al. 2016] that each essential normal curve can be given a canonical orientation which depends only on which normal coordinate is zero. This in turn can be viewed as an orientation of the edges of the corresponding figure-8 graph as shown in Figure 5.

In particular, the oriented intersection numbers between an oriented normal curve and two oriented edges e_v and e_h of the triangulation $\mathscr{T}_{\partial}$ can be read off from the normal coordinates. We call them *oriented edge weights*. Hence the intersection number of two essential normal curves can be computed as a determinant in these oriented edge weights. With respect to the labelling shown in Figure 5, the normal coordinates $x = (x_1, x_2, x_3)$ result in the oriented edge weights $(x_2 + x_3, x_3)$ if $x_1 = 0$, $(x_3, x_1 + x_3)$ if $x_2 = 0$ and $(x_2, -x_1)$ if $x_3 = 0$. The intersection number between two oriented essential normal curves is then the determinant of the 2×2 matrix with these oriented edge weights as columns. For example, if $x = (0, x_2, x_3)$ and $y = (y_1, 0, y_3)$, then the intersection number is $(x_2 + x_3)(y_1 + y_3) - x_3 y_3 = x_2 y_1 + x_2 y_3 + x_3 y_1$.

The upshot of this discussion is that given a normal curve with respect to $\mathcal{T}_\partial$ that represents the meridian, we have an algorithm to determine whether a normal surface is a spanning surface.

Efficient suitable triangulations One of the tools used in [Burton and Ozlen 2012] is the following concept. We say that $\mathcal{T}$ is an *efficient suitable triangulation* of M if there is no normal 2–sphere with respect to $\mathcal{T}$, the induced triangulation $\mathcal{T}_\partial$ of the boundary has a single vertex, and the geometric meridian is isotopic with a boundary edge. As in [loc. cit.], it follows easily from [Jaco and Rubinstein 2003, Proposition 5.15 and Theorem 5.20] that if K is a nontrivial knot in N, then M has an efficient suitable triangulation and that this can be constructed algorithmically from any triangulation of M. Namely, one first constructs a 0–efficient triangulation of M. If one of the edges in the boundary is isotopic with the geometric meridian, then we are done. Otherwise, we layer tetrahedra on the triangulation until one of the edges is the geometric meridian. A minimal layering sequence can be determined from the Farey tessellation. Now a 0–efficient triangulation has no normal 2–spheres. Consider a step in the layering procedure. If the triangulation before layering a tetrahedron on the boundary has no normal 2–sphere, then so does the triangulation after layering a tetrahedron, since any normal surface that meets the new tetrahedron in a disc has nonempty boundary. So it follows inductively that an efficient suitable triangulation can be obtained from a 0–efficient triangulation.

Normalisation An excellent discussion of the procedure that constructs normal surfaces from properly embedded surfaces transverse to a triangulation can be found in Matveev's book [2007, Section 3.3.3]. See also [Burton and Ozlen 2012] for a similar discussion to what follows. Normalisation moves are isotopies, removal of trivial components, compressions along circles of intersection of the surface with triangles of the triangulation, or boundary compressions along discs that have part of their boundary in the interior of boundary edges of the triangulation.

Suppose S is a properly embedded surface in M that is incompressible and transverse to the triangulation. Since M is irreducible, any normalisation move that would result in a 2–sphere or properly embedded disc split off from S can be avoided by a suitable isotopy of S that is supported in a regular neighbourhood of a ball bounded by the 2–sphere or cobounded by the disc. In particular, this isotopy removes some intersection points of S with the edges or some intersection circles of S with the triangles (possibly both). Instead of a normalisation move of this type, we perform the associated isotopy. If S can be transformed to a normal surface in this way, then we say that S *normalises by isotopies*.

The only normalisation moves that cannot be replaced by an isotopy are boundary compressions where the boundary compression disc D is contained in a 3–simplex Δ of the triangulation and $\partial D = \gamma \cup \gamma'$ with the property that γ is an arc contained in the interior of an edge $e \subset \partial M$ of Δ and γ' is an arc contained in S. Moreover, $D \setminus \gamma$ is contained in the interior of Δ. Note that such a boundary compression decreases $S \cap e$ by two points and leaves the intersections with all other boundary edges unchanged. The normalisation procedure may involve multiple boundary compressions. The upshot of this discussion is

that if S meets one boundary edge in a single point, then this is also true of any surface obtained from S by applying the normalisation procedure (with or without the modification of using isotopies).

3.2 Crosscap number via suitable triangulations

Theorem 1, which is proved in this section, is a result about the crosscap number of certain knots in closed 3–manifolds, which, in particular, gives an algorithm to compute the crosscap number of an arbitrary knot in $\mathbb{S}^3$. The result requires the use of efficient suitable triangulations, as defined in the previous section. Theorem 3 in Section 3.3 is the equivalent result for arbitrary 0–efficient triangulations. Both results share Lemma 19, which is why forward references to Theorem 3 and Section 3.3 appear in this section.

The proof of Theorem 1 shows that the following algorithm to compute the crosscap number is correct:

Algorithm 18 Compute the crosscap number of the knot K in the 3–manifold M, satisfying the conditions of Theorem 1.

Input:
(1) $\mathcal{T}$, an efficient suitable triangulation of M with boundary $\mathcal{T}_\partial$
(2) $\mathfrak{m}$, a meridian of K represented by an edge of $\mathcal{T}_\partial$
compute S_0, the set of fundamental surfaces of $\mathcal{T}$
compute $S_1 = \{S \in S_0 \mid \partial S \cap \mathfrak{m} = 1\}$, the set of spanning fundamental surfaces
compute $x = \max\{\chi(S) \mid S \in S_1\}$, maximum Euler characteristic of all spanning fundamental surfaces
if $x = 1$ **then**
    return $c(K) = 0$
end if
if $x = 0$ **then**
    return $c(K) = 1$
end if
compute $S_2 = \{S \in S_1 \mid \chi(S) = x\}$, the set of maximal Euler characteristic spanning fundamental surfaces
if S_2 contains nonorientable surface **then**
    return $c(K) = 1 - x$
else
    return $c(K) = 2 - x$
end if

Proof of Theorem 1 Let S_o be an orientable spanning surface of maximal Euler characteristic and S_n be a nonorientable spanning surface of maximal Euler characteristic. Then $c(K) = \min(1 - \chi(S_n), 2 - \chi(S_o))$. If there is no orientable spanning surface, then $c(K) = 1 - \chi(S_n)$. The definitions imply that $c(K) \leq \min(A, B)$, since any orientable spanning surface can be turned into a nonorientable spanning surface by adding a saddle appropriately.

Both S_o and S_n are incompressible due to the maximality condition on the Euler characteristic.

Since ∂M is a torus and S_o is orientable, this also implies that S_o is ∂–incompressible. We may therefore assume that S_o is normal in M. By [Jaco and Sedgwick 2003, Corollary 3.8], two compatible normal surfaces with nonempty boundary either have the same slope (and hence their sum has at least two boundary curves) or complementary boundary curves (and hence their sum has boundary containing a trivial curve). Therefore only one of the fundamental surface summands, G_o, yielding S_o has nonempty boundary and $\partial S_o = \partial G_o$. Since Euler characteristic is additive and there are no normal 2–spheres, $\chi(S_o) \leq \chi(G_o)$. Since $\partial S_o = \partial G_o$, G_o is also a (not necessarily orientable) spanning surface.

If $\chi(S_n) < \chi(S_o)$, then G_o must be orientable.[1] Also, $\chi(S_n) < \chi(S_o)$ implies $c(K) = 2 - \chi(S_o) \leq 1 - \chi(S_n)$. Since G_o is an orientable spanning surface, $\chi(G_o) = \chi(S_o)$, and therefore $c(K) = 2 - \chi(G_o) \geq B$. Thus $c(K) = B$.[2]

Hence assume $\chi(S_n) \geq \chi(S_o)$. In this case $c(K) = 1 - \chi(S_n) < 2 - \chi(S_o)$.

Now S_n may not be ∂–incompressible. We use the following argument from [Burton and Ozlen 2012]. Since ∂S_n has algebraic intersection number one with the geometric meridian $\mathfrak{m}_g$, we may isotope S_n in M so that it meets the edge in $\mathscr{T}_\partial$ that represents $\mathfrak{m}_g$ in exactly one point. We now isotope S_n so that it is transverse to the triangulation. This still meets $\mathfrak{m}_g$ in exactly one point. As noted above, any surface obtained by applying the normalisation procedure to S_n results in a surface meeting $\mathfrak{m}_g$ in exactly one point, and hence a spanning surface.

In particular, any nontrivial boundary compression involved in putting S_n into normal form results in a spanning surface of larger Euler characteristic. Since $\chi(S_n) \geq \chi(S_o)$, this surface must again be nonorientable, which contradicts the maximality of the Euler characteristic of S_n. Hence S_n can be normalised by isotopies. As above, we see that only one of the fundamental surface summands, G_n, yielding S_n has nonempty boundary. Since Euler characteristic is additive, $\chi(S_n) \leq \chi(G_n)$. Since $\partial S_n = \partial G_n$, G_n is also a spanning surface.

We now have the following cases:

If $\chi(S_n) > \chi(S_o)$, then G_n must be nonorientable. This forces $\chi(G_n) = \chi(S_n)$, and we have $c(K) \geq A$, which implies $c(K) = A < B$.

If $\chi(S_n) = \chi(S_o)$, then $\chi(G_o) = \chi(S_n) = \chi(S_o) = \chi(G_n)$. The proof is continued with Lemma 19, where it is shown that there is at least one nonorientable fundamental spanning surface with this maximal Euler characteristic. Hence $c(K) = A < B$.   □

Lemma 19 *Suppose M and $\mathscr{T}$ are as in the hypothesis of Theorem 1 or Theorem 3. Also assume, in the case of Theorem 3, that $c(K) < Z$.*

---

[1] See the proof of Theorem 5 for an argument that G_o is always orientable if S_o is of least weight in its isotopy class.

[2] Note that in this case, we either have $c(K) = B < A$ or $c(K) = B = A$. See Section 6 for examples of both cases.

Suppose S_o is an orientable spanning surface of maximal Euler characteristic, and S_n is a nonorientable spanning surface of maximal Euler characteristic. If $\chi(S_n) = \chi(S_o)$, then there is at least one nonorientable fundamental spanning surface F with $\chi(F) = \chi(S_n) = \chi(S_o)$.

Proof We prove this by contradiction. We know that every nonorientable spanning surface of maximal Euler characteristic is isotopic to a normal surface (since we either assume that the triangulation is suitable or that $c(K) < Z$). We also know (from the last paragraph in the proof of both Theorems 1 and 3) that there is at least one fundamental spanning surface of Euler characteristic $\chi(S_n) = \chi(S_o)$. Suppose every such fundamental spanning surface is orientable, and hence S_n is isotopic to a nontrivial Haken sum S of normal surfaces. We then show that this implies that S must be orientable as well, a contradiction.

Here is the outline of our proof:

(1) **Setup** Let S be a nonorientable normal spanning surface with $\chi(S) = \chi(S_n)$ that is of least weight.

(2) **Setting up the Haken sum for S** We show that $S = R + T$, where T is a torus, R is connected (incompressible), and no patch is a disc.

(3) **R meets T an even number of times** We show that $R \cap T$ is an even number of essential curves on T.

(4) **All components of $R \setminus T$ above T are annuli** We show that any patch of R in the component of $M \setminus T$ that does not contain ∂M must be a (separating) annulus.

(5) **$R + T$ is orientable** We use the above statements to conclude that $R + T$ must be orientable.

Setup Suppose $x = \chi(S_n) = \chi(S_o)$ is the maximum Euler characteristic of a fundamental spanning surface. We assume that all fundamental spanning surfaces with maximum Euler characteristic are orientable. Since K is nontrivial, none of these is a disc. Since there are no orientable spanning surfaces of Euler characteristic 0, we can assume for the remainder of the proof that all orientable fundamental spanning surfaces have strictly negative Euler characteristic.

We know that a nonorientable *normal* spanning surface S with $\chi(S) = x$ exists. Furthermore, we may assume that S has least weight amongst all nonorientable normal spanning surfaces with Euler characteristic x.

By hypothesis, S is not fundamental.

Setting up the Haken sum for S By Proposition 12, two compatible normal surfaces with nonempty boundary either have the same slope (so their sum has at least two boundary curves) or complementary boundary curves (so their sum has boundary containing a trivial curve). Hence only one of the fundamental surface summands, F, yielding S has nonempty boundary and $\partial F = \partial S$. In particular, F is a spanning surface for K and has negative Euler characteristic.

By hypothesis, $\mathscr{T}$ does not admit any normal spheres, and since M is irreducible with nonempty boundary a torus it does not admit any embedded $\mathbb{R}P^2$. Hence every surface in the Haken sum giving S has nonpositive Euler characteristic. Since F is a spanning surface and Euler characteristic is additive in Haken

sums, the maximality of Euler characteristic x amongst all spanning surfaces forces $\chi(F) = x$, and all other summands have Euler characteristic zero. Since F is a fundamental spanning surface and $\chi(F) = x$, it follows that F is orientable. Since M contains no Klein bottles, all other summands are fundamental tori.

Note that removing any normal torus from any Haken sum with result S must produce an orientable surface, since S is least-weight.

Let R be such an orientable surface, and let T be the missing torus such that $R + T = S$. There are potentially many such decompositions, and we claim that for at least one of them R is connected. Assume that we have $R + T = S$ with R disconnected. Then R must consist of one spanning surface of maximal Euler characteristic, and a number of tori. Since $R + T = S$ is connected, all of these extra torus components of R must intersect T. Pick one of these tori, denote it by T', and use the remaining components of R to form the Haken sum $R' = (R \setminus T') + T$. By construction, $S = R' + T'$ and $R' \cap T' \subset R \cap T$ is a proper subset. Iterating this process eventually yields a decomposition $S = R + T$ with R connected.

Since R is an orientable spanning surface of maximal Euler characteristic, it is incompressible and boundary incompressible.

As is customary when talking about *Haken sums*, we call the connected components of $(R \cup T) \setminus \nu(R \cap T)$ *patches*. Every patch is a compact orientable subsurface of S with nonempty boundary. Since both R and T are orientable, the patches are connected on S by annuli contained on the frontier of $\nu(R \cap T)$. These are called the *exchange annuli*. The core curve of an exchange annulus is called a *trace curve*. Note that a trace curve corresponds to the boundary curves of patches on both R and T that are joined by the corresponding annulus.

The complementary annuli on the frontier of $\nu(R \cap T)$ are the *irregular annuli*. Let $\gamma \subset R \cap T$. Attaching the exchange annuli to the patches is called a *regular exchange* at γ, and attaching the irregular annuli to the patches is called an *irregular exchange*.

Amongst all ways to write $S = R + T$, where R is an orientable normal spanning surface of maximal Euler characteristic and T is a normal torus, we assume we have chosen one with the least number of patches. Note that $|R \cap T| \neq \varnothing$ as S is nonorientable.

No patch of $S = R + T$ is a disc This follows almost verbatim from the proof of [Jaco and Oertel 1984, Lemma 2.1]. We repeat a slightly adapted version of the beginning of the proof here, as our setup is slightly different. However, the endgame remains the same.

Suppose $\gamma \subset R \cap T$ and denote the two associated trace curves by γ' and γ''. Suppose γ' bounds a patch D' that is a disc on S. Since a disc is 2–sided in M, there is an embedded disc $D \subset M$ with interior disjoint from S and boundary curve γ''. Since S is incompressible, this implies that γ'' also bounds a disc D'' on S. Since D' is a patch, $D'' \nsubseteq D'$. We claim that D'' is not a patch.

If $D' \subset D''$, then D'' is clearly not a patch.
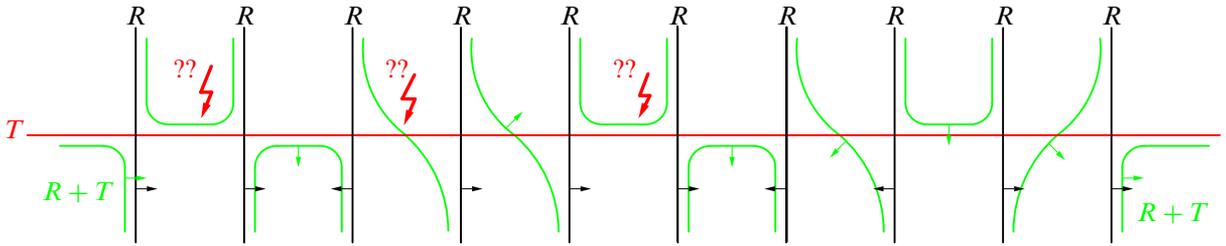
Figure 6: Cross section of the intersection of T and R together with the transverse orientation of R and marked potential obstructions to orientability of the resolution realised by the Haken sum $S = R + T$.

Hence assume that $D' \nsubseteq D''$. Then $D' \cap D'' = \varnothing$. Perform an irregular exchange along γ and regular exchanges along all other curves in $R + T$. If an irregular annulus does not join D' and D'', then we can construct a 2–sphere in M that separates S, a contradiction since S is connected. Hence the union of D', D'' and an irregular annulus is an embedded 2–sphere in M. Now if both D' and D'' are patches, then we may perform a regular exchange along only γ and no other component of $R \cap T$ to obtain new normal surfaces R' and T' with $S = R' + T'$. Moreover, R' is isotopic with R and T' is isotopic with T. So $R' + T'$ still satisfies our hypothesis on the Haken sum that R' is an orientable normal spanning surface of maximal Euler characteristic and T' is a normal torus. Now $R' + T'$ has two patches fewer than $R + T$, since D' and D'' are now contained in patches. It follows that D'' is not a patch.

So in either case, we establish that D'' is not a patch. To conclude, we know that for each patch D' that is a disc there is an associated disc D'' on S which is not a patch and with the property that the boundary curves of D' and D'' are associated with the same curve in $R \cap T$. Now the construction and argument given in the last four paragraphs in the proof of [loc. cit., Lemma 2.1] results in a surface of lower weight than S and isotopic with S. Hence no patch is a disc, and in particular, no patch is of positive Euler characteristic.

Irregular exchanges Recall that we have established that $R \cap T$ is a collection of essential curves on T slicing up T into a set of annuli. Any such situation can be fully reconstructed from the schematic picture shown in Figure 6, where the horizontal line represents the torus T, and the vertical lines represent R slicing through T. Every possible Haken sum is then described by an orientation at each intersection for how the Haken sum operates, and the transverse orientation for every sheet of R.

Suppose any nonempty subset of components of $R \cap T$ are resolved via irregular exchanges, whilst all other components are resolved via regular exchanges. This results in a surface that is not normal, has nonempty boundary and is not necessarily connected. However, it contains a component C that is a maximum Euler characteristic spanning surface. Since we have performed at least one irregular exchange, we know that C is not a normal surface and hence normalises to a normal surface of lower weight than S.

We claim that C is orientable. This is the only place where we have different arguments given the hypotheses of the theorems:

First assume the hypotheses of Theorem 1. Since $\partial C = \partial S$, the surface C is a maximum Euler characteristic spanning surface that meets the meridian edge in exactly one point and hence normalises by isotopies. In particular, the resulting normal surface is again a spanning surface. The minimality of the weight of S implies that this spanning surface (and therefore C) is orientable.

Next assume the hypotheses of Theorem 3 and that $c(K) < Z$. If C is nonorientable, then it normalises by isotopies. But this contradicts the minimality of S. Hence C is orientable.

This proves the claim. It follows that if we perform at least one irregular exchange, then there is a component C that is a maximum Euler characteristic spanning surface and that normalised to an orientable spanning surface of maximal Euler characteristic. We will repeatedly make use of this observation.

R meets T an even number of times Since $R + T$ is nonorientable and R and T are orientable, every orientation-reversing loop in $R + T$ must pass through patches of both R and T. Take one such loop $c \subset R + T$ that minimises the number of times it intersects the exchange annuli between patches of R and T. We claim that c intersects each exchange annulus in at most one arc from one boundary component to the other.

To see this, first note that if c intersects an exchange annulus in an arc going back to the same boundary component we can simply isotope it out of the exchange annulus. This is a contradiction to the assumption that c minimises intersections with exchange annuli.

If c intersects the same exchange annulus in more than one arc, take two such arcs that are next to each other and isotope them into a small disc containing a section of the exchange annulus and a piece of R and T on either side. Delete the two arcs meeting the disc and connect the four ends outside the exchange annulus to obtain a curve c' meeting the exchange annuli of $R + T$ fewer times than c (see Figure 7).

We show that one component of c' must still be orientation reversing. Referring to Figure 7, we start at end 1, which can be connected to ends 2, 3 or 4.

• If it is connected to end 2, we immediately obtain a contradiction to the assumption that c is connected; see Figure 7, first row.

• If it is connected to end 3, then end 2 must connect to end 4. Tracing transverse orientations through both arcs and connecting them in the disc leaves us with no, one or two orientation-reversing components of c'. We can now check that if neither or both components were orientation reversing, then c would have been orientation preserving, which is a contradiction. Hence exactly one of them is orientation reversing, a contradiction to the assumption that c intersects the exchange annuli of $R + T$ a minimum number of times; see Figure 7, middle row.

• If it is connected to end 4, then end 2 must be connected to end 3. Similarly to the case before, we can trace orientations through c', thereby checking all four choices. As before, in each case we obtain contradictions to either c being orientation reversing or c having minimal intersection with the exchange annuli of $R + T$; see Figure 7, bottom row.
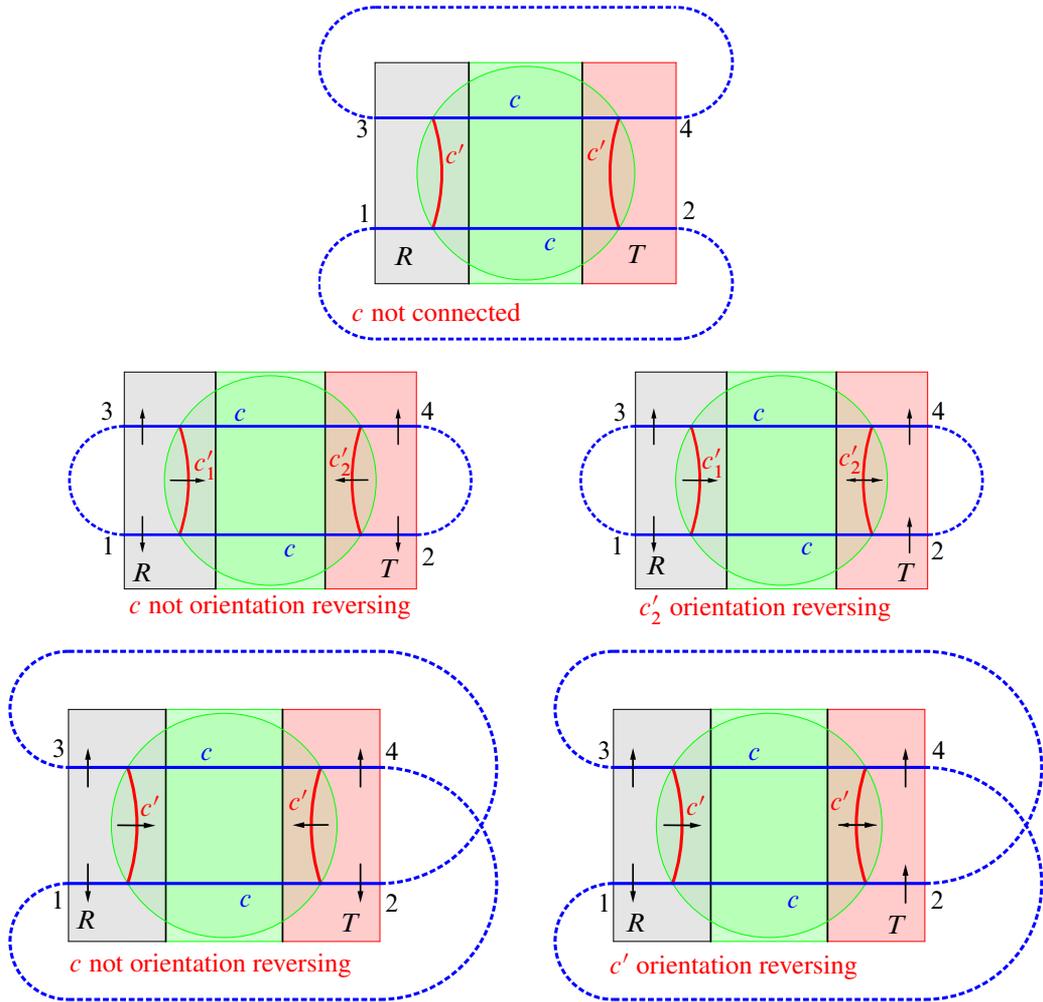
Figure 7: Top row: 1 is connected to 2 and c is not connected. Middle row: 1 is connected to 3. Two cases of 16 for choosing orientations on c near 1, 2, 3 and 4 are shown. The others follow by flipping R and T, top and bottom, and direction of all arrows. Bottom row: 1 is connected to 4. Again, only two out of 16 cases are shown and the others follow by symmetry.

Altogether it follows that c intersects every exchange annulus at most once.

Now assume that c is disjoint to the two exchange annuli coming from some component of $R \cap T$. If this is the case, we can make an irregular exchange along that component and regular exchanges along all other components of $R \cap T$. Then the resulting new surface still contains an orientation-reversing loop c, and hence is nonorientable — a contradiction to S being least-weight, as explained above.

Next, assume that c intersects both exchange annuli coming from some component of $R \cap T$. Consider the solid torus $D^2 \times S^1$ that is a regular neighbourhood of this component and contains the two exchange annuli. Choose an open regular neighbourhood of the two exchange annuli on S. There is an isotopy of c

on S that is the identity on the complement of this neighbourhood, and has the effect that the intersection of c with these exchange annuli is contained in some disc $D = D^2 \times \{x\}$. In D we now have a similar picture as before, and we refer to Figure 7. Performing an annular exchange yielding a new surface $(R + T)'$ cuts c open along two small arcs with ends 1, 2, 3 and 4. Connecting ends 1 and 3 and ends 2 and 4 through the newly added annuli yields another curve c' in the changed surface. An analysis of how orientations can be traced on c' results in an orientation-reversing loop, assuming that c was orientation reversing. It follows that $(R + T)'$ must be nonorientable, a contradiction to the assumption that $R + T$ is least-weight.

Altogether we conclude that c must meet exactly one of the two exchange annuli coming from a component of $R \cap T$ for every such component. Running through c, we meet patches of R or T and exchange annuli in an alternating fashion. Moreover, a patch of R, followed by an exchange annulus, must then be followed by a patch of T and vice versa (ie c runs through patches of R and T in an alternating fashion as well). In particular, c must run through an even number of exchange annuli. But since the number of exchange annuli meeting c is in bijection to the components of $R \cap T$, we conclude that the number of components of $R \cap T$ must be even.

All patches of S on one side of T are annuli   The Euler characteristic of S equals the sum of the Euler characteristics of all patches on S. Since no patch is a disc, the Euler characteristic of any patch is nonpositive. We already know that all patches of S contained on T are annuli. Since T is separating in M we say that the component of $M \setminus T$ containing ∂M is *below T* and that the other component is *above T*.

Since $|R \cap T|$ is even, we now perform (possibly irregular) exchanges on $R \cap T$ so that we obtain a new surface S' that is isotopic to a surface disjoint from T. This is achieved by alternating the exchanges so that patches on R above T are joined via annuli on T with patches above T, and patches below T with patches below T. This results in one surface, R', below T and one surface, T', above T. As above, $\chi(R') = \chi(S)$, and this forces all patches that are not contained on R' to be annuli.

$R + T$ is orientable   Every properly embedded annulus above T is separating in $M \setminus T$, since otherwise we have a nonseparating torus in M. Take an innermost annulus patch of S, ie a patch contained in a component C of $R \setminus T$ with the property that the frontier of C bounds an annulus on T that does not contain any components of $R \cap T$. There are four possibilities for how the Haken sum $R + T$ connects the patch on C with patches on T. One of them produces a separate connected torus component, and hence contradicts the connectedness (least-weight) assumption for $R + T$ (see Figure 8, top right). Two solutions give us the opportunity to resolve the Haken sum in the opposite way on one of the crossings, producing an extra torus component and not changing the nonorientability of the rest of the Haken sum (see Figure 8, bottom row). This leaves us with the last option, which is an exchange of annuli between the summands (see Figure 8, top left).

Given that all components of $R \setminus T$ on the side of M not containing ∂M are separating annuli, we can iterate this argument stating that the entire Haken sum can be resolved this way. But then $R + T$ is disconnected, which is a contradiction.
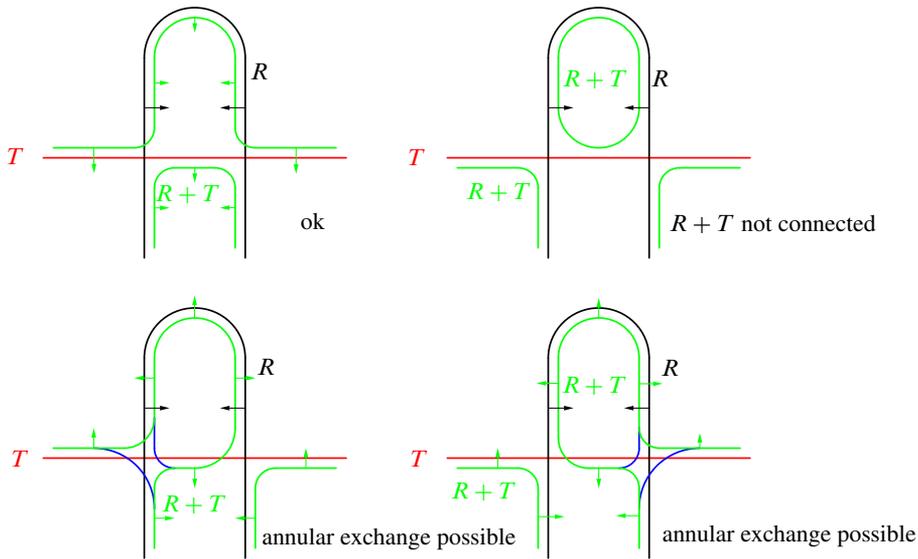
Figure 8: Resolving a separating annulus in all four possible ways.

We conclude that there is no nonorientable normal spanning surface S with $\chi(S) = \chi(S_o)$. This contradicts our hypothesis that $\chi(S_o) = \chi(S_n)$ and the hypothesis that S_n is normal. $\qquad\square$

3.3 Crosscap number via arbitrary 0–efficient triangulations

A trade-off in the previous algorithm is that one cannot apply it to arbitrary 0–efficient or minimal triangulations.

Even integral subtree With respect to the geometric framing $(\mathfrak{m}_g, \mathfrak{l}_g)$, construct the Farey tessellation. We claim that the dual 1–skeleton restricted to all ideal triangles that are labelled with the slopes of spanning surfaces is connected. These are precisely the ideal triangles with labels of the form $2k/1$, where $k \in \mathbb{Z}$. For each fixed k, these triangles correspond to an infinite path in the dual 1–skeleton. The path corresponding to $(2k-2)/1$ then connects to the path corresponding to $2k/1$ through a single arc to form what we call the *even integral subtree* of the dual 1–skeleton of the Farey tessellation, denoted by $\mathscr{F}_e$. A portion of this tree is shown in Figure 9.

In more detail, first suppose $k > 0$. There is a Farey triangle with vertices $1/0$, $2k/1$ and $(2k-1)/1$. This is since we can act by the element

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

on the base triangle with vertices $1/0$, $0/1$ and $-1/1$. Then flipping across the ideal edge $[1/0, (2k-1)/1]$ gives the triangle with vertices $1/0$, $(2k-1)/1$ and $(2k-2)/1$. Travelling along the path corresponding to $(2k-2)/1$, we arrive at the triangle with vertices $1/0$, $(2k-2)/1$ and $(2k-3)/1$. Hence inductively we arrive at the base triangle with even slope $0/1$.
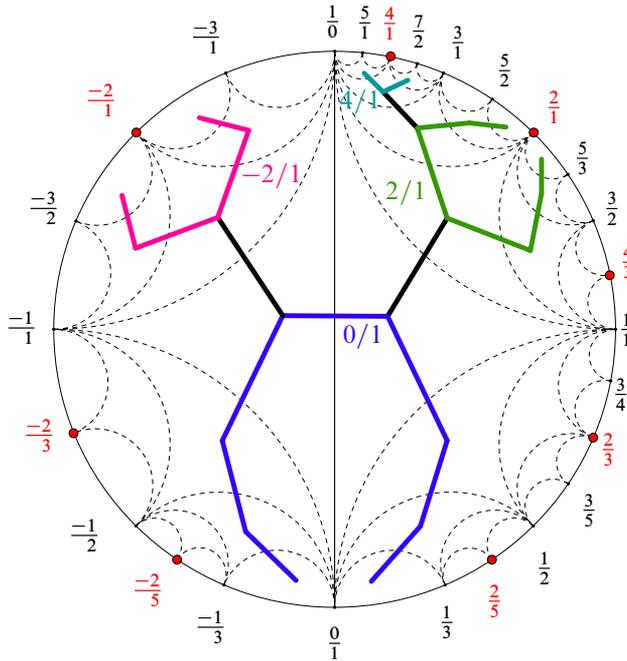
Figure 9: The even integral subtree $\mathscr{F}_e$ representing spanning slopes in the dual of the Farey tessellation. Every spanning slope is represented by an infinite path of a given colour, and paths are connected by a single saddle attachment (black arcs). Only a small portion of $\mathscr{F}_e$ is shown.

Now suppose $k < 0$. Then there is a Farey triangle with vertices $1/0$, $2k/1$ and $(2k + 1)/1$. This is obtained by acting by

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

on the triangle with vertices $1/0$, $0/1$ and $1/1$. Again, flipping across the ideal edge $[1/0, (2k + 1)/1]$ gives a triangle with even slope $(2k + 2)/1$. So inductively we arrive at the ideal triangle with vertices $1/0$, $0/1$ and $-1/1$, which shares an edge with the base triangle. This completes the proof that the even integral subtree is connected.

For any slope p/q, we define its *even integral subtree distance* $d(p/q, \mathscr{F}_e)$ to be the number of edges in the shortest path between $\tau(p/q)$ and $\mathscr{F}_e$ that have endpoints in triangles of different slopes.

Proof of Theorem 3 It follows from the definitions that $c(K) \leq A$ and $c(K) \leq B$. Note that if S is a fundamental nonspanning surface for K with connected essential boundary, then adding $d(\partial S, \mathscr{F}_e)$ saddles according to the corresponding shortest path in the Farey tessellation gives a nonorientable spanning surface of Euler characteristic $\chi(S) - d(\partial S, \mathscr{F}_e)$ for K. Hence $c(K) \leq Z$. So $c(K) \leq \min(A, B, Z)$, and we need to show equality.

Let S_o be an orientable spanning surface of maximal Euler characteristic, and let S_n be a nonorientable spanning surface of maximal Euler characteristic. We have $c(K) = \min(1 - \chi(S_n), 2 - \chi(S_o))$.

As in the proof of Theorem 1, S_o is incompressible and ∂–incompressible, and hence may be normalised using isotopies. As before, there is a fundamental surface G_o with $\chi(S_o) \leq \chi(G_o)$ and $\partial S_o = \partial G_o$. Hence G_o is also a spanning surface. If $\chi(S_n) < \chi(S_o)$, then G_o must be orientable. This forces $\chi(G_o) = \chi(S_o)$, and we have $c(K) = B$.

Hence assume $\chi(S_n) \geq \chi(S_o)$. In this case $c(K) = 1 - \chi(S_n) < 2 - \chi(S_o) \leq B$.

Note that S_n is incompressible, but S_n may not be ∂–incompressible.

Normalisation of S_n may involve a finite number of nontrivial boundary compressions resulting in surfaces that are not spanning surfaces. Each nontrivial boundary compression increases Euler characteristic by one. Suppose S_n normalises to the normal surface S'_n, and, topologically, the latter is obtained from the former by performing $k > 0$ nontrivial boundary compressions. Then $\chi(S_n) = \chi(S'_n) - k$. Now if S'_n is not fundamental, then we obtain a fundamental surface G_n with $\partial G_n = \partial S'_n$ and $\chi(S'_n) \leq \chi(G_n)$. The maximality of $\chi(S_n)$ implies that $\chi(S'_n) = \chi(G_n)$ and $k = d(\partial G_n, \mathscr{F}_e)$ since any surface obtained by adding saddles to G_n is nonorientable (regardless of whether G_n is orientable or not). Hence

$$c(K) = 1 - \chi(S_n) = 1 - \chi(G_n) + d(\partial G_n, \mathscr{F}_e) \geq Z,$$

and so $c(K) = Z$.

We may therefore assume that $c(K) < \min(B, Z)$. So $\chi(S_n) \geq \chi(S_o)$ and *every* nonorientable fundamental spanning surface with maximal Euler characteristic $\chi(S_n)$ normalises by isotopies. As above, we obtain a fundamental surface G_n with $\partial G_n = \partial S_n$ and $\chi(S_n) \leq \chi(G_n)$.

If $\chi(S_n) > \chi(S_o)$, then G_n must be nonorientable. This forces $\chi(G_n) = \chi(S_n)$, and we have $c(K) \geq A$, which implies $c(K) = A$.

If $\chi(S_n) = \chi(S_o)$, then $\chi(G_o) = \chi(S_n) = \chi(S_o) = \chi(G_n)$. The proof is continued with Lemma 19, where it is shown that there is at least one nonorientable fundamental spanning surface with this maximal Euler characteristic. Hence $c(K) = A$. ☐

Remark 20 It is known from work by Clark [1978] that a knot in the 3–sphere has crosscap number zero or one if and only if it is a $(2, 2k+1)$–cable of a knot for $k \in \mathbb{Z}$. Hence, one corollary of Theorem 3 is that a given 0–efficient triangulation of a nontrivial knot complement in the 3–sphere is that of a $(2, 2k+1)$–cable of a knot if and only if one of the fundamental surfaces is a Möbius strip.

4 Genus of knots

In our setting it is not difficult to recover a special case of a more general result of Schubert [1961] (which was originally proved in the context of normal surfaces with respect to handle decompositions). Namely, there is an algorithm to determine the genus of a knot using normal surface theory.

Proof of Theorem 5 Suppose S_o is an orientable spanning surface of maximal Euler characteristic. Since ∂M is a torus and S_o is orientable, this also implies that S_o is ∂–incompressible. We may therefore

assume that S_o is normal in M. Amongst all maximal Euler characteristic orientable normal spanning surfaces, we choose a surface S of least weight.

By [Jaco and Sedgwick 2003, Corollary 3.8], two compatible normal surfaces with nonempty boundary either have the same slope (and hence their sum has at least two boundary curves) or complementary boundary curves (and hence their sum has boundary containing a trivial curve). Therefore only one of the fundamental surface summands, F, yielding S has nonempty boundary and $\partial S = \partial F$.

Since Euler characteristic is additive and there are no normal 2–spheres, $\chi(S) \leq \chi(F)$. We also note that the weight of F is strictly less than the weight of S unless $S = F$.

These two observations imply that if F is orientable then $S = F$, and hence S is fundamental.

Suppose that F is nonorientable. In this case, $S = F + G$ is a nontrivial Haken sum with $G \neq \varnothing$ a closed normal surface. We give all patches of $F + G$ the induced orientation from S. Since F is nonorientable, there is a closed curve γ in $F \cap G$ where the induced orientations from S on the two patches on F meeting in γ do not agree. Since the Haken sum is orientable, it is also the case that the induced orientations on G do not agree. It follows that if one performs an irregular exchange at γ and regular exchanges at all other intersection curves, then one obtains an orientable spanning surface with the same Euler characteristic (and boundary) as S but which is not normal. Hence a normalisation of this surface will have lower weight than S. This is a contradiction, and so F is indeed orientable. $\qquad \square$

5 Quadrilateral space

We now provide some results that allow us to obtain minimising slopes and crosscap numbers of knots using computations in quadrilateral space.

We begin with some general observations that will then be adapted under varying hypotheses. We assume that M is an orientable compact irreducible 3–manifold with ∂M a single incompressible torus, and that $\mathcal{T}$ is a 0–efficient triangulation of M.

For a normal surface F, denote by $[F]_Q$ the normal Q–coordinates of F. If F is not a vertex linking disc, then $[F]_Q \neq 0$. If $[F]_Q \neq 0$ is not fundamental, then we write $[F]_Q = \sum[F_i]_Q$, where the $[F_i]_Q$ are fundamental normal Q–coordinates (with possible repetitions), and each of the corresponding normal surfaces F_i is connected and not a vertex link. Such an F_i is called a *Q–fundamental surface*. With respect to standard coordinates, $\{F_i\}$ is a compatible set of normal surfaces since triangle coordinates do not affect compatibility, and each F_i is a fundamental normal surface that is not a vertex linking disc.

Let D be a vertex linking disc. Then $F + kD = \sum F_i$ as a Haken sum of normal surfaces. The boundary of F_i may consist of essential curves and trivial curves, only consist of trivial curves, or be empty. Since the F_i are compatible normal surfaces, [Jaco and Sedgwick 2003, Corollary 3.8] implies that essential curves of at most two different slopes appear in the Haken sum, namely the slope of ∂F and its complementary slope (which depends on the triangulation of the boundary).

Since the triangulation is 0–efficient, $\chi(F_i) \leq 0$. Let $\{F_i\} = \{G_j\} \cup \{H_n\}$ be a partition into two nonempty sets. Then there are normal surfaces G and H with the property that none of their connected components is a vertex linking disc, and integers k' and k'' such that $G + k'D = \sum G_j$ and $H + k''D = \sum H_n$. Now

$$F + kD = \sum F_i = \sum G_j + \sum H_n = G + H + (k' + k'')D,$$

since vertex linking discs can be isotoped to be disjoint from a Haken sum. Similarly $k \geq k' + k''$. Note that

$$\chi(G) = \chi(F) + (k - k' - k'') - \chi(H) \geq \chi(F),$$

since $\chi(H) = \sum \chi(H_n) \leq 0$. We write $k = \tilde{k} + \hat{k}$, where $\hat{k}$ is the total number of trivial boundary components of the F_i. As in [loc. cit.], let $\mu(\partial F)$ be the maximal normal arc coordinate of the slope ∂F. We have that $\tilde{k}$ equals $\mu(\partial F)$ times the total number of essential boundary components of complementary slope in the F_i. To see this, note that forming the Haken sum of normal surfaces with boundary slopes ∂F and its complementary slope produces $\mu(\partial F)$ copies of the trivial curve. In particular, $\tilde{k}$ is a multiple of $\mu(\partial F)$, and hence either $\tilde{k} = 0$ or $\tilde{k} \geq \mu(\partial F)$.

Suppose ∂F is a single essential boundary curve. Since $\partial F + k \partial D$ consists of $k + 1 = (\tilde{k} + 1) + \hat{k}$ curves, [loc. cit., Corollary 3.8] implies that the essential boundary curves of the surfaces $\{F_i\}$ are $1 + \tilde{k}/\mu(\partial F)$ connected curves of the slope of F and $\tilde{k}/\mu(\partial F)$ connected curves of the complementary slope.

In particular, if $k = 0$, then there is exactly one surface, say F_1, with $\partial F_1 \neq \varnothing$. Hence we have F_1 fundamental, $\partial F_1 = \partial F$ a connected essential curve, $\chi(F_1) \geq \chi(F)$ (by choosing $G = F_1$), and since there are no vertex linking discs in the sum, F_1 has lower weight than F unless $F = F_1$. Also, F_1 has lower Q–weight than F unless $F = F_1$. Here *weight* still refers to the number of intersections of a normal surface with the 1–skeleton, and *Q–weight* is the total number of quadrilateral discs.

5.1 Minimising slopes

Proposition 21 *Let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Suppose $\mathscr{T}$ is a 0–efficient triangulation of M.*

Let S be a connected surface of maximal Euler characteristic amongst all properly embedded surfaces in M with boundary a single essential curve on ∂M. Then there is a Q–fundamental surface F with $\partial F = \partial S$ and $\chi(F) = \chi(S)$.

Proof Suppose S is a surface of maximal Euler characteristic amongst all properly embedded surfaces with boundary a single essential curve in M. Then S must be incompressible and ∂–incompressible, and hence normalises by isotopies. Amongst all normal surfaces with a single essential boundary component of the same slope as S and the same Euler characteristic as S, choose one of least weight. Call this surface F (noting that F may not be isotopic with S) and apply the preliminary observations. In particular, since F is of least weight, either $[F]_Q$ is fundamental or $k > 0$.

So suppose $k > 0$. If $\tilde{k}/\mu(\partial F)$ is odd, let $\{G_j\}$ be the subset of surfaces in $\{F_i\}$ with slope complementary to ∂F. If $\tilde{k}/\mu(\partial F)$ is even (and hence $1 + \tilde{k}/\mu(\partial F)$ is odd), let $\{G_j\}$ be the subset of surfaces in $\{F_i\}$

with the same slope as ∂F. It follows that the surfaces in $\{G_j\}$ all have the same slope, and possibly some trivial curves as boundary. Thus $G + \hat{k}'D = \sum G_j$, ∂G consists of an odd number of essential curves (and may have some trivial curves), and $\hat{k}'$ is bounded above by the total number of trivial curves in the boundaries of the G_j.

In particular, we may attach annuli to pairs of essential boundary components of G so that (after a small isotopy) we obtain a properly embedded surface G' with a single essential boundary component and $\chi(G') = \chi(G)$. Note that G' may not be connected. Denote by G'' the component of G' with boundary containing the essential curve. Since the triangulation is 0–efficient and no component of G' is a vertex linking disc, each component of G' has nonpositive Euler characteristic, and hence $\chi(G'') \geq \chi(G')$. If G'' has any boundary components that are trivial, we cap these off with disc, and denote the resulting surface again by G''. This still satisfies $\chi(G'') \geq \chi(G')$.

Similarly $\sum H_n$ has boundary a family of parallel essential curves, and hence $H + \hat{k}''D = \sum H_n$, where $\hat{k}''$ is bounded above by the total number of trivial curves in the boundaries of the H_n. We have $\hat{k}' + \hat{k}'' \leq \hat{k}$, which implies

$$F + (\tilde{k} + \hat{k})D = F + kD = \sum F_i = \sum G_j + \sum H_n = G + H + (\hat{k}' + \hat{k}'')D.$$

If $\tilde{k} > 0$, then $\chi(G'') \geq \chi(G') = \chi(G) = \chi(F) + (\tilde{k} + \hat{k} - \hat{k}' + \hat{k}'') - \chi(H) > \chi(F) = \chi(S)$, contradicting the maximality of the Euler characteristic of S amongst all surface with boundary a single essential curve.

Hence $\tilde{k} = 0$. But then there is a unique surface in $\{F_i\}$ with an essential curve in its boundary. Without loss of generality, assume this is F_1. If F_1 only has one boundary component, then $\chi(F_1) \geq \chi(F)$ and either the weight of F_1 is less than that of F (which would be a contradiction) or $F = F_1$ is Q–fundamental. If F_1 has more than one boundary component, then the other boundary components are trivial, and hence we may cap them off with discs, obtaining a surface F' with $\partial F' = \partial F$ and $\chi(F') > \chi(F) = \chi(S)$, contradicting the maximality of the Euler characteristic of S. Hence $[F]_Q$ is fundamental. $\square$

Corollary 22 *Let M be an orientable compact irreducible 3–manifold with ∂M a single incompressible torus. Suppose $\mathcal{T}$ is a 0–efficient triangulation of M. Then α is a minimising slope for M if and only if there is a Q–fundamental surface F of $\mathcal{T}$ with $[\partial F] = \alpha$ and $\chi(F) = -\|M\|$.*

The above corollary gives an algorithm to compute $\|M\|$ and the set of all minimising slopes from the Q–fundamental solutions. However, in the presence of incompressible and ∂–incompressible surfaces at slopes other than the minimising slopes, we only obtain an upper bound on the norm of any slope if only the Q–fundamental solutions and not all fundamental solutions are computed.

5.2 Crosscap number

This section gives a proof of [Theorem 6]. We organise the proof in three stages. It follows from the previous section that the crosscap number can be computed from the Q–fundamental solutions if a spanning slope is a minimising slope for M. This is immediate in the case where a nonorientable surface

achieves the minimising slope (Corollary 23), and requires a little more effort when all these surfaces are orientable (Proposition 24). The proof is then completed by showing that we can always compute the crosscap number from the Q–fundamental solutions that are spanning surfaces.

Corollary 23 Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathscr{T}$ be a 0–efficient triangulation of M and suppose that the coordinates for a meridian for K on the induced triangulation $\mathscr{T}_\partial$ of ∂M are given.

Suppose that amongst the Q–fundamental surfaces with a single boundary component, the maximal Euler characteristic is achieved by a nonorientable spanning surface S for K. Then $c(K) = 1 - \chi(S)$.

Proof Suppose S_o is an orientable spanning surface of maximal Euler characteristic, and S_n is a nonorientable spanning surface of maximal Euler characteristic. Then $c(K) = \min(1 - \chi(S_n), 2 - \chi(S_o))$.

It follows from Proposition 21 that there is no spanning surface in M of larger Euler characteristic than the surface S in the hypothesis. Hence $\chi(S_o) \leq \chi(S) = \chi(S_n)$, and therefore $c(K) = 1 - \chi(S)$. □

In order to break up the proof of our main theorem, we offer the following improvement to the previous corollary in the context of a 0–efficient and suitable triangulation. This result is an auxiliary step towards our main result Theorem 6.

Proposition 24 Let M be the exterior of a nontrivial knot K in a closed 3–manifold N with $[K] = 0 \in H_1(N; \mathbb{Z}_2)$. Suppose that M is irreducible and contains no embedded nonseparating torus and no embedded Klein bottle. Let $\mathscr{T}$ be a 0–efficient suitable triangulation of M.

Suppose that amongst the Q–fundamental surfaces with a single boundary component, the maximal Euler characteristic is achieved by a spanning surface S for K. Then $c(K) = \min(A', B')$, where

$$A' = \min\{1 - \chi(S) \mid S \text{ is a nonorientable } Q\text{–fundamental spanning surface for } K\},$$
$$B' = \min\{2 - \chi(S) \mid S \text{ is an orientable } Q\text{–fundamental spanning surface for } K\}.$$

Proof Suppose S_o is an orientable spanning surface of maximal Euler characteristic, and S_n is a nonorientable spanning surface of maximal Euler characteristic. The surface S_o (if it exists) is isotopic to a normal surface. Since the triangulation is suitable, the same is true for S_n. We may therefore assume that S_n and S_o are least-weight normal representatives amongst all normal spanning surfaces in the same orientability class and with maximal Euler characteristic.

In the preliminary observation, let F be a normal spanning surface of maximal Euler characteristic. First suppose $\tilde{k} > 0$. We let $\{G_j\}$ be the subset of surfaces in $\{F_i\}$ with boundary curves of the same slope as F, and $\{H_n\}$ be the complementary set. As in the proof of Proposition 21, we write $G + \hat{k}'D = \sum G_j$, where $\hat{k}'$ is bounded above by the total number of trivial curves in the boundaries of the G_j and G does not contain any vertex linking discs. Hence ∂G consists of $\tilde{k}/\mu(\partial F) + 1$ essential curves and some trivial curves.

Similarly $\sum H_n$ has boundary a family of $\tilde{k}/\mu(\partial F)$ parallel essential curves of complementary slope and some trivial curves, and we write $H + \hat{k}'' D = \sum H_n$, where no component of H is a vertex linking disc and $\hat{k}''$ is bounded above by the total number of trivial curves in the boundaries of the H_n. This implies

$$F + (\tilde{k} + \hat{k})D = \sum F_i = \sum G_j + \sum H_n = G + H + (\hat{k}' + \hat{k}'')D.$$

Since $\hat{k} \geq \hat{k}' + \hat{k}''$ and we assume $\tilde{k} > 0$, we have $\chi(G) = \chi(F) + (\tilde{k} + \hat{k} - \hat{k}' - \hat{k}'') - \chi(H) > \chi(F)$, and similarly $\chi(H) > \chi(F)$. Now either G or H has an odd number of essential boundary components, and hence a connected component X with an odd number of essential boundary components. By capping off trivial boundary components of X by discs and connecting essential boundary components in pairs, we obtain a properly embedded surface X' with boundary a single essential simple closed curve, and $\chi(X') \geq \chi(X) > \chi(S)$. It follows that X' is not a spanning surface since F is a maximal Euler characteristic spanning surface. For future reference we note that

$$\chi(X') \geq \chi(X) \geq \chi(H) \geq \chi(F) + (\tilde{k} + \hat{k} - \hat{k}' - \hat{k}'') - \chi(G) \geq \chi(F) + \tilde{k} - \chi(G) \geq \chi(F) + \mu(\partial F) > \chi(F),$$

and $\partial X'$ is a single curve of complementary slope to ∂F. We will analyse this in detail in the proof of Theorem 6.

Note that Proposition 21 implies that there is a Q–fundamental surface Y of larger Euler characteristic than $\chi(F)$ and with a single boundary curve (which is possibly not of complementary slope). The existence of Y contradicts our hypothesis that amongst the Q–fundamental surfaces with a single boundary component, the maximal Euler characteristic is achieved by a spanning surface S for K.

Hence $\tilde{k} = 0$ and

$$F + \hat{k} D = \sum F_i$$

We may assume that the boundary curves of the F_i are pairwise disjoint since they are a single essential curve and a finite number of trivial curves. There is exactly one surface, say F_1, with $\partial F \subseteq \partial F_1$. Now

$$\chi(F_1) = \chi(F) + \hat{k} - \sum_{i \geq 2} F_i \geq \chi(F) + \hat{k}.$$

If $\partial F = \partial F_1$, then F_1 is a spanning surface. Since F is a spanning surface of maximal Euler characteristic, this implies $\hat{k} = 0$ and $\chi(F_1) = \chi(F)$. If ∂F_1 also contains trivial curves, then we may cap these off with discs to obtain a spanning surface F_1' with $\chi(F_1') > \chi(F_1) \geq \chi(F)$, which is a contradiction.

Hence $\hat{k} = 0$ and we have $\partial F_1 = \partial F$ and $\chi(F_1) = \chi(F)$. So for every normal spanning surface F of maximal Euler characteristic

$$F = \sum F_i,$$

where the F_i are Q–fundamental and, without loss of generality, F_1 is a normal spanning surface of maximal Euler characteristic. In particular, for each $i \geq 2$, we have $\chi(F_i) = 0$ and $\partial F_i = \varnothing$.

If $\chi(S_n) > \chi(S_o)$, then F_1 is nonorientable and $c(K) = 1 - \chi(F_1) = A' < B'$.

Similarly if $\chi(S_o) > \chi(S_n)$, then F_1 is orientable and $c(K) = 2 - \chi(F_1) = B' \leq A'$.

Hence suppose $\chi(S_o) = \chi(S_n)$, and let $F = S_n$. Again, if F_1 is nonorientable, then $c(K) = 1 - \chi(F_1) = A'$. Therefore suppose that amongst all Q–fundamental spanning surfaces there is no nonorientable surface with Euler characteristic equal to $\chi(S_n)$. In particular

$$S_n = F_1 + \sum_{i \geq 2} F_i,$$

where F_1 is orientable and $\sum_{i \geq 2} F_i$ is a closed surface of Euler characteristic zero, and hence a union of separating tori. We are therefore in the setting of the proof of Lemma 19. The arguments therein hinge on S_n being of least weight and equalling a Haken sum of the form $F_1 + \sum_{i \geq 2} F_i$, but do not depend on whether the F_i are fundamental. Hence we obtain a contradiction, and there must be a nonorientable Q–fundamental spanning surface F' with Euler characteristic equal to $\chi(S_n)$. Therefore $c(K) = 1 - \chi(F') = A'$ and we are done. □

Proof of Theorem 6 There is only one place in the proof of Proposition 24 where we used the hypothesis that, amongst the Q–fundamental surfaces with a single boundary component, the maximal Euler characteristic is achieved by a spanning surface S for K.

Hence suppose F is a spanning surface of maximal Euler characteristic, and that there is a nonspanning surface X' with a single boundary curve of complementary slope to ∂F and satisfying

$$\chi(X') \geq \chi(F) + \tilde{k} \geq \chi(F) + \mu(\partial F).$$

Let $\gamma = \partial F$ and $\gamma^{\perp} = \partial X'$. Since F is a spanning surface of maximal Euler characteristic, we have

$$\chi(F) \geq \chi(X') - d(\gamma^{\perp}, \mathscr{F}_e)$$

Hence

$$\chi(F) + d(\gamma^{\perp}, \mathscr{F}_e) \geq \chi(X') \geq \chi(F) + \mu(\gamma),$$

and so

(5-1) $$d(\gamma^{\perp}, \mathscr{F}_e) \geq \mu(\gamma).$$

This reduces our proof to a calculation in the Farey tessellation, with the aim of obtaining a contradiction to the above inequality. The boundary slope of F is $\gamma = \mathfrak{m}_g^{2m} \mathfrak{l}_g$ for some $m \in \mathbb{Z}$. The complementary slope $\gamma^{\perp}$ depends on the boundary pattern of the triangulation of ∂M.

There is $p \geq 0$ such that the oriented boundary edges represent the classes $\mathfrak{m}_g$, $\mathfrak{m}_g^p \mathfrak{l}_g$ and $\mathfrak{m}_g^{p+1} \mathfrak{l}_g$. Hence the signed edge weights of γ with the three edges are

$$\langle \mathfrak{m}_g, \gamma \rangle = 1, \quad \langle \mathfrak{m}_g^p \mathfrak{l}_g, \gamma \rangle = p - 2m \quad \text{and} \quad \langle \mathfrak{m}_g^{p+1} \mathfrak{l}_g, \gamma \rangle = p + 1 - 2m.$$

This determines the signed edge weights of γ with respect to the framing, as shown in Figure 10. Using the convention from Figure 5, we can compute the normal arc coordinates of γ from this information. We then compute the normal arc coordinates of $\gamma^{\perp}$, and hence the slope of $\gamma^{\perp}$ with respect to our framing.
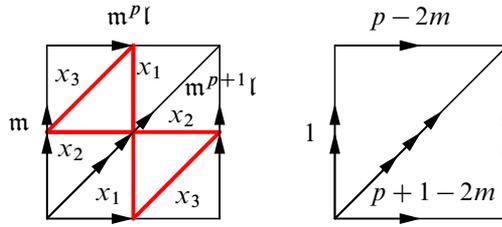
Figure 10: The boundary pattern and arc coordinates, and signed edge weights of γ.

Since $\gamma^\perp$ is not a spanning slope, we show below that $p/1 < \gamma^\perp < (p+1)/1$. Now exactly one of $p/1$ or $(p+1)/1$ is a spanning slope. This implies that we can compute $d(\gamma^\perp, \mathscr{F}_e)$ as the saddle distance of $\gamma^\perp$ to this spanning slope. Our proof is completed by showing that in each case, (5-1) cannot be satisfied.

We remark that at this point one could change the framing to simplify some of the notation, but we choose not to, as it does not simplify the argument.

Case 1 First suppose that $p - 2m \geq 0$. Then the normal arc coordinates of γ are $(p - 2m, 1, 0)$, and so $\mu(\gamma) = \max(1, p - 2m)$.

If $p - 2m = 0$, then the complementary slope has normal arc coordinates $(1, 0, 1)$, and hence $\gamma^\perp = \mathfrak{m}_g^{p+2} \mathfrak{l}_g$ is a spanning slope. This is a contradiction.

If $p - 2m = 1$, then the complementary slope has normal arc coordinates $(0, 0, 1)$, and hence $\gamma^\perp = \mathfrak{m}_g^{p+1} \mathfrak{l}_g$, which again contradicts $\gamma^\perp$ not being the slope of a spanning surface.

Hence $p - 2m > 1$ and so $\mu(\gamma) = p - 2m$. Then the complementary slope has normal arc coordinates $(0, p - 2m - 1, p - 2m)$, and hence satisfies

$$\langle \mathfrak{m}_g, \gamma^\perp \rangle = 2p - 4m - 1, \quad \langle \mathfrak{m}_g^p \mathfrak{l}_g, \gamma^\perp \rangle = 2m - p \quad \text{and} \quad \langle \mathfrak{m}_g^{p+1} \mathfrak{l}_g, \gamma^\perp \rangle = p - 2m - 1,$$

and so (switching to additive notation) we have

$$\gamma^\perp = (2p^2 - 2m(1 + 2p))\mathfrak{m}_g + (2p - 1 - 4m)\mathfrak{l}_g.$$
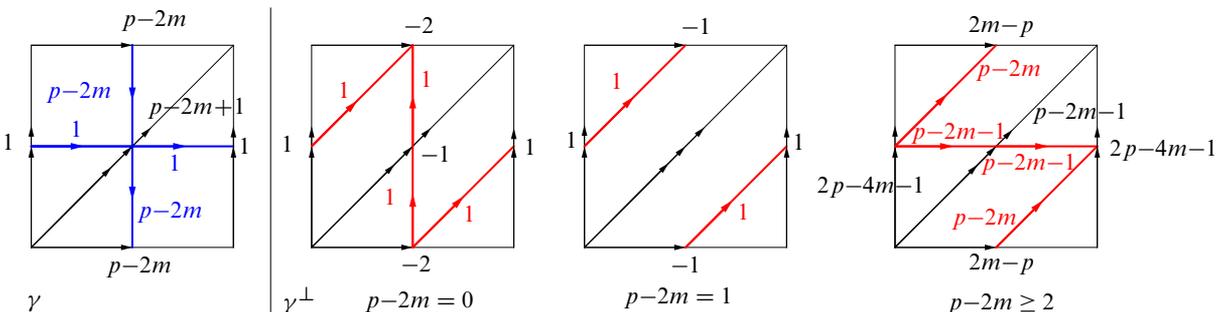
Figure 11: Normal coordinates of γ and its complementary curve in the first case.

To determine the distance to the even slope tree, we flesh out a part of the Farey tessellation. The continued fraction expansion determines a path of edges to $\gamma^\perp$ in the Farey tessellation. We compute:

$$\frac{2p^2 - 2m(1 + 2p)}{2p - 1 - 4m} = p + \frac{1}{2 + 1/(2m - p)} = p + [2, 2m - p].$$

Note that, since $p/1 \oplus (p+1)/1 = (2p+1)/2$ and

$$\det \begin{pmatrix} p & 2p+1 \\ 1 & 2 \end{pmatrix} = -1,$$

there is a triangle τ with vertices $p/1, (2p+1)/2$ and $(p+1)/1$. Let

$$\gamma_j = \frac{p+1}{1} \oplus j \frac{2p+1}{2},$$

where $j \in \{0, \ldots, j_0\}$ and $j_0 = p - 2m - 2$. Note that

$$\frac{p+1}{1} = \gamma_0 > \gamma_1 > \cdots > \gamma_{j_0} > \gamma^\perp > \frac{2p+1}{2} > \frac{p}{1}.$$

We have

$$\gamma_{j_0} = \frac{2p^2 - (2m+1)(1+2p)}{2p - 3 - 4m}.$$

Then since

$$\det \begin{pmatrix} 2p^2 - (2m+1)(1+2p) & 2p+1 \\ 2p - 3 - 4m & 2 \end{pmatrix} = 1$$

and $\gamma_{j_0} \oplus (2p+1)/2 = \gamma^\perp$, there is a triangle τ' in the Farey triangulation with vertices $\gamma_{j_0}, \gamma^\perp$ and $(2p+1)/2$. It follows from the expression for γ_j, that there are $j_0 = p - 2m - 2$ triangles between the triangles τ and τ' with pivot around the common vertex $(2p+1)/2$.
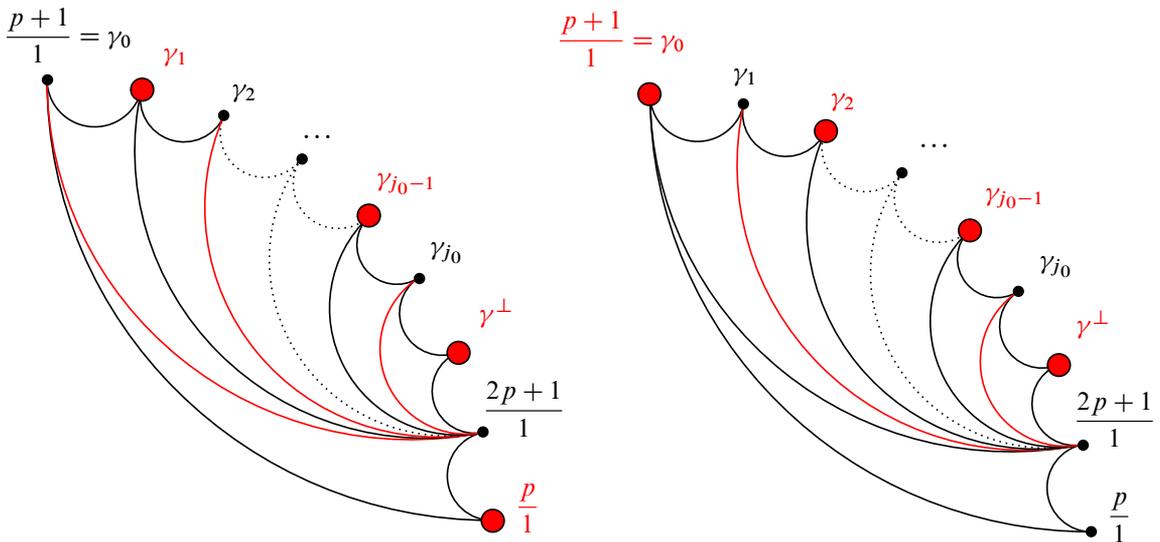
Figure 12: Relevant part of the Farey tessellation in the first case (not drawn to scale) if p is even (left) and if p is odd (right). Even slopes and arcs adding saddles are marked in red.

Figure 13: Normal coordinates of γ and its complementary curve in the second case.

If p is even $d(\gamma^\perp, \mathscr{F}_e) = j_0/2 + 1 = (p - 2m)/2$. Hence

$$\frac{p - 2m}{2} = d(\gamma^\perp, \mathscr{F}_e) \geq \mu(\gamma) = p - 2m.$$

This is impossible since $p - 2m > 1$.

If p is odd $d(\gamma^\perp, \mathscr{F}_e) = (j_0 - 1)/2 + 1 = (p - 2m - 1)/2$. Hence

$$\frac{p - 2m - 1}{2} = d(\gamma^\perp, \mathscr{F}_e) \geq \mu(\gamma) = p - 2m.$$

This is also impossible since $p - 2m > 1$.

Case 2 Suppose that $p - 2m < 0$, or equivalently $2m - p \geq 1$. Then the normal arc coordinates of γ are $(2m - p - 1, 0, 1)$, and so $\mu(\gamma) = \max(1, 2m - p - 1)$.

If $2m - p = 1$ then $\gamma^\perp = \mathfrak{m}_g^{p+1} \mathfrak{l}_g$ is a spanning slope. This is a contradiction.

If $2m - p = 2$ then $\gamma^\perp = \mathfrak{m}_g^{p} \mathfrak{l}_g$ is a spanning slope. This is a contradiction.

Hence $2m - p \geq 3$ and so $\mu(\gamma) = 2m - p - 1$. Then the complementary slope has normal arc coordinates $(0, 2m - p - 1, 2m - p - 2)$, and hence we have

$$\langle \mathfrak{m}_g, \gamma^\perp \rangle = 4m - 2p - 3, \quad \langle \mathfrak{m}_g^{p} \mathfrak{l}_g, \gamma^\perp \rangle = p + 2 - 2m \quad \text{and} \quad \langle \mathfrak{m}_g^{p+1} \mathfrak{l}_g, \gamma^\perp \rangle = 2m - p - 1.$$

This gives (again shown in additive notation)

$$\gamma^\perp = (2m(1 + 2p) - 2(p + 1)^2)\mathfrak{m}_g + (4m - 2p - 3)\mathfrak{l}_g.$$

Now

$$\gamma^\perp = p + 1 + \frac{1}{-2 + 1/(2m - p - 1)} = p + 1 + [-2, 2m - p - 1].$$

Let

$$\gamma_j = \frac{p}{1} \oplus j \frac{2p + 1}{2}$$

for $j \in \{0, \ldots, j_0\}$, where $j_0 = 2m - p - 3$. In particular,

$$\gamma_{j_0} = \frac{4mp + 2m - 6p - 2p^2 - 3}{4m - 2p - 5}$$

and we have

$$\frac{p+1}{1} > \frac{2p+1}{2} > \gamma^\perp > \gamma_{j_0} > \cdots > \gamma_2 > \gamma_1 > \gamma_0 = \frac{p}{1}.$$

We observe that $\gamma^\perp = \gamma_{j_0} \oplus (2p+1)/2$ and that

$$\det \begin{pmatrix} 4mp + 2m - 6p - 2p^2 - 3 & 2p+1 \\ 4m - 2p - 5 & 2 \end{pmatrix} = -1,$$

and so there is a triangle in the Farey tessellation with vertices $\gamma^\perp$, γ_{j_0} and $(2p+1)/2$. As above, this allows us to compute the slope distance of $\gamma^\perp$ to the nearest spanning slope from the triangles pivoting about $(2p+1)/2$.

If p is even $d(\gamma^\perp, \mathcal{F}_e) = (j_0 - 1)/2 + 1 = (2m - p - 2)/2$. Hence

$$(2m - p - 2)/2 = d(\gamma^\perp, \mathcal{F}_e) \geq \mu(\gamma) = 2m - p - 1.$$

This is impossible since $2m - p \geq 3$.

If p is odd $d(\gamma^\perp, \mathcal{F}_e) = j_0/2 + 1 = (2m - p - 1)/2$. Hence

$$\frac{2m - p - 1}{2} = d(\gamma^\perp, \mathcal{F}_e) \geq \mu(\gamma) = 2m - p - 1.$$

This is also impossible since $2m - p \geq 3$.

Since in each case the existence of X' with complementary slope to ∂F gives a contradiction to the maximality of $\chi(F)$ amongst all spanning surfaces, this completes the proof. $\qquad\square$
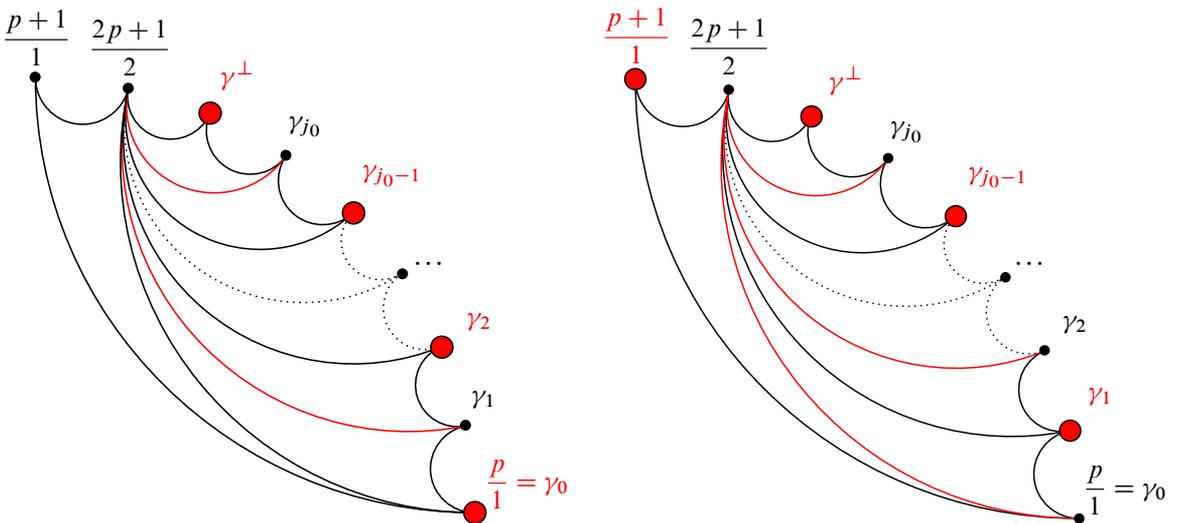
Figure 14: Relevant part of the Farey tessellation in the second case (not drawn to scale) if p is even (left) and if p is odd (right). Even slopes and arcs adding saddles are marked in red.

6 Implementation and computational results

According to the KnotInfo database [Livingston and Moore 2021], crosscap numbers are known for all knots with fewer than 10 crossings due to work of [Burton and Ozlen 2012; Kindred 2020; Murakami and Yasuhara 1995; Hirasawa and Teragaito 2006; Teragaito 2004]. But there are 5 knots with 10, 96 knots with 11, and 668 knots with 12 crossings for which only bounds have been known for the crosscap number.

Here we present 196 crosscap numbers of knots, for which crosscap numbers were previously unknown. This includes all five such 10–crossing knots, 45 of the missing 96 crosscap numbers for 11–crossing knots, and 146 of the missing 668 crosscap numbers for 12–crossing knots. As a result, crosscap numbers for all knots up to and including ten crossings are now known.

6.1 Implementation

Our implementation uses out-of-the-box Regina functions. It is based on Proposition 24, rather than the stronger Theorem 6, because we were also interested in the Euler characteristic of nonspanning surfaces with connected boundary.

Algorithm 25 Compute crosscap numbers using Q–coordinates and Proposition 24.

> Input:
> (1) $\mathscr{T}$, a 0–efficient suitable triangulation of M with boundary $\mathscr{T}_\partial$
> (2) $\mathfrak{m}$, a meridian of K represented by an edge of $\mathscr{T}_\partial$
> compute S_0, the set of Q–fundamental surfaces of $\mathscr{T}$
> compute $S_1 = \{S \in S_0 \mid \partial S \neq \varnothing \text{ is connected and nontrivial}\}$, Q–fundamental surfaces with single essential boundary component
> compute $B' = \min\{2 - \chi(S) \mid S \in S_1 \text{ orientable}, |\partial S \cap \mathfrak{m}| = 1\}$
> compute $A' = \min\{1 - \chi(S) \mid S \in S_1 \text{ nonorientable}, |\partial S \cap \mathfrak{m}| = 1\}$
> compute $N' = \min\{1 - \chi(S) \mid S \in S_1, |\partial S \cap \mathfrak{m}| > 1\}$
> **if** $N' < \min(A', B')$ **then**
> > **return** cannot determine crosscap number
>
> **else**
> > **return** $\min(A', B')$
>
> **end if**

The main computational effort is in Regina's enumeration algorithm for Q–fundamental surfaces [Burton et al. 1999–2024], which in turn runs a Hilbert basis enumeration on a high-dimensional polytope. The verification of the correctness of the input also takes up significant — but smaller amounts of — computational resources. The first verification is the test for 0–efficiency of the triangulation. The second verifies the meridian edge. Here we perform a Dehn surgery along this edge, and then use Regina's 3–sphere recognition routine to check that the resulting 3–manifold is indeed the 3–sphere.

DT	$c(K)$	nOr	or	nSp	DT	$c(K)$	nOr	or	nSp	DT	$c(K)$	nOr	or	nSp
10_{157}	4	−3	−5	−7	10_{159}	4	−3	−5	−4	10_{164}	4	−3	−3	−6
10_{158}	4	−3	−5	−6	10_{163}	4	−3	−5	−6					
$11n_2$	4	−3	−5	−6	$11n_{59}$	4	−3	−5	−5	$11n_{120}$	4	−3	−7	−6
$11n_3$	4	−3	−3	−5	$11n_{75}$	4	−3	−5	−6	$11n_{121}$	4	−3	−5	−6
$11n_4$	4	−3	−5	−6	$11n_{76}$	3	−2	−7	−6	$11n_{123}$	4	−3	−3	−7
$11n_7$	4	−3	−5	−6	$11n_{77}$	4	−3	−7	−5	$11n_{124}$	4	−3	−5	−6
$11n_{11}$	4	−3	−5	−7	$11n_{78}$	3	−2	−7	−5	$11n_{130}$	4	−3	−5	−8
$11n_{22}$	4	−3	−5	−6	$11n_{83}$	4	−3	−3	−6	$11n_{134}$	4	−3	−3	−4
$11n_{25}$	4	−3	−5	−6	$11n_{86}$	4	−3	−5	−6	$11n_{137}$	4	−3	−5	−5
$11n_{29}$	4	−3	−3	−5	$11n_{87}$	4	−3	−5	−7	$11n_{158}$	4	−3	−7	−5
$11n_{33}$	4	−3	−5	−6	$11n_{89}$	4	−3	−5	−5	$11n_{162}$	4	−3	−3	−6
$11n_{39}$	4	−3	−3	−5	$11n_{93}$	4	−3	−5	−5	$11n_{164}$	4	−3	−5	−6
$11n_{45}$	4	−3	−5	−6	$11n_{100}$	4	−3	−3	−5	$11n_{170}$	4	−3	−3	−6
$11n_{47}$	4	−3	−7	−6	$11n_{109}$	4	−3	−5	−6	$11n_{172}$	4	−3	−5	−5
$11n_{52}$	4	−3	−5	−7	$11n_{112}$	4	−3	−5	−6	$11n_{173}$	4	−3	−7	−6
$11n_{54}$	4	−3	−5	−5	$11n_{114}$	4	−3	−3	−7	$11n_{175}$	4	−3	−5	−5
$11n_{55}$	4	−3	−5	−6	$11n_{117}$	3	−2	−3	−4	$11n_{180}$	4	−3	−5	−6

Table 1: New 10– and 11–crossing crosscap numbers.

6.2 Computational results

Supporting material for [Burton and Ozlen 2012] contains a list of triangulations of all knot complements up to 12 crossings that are 0–efficient, with real boundary and one of the boundary edges running parallel to the meridian $\mathfrak{m}$ (ie 0–efficient suitable triangulations). This list is available from the webpage of the first author. Using this list of triangulations, we applied the implementation outlined in Section 6.1.

The results are summarised in Tables 1 and 2. Here "nOr" is the maximal Euler characteristic of a nonorientable Q–fundamental spanning surface, "or" that of an orientable Q–fundamental spanning surface and "nSp" that of a Q–fundamental nonspanning surface with single boundary component. As an additional check, we also ran our algorithm for a larger collection of knots for which computations are feasible.

We ran our computations on a machine with 2×24 Intel Xeon Gold6240R processors and 192GB of memory. Computations were feasible for triangulations of up to 30 tetrahedra (on a standard laptop, triangulations up to 27 tetrahedra can still be handled). We used roughly six months of CPU time to obtain the data in Tables 1 and 2. We only tried Regina's default choice of Hilbert basis algorithm.

6.3 Surfaces realising crosscap number

One interesting aspect of our algorithms is the relationship between our numbers A and B from Theorem 1 and A' and B' from Theorem 6. We have $\min(A, B) = c(K) = \min(A', B')$. Since Q–fundamental surfaces are fundamental surfaces, $A \le A'$ and $B \le B'$.

DT	$c(K)$	nOr	or	nSp	DT	$c(K)$	nOr	or	nSp	DT	$c(K)$	nOr	or	nSp
$12n_1$	5	−4	−5	−6	$12n_{195}$	3	−2	−5	−6	$12n_{423}$	4	−3	−5	−7
$12n_7$	4	−3	−5	−5	$12n_{203}$	4	−3	−5	−5	$12n_{425}$	5	−4	−7	−6
$12n_8$	4	−3	−7	−6	$12n_{210}$	4	−3	−5	−6	$12n_{430}$	4	−3	−3	−4
$12n_{10}$	4	−3	−5	−6	$12n_{211}$	4	−3	−5	−7	$12n_{437}$	4	−3	−5	−6
$12n_{11}$	4	−3	−3	−6	$12n_{215}$	3	−2	−5	−6	$12n_{442}$	4	−3	−3	−5
$12n_{16}$	3	−2	−7	−5	$12n_{220}$	4	−3	−7	−6	$12n_{452}$	4	−3	−3	−6
$12n_{19}$	3	−2	−5	−6	$12n_{225}$	4	−3	−3	−8	$12n_{469}$	3	−2	−5	−6
$12n_{20}$	5	−4	−5	−6	$12n_{229}$	4	−3	−7	−5	$12n_{476}$	4	−3	−5	−5
$12n_{24}$	4	−3	−5	−6	$12n_{230}$	3	−2	−5	−5	$12n_{479}$	4	−3	−3	−7
$12n_{38}$	5	−4	−5	−7	$12n_{237}$	4	−3	−5	−7	$12n_{484}$	4	−3	−5	−6
$12n_{40}$	5	−4	−7	−7	$12n_{241}$	4	−3	−5	−5	$12n_{494}$	4	−3	−7	−5
$12n_{42}$	4	−3	−5	−5	$12n_{247}$	3	−2	−3	−5	$12n_{495}$	4	−3	−5	−7
$12n_{43}$	4	−3	−7	−7	$12n_{257}$	3	−2	−5	−5	$12n_{509}$	4	−3	−7	−5
$12n_{45}$	5	−4	−5	−6	$12n_{261}$	4	−3	−7	−6	$12n_{526}$	4	−3	−7	−7
$12n_{51}$	4	−3	−3	−6	$12n_{271}$	4	−3	−5	−6	$12n_{535}$	4	−3	−3	−6
$12n_{53}$	4	−3	−3	−7	$12n_{274}$	4	−3	−3	−6	$12n_{547}$	4	−3	−3	−7
$12n_{56}$	4	−3	−5	−6	$12n_{276}$	4	−3	−5	−6	$12n_{552}$	3	−2	−5	−6
$12n_{63}$	4	−3	−5	−6	$12n_{278}$	4	−3	−3	−7	$12n_{554}$	4	−3	−3	−7
$12n_{64}$	4	−3	−7	−5	$12n_{279}$	4	−3	−3	−7	$12n_{566}$	4	−3	−3	−5
$12n_{67}$	4	−3	−7	−6	$12n_{280}$	4	−3	−5	−6	$12n_{572}$	4	−3	−5	−5
$12n_{68}$	4	−3	−7	−6	$12n_{285}$	4	−3	−5	−6	$12n_{573}$	4	−3	−5	−5
$12n_{71}$	4	−3	−7	−4	$12n_{290}$	4	−3	−5	−7	$12n_{580}$	4	−3	−3	−6
$12n_{73}$	4	−3	−5	−6	$12n_{304}$	4	−3	−5	−6	$12n_{585}$	4	−3	−5	−6
$12n_{74}$	4	−3	−7	−6	$12n_{308}$	4	−3	−5	−5	$12n_{601}$	4	−3	−5	−6
$12n_{78}$	4	−3	−3	−6	$12n_{311}$	4	−3	−3	−6	$12n_{605}$	4	−3	−7	−6
$12n_{82}$	4	−3	−5	−7	$12n_{312}$	4	−3	−5	−7	$12n_{607}$	3	−2	−5	−6
$12n_{84}$	4	−3	−5	−7	$12n_{324}$	4	−3	−3	−7	$12n_{610}$	4	−3	−7	−7
$12n_{89}$	4	−3	−7	−7	$12n_{327}$	4	−3	−7	−6	$12n_{623}$	4	−3	−7	−6
$12n_{93}$	4	−3	−7	−6	$12n_{331}$	3	−2	−5	−4	$12n_{630}$	4	−3	−5	−5
$12n_{97}$	4	−3	−5	−6	$12n_{334}$	4	−3	−3	−5	$12n_{641}$	4	−3	−7	−6
$12n_{104}$	4	−3	−7	−6	$12n_{341}$	4	−3	−5	−6	$12n_{642}$	4	−3	−3	−4
$12n_{106}$	4	−3	−7	−5	$12n_{342}$	4	−3	−3	−5	$12n_{643}$	3	−2	−5	−7
$12n_{116}$	4	−3	−5	−6	$12n_{343}$	4	−3	−3	−5	$12n_{650}$	4	−3	−3	−5
$12n_{124}$	4	−3	−3	−6	$12n_{345}$	4	−3	−7	−6	$12n_{674}$	4	−3	−7	−6
$12n_{129}$	4	−3	−5	−6	$12n_{354}$	4	−3	−5	−5	$12n_{688}$	4	−3	−9	−6
$12n_{134}$	4	−3	−7	−5	$12n_{355}$	3	−2	−5	−5	$12n_{699}$	4	−3	−3	−6
$12n_{146}$	4	−3	−3	−6	$12n_{359}$	4	−3	−3	−7	$12n_{709}$	4	−3	−7	−5
$12n_{150}$	4	−3	−7	−6	$12n_{360}$	4	−3	−3	−6	$12n_{718}$	4	−3	−5	−6
$12n_{152}$	4	−3	−5	−6	$12n_{362}$	4	−3	−5	−6	$12n_{719}$	4	−3	−5	−6
$12n_{154}$	4	−3	−5	−5	$12n_{366}$	4	−3	−5	−6	$12n_{726}$	4	−3	−3	−5
$12n_{160}$	4	−3	−5	−6	$12n_{377}$	4	−3	−5	−5	$12n_{730}$	4	−3	−5	−6
$12n_{162}$	4	−3	−5	−5	$12n_{379}$	4	−3	−5	−6	$12n_{739}$	3	−2	−7	−5
$12n_{170}$	4	−3	−3	−6	$12n_{381}$	4	−3	−3	−6	$12n_{764}$	4	−3	−5	−5
$12n_{179}$	4	−3	−5	−7	$12n_{383}$	4	−3	−3	−5	$12n_{797}$	4	−3	−3	−6
$12n_{185}$	4	−3	−7	−5	$12n_{388}$	4	−3	−5	−5	$12n_{808}$	4	−3	−5	−5
$12n_{187}$	4	−3	−7	−7	$12n_{390}$	4	−3	−5	−6	$12n_{824}$	4	−3	−5	−7
$12n_{188}$	4	−3	−7	−6	$12n_{397}$	4	−3	−5	−6	$12n_{870}$	3	−2	−5	−5
$12n_{190}$	4	−3	−7	−5	$12n_{407}$	3	−2	−5	−5	$12n_{884}$	4	−3	−3	−7
$12n_{193}$	3	−2	−5	−6	$12n_{418}$	3	−2	−7	−5					

Table 2: New 12–crossing crosscap numbers.

Our computations give rise to the following observations:

(1) **$A' < B'$ and $A' = A \leq B \leq B'$** This is the most common case for small crossing knots. Smallest examples are the trefoil and the figure-8 knot. Moreover, the gap between A' and B' can be arbitrarily large, as can be seen from the torus knots $T(2, 2k + 1)$ for $k \geq 1$: the crosscap number of these knots is 1, whereas the knot genus is k. In other words, we have $A' = A = 1 < 2k + 1 = B \leq B'$.

(2) **$A' = B'$ and $A' = A = B = B'$** Here the crosscap number is realised by a ∂–compressible surface obtained from a minimum-genus Seifert surface with a Möbius band attached, but the existence of a ∂–incompressible nonorientable spanning surface realising it is not excluded. Smallest knots with this property are $7a_6$ (7_4) and $8a_{18}$ (8_3). All such knots must necessarily have genus k and crosscap number $2k + 1$. It is worthwhile to mention that $7a_6$ is known not to admit a ∂–incompressible nonorientable spanning surface realising the crosscap number; see [Ichihara et al. 2002].

(3) **$A' > B'$ and $A' \geq A \geq B = B'$** Interesting examples from our calculations are:

  (a) The case $A' > A = B' = B$ occurred for the 11–crossing knot $11a_{362}$ of genus one and crosscap number three, where $A' = 4 > 3 = A = B = B'$ for the suitable triangulation with Regina isomorphism signature

    ```
    uLLvMPvwMwAMQkcacfgihjmklnnrqstrqrtnkvjhavkbveekgjxfcvp.
    ```

  In standard coordinates, fundamental normal surfaces realising the spanning punctured torus T and a spanning nonorientable surface S of Euler characteristic -2 have complementary boundary slopes, and $\mu(\partial T) = 1 = \mu(\partial S)$. In quadrilateral coordinates $S + D = A + T$, where A is a fundamental boundary parallel annulus with boundary curves parallel to ∂S, and D is the vertex linking disk. In particular, using the notation from Section 5, $\tilde{k} = 1$, $\hat{k} = 0$, $F_i = A$, $G_j = S$ and $\hat{k}' = \hat{k}'' = 0$.

  (b) The case $A' = A > B = B'$ occurred for the 11–crossing knot $11n_{139}$. This knot has genus one and crosscap number three, but $A' = A = 4 > 3 = B = B'$.

(4) For 112 knots we computed A, A', B and B', and observed $B = B'$.

References

[Adams and Kindred 2013] **C Adams**, **T Kindred**, *A classification of spanning surfaces for alternating links*, Algebr. Geom. Topol. 13 (2013) 2967–3007 MR Zbl

[Bachman et al. 2016] **D Bachman**, **R Derby-Talbot**, **E Sedgwick**, *Heegaard structure respects complicated JSJ decompositions*, Math. Ann. 365 (2016) 1137–1154 MR Zbl

[Burton and Ozlen 2012] **B A Burton**, **M Ozlen**, *Computing the crosscap number of a knot using integer programming and normal surfaces*, ACM Trans. Math. Software 39 (2012) art. id. 4 MR Zbl

[Burton and Tillmann 2018] **B A Burton**, **S Tillmann**, *Computing closed essential surfaces in 3–manifolds*, preprint (2018) arXiv 1812.11686

[Burton et al. 1999–2024]   **B A Burton**, **R Budney**, **W Pettersson**, et al., *Regina*: *software for low-dimensional topology* (1999–2024) Available at https://regina-normal.github.io/

[Clark 1978]   **B E Clark**, *Crosscaps and knots*, Int. J. Math. Math. Sci. 1 (1978) 113–123 MR Zbl

[Hatcher 1982]   **A E Hatcher**, *On the boundary curves of incompressible surfaces*, Pacific J. Math. 99 (1982) 373–377 MR Zbl

[Hatcher 2023]   **A E Hatcher**, *Notes on basic* 3*–manifold topology*, preprint (2023) Available at https://pi.math.cornell.edu/~hatcher/3M/3Mfds.pdf

[Hirasawa and Teragaito 2006]   **M Hirasawa**, **M Teragaito**, *Crosscap numbers of 2–bridge knots*, Topology 45 (2006) 513–530 MR Zbl

[Howie 2021]   **J Howie**, *Geography of spanning surfaces*, workshop talk, Banff International Research Station (2021) Available at https://www.birs.ca/events/2021/5-day-workshops/21w5094/videos/watch/202106221610-Howie.html

[Ichihara et al. 2002]   **K Ichihara**, **M Ohtouge**, **M Teragaito**, *Boundary slopes of non-orientable Seifert surfaces for knots*, Topology Appl. 122 (2002) 467–478 MR Zbl

[Ito and Takimura 2018]   **N Ito**, **Y Takimura**, *Crosscap number and knot projections*, Int. J. Math. 29 (2018) art. id. 1850084 MR Zbl

[Ito and Takimura 2020a]   **N Ito**, **Y Takimura**, *Crosscap number of knots and volume bounds*, Int. J. Math. 31 (2020) art. id. 2050111 MR Zbl

[Ito and Takimura 2020b]   **N Ito**, **Y Takimura**, *A lower bound of crosscap numbers of alternating knots*, J. Knot Theory Ramifications 29 (2020) art. id. 1950092 MR Zbl

[Jaco and Oertel 1984]   **W Jaco**, **U Oertel**, *An algorithm to decide if a 3–manifold is a Haken manifold*, Topology 23 (1984) 195–209 MR Zbl

[Jaco and Rubinstein 2003]   **W Jaco**, **J H Rubinstein**, 0*–efficient triangulations of 3–manifolds*, J. Differential Geom. 65 (2003) 61–168 MR Zbl

[Jaco and Sedgwick 2003]   **W Jaco**, **E Sedgwick**, *Decision problems in the space of Dehn fillings*, Topology 42 (2003) 845–906 MR Zbl

[Jaco et al. 2009]   **W Jaco**, **H Rubinstein**, **S Tillmann**, *Minimal triangulations for an infinite family of lens spaces*, J. Topol. 2 (2009) 157–180 MR Zbl

[Jaco et al. 2020a]   **W Jaco**, **H Rubinstein**, **J Spreer**, **S Tillmann**, *On minimal ideal triangulations of cusped hyperbolic 3–manifolds*, J. Topol. 13 (2020) 308–342 MR Zbl

[Jaco et al. 2020b]   **W Jaco**, **J H Rubinstein**, **J Spreer**, **S Tillmann**, $\mathbb{Z}_2$*–Thurston norm and complexity of 3–manifolds, II*, Algebr. Geom. Topol. 20 (2020) 503–529 MR Zbl

[Kalfagianni and Lee 2016]   **E Kalfagianni**, **C R S Lee**, *Crosscap numbers and the Jones polynomial*, Adv. Math. 286 (2016) 308–337 MR Zbl

[Kindred 2020]   **T Kindred**, *Crosscap numbers of alternating knots via unknotting splices*, Int. J. Math. 31 (2020) art. id. 2050057 MR Zbl

[Livingston and Moore 2021]   **C Livingston**, **A H Moore**, *Crosscap number*, from KnotInfo: table of knot invariants (2021) Available at https://knotinfo.math.indiana.edu/descriptions/crosscap_number.html

[Matveev 2007]   **S Matveev**, *Algorithmic topology and classification of 3–manifolds*, 2nd edition, Algor. Computat. Math. 9, Springer (2007) MR Zbl

[Murakami and Yasuhara 1995]   **H Murakami**, **A Yasuhara**, *Crosscap number of a knot*, Pacific J. Math. 171 (1995) 261–273   MR Zbl

[Schubert 1961]   **H Schubert**, *Bestimmung der Primfaktorzerlegung von Verkettungen*, Math. Z. 76 (1961) 116–148   MR Zbl

[Stallings 1961]   **J Stallings**, *On fibering certain 3–manifolds*, from "Topology of 3–manifolds and related topics", Prentice-Hall, Englewood Cliffs, NJ (1961) 95–100   MR Zbl

[Teragaito 2004]   **M Teragaito**, *Crosscap numbers of torus knots*, Topology Appl. 138 (2004) 219–238   MR Zbl

[Tillmann 2008]   **S Tillmann**, *Normal surfaces in topologically finite 3–manifolds*, Enseign. Math. 54 (2008) 329–380   MR Zbl

[Tollefson 1998]   **J L Tollefson**, *Normal surface Q–theory*, Pacific J. Math. 183 (1998) 359–374   MR Zbl

Department of Mathematics, Oklahoma State University
Stillwater, OK, United States

School of Mathematics and Statistics, The University of Melbourne
Melbourne VIC, Australia

School of Mathematics and Statistics, The University of Sydney
Sydney NSW, Australia

School of Mathematics and Statistics, The University of Sydney
Sydney NSW, Australia

jaco@math.okstate.edu,   joachim@unimelb.edu.au,   jonathan.spreer@sydney.edu.au,
stephan.tillmann@sydney.edu.au

A cubical Rips construction

MACARENA ARENAS

Given a finitely presented group Q and a compact special cube complex X with nonelementary hyperbolic fundamental group, we produce a nonelementary torsion-free cocompactly cubulated hyperbolic group Γ that surjects onto Q, with kernel isomorphic to a quotient of $G = \pi_1 X$ and such that $\max\{\mathrm{cd}(G), 2\} \geq \mathrm{cd}(\Gamma) \geq \mathrm{cd}(G) - 1$.

1 Introduction

The Rips exact sequence, first introduced by Rips [1982], is a useful tool for producing examples of groups satisfying combinations of properties that are not obviously compatible. It works by taking as input an arbitrary finitely presented group Q, and producing as output a hyperbolic group Γ that maps onto Q with finitely generated kernel. The "output group" Γ is crafted by adding generators and relations to a presentation of Q in such a way that these relations create enough "noise" in the presentation to ensure hyperbolicity. One can then lift pathological properties of Q to (some subgroup of) Γ. For instance, Rips used his construction to produce the first examples of incoherent hyperbolic groups, hyperbolic groups with unsolvable generalised word problem, hyperbolic groups having finitely generated subgroups whose intersection is not finitely generated, and hyperbolic groups containing infinite ascending chains of r–generated groups.

Our purpose here is to present a new variation of the Rips exact sequence. Our main result is:

Theorem 1.1 (Theorem 4.1) *Let Q be a finitely presented group and G be the fundamental group of a compact special* (in the sense of [Haglund and Wise 2008]) *cube complex X. If G is hyperbolic and nonelementary*, *then there is a short exact sequence*

$$1 \to N \to \Gamma \to Q \to 1,$$

where:

  (i) Γ *is a hyperbolic cocompactly cubulated group.*

 (ii) $N \cong G/K$ *for some $K < G$.*

(iii) $\max\{\mathrm{cd}(G), 2\} \geq \mathrm{cd}(\Gamma) \geq \mathrm{cd}(G) - 1$. *In particular, Γ is torsion-free.*

Remark 1.2 By Agol's theorem [2013], the group Γ obtained in Theorem 4.1 is in fact virtually special.

Many variations of Rips' original construction have been produced over the years by a number of authors, including Arzhantseva and Steenbock [2023], Barnard, Brady and Dani [Barnard et al. 2007], Baumslag, Bridson, Miller and Short [Baumslag et al. 2000], Belegradek and Osin [2008], Bridson and Haefliger [1999], Bumagin and Wise [2005], Haglund and Wise [2008], Ollivier and Wise [2007], and Wise [2003; 1998]. Below is a sample of their corollaries:

- There exist non-Hopfian groups with Kazhdan's property (T) [Ollivier and Wise 2007].

- Every countable group embeds in the outer automorphism group of a group with Kazhdan's property (T) [Ollivier and Wise 2007; Belegradek and Osin 2008].

- Every finitely presented group embeds in the outer automorphism group of a finitely generated residually finite group [Wise 2003].

- There exists an incoherent group that is the fundamental group of a compact negatively curved 2–complex [Wise 1998].

- There exist hyperbolic special groups that contain (nonquasiconvex) nonseparable subgroups [Haglund and Wise 2008].

- Properties (T) and FA are not recursively recognisable among hyperbolic groups [Belegradek and Osin 2008].

The groups in Rips' original constructions are cubulable by [Wise 2004], as are the groups in [Haglund and Wise 2008]; on the other extreme, the groups produced in [Ollivier and Wise 2007] and in [Belegradek and Osin 2008] can have property (T), and so will not be cubulable in general; see [Niblo and Reeves 1997].

A notable limitation of all available Rips-type techniques is that the hyperbolic group Γ surjecting onto Q will have cohomological dimension at most equal to 2. This is unsurprising, since, in a precise sense, "most" hyperbolic groups are 2–dimensional; see [Gromov 1993; Ollivier 2005]. Moreover, examples of hyperbolic groups having large cohomological dimension are scarce: Gromov [1987] conjectured that all constructions of high-dimensional hyperbolic groups must utilise number-theoretic techniques, and later on, Bestvina [2000] made this precise by asking whether for every $K > 0$ there is an $N > 0$ such that all hyperbolic groups of (virtual) cohomological dimension $\geq N$ contain an arithmetic lattice of dimension $\geq K$. Both of these questions have been answered in the negative by work of a number of people, including Mosher and Sageev [1997], Januszkiewicz and Świątkowski [2003], and later on Fujiwara and Manning [2010] and Osajda [2013], but flexible constructions are still difficult to come by.

Theorem 4.1 produces cocompactly cubulated hyperbolic groups containing quotients of arbitrary special hyperbolic groups. While our construction particularises to Rips' original result, it can also produce groups with large cohomological dimension. Thus, it serves to exhibit a collection of examples of hyperbolic groups that is new and largely disjoint from that produced by all other Rips-type theorems.

Most versions of Rips' construction, including the original, rely on some form of small cancellation. This is what imposes a bound on the dimension of the groups thus obtained: group presentations are inherently 2–dimensional objects, and one can prove that the presentation complexes associated to (classical and graphical) small cancellation presentations are aspherical.

We rely instead on *cubical* presentations and *cubical* small cancellation theory, which are intrinsically higher-dimensional. Roughly speaking, cubical small cancellation theory considers higher-dimensional analogues of group presentations: pieces in this setting are overlaps between higher-dimensional subcomplexes, and cubical small cancellation conditions measure these overlaps. This viewpoint allows for the use of nonpositively curved cube complexes and their machinery, and has proved fruitful in many contexts, most notably in Agol's proofs [2013; 2008] of the virtual Haken and virtual fibred conjectures, which build on work of Wise [2021] and his collaborators [Bergeron and Wise 2012; Haglund and Wise 2012; Hsu and Wise 2015].

While many groups have convenient cubical presentations, producing these, or proving that they do satisfy useful cubical small cancellation conditions, is difficult in general. Some examples are discussed in [Wise 2021; Arzhantseva and Hagen 2022; Jankiewicz and Wise 2022]. Other than these, we are not aware of instances where explicit examples of nontrivial cubical small cancellation presentations are given, nor of many results producing families of examples with some given list of properties. This note can be viewed as one such construction, and can be used to produce explicit examples that are of a fundamentally different nature to those already available.

Structure In Section 2 we present the necessary background on hyperbolicity, quasiconvexity and cubical small cancellation theory. In Section 3 we state and prove Theorem 3.2, which is the main technical result, and also state and prove some auxiliary lemmas. In Section 4 we give the proof of Theorem 4.1. Finally, in Section 5 we review some standard material on the cohomological dimension of groups, and analyse the cohomological dimension of Γ.

2 Background

We utilise the following theorem of Arzhantseva [2001]:

Theorem 2.1 *Let G be a nonelementary torsion-free hyperbolic group and H a quasiconvex subgroup of G of infinite index. Then there exist infinitely many $g \in G$ for which the subgroup $\langle H, g \rangle$ is isomorphic to $H * \langle g \rangle$ and is quasiconvex in G.*

Since cyclic subgroups of a hyperbolic group G are necessarily quasiconvex, one can repeatedly apply Theorem 2.1 to produce quasiconvex free subgroups of any finite rank:

Corollary 2.2 *If G is a nonelementary torsion-free hyperbolic group, then for every $n \in \mathbb{N}$ there exists a quasiconvex subgroup $F_n < G$.*

Recall that for a graph B, a subgraph $A \subset B$ is *full* if, whenever vertices $a_1, a_2 \in A$ are joined by an edge e of B, $e \subset A$. In other words, A is the subgraph of B induced by A^0. A map $X \to Y$ between cell complexes is *combinatorial* if it maps cells to cells of the same dimension. An *immersion* is a local injection.

Definition 2.3 A *local isometry* $\varphi \colon Y \to X$ between nonpositively curved cube complexes is a combinatorial map such that, for each $y \in Y^0$ and $x = \varphi(y)$, the induced map $\varphi \colon \mathrm{link}(y) \to \mathrm{link}(x)$ is an injection of a full subgraph.

A more visual way to think about local isometries is the following: an immersion φ is a local isometry if, whenever two edges $\varphi(e)$ and $\varphi(f)$ form the corner of a square in X, we have that e and f already formed the corner of a square in Y.

A key property of local isometries is that they are π_1–injective. It is then natural to ask *which* subgroups of the fundamental group of a nonpositively curved cube complex can be realised by local isometries of compact nonpositively curved cube complexes. In the setting of nonpositively curved cube complexes with hyperbolic fundamental group, one large class of subgroups having this property is that of quasiconvex subgroups. This is proved in [Haglund 2008], and collected in [Wise 2021, Proposition 2.31 and Lemma 2.38], as presented below.

Definition 2.4 A subspace $\widetilde{Y} \subset \widetilde{X}$ is *superconvex* if it is convex and for every bi-infinite geodesic line L, if $L \subset N_r(\widetilde{Y})$ for some $r > 0$ then $L \subset \widetilde{Y}$. A map $Y \to X$ is *superconvex* if the induced map between universal covers $\widetilde{Y} \to \widetilde{X}$ is an embedding onto a superconvex space.

Proposition 2.5 *Let X be a compact nonpositively curved cube complex with $\pi_1 X$ hyperbolic. Let $H < \pi_1 X$ be a quasiconvex subgroup and let $C \subset \widetilde{X}$ be a compact subspace. Then there exists a superconvex H–cocompact subspace $\widetilde{Y} \subset \widetilde{X}$ with $C \subset \widetilde{Y}$.*

Proposition 2.6 *Let X be a compact nonpositively curved cube complex with $\pi_1 X$ hyperbolic. Let $H < \pi_1 X$ be a quasiconvex subgroup. Then there exists a local isometry $Y \to X$ with $\pi_1 Y = H$.*

2.1 Cubical small cancellation theory

Cubical presentations, cubical small cancellation theory, and many related notions were introduced in [Wise 2021]. We recall them below.

Definition 2.7 A *cubical presentation* $\langle X \mid \{Y_i\} \rangle$ consists of a connected nonpositively curved cube complex X together with a collection of local isometries of connected nonpositively curved cube complexes $Y_i \xrightarrow{\varphi_i} X$. In this setting, we shall think of X as a "generator" and of the $Y_i \to X$ as "relators". The fundamental group of a cubical presentation is defined as $\pi_1 X / \langle\!\langle \{\pi_1 Y_i\} \rangle\!\rangle$.

Associated to a cubical presentation $\langle X \mid \{\varphi_i : Y_i \to X\}\rangle$ there is a *coned-off space* X^* obtained from $(X \cup \{Y_i \times [0,1]\})/\{(y_i, 1) \sim \varphi_i(y_i)\}$ by collapsing each $Y_i \times \{0\}$ to a point. By the Seifert–Van Kampen theorem, the group $\pi_1 X / \langle\!\langle \{\pi_1 Y_i\}\rangle\!\rangle$ is isomorphic to $\pi_1 X^*$. Thus, the coned-off space is a presentation complex of sorts for $\langle X \mid \{Y_i\}\rangle$. In practice, when discussing cubical presentations, we often have in mind the coned-off space X^*, rather than the abstract cubical presentation.

Remark 2.8 A group presentation $\langle a_1, \ldots, a_s \mid r_1, \ldots, r_m\rangle$ can be interpreted cubically by letting X be a bouquet of s circles and letting each Y_i map to the path determined by r_i. On the other extreme, for every nonpositively curved cube complex X there is a "free" cubical presentation $X^* = \langle X \mid \ \rangle$ with fundamental group $\pi_1 X = \pi_1 X^*$.

In the cubical setting, there are two types of pieces: *wall-pieces* and *cone-pieces*. Cone-pieces are very much like pieces in the classical sense — they measure overlaps between relators in the presentation. On the other hand, wall-pieces measure the overlaps between cone-cells and *rectangles* (hyperplane carriers) — wall-pieces are always trivial in the classical case, since the square part of X^* coincides with the 1–skeleton of the presentation complex.

Definition 2.9 (elevations) Let $Y \to X$ be a map and $\widehat{X} \to X$ a covering map. An *elevation* $\widehat{Y} \to \widehat{X}$ is a map such that

(i) the composition $\widehat{Y} \to Y \to X$ equals $\widehat{Y} \to \widehat{X} \to X$, and

(ii) assuming all maps involved are basepoint preserving, $\pi_1 \widehat{Y}$ equals the preimage of $\pi_1 \widehat{X}$ in $\pi_1 Y$.

Notation 2.10 In the entirety of this text, a path $\sigma \to X$ is assumed to be a combinatorial path mapping to the 1–skeleton of X.

Definition 2.11 (pieces) Let $\langle X \mid \{Y_i\}\rangle$ be a cubical presentation. An *abstract contiguous cone-piece* of Y_j in Y_i is an intersection $\widetilde{Y}_j \cap \widetilde{Y}_i$ where either $i \neq j$, or where $i = j$ but $\widetilde{Y}_j \neq \widetilde{Y}_i$. A *cone-piece* of Y_j in Y_i is a path $p \to P$ in an abstract contiguous cone-piece of Y_j in Y_i. An *abstract contiguous wall-piece* of Y_i is an intersection $N(H) \cap \widetilde{Y}_i$ where $N(H)$ is the carrier of a hyperplane H that is disjoint from $\widetilde{Y}_i$. To avoid having to deal with empty pieces, we shall assume that H is dual to an edge with an endpoint on $\widetilde{Y}_i$. A *wall-piece* of Y_i is a path $p \to P$ in an abstract contiguous wall-piece of Y_i.[1]

A *piece* is either a cone-piece or a wall-piece.

Remark 2.12 In [Definition 2.11](#), two lifts of a cone Y are considered identical if they differ by an element of $\mathrm{Stab}_{\pi_1 X}(\widetilde{Y})$. This is in keeping with the conventions of classical small cancellation theory, where overlaps between a relator and any of its cyclic permutations are not regarded as pieces. This hypothesis lets us replace relators by their proper powers to achieve good small cancellation conditions in some cases.

---

[1]The "abstract contiguous cone-piece" and "abstract contiguous wall-piece" terminology comes from the fact that it is also a priori necessary to consider "noncontiguous cone-pieces" and "noncontiguous wall-pieces". However, [Wise 2021, Lemma 3.7] shows that one can limit oneself to the analysis of contiguous pieces.

The $C(p)$ and $C'(1/p)$ conditions are now defined as in the classical case (making no distinction between the two types of pieces when counting them). Namely:

Definition 2.13 A cubical presentation X^* satisfies the $C(p)$ *small cancellation condition* if no essential closed path $\sigma \to X^*$ is the concatenation of fewer than p pieces, and the $C'(1/p)$ *condition* if, whenever $\mu \to X^*$ is a piece in an essential closed path $\sigma \to X^*$, we have that $|\mu| < (1/p)|\sigma|$, where $|\mu|$ is the distance between endpoints of $\tilde{\mu} \subset \widetilde{X}$.

As in the classical case, if the fundamental group of X in a cubical presentation $X^* = \langle X \mid \{Y_i\} \rangle$ is hyperbolic, sufficiently good small cancellation conditions lead to hyperbolicity. The following form of [Wise 2021, Theorem 4.7] follows immediately from the fact that a cubical presentation that is $C'(1/\alpha)$ for $\alpha \geq 12$ can be endowed with a nonpositively curved angling rule that satisfies the short innerpaths condition when $\alpha \geq 14$ [Wise 2021, Theorem 3.32 and Lemma 3.70]:

Theorem 2.14 Let X^* be a cubical presentation satisfying the $C'(1/p)$ small cancellation condition for $p \geq \frac{1}{14}$. Suppose $\pi_1 X$ is hyperbolic and X^* is compact. Then $\pi_1 X^*$ is hyperbolic.

Definition 2.15 A collection $\{H_1, \ldots, H_r\}$ of subgroups of a group G is *malnormal* provided that $H_i^g \cap H_j = 1$ unless $i = j$ and $g \in H_i$.

Compactness, malnormality and superconvexity will together guarantee the existence of a uniform bound on the size of both cone-pieces and wall-pieces. This is the content of [Wise 2021, Lemmas 2.40 and 3.52], which we recall below:

Lemma 2.16 Let X be a nonpositively curved cube complex with $\pi_1 X$ hyperbolic. For $1 \leq i \leq r$, let $Y_i \to X$ be a local isometry with Y_i compact, and assume the collection $\{\pi_1 Y_1, \ldots, \pi_1 Y_r\}$ is malnormal. Then there is a uniform upper bound D on the diameters of intersections $g\widetilde{Y}_i \cap h\widetilde{Y}_j$ between distinct $(\pi_1 X)$–translates of their universal covers in $\widetilde{X}$.

Lemma 2.17 Let Y be a superconvex cocompact subcomplex of a CAT(0) cube complex X. There exists $D \geq 0$ such that for each $n \geq 0$, if $I_1 \times I_n \to X$ is a combinatorial strip whose base $0 \times I_n$ lies in Y and such that $d((0,0),(0,n)) \geq D$, then $I_1 \times I_n$ lies in Y.

Recall that a *wallspace* is a set X together with a collection of *walls* $\{W_i\}_{i \in I} = \mathcal{W}$ where $W_i = \{\overleftarrow{W}_i, \overrightarrow{W}_i\}$ and $\overleftarrow{W}_i, \overrightarrow{W}_i \subset X$ for each $i \in I$, and such that

  (i) $\overleftarrow{W}_i \cup \overrightarrow{W}_i = X$ and
  (ii) $\overleftarrow{W}_i \cap \overrightarrow{W}_i = \varnothing$.

Moreover, $\mathcal{W}$ satisfies a *finiteness property*: for every $p, q \in X$, the number of walls separating p and q, denoted by $\#_{\mathcal{W}}(p, q)$, is finite. The $\overleftarrow{W}_i$ and $\overrightarrow{W}_i$ above are the *half-spaces* of W_i.

Once we have specified a cubical presentation, we will cubulate its fundamental group $\pi_1 X^*$ via Sageev's construction, which produces a CAT(0) cube complex that is dual to a wallspace. We will assume the reader is familiar with this procedure. Good references include [Sageev 1995; Chatterji and Niblo 2005; Bestvina et al. 2014]. Cocompactness of the action on the dual cube complex will readily follow from Proposition 2.18, which is a well-known result of Sageev [1997]. Properness is more delicate, and will follow from Theorem 2.27 once we know that $\pi_1 X^*$ is hyperbolic, since in that case $\pi_1 X^*$ has no infinite torsion subgroups.

Proposition 2.18 *Let G be hyperbolic and $\{H_1, \ldots, H_n\}$ be a collection of quasiconvex subgroups. Then the action of G on the dual CAT(0) cube complex C is cocompact.*

Before stating Theorem 2.27, we need some definitions:

Definition 2.19 Let $Y \to X$ be a local isometry. $\mathrm{Aut}_X(Y)$ is the group of automorphisms $\psi \colon Y \to Y$ such that the diagram below is commutative:

$$Y \xrightarrow{\ \psi\ } Y$$
$$\searrow \quad \downarrow$$
$$X$$

If Y is simply connected, then $\mathrm{Aut}_X(Y)$ is equal to $\mathrm{Stab}_{\pi_1 X}(Y)$.

Definition 2.20 A cubical presentation $\langle X \mid Y_i \rangle$ satisfies the $B(6)$ condition if it satisfies:

(i) **Small cancellation** $\langle X \mid Y_i \rangle$ satisfies the $C'(1/\alpha)$ condition for $\alpha \geq 14$.

(ii) **Wallspace cones** Each Y_i is a wallspace where each wall in Y_i is the union $\bigsqcup U_j$ of a collection of disjoint embedded 2–sided hyperplanes in Y_i, and there is an embedding $\bigsqcup N(U_j) \to Y_i$ of the disjoint union of their carriers into Y_i. Each such collection separates Y_i. Each hyperplane in Y_i lies in a unique wall.

(iii) **Hyperplane convexity** If $P \to Y_i$ is a path that starts and ends on vertices on 1–cells dual to a hyperplane U of Y_i and P is the concatenation of at most seven pieces, then P is path homotopic in Y_i to a path $P \to N(U) \to Y_i$.

(iv) **Wall convexity** Let S be a path in Y_i that starts and ends with 1–cells dual to the same wall of Y_i. If S is the concatenation of at most seven pieces, then S is path-homotopic into the carrier of a hyperplane of that wall.

(v) **Equivariance** The wallspace structure on each cone Y is preserved by $\mathrm{Aut}_X(Y)$.

Historical Remark 2.21 In the setting of classical small cancellation theory, the $B(2n)$ condition was defined in [Wise 2004]. Specifically, the "classical" $B(2n)$ condition states that for each 2–cell R in a 2–complex X, and for each path $S \to \partial R$ which is the concatenation of at most n pieces in X, we have $|S| \leq \frac{1}{2}|\partial R|$. The classical $B(2n)$ condition is intermediate to the $C'(1/(2n))$ and $C(2n)$ conditions in

the sense that $C'(1/(2n)) \implies B(2n) \implies C(2n)$. While not a perfect parallel with the notion considered here, this notation is meant to suggest the fact that, in the classical setting, the $B(6)$ condition is sufficient to guarantee the existence of a wallspace structure on X that leads to cocompact cubulability.

The $B(6)$ condition is extremely useful because it facilitates producing a wallspace structure on the coned-off space X^* by starting *only* with a wallspace structure (satisfying some extra conditions) on each of the cones. This is done by defining an equivalence relation $\sim$ on the hyperplanes of $\widetilde{X}^*$ as explained below.

Definition 2.22 Let U and U' be hyperplanes in $\widetilde{X}^*$. Then $U \sim U'$ provided that, for some translate of some cone Y_i in $\widetilde{X}^*$, the intersections $U \cap Y_i$ and $U' \cap Y_i$ lie in the same wall of Y_i. A *wall* of $\widetilde{X}^*$ is a collection of hyperplanes of $\widetilde{X}^*$ corresponding to an equivalence class.

That this equivalence relation does in fact define a wallspace structure on X^* when the $B(6)$ condition is satisfied is the content of [Wise 2021, Section 5.f].

Definition 2.23 A hyperplane U is *m–proximate* to a 0–cube v if there is a path $P = P_1, \ldots, P_m$ such that each P_i is either a single edge or a piece, v is the initial vertex of P_1 and U is dual to an edge in P_m. A wall is *m–proximate* to v if it has a hyperplane that is m–proximate to v. A hyperplane is *m–far* from a 0–cube if it is not m'–proximate to it for any $m' \leq m$.

Definition 2.24 A hyperplane U of a cone Y is *piecefully convex* if, for any path $\tau\rho \to Y$ with endpoints on $N(U)$, if τ is a geodesic and ρ is trivial or lies in a piece of Y containing an edge dual to U, then $\tau\rho$ is path-homotopic in Y to a path $\mu \to N(U)$.

The following is noted in [Wise 2021, Remark 5.43]. We write $\widetilde{N}(U) := \widetilde{N(U)}$.

Proposition 2.25 *Let K be the maximal diameter of any piece of Y_i in X^*. Then a hyperplane U of Y_i is piecefully convex provided that its carrier $N(U)$ satisfies $d_{\widetilde{Y}_i}(g\widetilde{N}(U), \widetilde{N}(U)) > K$ for any translate $g\widetilde{N}(U) \neq \widetilde{N}(U) \subset \widetilde{Y}_i$.*

Definition 2.26 (cut by a wall) Let $g \in G$ be an element acting on $\widetilde{X}$. An *axis* $\mathbb{R}_g$ for g is a g–invariant copy of $\mathbb{R}$ in $\widetilde{X}$. An element g is *cut by a wall* W if $g^n W \cap \mathbb{R}_g = \{n\}$ for all $n \in \mathbb{Z}$.

The theorem below is a restatement of [Wise 2021, Theorem 5.44] with [Wise 2021, Corollary 5.45], and the fact that the short innerpaths condition is satisfied when $C'(1/\alpha)$ holds for $\alpha \geq 14$.

Theorem 2.27 *Suppose $X^* = \langle X \mid \{Y_i\} \rangle$ satisfies the following hypotheses*:

(i) X^* *satisfies the $B(6)$ condition.*

(ii) *Each hyperplane U of each cone Y_i is piecefully convex.*

(iii) *Let $k \to Y \in \{Y_i\}$ be a geodesic with endpoints p and q. Let U_1 and U_1' be distinct hyperplanes in the same wall w_1 of Y. Suppose k traverses a 1–cell dual to U_1, and either U_1' is 1–proximate to q*

or k traverses a 1–cell dual to U_1'. Then there is a wall w_2 in Y that separates p and q but is not 2–proximate to p or q.

(iv) *Each infinite-order element of* $\mathrm{Aut}(Y_i)$ *is cut by a wall.*

Then the action of $\pi_1 X^*$ on C has torsion stabilisers.

3 Cubical noise

In the classical setting, there are two essentially distinct strategies for producing group presentations satisfying good small cancellation conditions: taking large enough powers of the relators, and adding "noise" to the presentation by multiplying each relator by a sufficiently long, suitably chosen word. In the cubical setting, taking powers of relators translates to taking finite-degree covers of the cycles that represent the relators, and this method generalises to taking finite-degree covers of cube complexes with more complicated fundamental groups. This is the line of enquiry that has been most explored, and for which there exist useful theorems producing cubical small cancellation. This will, however, not be suitable for our applications, because once a cubical presentation $\langle X \mid \{\widehat{Y}_i \to X\} \rangle$ has been obtained by taking covers, any modifications of the cones (other than taking further covers) will dramatically affect the size of pieces, and possibly invalidate whatever small cancellation conclusions had been attained. Thus, we instead prove a cubical small cancellation theorem that builds on the idea of adding noise to a presentation. The procedure we describe will be more stable, in the sense that slightly perturbing the choice of cones will not affect the small cancellation conclusions.

Remark 3.1 We state and prove Theorem 3.2 in more generality than is needed for later applications. In practice the reader may take Y to be a bouquet of finitely many circles, as this is all that is required for the proof of Theorem 4.1. It is also worth noting that while the statement of Theorem 3.2(iii) requires that X and Y be special, the proof only uses that hyperplanes are embedded and 2–sided.

Theorem 3.2 *Let* X *and* Y *be compact nonpositively curved cube complexes with hyperbolic fundamental groups and let* $\mathcal{H}$ *be the set of hyperplanes of* X. *Let* $\{H_1, \ldots, H_r\}$ *be a malnormal collection of free nonabelian quasiconvex subgroups of* $\pi_1 X$, *and suppose that* $H_i \cap \mathrm{Stab}(\widetilde{U})$ *is trivial or equal to* H_i *for all* $U \in \mathcal{H}$. *Let* $\gamma_1 \to X \vee Y, \ldots, \gamma_r \to X \vee Y$ *be closed essential paths based at the wedge point and let* $y_1, \ldots, y_r$ *be the words in* $\pi_1 X * \pi_1 Y$ *represented by the* γ_i. *Then for each* $\alpha \geq 1$ *there are cyclic subgroups* $\langle w_i \rangle \subset H_i * \langle y_i \rangle$ *such that* $w_i = w_i' y_i$ *where* $w_i' \in H_i$ *for each* $i \in \{1, \ldots, r\}$ *and:*

(i) *The group* $\pi_1 X * \pi_1 Y / \langle\!\langle w_1, \ldots, w_r \rangle\!\rangle$ *has a cubical presentation satisfying the* $C'(1/\alpha)$ *condition.*

(ii) *If* $\alpha \geq 14$, *then the group* $\pi_1 X * \pi_1 Y / \langle\!\langle w_1, \ldots, w_r \rangle\!\rangle$ *is hyperbolic.*

(iii) *If* X *and* Y *are special, there is an* $\alpha_0 \geq 14$ *such that if* $\alpha \geq \alpha_0$, *then* $\pi_1 X * \pi_1 Y / \langle\!\langle w_1, \ldots, w_r \rangle\!\rangle$ *acts properly and cocompactly on a* $\mathrm{CAT}(0)$ *cube complex.*

Remark 3.3 The reader might wish to compare Theorem 3.2 with [Wise 2021, Corollary 5.48], which is the analogous result for finite-degree coverings, and whose proof informs the proof below.

Proof Obtaining small cancellation Since $H_1, \ldots, H_r$ are quasiconvex subgroups of $\pi_1 X * \pi_1 Y$ and $\pi_1 X * \pi_1 Y$ is hyperbolic, by Proposition 2.6 there are based local isometries $C_1 \to X \vee Y, \ldots, C_r \to X \vee Y$ of superconvex subcomplexes with $\pi_1 C_i \cong H_i$ for each $i \in \{1, \ldots, r\}$. So there is a cubical presentation $(X \vee Y)^* = \langle X \vee Y \mid \{C_i \to X \vee Y\}_{i=1}^r \rangle$ with fundamental group $\pi_1 X * \pi_1 Y / \langle\!\langle \pi_1 C_1, \ldots, \pi_1 C_r \rangle\!\rangle$. By Lemmas 2.16 and 2.17, malnormality of $\{H_1, \ldots, H_r\}$ and superconvexity of the $\{C_1, \ldots, C_r\}$ ensures that there is a uniform upper bound K on the diameter of pieces.

By hyperbolicity, any cyclic subgroup of $\pi_1 X * \pi_1 Y$ is quasiconvex. So for any choice of cyclic subgroups $\langle w_i \rangle < \pi_1 X * \pi_1 Y$ with $i \in \{1, \ldots, r\}$ there are local isometries $W_i \to X \vee Y$ with $\pi_1 W_i \cong \langle w_i \rangle$.

Let $\alpha \geq 1$, and choose each $\sigma_i \to X \vee Y$ so that $\sigma_i = \sigma_i' \gamma_i$ where

(i) σ_i' is a based closed path in $C_i \subset X$,

(ii) σ_i' is not a proper power and does not contain subpaths of length $\geq K\alpha$ that are proper powers,

(iii) $\sigma_i' \gamma_i$ does not have any backtracks, and

(iv) the W_i corresponding to $\langle w_i \rangle = \langle \sigma_i \rangle$ has diameter $\|W_i\| \geq K\alpha^2$.

For instance, one can choose σ_i' to be of the form $\sigma_i' = \lambda_1 \lambda_2 \lambda_1 \lambda_2^2 \cdots \lambda_1 \lambda_2^{K\alpha}$ where λ_1 and λ_2 are paths representing distinct generators of the corresponding $H_i < \pi_1 X$. Without loss of generality, we can assume that λ_1, λ_2 and $\gamma_1, \ldots, \gamma_r$ have minimal length in their homotopy classes, and therefore that none of the λ_i or γ_i have any backtracks, so any backtracks in $\sigma_i' \gamma_i$ arise from cancellation between σ_i' and γ_i. If any cancellation happens, we can rechoose λ_1 and λ_2 to eliminate it (for instance, by shortening λ_1 and λ_2).

Pieces in each C_i have size bounded by K, and each $W_i \to X \vee Y$ factors through the corresponding C_i, so the size of pieces between different cone-cells or between cone-cells and rectangles is bounded by K; the size of pieces between a cone-cell and itself is bounded by $K\alpha - 1$ and $\|W_i\| \geq (K\alpha)! + K\alpha \geq K\alpha^2$, so $\langle X \vee Y \mid \{W_i \to X \vee Y\}_{i=1}^r \rangle$ satisfies the $C'(1/\alpha)$ condition.

Obtaining hyperbolicity As explained above, the $\langle w_i \rangle$ can be chosen so that

$$(X \vee Y)^* = \langle X \vee Y \mid \{W_i \to X \vee Y\}_{i=1}^r \rangle$$

satisfies the $C'(1/\alpha)$ condition for $\alpha \geq 14$. Since $\pi_1 X * \pi_1 Y$ is hyperbolic and $(X \vee Y)^*$ is compact, Theorem 2.14 then implies that $(X \vee Y)^*$ is hyperbolic.

Obtaining cocompact cubulability Define a wallspace structure on $(X \vee Y)^*$ as follows. By subdividing $X \vee Y$, we may assume that each W_i has an even number of hyperplanes cutting the generator of $\langle w_i \rangle$. The specialness hypothesis ensures that all hyperplanes of $X \vee Y$ and of each W_i are embedded and 2–sided. Moreover, since each W_i has cyclic fundamental group and $H_i \cap \mathrm{Stab}(\widetilde{U})$ is trivial or equal to $\mathrm{Stab}(\widetilde{U})$ for each $\widetilde{U}$, all the hyperplanes of each W_i are contractible or have the homotopy type of a circle representing the generator of $\pi_1 W_i$. Hence, we can define a wallspace structure on each of the cones by defining a wall to be either a single hyperplane if the hyperplane does not cut the generator of the corresponding $\langle w_i \rangle$, or by defining a wall to be an equivalence class consisting of two antipodal
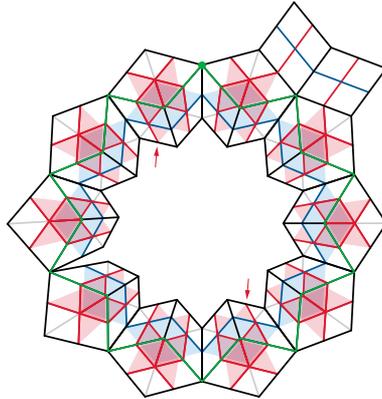
Figure 1: A potential cone-cell and its hyperplanes. The generator of its fundamental group is drawn in green, the hyperplanes that cross it are drawn in red, and the hyperplanes that do not cross it are drawn in blue. A pair of antipodal hyperplanes is indicated.

hyperplanes cutting the generator. Concretely, if the generator of $\langle w_i \rangle$ is a cycle $\sigma \to X \vee Y$ of length $2n$, then letting $\sigma = e_1 \cdots e_{2n}$, hyperplanes U and U' are in the same equivalence class if and only if U is dual to e_j and U' is dual to $e_{j+n} \pmod{n}$ for some $j \in \{1, \ldots, 2n\}$. These choices are exemplified in Figure 1.

We now check condition (i) of Theorem 2.27, which will allow us to extend the wallspace structure on the cones to a wallspace structure for $(X \vee Y)^*$ using the equivalence relation in Definition 2.22.

Choose the $\langle w_i \rangle$ so that $(X \vee Y)^* = \langle X \vee Y \mid \{W_i \to X \vee Y\}_{i=1}^r \rangle$ satisfies the $C'(1/\alpha)$ condition for $\alpha \geq 14$. With the choice of walls described above, each cone is a wallspace satisfying condition (ii) of Definition 2.20. The $C'\left(\frac{1}{14}\right)$ condition is also sufficient to ensure that condition (iii) is met. Indeed, the only way for a path with endpoints on the carrier of a hyperplane U to not be homotopic into the carrier of the hyperplane is if the path is homotopic into a power of the generator of W_i, and such a path would have to traverse at least 14 pieces. Moreover, condition (iv) is met by rechoosing the cyclic subgroups so that the cubical presentation satisfies $C'\left(\frac{1}{16}\right)$. To wit, since pairs (U, U') of hyperplanes lying on the same wall W are antipodal and $(X \vee Y)^*$ satisfies $C'(1/\alpha)$, the number of pieces in a path $\sigma \to C_i$ with endpoints on distinct hyperplanes of W is at least $\frac{1}{2}\alpha$, and so choosing $\alpha \geq 16$ ensures that such a path traverses at least 8 pieces. The choice of wallspace on each cone also ensures that condition (v) is met: any automorphism of $X \vee Y$ sends a wall not cutting a generator to a wall not cutting a generator, and a wall cutting a generator to a wall cutting a generator.

Thus, condition (i) of Theorem 2.27 is satisfied. Since each wall arises from a quasiconvex subgroup, Proposition 2.18 ensures cocompactness of the action on the dual cube complex. To ensure properness of the action, we check the rest of the conditions of Theorem 2.27.

Similar modifications to X^* will ensure that conditions (ii) and (iii) of Theorem 2.27 are met. For condition (ii), by Proposition 2.25, it suffices to ensure that $d_{\widetilde{W}_i}(g\widetilde{N}(U), \widetilde{N}(U)) > K$ for any translate $g\widetilde{N}(U) \neq \widetilde{N}(U) \subset \widetilde{W}_i$. Since each piece of W_i contains at least one edge, this can be guaranteed by

rechoosing the w_i so that X^* satisfies the $C'(1/K')$ condition, where $K' = \max\{K, 16\}$. Condition (iii) also follows, because any two hyperplanes in the same wall are at least 8–far, so there is a hyperplane V crossing the generator of $\langle w_i \rangle$ that is 2–far from both U and U', and one can ensure that the antipodal hyperplane V' is also 2–far from both U and U' by rechoosing X^* so that it satisfies the $C'(1/K'')$ condition, where $K'' = \max\{2K, 16\}$.

Finally, the choice of walls implies that condition (iv) also holds: since $\pi_1(W_i)$ is cyclic for each $i \in I$, every element $g \in \mathrm{Aut}(W_i)$ has an axis, which is cut by a wall of X^* crossing the generator of $\pi_1(W_i)$. □

Definition 3.4 (height) The *height* of $H \leq G$ is the maximal $n \in \mathbb{N}$ such that there exist distinct cosets $g_1 H, \ldots, g_n H$ for which $H^{g_1} \cap \cdots \cap H^{g_n}$ is infinite.

In [Gitik et al. 1998] it was proven that:

Theorem 3.5 *Quasiconvex subgroups of hyperbolic groups have finite height.*

Definition 3.6 The *commensurator* $C_G(H)$ of a subgroup H of G is the set

$$C_G(H) = \{g \in G \mid [H : H^g \cap H] < \infty \text{ and } [H^g : H^g \cap H] < \infty\}.$$

Remark 3.7 If G is hyperbolic and the subgroup H is infinite and quasiconvex, then $[C_G(H) : H] < \infty$ by [Kapovich and Short 1996]. In particular, $C_G(H)$ is also a quasiconvex subgroup of G, and if G is torsion-free and H is free and nonabelian then so is $C_G(H)$.

The following result will be used in the proof of Lemma 3.9:

Lemma 3.8 [Wise 2021, Lemma 8.6] *Let G be hyperbolic and torsion-free and let $H_1, \ldots, H_r$ be a collection of quasiconvex subgroups of G. Let $K_1, \ldots, K_s$ be representatives of the finitely many distinct conjugacy classes of subgroups consisting of intersections of collections of distinct conjugates of $H_1, \ldots, H_r$ in G that are maximal with respect to having infinite intersection. Then $\{C_G(K_1), \ldots, C_G(K_s)\}$ is a malnormal collection of subgroups of G.*

The ensuing lemma surely exists in some form in the literature, but we include a proof for completeness:

Lemma 3.9 *Let G be a nonelementary torsion-free hyperbolic group. For every $k \in \mathbb{N}$, G contains a malnormal collection $\{H_1, \ldots, H_k\}$ of infinite-index quasiconvex nonabelian free subgroups.*

Before proving Lemma 3.9, we make a few observations about malnormal subgroups of free groups. Recall that a subgroup $H < G$ is *isolated* if $g \in H$ whenever $g^n \in H$ for $g \in G$; a subgroup $H < G$ is *malnormal on generators* if $a_i^g \notin H$ for any $i \in \{1, \ldots, n\}$, for any generating set $\{a_1, \ldots, a_n\}$ for H, $g \in G$ and $g \notin H$.

Lemma 3.10 [Fine et al. 2002, Lemma 1] *Let F be a free group and $H \subset F$ a 2–generator subgroup. Then H is malnormal if and only if H is isolated and malnormal on generators.*

Claim 3.11 *Let F be a finite-rank nonabelian free group and let $\{h_1, \ldots, h_k\}$ be a finite collection of nontrivial elements of F. Then there is a subgroup $J < F$ for which $\{J, \langle h_i \rangle\}$ is a malnormal collection for each $i \in \{1, \ldots, k\}$.*

Proof Assume that a basis is given for F, and, abusing notation, also write $h_1, \ldots, h_k$ to denote reduced words on the basis representing the finite set of elements $h_1, \ldots, h_k$. We may also assume that $h_i \neq h_j^m$ whenever $m \in \mathbb{Z} - \{0\}$ and $i \neq j$. Let $J = \langle a, b \rangle$ where $a = h_1 \beta_1 \cdots h_k \beta_k$ and $b = \beta_1' h_1 \cdots \beta_k' h_k$, and where each β_i and β_i' is a reduced word on the basis satisfying that, for each $i \in \{1, \ldots, k\}$:

(i) $\beta_i \neq \beta_j^m$, $\beta_i' \neq (\beta_j')^m$ and $\beta_i \neq (\beta_j')^m$ whenever $i \neq j$ and $m \in \mathbb{Z} - \{0\}$.

(ii) No β_i or β_i' is a product of h_i or their inverses.

(iii) For each β_i, its first letter is not equal to the last letter of h_i, and its last letter is not equal to the first letter of h_{i+1} (modulo k). Similarly, for each β_i', its last letter is not equal to the first letter of h_i, and its first letter is not equal to the last letter of h_{i+1} (modulo k). Moreover, the last letter of β_k is not equal to the inverse of the last letter of h_k, the first letter of β_1' is not equal to the inverse of the first letter of h_1, and the last letter of β_k is not equal to the first letter of β_1'.

As there is no cancellation, J is a rank-2 free group. By Lemma 3.10, to prove that J is malnormal it suffices to show that J is isolated and malnormal on generators. The choice of the β_i and β_i' implies in particular that a and b are not proper powers, and this implies in turn that J is isolated, since F is free. We now show that J is malnormal on generators. Consider a conjugate $a^g = (h_1 \beta_1 \cdots h_k \beta_k)^g$ where $g \notin J$ (the case of b^g is analogous). Since $g \notin J$, g cannot be written as a nontrivial product of powers of a and b and their inverses. If g cannot be written as a subword of a product of a's and b's and their inverses, then a^g cannot be an element of J as there will be no cancellation. The choice of a and b implies that no cyclic permutation of a, b, their product or their proper powers lies in J, so no conjugate of a by a subword of a product of a's and b's and their inverses lies in J.

Finally, consider a conjugate h_i^g of h_i. If $\langle h_i^g \rangle \cap J$ is nontrivial, then since F is free, it follows that g must be a subword of some $j \in J$, and even in this case h_i^g can only be a (nontrivial) cyclic permutation of a, b, their product or their proper powers, but no such cyclic permutation is an element of J, so J intersects all conjugates of the h_i trivially. $\qquad \square$

Proof of Lemma 3.9 It suffices to show that G contains a malnormal quasiconvex free subgroup J of arbitrarily high rank, for then if $J = \langle a_1, \ldots, a_k, b_1, \ldots, b_k \rangle$, the collection $\{H_1, \ldots, H_k\}$ where $H_i = \langle a_i, b_i \rangle$ is also malnormal and quasiconvex. For this, it suffices to show that G contains a malnormal quasiconvex free group J of *some* rank ≥ 2. Indeed, for any $n \in \mathbb{Z}$, J contains infinitely many subgroups of rank n, all of which are malnormal and quasiconvex.

By Theorem 2.1, G contains a free nonabelian quasiconvex subgroup J_0; we may assume further that J_0 has infinite index in G. Let $\mathcal{J}_0$ be the lattice of infinite intersections of conjugates of J_0. This lattice is

finite by Theorem 3.5. If $\mathcal{J}_0$ contains a nonabelian free group J_1 then we replace J_0 with J_1, and we can repeat this process a finite number of times until we either reach a maximal intersection of conjugates of some J_i that is itself free nonabelian or until all subgroups in the lattice $\mathcal{J}_i$ are cyclic. In the former case, the commensurator $C_G(J_i)$ is malnormal and quasiconvex by Lemma 3.8. In the latter case, by Claim 3.11, J_i contains a free nonabelian subgroup J that forms a malnormal collection with each of these cyclic subgroups, and hence J is malnormal in G. □

Lemma 3.9 can be improved to control intersections with quasiconvex subgroups:

Corollary 3.12 *Let G be a nonelementary torsion-free hyperbolic group and let $\{S_1, \ldots, S_l\}$ be a collection of quasiconvex subgroups of G. Then the collection $\{H_1, \ldots, H_k\}$ from the conclusion of Lemma 3.9 can be chosen so that $H_i \cap S_j$ is either trivial or equal to H_i for each $i \in \{1, \ldots, k\}$ and each $j \in \{1, \ldots, l\}$.*

Remark 3.13 In particular, if G is the fundamental group of a compact nonpositively curved cube complex X, and $\mathcal{H}$ is the set of hyperplanes of X, then $\{H_1, \ldots, H_k\}$ can be chosen so that $H_i \cap \mathrm{Stab}(\widetilde{U})$ is either trivial or equal to H_i for each $U \in \mathcal{H}$ and each $i \in \{1, \ldots, k\}$. Indeed, since X is compact, it has finitely many hyperplanes, and hence finitely many hyperplane stabilisers. Each hyperplane stabiliser is quasi-isometrically embedded and $\pi_1 X$ is hyperbolic, so each hyperplane stabiliser is quasiconvex.

Proof of Corollary 3.12 It suffices to prove the result for a single malnormal quasiconvex nonabelian subgroup $J \leq G$. Indeed, as explained in the first paragraph of the proof of Lemma 3.9, the subgroups in the malnormal collection $\{H_1, \ldots, H_k\}$ are produced as subgroups of a single nonabelian free subgroup J, so if we ensure that $J \cap S_j$ is either trivial or equal to J for each $j \in \{1, \ldots, l\}$, then this will also be the case for $\{H_1, \ldots, H_k\}$.

We now proceed by induction on l. Assume that $l = 1$ and consider the intersection $J \cap S_1$, where J is as provided in Lemma 3.9. If this intersection is trivial, then there is nothing to show, so suppose that $K_1 := J \cap S_1$ is nontrivial. So K_1 is either cyclic or free of rank ≥ 2. If K_1 is cyclic, say generated by k_1, then as J is free, by Claim 3.11 there exists a $J' \leq J$ such that $\{J', \langle k_1 \rangle\}$ is malnormal, so in particular $J' \cap S_1$ is trivial; if K_1 is free of rank ≥ 2, then since K_1 is quasiconvex, Lemma 3.9 implies that there exists a quasiconvex, nonabelian free subgroup $J'' \leq K_1$ that is malnormal in G, so that $J'' \cap S_1 = J''$.

Now assume that the result holds for $m = l - 1$ and let $\{S_1, \ldots, S_l\}$ be a collection of quasiconvex free nonabelian subgroups. Then by the induction hypothesis and Lemma 3.9, there is a quasiconvex nonabelian subgroup $J < G$ such that $J \cap S_i$ is trivial or equal to J for each $i \in \{1, \ldots, m\}$. As before, if $K_l := J \cap S_l = \{1\}$ then there is nothing to show, if K_l is cyclic then by Claim 3.11 there exists a $J' \leq J$ such that $J' \cap S_l$ is trivial, and since $J' \leq J$ then it is still the case that $J' \cap S_i$ is trivial or equal to J' for each $i \in \{1, \ldots, m\}$. Finally, if K_l is nonabelian then since it is quasiconvex, Lemma 3.9 produces a new J'' inside K_l for which $J'' \cap S_l = J''$. Since $J'' < J$, then $J'' \cap S_i$ is trivial or equal to J'' for each $i \in \{1, \ldots, m\}$, and the result follows. □

4 Main theorem

Theorem 4.1 *Let Q be a finitely presented group and G be the fundamental group of a compact special cube complex X. If G is hyperbolic and nonelementary, then there is a short exact sequence*

(1) $$1 \to N \to \Gamma \to Q \to 1$$

where:

(i) Γ *is a hyperbolic cocompactly cubulated group.*

(ii) $N \cong G/K$ *for some $K < G$.*

(iii) $\max\{\operatorname{cd}(G), 2\} \geq \operatorname{cd}(\Gamma) \geq \operatorname{cd}(G) - 1$. *In particular, Γ is torsion-free.*

In what follows, we prove all parts of Theorem 4.1 except for (iii), which we explain in the next section.

Remark 4.2 Cocompact cubulability of Γ hinges on the specialness of X as this is the hypothesis that is used in Theorem 3.2. However, as noted in Remark 3.1, in reality all that is needed is for the hyperplanes of X to be embedded and 2–sided.

Proof Choose a finite presentation $\langle a_1, \ldots, a_s \mid r_1, \ldots, r_k \rangle$ for Q, let B be a bouquet of s circles $a_1, \ldots, a_s$, and let X be a compact nonpositively curved cube complex with $\pi_1 X = G = \langle x_1, \ldots, x_m \rangle$. By Lemma 3.9 and Remark 3.13, there is a malnormal collection $\{H_l\} \cup \{H'_{ij}\} \cup \{H''_{ij}\}$ of quasiconvex free subgroups of rank ≥ 2 of $\pi_1(X \vee B)$, so we can apply Theorem 3.2 to X and B, where the y_i are given by

(i) $y_l = r_l$ for each $1 \leq l \leq k$,

(ii) $y'_{ij} = a_i x_j a_i^{-1}$ for each $1 \leq i \leq s$ and $1 \leq j \leq m$,

(iii) $y''_{ij} = a_i^{-1} x_j a_i$ for each $1 \leq i \leq s$ and $1 \leq j \leq m$.

Hence, there are words $w_1, \ldots, w_l, w'_{1,1}, \ldots, w'_{ij}, w''_{1,1}, \ldots, w''_{ij} \in \pi_1 X$ for which the group

$$\Gamma = G * \pi_1 B / \langle\langle \{r_l w_l\}_{l=1}^k, \{a_i x_j a_i^{-1} w'_{ij}\}_{i=1,j=1}^{s,m}, \{a_i^{-1} x_j a_i w''_{ij}\}_{i=1,j=1}^{s,m} \rangle\rangle$$

is hyperbolic and acts properly and cocompactly on a CAT(0) cube complex.

There is a homomorphism $\Gamma \xrightarrow{\phi} \pi_1 B$ that sends every generator of $\pi_1 X$ to 1. Hence the relations $\{r_l w_l = 1\}_l$ map exactly to the relations $\{r_l = 1\}_l$ in $\pi_1 B$, and so we see that the image of the homomorphism is Q.

The relations $\{a_i x_j a_i^{-1} w'_{ij} = 1\}_{i \in S, j \in M}$, and $\{a_i^{-1} x_j a_i w''_{ij} = 1\}_{i \in S, j \in M}$ ensure that $\langle x_1, \ldots, x_m \rangle$ is a normal subgroup of Γ, so $N = \operatorname{Ker} \phi = \langle x_1, \ldots, x_m \rangle$ and $N \cong \pi_1 X / K$ for some subgroup $K < G$. $\square$

5 Cohomological dimension

We briefly recall some standard facts about the cohomological dimension of groups. We refer the reader to [Brown 1982] for more details and proofs.

A *resolution* for a module M over a ring R is a long exact sequence of R–modules

$$\cdots \to M_n \to M_{n-1} \to \cdots \to M_1 \to M_0 \to M \to 0.$$

A resolution is *finite* if only finitely many of the M_i are nonzero. The *length* of a finite resolution is the maximum integer n such that M_n is nonzero. A resolution is *projective* if each M_i is a projective module.

Definition 5.1 The *cohomological dimension* $\mathrm{cd}(G)$ of a group G is the length of the shortest projective resolution of $\mathbb{Z}$ as a trivial $\mathbb{Z}G$–module.

There is a natural topological analogue of cohomological dimension:

Definition 5.2 The *geometric dimension* $\mathrm{gd}(G)$ of a group G is the least dimension of a classifying space for G.

Remark 5.3 The cellular chain complex of a classifying space for a group G yields a free (in particular, projective) resolution of $\mathbb{Z}$ over $\mathbb{Z}G$, the length of which is equal to the dimension of the classifying space. This implies immediately that $\mathrm{cd}(G) \leq \mathrm{gd}(G)$ for any group G. In particular, if G is free, then $\mathrm{cd}(G) = 1$.

Remark 5.4 The universal covers of nonpositively curved cube complexes are CAT(0) spaces, and hence are contractible, so every nonpositively curved cube complex X is a classifying space for its fundamental group. Therefore, if $G = \pi_1 X$ for a compact nonpositively curved cube complex X, then the cohomological dimension of G is bounded above by the dimension of X.

Proposition 5.5 *The following hold for any group G:*

  (i)   *If $G' < G$ then $\mathrm{cd}(G') \leq \mathrm{cd}(G)$, and equality holds provided that $\mathrm{cd}(G) \leq \infty$ and $[G : G'] < \infty$.*

  (ii) *If $1 \to G' \to G \to G'' \to 1$ is exact, then $\mathrm{cd}(G) \leq \mathrm{cd}(G') + \mathrm{cd}(G'')$.*

  (iii) *If $G = G_1 * G_2$, then $\mathrm{cd}(G) = \max\{\mathrm{cd}(G_1), \mathrm{cd}(G_2)\}$.*

The following result is a consequence of Corollary 5.11, stated below.

Proposition 5.6 *Let G, Q, B and Γ be as in Theorem 4.1 and let $q \colon G * \pi_1 B \to \Gamma$ be the natural quotient. Then $\mathrm{Ker}(q)$ is free.*

Proposition 5.7 *Γ can be chosen so that $\mathrm{cd}(\Gamma) \geq \mathrm{cd}(G) - 1$.*

Proof There is a short exact sequence

$$1 \to \mathrm{Ker}(q) \to G * \pi_1 B \to \Gamma \to 1.$$

Since $\mathrm{Ker}(q)$ is a free group and $\mathrm{cd}(G * \pi_1 B) = \max\{\mathrm{cd}(G), 1\} = \mathrm{cd}(G)$, we have that $\mathrm{cd}(G) \leq \mathrm{cd}(\Gamma) + \mathrm{cd}(\mathrm{Ker}(q)) = \mathrm{cd}(\Gamma) + 1$. □

The torsion-freeness of Γ will follow from having a finite upper bound on its cohomological dimension:

Proposition 5.8 Γ *can be chosen so that* $\max\{\mathrm{cd}(G), 2\} \geq \mathrm{cd}(\Gamma)$.

Before proving Proposition 5.8, we state some auxiliary results:

Definition 5.9 (the Cohen–Lyndon property) Let G be a group, $\{H_i\}_{i \in I}$ a family of subgroups and $N_i \lhd H_i$ for each $i \in I$. The triple $(G, \{H_i\}, \{N_i\})$ has the *Cohen–Lyndon property* if for each $i \in I$ there exists a left transversal T_i of $H_i \langle\!\langle \bigcup_{i \in I} N_i \rangle\!\rangle$ in G such that $\langle\!\langle \bigcup_{i \in I} N_i \rangle\!\rangle$ is the free product of the subgroups N_i^t for $t \in T_i$, so

$$\left\langle\!\!\!\left\langle \bigcup_{i \in I} N_i \right\rangle\!\!\!\right\rangle = \operatorname*{\LARGE *}_{i \in I, t \in T_i} N_i^t.$$

The Cohen–Lyndon property was first defined and studied in [Cohen and Lyndon 1963], where it was proven to hold for triples $(F, C, \langle c \rangle)$ where F is free, C is a maximal cyclic subgroup of F and $c \in C - \{1\}$. This was later generalised in [Edjvet and Howie 1987] to the setting of free products of locally indicable groups. Most remarkably, it was recently proven in [Sun 2020] that triples $(G, \{H_i\}, \{N_i\})$ have the Cohen–Lyndon property when the H_i are "hyperbolically embedded" subgroups of G and the N_i avoids a finite set of "bad" elements depending only on the H_i. We will not define hyperbolically embedded subgroups here, and instead state only the particular case of the theorem that is required for our applications:

Theorem 5.10 [Sun 2020] *Let G be hyperbolic, $\{H_i\}$ be malnormal and quasiconvex subgroups of G, and $N_i \lhd H_i$ for each i. Then there exists a finite set of elements $\{g_1, \ldots, g_n\} \in \bigcup_i H_i - \{1\}$ such that the triple $(G, \{H_i\}, \{N_i\})$ has the Cohen–Lyndon property provided that $N_i \cap \{g_1, \ldots, g_n\} = \varnothing$ for all i.*

To simplify notation, let $\{H_l\} \cup \{H'_{ij}\} \cup \{H''_{ij}\} := \boldsymbol{H}$ and write $H_\iota \in \boldsymbol{H}$. Let $S \subset \boldsymbol{H}$ be a finite set. It is clear from the constructions in Theorems 4.1 and 3.2 (say, by applying Claim 3.11 to S before producing the cyclic subgroups in the proof of Theorem 3.2) that the elements $\{r_l w_l\}_{l=1}^{k}$, $\{a_i x_j a_i^{-1} w'_{ij}\}_{i=1,j=1}^{s,m}$ and $\{a_i^{-1} x_j a_i w''_{ij}\}_{i=1,j=1}^{s,m}$ can be chosen so that for each

$$c_\iota \in \{r_l w_l\}_{l=1}^{k} \cup \{a_i x_j a_i^{-1} w'_{ij}\}_{i=1,j=1}^{s,m} \cup \{a_i^{-1} x_j a_i w''_{ij}\}_{i=1,j=1}^{s,m},$$

where $c_\iota \in H_\iota$, the intersection $\langle\!\langle c_\iota \rangle\!\rangle_{H_\iota} \cap S$ is empty, and hence:

Corollary 5.11 Γ *can be chosen so that the triple $(G *_{\pi_1} B, \{H_\iota\}, \{\langle\!\langle c_\iota \rangle\!\rangle_{H_\iota}\})$ has the Cohen–Lyndon property.*

The following is proven in [Petrosyan and Sun 2024]:

Proposition 5.12 *If $(G, \{H_i\}, \{N_i\})$ has the Cohen–Lyndon property, then*

$$\mathrm{cd}(G/\langle\!\langle \cup_{i \in I} N_i \rangle\!\rangle) \leq \max\{\mathrm{cd}(G), \sup\{\mathrm{cd}(H_i) + 1\}, \sup\{\mathrm{cd}(H_i/N_i)\}\}.$$

Definition 5.13 A *graphical* small cancellation presentation is a 1–dimensional cubical small cancellation presentation, namely, a cubical presentation $X^* = \langle X \mid \{Y_i\} \rangle$ where X is a graph and $Y_i \to X$ are graph immersions.

In the particular setting of graphical presentations, it is well known that the coned-off space X^* is aspherical. Concretely, the following theorem holds. A proof is given in [Gruber 2015], though we caution the reader that the language utilised there differs from ours.

Theorem 5.14 *Let $X^* = \langle X \mid \{Y_i\} \rangle$ be a $C(6)$–graphical small cancellation presentation. Then X^* is aspherical.*

Remark 5.15 As in Corollary 5.11, we simplify the notation so that $\{H_l\} \cup \{H'_{ij}\} \cup \{H''_{ij}\} := \mathbf{H}$, for $c_\iota \in H_\iota$, equals the corresponding $r_l w_l$, $a_i x_j a_i^{-1} w'_{ij}$ or $a_i^{-1} x_j a_i w''_{ij}$, and $C_\iota \to X \vee B$ and $W_\iota \to X \vee B$ are the corresponding local isometries defined in the proof of Theorem 3.2 having $\pi_1(C_\iota) = H_\iota$ and $\pi_1(W_\iota) = \langle c_\iota \rangle$ for each ι.

Proof of Proposition 5.8 By the Cohen–Lyndon property for $(G *_{\pi_1} B, \{H_\iota\}, \{\langle\!\langle c_\iota \rangle\!\rangle_{H_\iota}\})$, we have that $\operatorname{cd}\!\big(G *_{\pi_1} B / \langle\!\langle \bigcup_\iota c_\iota \rangle\!\rangle\big) \leq \max\{\operatorname{cd}(G) *_{\pi_1} B, \sup\{\operatorname{cd}(H_\iota) + 1\}, \sup\{\operatorname{cd}(H_\iota / \langle\!\langle c_\iota \rangle\!\rangle_{H_\iota})\}\}$, and since each H_ι is free, $\operatorname{cd}(H_\iota) = 1$ for all $H_\iota \in \mathbf{H}$. We claim that each of the quotients $H_\iota / \langle\!\langle c_\iota \rangle\!\rangle_{H_\iota}$ has a $C(6)$–graphical small cancellation presentation, and so $\operatorname{cd}(H_\iota / \langle\!\langle c_\iota \rangle\!\rangle_{H_\iota}) \leq 2$ for all $H_\iota \in \mathbf{H}$.

To see this, consider the cubical presentation $\langle X \vee B \mid \{W_\iota \to X \vee B\}_\iota \rangle$ constructed in the proof of Theorem 4.1. As explained in the proof of Theorem 3.2, each H_ι is carried by a local isometry $C_\iota \to X \vee B$. Since each C_ι is itself a nonpositively curved cube complex and $\pi_1 C_\iota \cong H_\iota$ is free, each C_ι is homotopy equivalent to a graph $\overline{C}_\iota$ in $C_\iota^{(1)}$. Similarly, each W_ι is homotopy equivalent to a cycle $\overline{W}_\iota$ in W_ι, and we can further assume that each $\overline{W}_\iota$ lies in $\overline{C}_\iota$. The intersections between pieces of each $\overline{W}_\iota$ are contained in the corresponding intersections between pieces of W_ι, and the length of each $\overline{W}_\iota$ is bounded below by the diameter of the corresponding W_ι, so each $\overline{W}_\iota$ has at least as many pieces as the corresponding W_ι. Since the W_ι are chosen to satisfy at least $C'\big(\frac{1}{16}\big)$, and in particular $C'\big(\frac{1}{16}\big) \implies C(6)$, each $\langle \overline{C}_\iota \mid \overline{W}_\iota \to \overline{C}_\iota \rangle$ satisfies the $C(6)$ condition, and the proof is complete. □

This finishes the proof of Theorem 4.1.

References

[Agol 2008]   **I Agol**, *Criteria for virtual fibering*, J. Topol. 1 (2008) 269–284 MR Zbl

[Agol 2013]   **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR Zbl

[Arzhantseva 2001]   **G N Arzhantseva**, *On quasiconvex subgroups of word hyperbolic groups*, Geom. Dedicata 87 (2001) 191–208 MR Zbl

[Arzhantseva and Hagen 2022]   **G N Arzhantseva**, **M F Hagen**, *Acylindrical hyperbolicity of cubical small cancellation groups*, Algebr. Geom. Topol. 22 (2022) 2007–2078 MR Zbl

[Arzhantseva and Steenbock 2023]   **G Arzhantseva**, **M Steenbock**, *Rips construction without unique product*, Pacific J. Math. 322 (2023) 1–9 MR Zbl

[Barnard et al. 2007] **J Barnard**, **N Brady**, **P Dani**, *Super-exponential distortion of subgroups of* CAT(−1) *groups*, Algebr. Geom. Topol. 7 (2007) 301–308 MR Zbl

[Baumslag et al. 2000] **G Baumslag**, **M R Bridson**, **C F Miller, III**, **H Short**, *Fibre products, non-positive curvature, and decision problems*, Comment. Math. Helv. 75 (2000) 457–477 MR Zbl

[Belegradek and Osin 2008] **I Belegradek**, **D Osin**, *Rips construction and Kazhdan property (T)*, Groups Geom. Dyn. 2 (2008) 1–12 MR Zbl

[Bergeron and Wise 2012] **N Bergeron**, **D T Wise**, *A boundary criterion for cubulation*, Amer. J. Math. 134 (2012) 843–859 MR Zbl

[Bestvina 2000] **M Bestvina**, *Questions in geometric group theory*, online reference (2000) Available at https://www.math.utah.edu/~bestvina/eprints/questions-updated.pdf

[Bestvina et al. 2014] **M Bestvina**, **M Sageev**, **K Vogtmann** (editors), *Geometric group theory*, IAS/Park City Math. Ser. 21, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[Bridson and Haefliger 1999] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[Brown 1982] **K S Brown**, *Cohomology of groups*, Grad. Texts in Math. 87, Springer (1982) MR Zbl

[Bumagin and Wise 2005] **I Bumagin**, **D T Wise**, *Every group is an outer automorphism group of a finitely generated group*, J. Pure Appl. Algebra 200 (2005) 137–147 MR Zbl

[Chatterji and Niblo 2005] **I Chatterji**, **G Niblo**, *From wall spaces to* CAT(0) *cube complexes*, Int. J. Algebra Comput. 15 (2005) 875–885 MR Zbl

[Cohen and Lyndon 1963] **D E Cohen**, **R C Lyndon**, *Free bases for normal subgroups of free groups*, Trans. Amer. Math. Soc. 108 (1963) 526–537 MR Zbl

[Edjvet and Howie 1987] **M Edjvet**, **J Howie**, *A Cohen–Lyndon theorem for free products of locally indicable groups*, J. Pure Appl. Algebra 45 (1987) 41–44 MR Zbl

[Fine et al. 2002] **B Fine**, **A Myasnikov**, **G Rosenberger**, *Malnormal subgroups of free groups*, Comm. Algebra 30 (2002) 4155–4164 MR Zbl

[Fujiwara and Manning 2010] **K Fujiwara**, **J F Manning**, CAT(0) *and* CAT(−1) *fillings of hyperbolic manifolds*, J. Differential Geom. 85 (2010) 229–269 MR Zbl

[Gitik et al. 1998] **R Gitik**, **M Mitra**, **E Rips**, **M Sageev**, *Widths of subgroups*, Trans. Amer. Math. Soc. 350 (1998) 321–329 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

[Gromov 1993] **M Gromov**, *Asymptotic invariants of infinite groups*, from "Geometric group theory, II" (G A Niblo, M A Roller, editors), Lond. Math. Soc. Lect. Note Ser. 182, Cambridge Univ. Press (1993) 1–295 MR Zbl

[Gruber 2015] **D Gruber**, *Groups with graphical $C(6)$ and $C(7)$ small cancellation presentations*, Trans. Amer. Math. Soc. 367 (2015) 2051–2078 MR Zbl

[Haglund 2008] **F Haglund**, *Finite index subgroups of graph products*, Geom. Dedicata 135 (2008) 167–209 MR Zbl

[Haglund and Wise 2008] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR Zbl

[Haglund and Wise 2012]   **F Haglund**, **D T Wise**, *A combination theorem for special cube complexes*, Ann. of Math. 176 (2012) 1427–1482   MR Zbl

[Hsu and Wise 2015]   **T Hsu**, **D T Wise**, *Cubulating malnormal amalgams*, Invent. Math. 199 (2015) 293–331   MR Zbl

[Jankiewicz and Wise 2022]   **K Jankiewicz**, **D T Wise**, *Cubulating small cancellation free products*, Indiana Univ. Math. J. 71 (2022) 1397–1409   MR Zbl

[Januszkiewicz and Świątkowski 2003]   **T Januszkiewicz**, **J Świątkowski**, *Hyperbolic Coxeter groups of large dimension*, Comment. Math. Helv. 78 (2003) 555–583   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

[Mosher and Sageev 1997]   **L Mosher**, **M Sageev**, *Nonmanifold hyperbolic groups of high cohomological dimension*, preprint (1997)

[Niblo and Reeves 1997]   **G Niblo**, **L Reeves**, *Groups acting on* CAT(0) *cube complexes*, Geom. Topol. 1 (1997) 1–7   MR Zbl

[Ollivier 2005]   **Y Ollivier**, *A January* 2005 *invitation to random groups*, Ensaios Mat. 10, Soc. Brasil. Mat., Rio de Janeiro (2005)   MR Zbl

[Ollivier and Wise 2007]   **Y Ollivier**, **D T Wise**, *Kazhdan groups with infinite outer automorphism group*, Trans. Amer. Math. Soc. 359 (2007) 1959–1976   MR Zbl

[Osajda 2013]   **D Osajda**, *A construction of hyperbolic Coxeter groups*, Comment. Math. Helv. 88 (2013) 353–367   MR Zbl

[Petrosyan and Sun 2024]   **N Petrosyan**, **B Sun**, *Cohomology of group theoretic Dehn fillings, II*, Adv. Math. 437 (2024) art. id. 109412   MR Zbl

[Rips 1982]   **E Rips**, *Subgroups of small cancellation groups*, Bull. Lond. Math. Soc. 14 (1982) 45–47   MR Zbl

[Sageev 1995]   **M Sageev**, *Ends of group pairs and non-positively curved cube complexes*, Proc. Lond. Math. Soc. 71 (1995) 585–617   MR Zbl

[Sageev 1997]   **M Sageev**, *Codimension-*1 *subgroups and splittings of groups*, J. Algebra 189 (1997) 377–389   MR Zbl

[Sun 2020]   **B Sun**, *Cohomology of group theoretic Dehn fillings, I: Cohen–Lyndon type theorems*, J. Algebra 542 (2020) 277–307   MR Zbl

[Wise 1998]   **D T Wise**, *Incoherent negatively curved groups*, Proc. Amer. Math. Soc. 126 (1998) 957–964   MR Zbl

[Wise 2003]   **D T Wise**, *A residually finite version of Rips's construction*, Bull. Lond. Math. Soc. 35 (2003) 23–29   MR Zbl

[Wise 2004]   **D T Wise**, *Cubulating small cancellation groups*, Geom. Funct. Anal. 14 (2004) 150–214   MR Zbl

[Wise 2021]   **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, Ann. of Math. Stud. 209, Princeton Univ. Press (2021)   MR Zbl

DPMMS, Centre for Mathematical Sciences, University of Cambridge
Cambridge, United Kingdom

mcr59@cam.ac.uk

Multipath cohomology of directed graphs

LUIGI CAPUTI

CARLO COLLARI

SABINO DI TRANI

This work is part of a series of papers focusing on multipath cohomology of directed graphs. Multipath cohomology is defined as the (poset) homology of the path poset — ie the poset of disjoint simple paths in a graph — with respect to a certain functor. This construction is essentially equivalent, albeit more computable, to taking the higher limits of said functor on (a certain modification of) the path poset. We investigate the functorial properties of multipath cohomology. We provide a number of sample computations, show that multipath cohomology does not vanish on trees, and show that, when evaluated at the coherently oriented polygon, it recovers Hochschild homology. Finally, we use the same techniques employed to study the functoriality to investigate the connection with the chromatic homology of (undirected) graphs introduced by L Helme-Guizon and Y Rong.

1 Introduction

Directed graphs are ubiquitous objects in mathematics and science in general. Due to their simplicity and flexibility, (directed) graphs find application in a wide range of fields: physics, computer science, complex systems, engineering, biology, neuroscience, medicine, robotics, etc, encompassing and embracing most scientific domains. Extracting topological and combinatorial information from directed graphs is, therefore, not only interesting, but also particularly important from different perspectives.

Cohomological invariants of directed graphs have been extensively studied in the last decades, with prominent work in combinatorial topology — see [Wachs 2003; Kozlov 2008; Jonsson 2008] — and have deep connections with other areas of mathematics — see [Jonsson 2008, Chapter 1] for an overview. One of the common strategies is to construct suitable simplicial complexes — eg matching complexes, independence complexes, complexes of directed trees, etc [Jonsson 2008] — associated to a (directed) graph, and then to analyse the associated homology groups. In this work we follow a similar approach; we first represent a graph using a suitable poset, called the *path poset* [Turner and Wagner 2012], and then apply a cohomology theory of posets known as *poset homology* — see eg [Chandler 2019] — to get cohomological invariants of directed graphs. We call these invariants *multipath cohomology groups*, as they are constructed using the combinatorial information of multipaths (ie the elements of the path poset) in a directed graph.

Roughly speaking, poset homology associates to a poset P and a functor $\mathcal{F}$ on P a graded module. As, in our case, the poset P (and the functor $\mathcal{F}$) depend on a directed graph G, we obtain a (cohomology) theory for directed graphs. This idea is not novel; for instance Helme-Guizon and Rong [2005], using a different poset, defined chromatic homology. On a different note, one might use a different homology theory of posets, for instance, the classical functor homology groups, an approach pursued by Turner and Wagner [2012]. Comparisons between multipath cohomology, Helme-Guizon and Rong's chromatic homology, Turner–Wagner's homology, and other theories obtained using different (co)homologies for posets will be carried out in Sections 6 and 7.

Our main goal is to investigate structural and combinatorial properties of digraphs through the lenses of multipath cohomology. In this work, we are interested in the definition and general properties of multipath cohomology, such as functoriality, and in its relationship with similar theories. The investigation of combinatorial properties, and the computations of multipath cohomology groups for various families of directed graphs are the subject of [Caputi et al. 2023; 2024b] and forthcoming papers. We believe that the framework developed here can be helpful both in answering theoretical questions, as well as solving problems in the applied setting — see Section 8.

1.1 Other approaches

The development and investigation of homology theories for directed graphs (shortly, *digraphs*) is a very active research field. To define such theories, a first approach comes from the observation that a (directed) graph can naturally be seen as a topological space (a 1–dimensional CW–complex) on which ordinary homology can be applied. However, in this case, the homology groups in degree $i > 1$ would vanish. To sidestep this issue, there are various ways that can be pursued. For instance, one can define (higher-dimensional) simplicial complexes from a graph — eg by constructing the (directed) flag complex (also known as clique complex) [Ivashchenko 1994; Chen et al. 2001; Aharoni et al. 2005; Govc et al. 2021], or the matching complexes, independence complexes — see [Jonsson 2008] — and hence, compute their ordinary (simplicial) homology. Alternatively, one can construct the so-called path complex (see eg [Grigoryan et al. 2016]), whose homology is called path homology. In a third approach, one can associate to a digraph the so-called path algebra. Then, homology groups of digraphs can be defined as the Hochschild homology groups of the path algebras — see [Happel 1989; Caputi and Riihimäki 2024]. A pitfall of most of these homology theories is that they vanish when evaluated on trees; this hints to the fact that they might be discarding an important part of the combinatorics of the input digraph, including their directionality information.

Turner and Wagner [2012] move in a different direction. Given a graph G they consider its path poset $P(\mathsf{G})$, that is, the collection of all the unions of disjoint simple paths in G, partially ordered by inclusion. Since posets can be seen as categories in a natural way, one can apply homology with functor coefficients [Gabriel and Zisman 1967] to the path poset and obtain topological invariants of the directed graph. On the one hand, this homology is nontrivial. In particular, for a given algebra A, the Turner–Wagner homology of the coherently oriented polygon with n edges is isomorphic, up to degree n, to the Hochschild homology groups

of A — see [Turner and Wagner 2012, Theorem 1]. On the other hand, the homology of a category with coefficients in a functor is generally difficult to compute. Computations can be done with relative ease if one considers the constant functor, ie the functor which associates to each element of the path poset the base ring. However, in this case, the result is trivial since the path poset has a minimum: the empty multipath.

Close to Turner–Wagner homology sits the so-called chromatic homology, introduced in [Helme-Guizon and Rong 2005]. Chromatic homology is a homology theory for unoriented graphs, inspired by Khovanov homology [2000], and with the remarkable property that it categorifies the chromatic polynomial. Przytycki [2010] has shown that a version of the chromatic homology (further extended to incorporate the orientation in the case of linear and polygonal graphs) can recover (a truncation of) the Hochschild homology. This fact was later used to prove [Turner and Wagner 2012, Theorem 1] by showing that for coherently oriented polygons their homology is in fact isomorphic to Przytycki's version of chromatic homology.

In this work, inspired by the approaches of Turner and Wagner and of Helme-Guizon and Rong, we follow a certain modification of Turner and Wagner's functorial framework; instead of directly applying functors to the path poset $P(G)$ of G, we use poset homology [Chandler 2019], a suitable adaptation of Helme-Guizon and Rong's construction to this context. Alternatively, instead of the naive poset homology, one can use the so-called cellular cohomology, introduced in [Everitt and Turner 2015]. Cellular cohomology extends poset homology to arbitrary finite (ranked) posets; this yields, after some minor modifications on the path poset, an essentially equivalent theory — see Section 6. Nonetheless, the advantage of poset homology over cellular homology is its computability, which is essential in view of possible applications — see Question 91 and the computations developed in [Caputi et al. 2023].

1.2 Statement of results

We construct a cochain complex $(C^*_{\mathcal{F}}(P), d^*)$; this depends on the datum of a poset P associated to a graph G, and a covariant functor $\mathcal{F}$, defined on (the category associated to) P with values in an additive category A. Roughly, $C^n_{\mathcal{F}}(P)$ is given by a directed sum of $\mathcal{F}(x)$ for all $x \in P$ of "level" n. The differential d^* is induced by the functor $\mathcal{F}$ applied to the covering relations in P. In Section 4.1, we specialise this construction to obtain multipath cohomology. First, we fix a ring R, an R–algebra A, and an (A, A)–bimodule M. Then the role of the poset P is played by the path-poset $P(G)$, and the part of the functor $\mathcal{F}$ is taken by $\mathcal{F}_{A,M}$. The latter assigns a tensor product of copies of M and A to each $H \in P(G)$. We finally denote by $(C^*_\mu(G; A, M), d^*)$ the cochain complex $(C^*_{\mathcal{F}_{A,M}}(P(G)), d^*)$. The main definition of multipath cohomology as homology of this complex is given in Definition 46.

Unless otherwise specified, for the rest of the Introduction we set $M = A$. In particular, we drop M from the notation of multipath cohomology, writing $H^*_\mu(G; A)$ instead of $H^*_\mu(G; A, M)$. Some computations of multipath cohomology, for $A = R = \mathbb{K}$ a field, are collected in Table 1.

A key property of cohomology theories is that they are functorial. One of the main results of this paper is that functoriality for multipath cohomology holds once we fix the number of vertices in our graphs.

Digraph G	$H^0_\mu(\texttt{G};\mathbb{K})$	$H^1_\mu(\texttt{G};\mathbb{K})$	$H^2_\mu(\texttt{G};\mathbb{K})$	$H^i_\mu(\texttt{G};\mathbb{K})$, $i > 2$
	$\mathbb{K}$	0	0	0
	0	0	0	0
	0	$\mathbb{K}$	0	0
	0	0	0	0
	0	$\mathbb{K}^2$	0	0
	0	0	$\mathbb{K}^2$	0

Table 1: Some digraphs and their respective multipath cohomologies.

Theorem 1 *Let R–$\mathbf{Alg}$ be the category of R–algebras, let $\mathbf{Digraph}(n)$ be the category of digraphs with n vertices, and let R–$\mathbf{Mod}^{\mathrm{gr}}$ be the category of graded R–modules. Then, multipath cohomology*

$$H_\mu(-;-)\colon \mathbf{Digraph}^{\mathrm{op}}(n) \times R\text{–}\mathbf{Alg} \to R\text{–}\mathbf{Mod}^{\mathrm{gr}}$$

is a bifunctor for all n.

We need to restrict to the category $\mathbf{Digraph}(n)$ for a purely technical reason; intuitively, the issue is due to the tensor products involved in the definition of multipath cohomology. More formally, the functor $\mathcal{F}_{A,A}$ is not a coefficients system — see Remark 59. This technical issue is solved when either we fix the number of vertices, or we take $A = R$. In the latter case, we have a stronger result (see Theorem 65).

Hochschild homology is a homology theory for pairs (A, M) with A an algebra and M an (A, A)–bimodule [Loday 1992]. It has been proven in [Przytycki 2010] that (a suitable modification of) the chromatic homology of the coherently oriented n–polygon recovers the Hochschild homology up to degree n. It turns out that multipath cohomology shares the same property — see Corollary 82.

A consequence of Corollary 82 and Theorem 1 is that, once one fixes a digraph G, the functor $H_\mu(\texttt{G};-)$, which is a functor between R–$\mathbf{Alg}$ and R–$\mathbf{Mod}^{\mathrm{gr}}$, can be seen as a homology theory of algebras. From this viewpoint, we can rephrase Corollary 82 by stating that the family of homologies for algebras $\{H_\mu(\texttt{G};-)\}_\texttt{G}$ generalises Hochschild homology (compare also with [Turner and Wagner 2012]). In light of these results, one can expect chromatic homology and multipath cohomology to be, in some sense, related. Despite being defined on different categories (of undirected and directed graphs, respectively) we obtain a short exact sequence featuring the two theories — see Proposition 84.

A great advantage of multipath cohomology is that it is amenable to computations. We postpone the (combinatorial) analysis, as well as the description of an algorithm to calculate the multipath cohomology of certain graphs, to [Caputi et al. 2023; 2024b] and forthcoming papers. In the present work, we limit our

computations to a restricted number of cases (with coefficients in a field $\mathbb{K} = R = A = M$); see Section 4.2 and Table 1. Such computations hint to the fact that multipath cohomology might be sensible to some combinatorial properties of graphs. We observe that multipath cohomology does not vanish, nor is it concentrated in degree 0, in the case of trees.

Another consequence of the computations collected in Table 1 is that chromatic, multipath, and Turner–Wagner (co)homologies are not isomorphic. We conclude with the following observation. Although not isomorphic "on the nose", Turner–Wagner and multipath cohomology are related, if $A = R$, by the universal coefficients short exact sequence. On the one hand, Turner–Wagner homology computes the higher colimits of the functor $\mathcal{F}_{R,R}$. On the other hand, multipath cohomology computes the associated higher limits; then, the short exact sequence gives the relation between the two, with a correcting Ext term. We refrain from giving a more detailed account of this case here, inviting the interested reader to Section 6.

Conventions

Typewriter font, eg G, H, etc, will be used to denote graphs (both directed and unoriented). Calligraphic font, eg $\mathcal{F}$, $\mathcal{G}$, etc, will be used to denote functors. Bold capital letters, eg $\boldsymbol{A}$, $\boldsymbol{C}$, etc will be used to denote categories. Depending on the context, $\boldsymbol{A}$ will denote an *Abelian*, or more generally, an *additive* category, and, for a given poset P, we will denote with the same letter in roman and bold, that is $\boldsymbol{P}$, its associated category. All rings are assumed to be unital and commutative, and algebras are assumed to be associative. Unless otherwise stated, R will denote a base ring, A will denote an R–algebra, M will denote an (A, A)–bimodule, and all tensor products $\otimes$ are assumed to be over the base ring. Given a (co)chain complex C^*, we will denote by $C^*[i]$ the shifted complex C^{*+i}.

Acknowledgements

2 Basic notions

In this section we introduce and provide the basic notions and conventions about graphs and posets that will be used throughout the paper. In particular, in Section 2.2, we introduce the path poset, one of the main ingredients in the construction of multipath homology (Section 4.1).

2.1 Digraphs and posets

We will only consider *finite* graphs and digraphs. For a set V, let $\wp(V)$ denote its power set. We recall the definitions of various types of graphs — see, for instance, [West 1996].

An *unoriented graph* G is a pair (V, E), where V is the set of *vertices* and E (the *edges*) are unordered pairs of distinct vertices. A graph is a *directed graph*, or a *digraph*, when the edges are ordered pairs, ie $E \subseteq V \times V \setminus \{(v, v) \mid v \in V\}$. Finally, a graph is *oriented* if it is a digraph and for any vertices v, w at most one among (v, w) and (w, v) are in E.

The main object of interest in this work is digraphs. Unless otherwise stated, we will refer to digraphs simply as *graphs*. When dealing with (un)oriented graphs the adjective "*(un)oriented*" will be explicitly stated. Two vertices v and w in a digraph can share at most two edges: (v, w) and (w, v). There are no multiple edges between two vertices in oriented and unoriented graphs.

For the sake of simplicity, we restrict to the case of digraphs. Everything in this paper can be carried out verbatim in the more general case of directed multigraphs.

By definition, an edge of a digraph is an ordered set of two distinct vertices, say $e = (v, w)$. The vertex v is called the *source* of e, while the vertex w is called the *target* of e. The source and target of an edge e will be denoted by $s(e)$ and $t(e)$, respectively. If a vertex v is either a source or a target of an edge e, we will say that e is incident to v.

Later we shall also deal with more than one graph at the time. In such cases, the sets of vertices and edges of a digraph G will be denoted by $V(\mathsf{G})$ and $E(\mathsf{G})$, respectively. A morphism of digraphs from G_1 to G_2 is a function $\phi \colon V(\mathsf{G}_1) \to V(\mathsf{G}_2)$ such that

$$e = (v, w) \in E(\mathsf{G}_1) \implies \phi(e) := (\phi(v), \phi(w)) \in E(\mathsf{G}_2).$$

A morphism of digraphs sends directed edges to directed edges; in particular, it does not allow collapsing — that is, $(v, w) \in E(\mathsf{G}_1) \Rightarrow \phi(v) \neq \phi(w)$. A morphism of digraphs is called *regular* if it is injective as a function; digraphs and regular morphisms of digraphs form a category that we denote by **Digraph**. If G_1 and G_2 are isomorphic in **Digraph**, we write $\mathsf{G}_1 \cong \mathsf{G}_2$.

A subgraph H of a graph G is a graph such that $V(\mathsf{H}) \subseteq V(\mathsf{G})$ and $E(\mathsf{H}) \subseteq E(\mathsf{G})$, and in such case we write $\mathsf{H} \leq \mathsf{G}$. If $\mathsf{H} \leq \mathsf{G}$ and $\mathsf{H} \neq \mathsf{G}$ we say that H is a proper subgraph of G, and write $\mathsf{H} < \mathsf{G}$. If $\mathsf{H} \leq \mathsf{G}$ and $V(\mathsf{H}) = V(\mathsf{G})$ we say that H is a *spanning subgraph* of G. Given a proper spanning subgraph $\mathsf{H} \leq \mathsf{G}$, we can always find an edge e in $E(\mathsf{G}) \setminus E(\mathsf{H})$. We use the following notation:

Notation 2 The spanning subgraph of G obtained from H by adding an edge e is denoted by $\mathsf{H} \cup e$.

Let S be a set. Recall that a (strict) partial order on S is a transitive binary relation $\lhd$ such that, for each $x, y \in S$, at most one of the following is true: $x \lhd y$, $y \lhd x$, or $x = y$. As a matter of notation, we will write $x \trianglelefteq y$ in place of "$x \lhd y$ or $x = y$".

Given a partial order, there is an associated covering relation given by $x \widetilde{\lhd} y$ if, and only if, $x \lhd y$ and there is no z such that $x \lhd z$, $z \lhd y$. A partial order can be also seen as the transitive closure[1] of its associated covering relation. Moreover, the associated covering relation is the smallest relation whose transitive closure is the given partial order. A partially ordered set, or simply *poset*, is a pair $(S, \lhd)$ consisting of a set S and a partial order $\lhd$ on S. A morphism of posets $f : (S, \lhd) \to (S', \lhd')$ is a monotonic map of sets, that is, a function $f : S \to S'$ such that $x \lhd y$ implies $f(x) \lhd' f(y)$. Posets and morphisms of posets form a category, which will be denoted by **Poset**.

Remark 3 Each poset $P = (S, \lhd)$ can be seen as a (small) category P in a straightforward manner; the set of objects of P is the set S, and the set of morphisms between x and y contains a single element if, and only if, $x \lhd y$ or $x = y$, and it is empty otherwise.

Let $P = (S, \lhd)$ be a poset. An element $m \in P$ is a maximal element if there are no elements of P strictly greater than m, ie if $m \unlhd s$ with $s \in P$, then $m = s$. A maximum of P is an element $M \in S$ which is greater than any other element, ie $s \unlhd M$ for all $s \in S$.

The following two facts are standard:

(M.i)   If $P = (S, \lhd)$ is a finite poset — that is, S is finite — then for each $s \in S$ there exists a maximal element $m \in S$ such that $s \unlhd m$.

(M.ii)   A poset has a unique maximal element if, and only if, said element is a maximum.

Minimal elements and minima are defined analogously by exchanging the role of s with m and M in the definitions of maximum and maximal elements, respectively. Moreover, the obvious translations of facts (M.i) and (M.ii) hold.

A poset is called a *Boolean poset* if it is isomorphic to the power set $\wp(S)$ — ie the set of all subsets — of a finite set S with partial order $\subset$ given by inclusion. The standard Boolean poset (of size 2^n) is by definition the poset $\mathbb{B}(n) = (\wp(\{0, 1, 2, \ldots, n-1\}), \subset)$.

Example 4 Let G be a (possibly unoriented) graph. Among others, we can specifically consider two posets: the *poset of subgraphs* $SG(G)$ and the *poset of spanning subgraphs* $SSG(G)$. The elements of these posets are all the subgraphs and all the spanning subgraphs of G, respectively. In both cases the order relation $<$ is given by the property of being a proper subgraph. The covering relation $\prec$ of $<$ in $SSG(G)$ is easily checked to be

$$H \prec H' \iff \text{there exists } e \in E(H') \setminus E(H) \text{ such that } H' = H \cup e.$$

Equivalently, $H \prec H'$ if, and only if, $E(H') \setminus E(H) = \{e\}$ and $E(H) \setminus E(H') = \varnothing$. The covering relation on $SG(G)$ is slightly different from $\prec$; we need to consider, in addition to the case above, also the case where $E(H') = E(H)$ and $V(H') = V(H) \cup \{v\}$ for a certain $v \notin V(H)$.

---

[1]The transitive closure of a relation $R \subset S \times S$ is a relation R' such that $(s, s') \in R'$ if, and only if, either $(s, s') \in R$, or there exists $s'' \in S$ such that $(s, s''), (s'', s') \in R'$.

Note that SSG(G) is a Boolean poset; in fact, we have natural isomorphisms of posets

$$(\mathrm{SSG}(\mathtt{G}), <) \cong (\wp(E(\mathtt{G})), \subset)$$

given by $\mathtt{H} \mapsto E(\mathtt{H})$. On the contrary, the poset $\mathrm{SG}(\mathtt{G})$ is generally not isomorphic to a Boolean poset; a counterexample is given by the 1–step graph — see Figure 1. However, $\mathrm{SG}(\mathtt{G})$ is a subposet of a Boolean poset, namely the poset $(\wp(V(\mathtt{G}) \cup E(\mathtt{G})), \subset)$.

Definition 5 Given a poset $(S, \lhd)$ a *subposet* is a subset $S' \subseteq S$ with the order $\lhd_{|S' \times S'}$ induced by $\lhd$. A subposet $(S', \lhd_{|S' \times S'})$ is called *downward closed* (resp *upward closed*) with respect to $(S, \lhd)$ if for every $h \in S$ such that $h \lhd h'$ (resp $h' \lhd h$) for some $h' \in S'$, we have $h \in S'$.

The poset of spanning subgraphs $\mathrm{SSG}(\mathtt{G})$ is a subposet of the subgraphs poset $\mathrm{SG}(\mathtt{G})$, but it is easily checked *not* to be downward closed. Nonetheless, it is upward closed.

Furthermore, observe the complement of an upward closed subposet is downward closed, and vice versa.

We conclude the subsection with the definition of two properties which will be essential to define multipath cohomology.

Definition 6 Let $(S, \lhd)$ be a poset and $(S', \lhd_{|S' \times S'})$ be a subposet of $(S, \lhd)$.

(1) We say that $(S, \lhd)$ is *squared* if for each triple $x, y, z \in S'$ such that z covers y and y covers x, there is a unique $y' \neq y$ such that z covers y' and y' covers x. Such elements x, y, y', and z will be called a *square* in S.

(2) We say that $(S', \lhd_{|S' \times S'})$ is *faithful* if the covering relation in S' induced by $\lhd_{|S' \times S'}$ is the restriction of the covering relation in S induced by $\lhd$.

Observe that square posets have also been called *thin posets* in the literature; see eg [Björner 1984, Section 4] or [Chandler 2019]. Prime examples of squared posets are Boolean posets.

Example 7 Downward and upward closed subposets are faithful. Furthermore, each downward or upward closed subposet of a squared poset is squared.

The following proposition is straightforward:

Proposition 8 *Let $(S, \lhd)$ be a poset. Given the subposets $S', S'' \subset S$, we have:*

(1) *If $S'' \subset S'$ is faithful and $S' \subset S$ is faithful, then S'' is faithful in S.*

(2) *If S' and S'' are faithful in S (resp squared), then $S' \cap S''$ is faithful in S (resp squared).*

2.2 Path posets

We now define one of the main ingredients in the construction of multipath cohomology: the path poset.

Let $\mathtt{G}$ be a graph and let $|\mathtt{G}|$ denote its geometric realisation as a CW–complex. A connected component of $\mathtt{G}$ is a subgraph $\mathtt{H}$ of $\mathtt{G}$ whose realisation $|\mathtt{H}|$ is connected. A *simple path* of $\mathtt{G}$ is a sequence of edges

$$v_0 \qquad v_1 \qquad \qquad v_{n-1} \qquad v_n$$
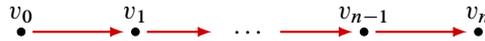
Figure 1: The n–step graph I_n.

$e_1, \ldots, e_n$ of G such that $s(e_{i+1}) = t(e_i)$ for $i = 1, \ldots, n-1$, and no vertex is encountered twice, ie if $s(e_i) = s(e_j)$ or $t(e_i) = t(e_j)$, then $i = j$, and it is not a cycle, ie $s(e_1) \neq t(e_n)$.

Remark 9 If a connected graph G admits an ordering of all its edges with respect to which it is a simple path, then it is isomorphic to the graph I_n shown in Figure 1. The explicit isomorphism is given by the morphism of digraphs $\phi: V(G) \to V(I_n)$, $s(e_i) \mapsto v_{i-1}$, $t(e_n) \mapsto v_n$.

We are interested in taking disjoint sets of simple paths; following [Turner and Wagner 2012], we call them multipaths.

Definition 10 A *multipath* of G is a spanning subgraph such that each connected component is either a vertex or its edges admit an ordering such that it is a simple path.

Remark 11 Every spanning subgraph of a multipath is still a multipath. In particular, the set of multipaths of a graph G — denoted by $P(G)$ — forms a downward closed subposet of $SSG(G)$ (with the induced order). Moreover, there is a unique minimum in both $P(G)$ and $SSG(G)$, which is the spanning subgraph with no edges.

With the definition of multipath in place we can present the main actor of the section.

Definition 12 The *path poset* of G is the poset $(P(G), <)$ associated to G, that is, the set of multipaths of G ordered by the relation of "being a subgraph".

When the partial order on $P(G)$ is not specified, we will always implicitly assume it to be the order relation $<$. Moreover, with abuse of notation, we will also write $P(G)$ instead of $(P(G), <)$.

We now provide some examples of path posets.

Example 13 Consider the coherently oriented linear graph I_n with n edges, illustrated in Figure 1. In this case all spanning subgraphs are multipaths, that is $P(I_n) = SSG(I_n)$. In particular, it follows that $(P(I_n), <)$ is a Boolean poset.

Example 14 Consider the coherently oriented polygonal graph P_n with $n+1$ edges, illustrated in Figure 2. Note that, according to our definition, also the digon P_1, which is shown explicitly in Figure 3, is a digraph. In this case all spanning subgraphs but the polygon itself are multipaths. Equivalently, we have $(P(P_n) \cup \{P_n\}, <) = (SSG(P_n), <)$. In particular, $(P(P_n), <)$ for $n \in \mathbb{N} \setminus \{0\}$ is *not* a Boolean poset (as it is missing the maximum).
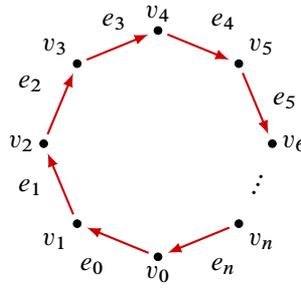
Figure 2: The coherently oriented polygonal graph P_n with a fixed ordering of vertices.
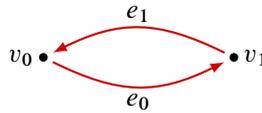
Figure 3: The digon graph P_1.

Recall that the symbol $\widetilde{\vartriangleleft}$ denotes a covering relation. In order to visually represent path posets associated to digraphs, we use the Hasse diagrams. The Hasse digraph $\texttt{Hasse}(S, \vartriangleleft)$ of a poset $(S, \vartriangleleft)$ is the graph whose vertices are the elements of S and such that (x, y) is an edge if, and only if, $x \widetilde{\vartriangleleft} y$. Note that the Hasse graph of a poset S completely encodes the covering relation of S and hence, by transitivity, the order relation.

Example 15 Consider the Y–shaped graphs in Figure 4. Their associated path posets, up to isomorphism, are shown in Figure 5; the figures show the covering relation in the posets or, alternatively, the Hasse digraph of the path poset. Note that the path poset of the graph in Figure 4(b) is isomorphic to the path poset of the graph in Figure 4(a); in fact, these two graphs are isomorphic up to reversing the orientation in all arcs of one of the two. However, the path poset of the graph in Figure 4(b) is not isomorphic to the path poset of the graph in Figure 4(c) (eg there are no multipaths of with two edges in the latter).

Example 16 Consider the H–shaped digraph of Figure 6. The associated path poset, which is illustrated in Figure 7, has multipaths with at most two edges.

The following remark will be essential in the functorial applications — see Section 5.

Remark 17 A morphism $f : G_1 \to G_2$ in **Digraph** (which is regular by definition) induces a morphism of posets $Pf : P(G_1) \to P(G_2)$; more precisely, to a multipath $H \subset P(G_1)$ we associate the spanning subgraph $Pf(H)$ of G_2 defined by $E(Pf(H)) = \{f(e) \mid e \in E(H)\}$. This association yields a functor $P : \textbf{Digraph} \to \textbf{Poset}$. Note that $Pf(P(G_1))$ is a faithful subposet of $P(G_2)$.

We conclude the section by noting that, in favourable cases, the path poset determines the graph.

Proposition 18 *Let* G *be a connected graph with* n *edges. If* $P(G)$ *has a maximum then we have that* $P(G) \cong \mathbb{B}(n)$ *and* $G \cong I_n$. *In particular, a connected graph has a Boolean path poset if, and only if, it is isomorphic to* I_n.
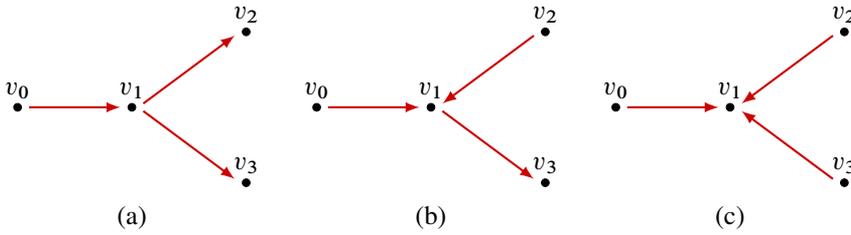
Figure 4: Three nonisomorphic Y–shaped digraphs.
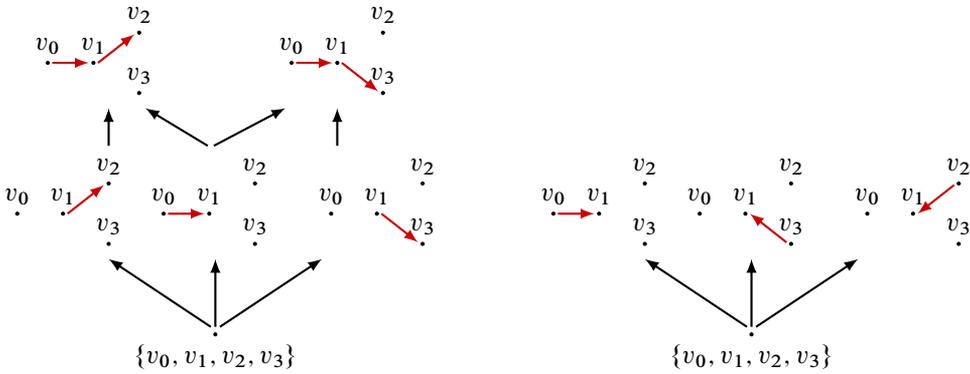
Figure 5: The path posets of the Y–shaped digraphs in Figure 4(a), left, and (c), right.
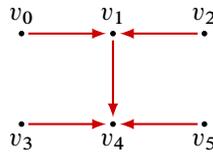
Figure 6: A figure H–type digraph.

Proof Denote by M the maximum, which is the unique maximal element, of $P(\mathtt{G})$. We shall now prove that $\mathtt{M} = \mathtt{G}$. That being true, $\mathtt{G}$ would be a connected graph admitting an ordering of the edges with respect to which is a simple path, since M is a connected multipath. The statement would then follow from Remark 9.

Since $P(\mathtt{G})$ is finite poset, for every $x \in P(\mathtt{G})$ there is a maximal element $m \in P(\mathtt{G})$ such that $x \leq m$. Assume, for the sake of contradiction, that $\mathtt{M} \neq \mathtt{G}$. Then, there exists an edge $e \in E(\mathtt{G}) \setminus E(\mathtt{M})$. Consider the (multi-)path e defined by $E(\mathtt{e}) = \{e\}$. Then, as stated above, we have a maximal multipath $\mathtt{M}'$ such that $\mathtt{e} \leq \mathtt{M}'$. In particular, $\mathtt{M} \neq \mathtt{M}'$; this is not possible as we have a unique maximal element in $P(\mathtt{G})$. □

We pointed out in Example 14 that the coherently oriented polygonal graphs have a path poset which is almost a Boolean poset; more precisely, the path poset of $\mathtt{P}_n$ is a Boolean poset minus the maximum.
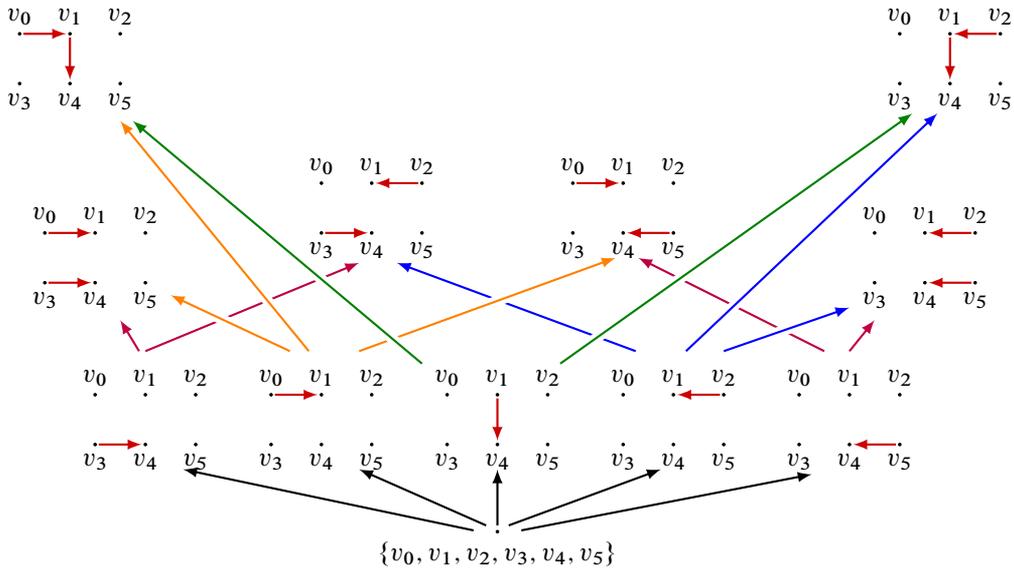
Figure 7: The path poset of the H–shaped digraph of Figure 6.
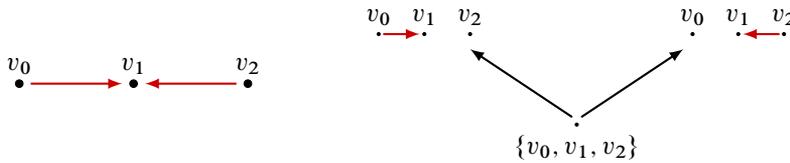
Figure 8: A noncoherent linear digraph with two edges and its path poset.
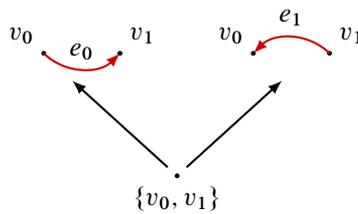
Figure 9: The path poset of the digon graph P_1 in Figure 3.

Example 19 Consider the digon graph P_1 illustrated in Figure 3. As depicted in Figure 9, its associated path poset consists of a minimum together with two elements corresponding to the two edges of the digon. It is easy to see that this poset is equivalent to the path poset associated to the linear digraph with two edges and noncoherent orientation illustrated in Figure 8.

We claim that, aside from the graph in Figure 8, the only connected graphs whose path poset is a Boolean poset minus its maximum are the coherently oriented polygonal graphs. The key observation to prove our claim is the following;

Remark 20 If G is a graph with n edges, and $P(\mathrm{G})$ is isomorphic to $\mathbb{B}(n)$ minus its maximum, then all the subgraphs of G but G itself must be multipaths. In fact, SSG(G) has exactly 2^n elements, which is the same number of elements in $\mathbb{B}(n)$. It follows that only one subgraph H of G does not belong to $P(\mathrm{G})$. Since $P(\mathrm{G})$ is downward closed in SSG(G), we must have $\mathrm{H} = \mathrm{G}$.

Proposition 21 *Let* G *be a connected graph with n edges. If $P(\mathrm{G})$ is isomorphic to $\mathbb{B}(n)$ minus its maximum, then $\mathrm{G} \cong \mathrm{P}_{n-1}$.*

Proof Take one of the n maximal elements in $P(\mathrm{G})$, say M. Note that $|E(\mathrm{M})| = n - 1$. Moreover, since M differs from G by a single edge, and G is connected, then M has at most two connected components.

Assume, for the sake of contradiction, that M is not connected. Each component of M is a simple path. It follows that G is either I_n — which is absurd by Proposition 18 — or contains the graph in Figure 8 (up to orientation reversal of both edges) as a subgraph. Note that the graph in Figure 8 cannot be G since $|E(\mathrm{G})| = n > 2$. It follows that any proper spanning subgraph containing a copy of the graph in Figure 8 is a subgraph different than G which belongs to SSG(G) but not to $P(\mathrm{G})$. This contradicts Remark 20.

From our argument above, it follows that M is connected, and thus isomorphic to I_n (since it is a multipath). So either $\mathrm{G} \cong \mathrm{P}_{n-1}$, $\mathrm{G} \cong \mathrm{I}_n$ (absurd), or again it contains a copy of the graph in Figure 8. The latter case can be excluded with the same argument as above. $\qquad\square$

3 Digraph (co)homologies

The goal of this section is to outline a rather general framework within which to define cohomology theories of directed graphs, using *poset homology* [Chandler 2019] as main tool — see also Remark 27. For the sake of being self-contained, and also for clarity, we provide a fairly detailed exposition of the construction of poset homology for posets of subgraphs. This is carried out in the first subsection. As the aforementioned construction might, a priori, depend upon the choice of a sign assignment on the considered posets, we further explore such dependence in the second subsection. We point out here that more general cohomology theories of posets can be used to obtain similar digraph cohomologies; we explore it in Section 6.

3.1 A poset homology

In this subsection we define, given a special type of poset coherently assigned to each digraph, and a choice of a sign assignment (see Definition 22), a cohomology theory for directed graphs; the cohomology theory depends on many choices and the functorial discussion is postponed to Section 5. The construction presented here has been inspired by [Turner and Wagner 2012], on one side, and by [Helme-Guizon and Rong 2005], on the other side. In the former paper homologies for digraphs were defined using the path poset functor — see Definition 12 and Remark 17; while in [Helme-Guizon and Rong 2005] a homology for nonoriented graphs was obtained via a construction similar to [Khovanov 2000]. Poset homology [Chandler 2019] interpolates between the two constructions.

Recall from Definition 6(1) that a square in a poset $(S, \lhd)$ is given by elements x, y, y', and z such that $y \neq y'$, $x \,\widetilde{\lhd}\, y \,\widetilde{\lhd}\, z$, and $x \,\widetilde{\lhd}\, y' \,\widetilde{\lhd}\, z$, where $\widetilde{\lhd}$ denotes the covering relation in S. Let $\mathbb{Z}_2$ be the cyclic group on two elements.

Definition 22 A *sign assignment* on a poset $(S, \lhd)$ is an assignment of elements $\epsilon_{x,y} \in \mathbb{Z}_2$ to each pair of elements $x, y \in S$ with $x \,\widetilde{\lhd}\, y$ such that the equation

(1) $$\epsilon_{x,y} + \epsilon_{y,z} \equiv \epsilon_{x,y'} + \epsilon_{y',z} + 1 \bmod 2$$

holds for each square $x \,\widetilde{\lhd}\, y$, $y' \,\widetilde{\lhd}\, z$.

Observe that the restriction of a sign assignment to a subposet is a sign assignment. In general the existence of a sign assignment on a given poset is not clear. However, for the spanning subgraphs poset — or, better, for Boolean posets — and their subposets, there is an easy sign assignment:

Example 23 Let $\mathtt{G}$ be a graph with a fixed total ordering $\lhd$ on the set of edges $E(\mathtt{G})$. Recall from Notation 2 and Example 4 that $\mathtt{H} \prec \mathtt{H}'$ in $\mathrm{SSG}(\mathtt{G})$ if, and only if, $\mathtt{H}' = \mathtt{H} \cup e$. Then, we can define a sign assignment on the poset $\mathrm{SSG}(\mathtt{G})$ as

$$\epsilon(\mathtt{H}, \mathtt{H}') := \#\{e' \in E(\mathtt{H}) \mid e' \lhd e\} \bmod 2,$$

where $\mathtt{H}' = \mathtt{H} \cup e$. The verification is straightforward, but the reader may consult eg [Khovanov 2000].

The following definition will be used to define the cochain complexes.

Definition 24 Let $P \subseteq \mathrm{SG}(\mathtt{G})$ be a faithful subposet. We define the *level* of an element $\mathtt{H} \in P$ as

$$\ell(\mathtt{H}) = \#E(\mathtt{H}) + \#V(\mathtt{H}) - \min\{\#E(\mathtt{H}') + \#V(\mathtt{H}') \mid \mathtt{H}' \in P\}.$$

Note that the level of an element $\mathtt{H} \in P \subseteq \mathrm{SG}(\mathtt{G})$ if P has a minimum is just the difference between the distances of $\mathtt{H}$ and the minimum of P, respectively, from the minimum of $\mathrm{SG}(\mathtt{G})$ in $\mathtt{Hasse}(\mathrm{SG}(\mathtt{G}))$. Note also that if $P = \mathrm{SG}(\mathtt{G})$, $P(\mathtt{G}) \subseteq \mathrm{SSG}(\mathtt{G})$ then ℓ is the number of edges.

Recall from Remark 3 that a poset $(S, \lhd)$ can be seen as a category $\boldsymbol{S}$ with set of objects S, and the set of morphisms between x and y containing a single element if and only if $x \lhd y$ or $x = y$.

Remark 25 Let $\boldsymbol{C}$ be a small category. For each square $x \,\widetilde{\lhd}\, y$, $y' \,\widetilde{\lhd}\, z$ in $(S, \lhd)$ and any covariant functor $\mathcal{F} \colon \boldsymbol{S} \to \boldsymbol{C}$, we have

$$\mathcal{F}(y \,\widetilde{\lhd}\, z) \circ \mathcal{F}(x \,\widetilde{\lhd}\, y) = \mathcal{F}(x \lhd z) = \mathcal{F}(y' \,\widetilde{\lhd}\, z) \circ \mathcal{F}(x \,\widetilde{\lhd}\, y').$$

In other words, all functors preserve the commutativity of the squares in $(S, \lhd)$.

Let $\boldsymbol{A}$ be an additive category, $P \subseteq \mathrm{SG}(\mathtt{G})$ squared and faithful and ϵ a sign assignment on P. Given a covariant functor $\mathcal{F} \colon \boldsymbol{P} \to \boldsymbol{A}$, we can define the cochain groups

$$C^n_{\mathcal{F}}(P) := \bigoplus_{\substack{\mathtt{H} \in P \\ \ell(\mathtt{H})=n}} \mathcal{F}(\mathtt{H})$$

and the differentials

$$d^n = d^n_{\mathcal{F}} := \sum_{\substack{\mathtt{H} \in P \\ \ell(\mathtt{H})=n}} \sum_{\substack{\mathtt{H}' \in P \\ \mathtt{H} \prec \mathtt{H}'}} (-1)^{\epsilon(\mathtt{H},\mathtt{H}')} \mathcal{F}(\mathtt{H} \prec \mathtt{H}').$$

Note that the differentials d^n, and therefore the cochain complexes, depend, a priori, upon the choice of the sign assignment ϵ. However, in the cases we are interested in, this choice does not affect the isomorphism type of the complexes — see Corollary 36. We will further discuss this topic in Section 3.2 below. We now give the proof that the defined complexes are indeed cochain complexes.

In the proof of the theorem below, P being squared plays an essential role; otherwise, the square of the differential might not be zero.

Theorem 26 *Let A be an additive category, $P \subseteq \mathrm{SG}(\mathtt{G})$ a squared and faithful poset, and ϵ a sign assignment on P. Then, for any $n \in \mathbb{N}$ and any covariant functor $\mathcal{F} \colon P \to A$, we have $d^n \circ d^{n-1} \equiv 0$. In particular, $(C^*_{\mathcal{F}}(P), d^*)$ is a cochain complex.*

Proof Fix a natural number n; then

$$C^n_{\mathcal{F}}(P) = \bigoplus_{\substack{\mathtt{H} \in P \\ \ell(\mathtt{H})=n}} \mathcal{F}(\mathtt{H}).$$

Let $\pi_{\mathtt{H}} \colon C^*_{\mathcal{F}}(P) \to \mathcal{F}(\mathtt{H})$ and $\iota_{\mathtt{H}} \colon \mathcal{F}(\mathtt{H}) \to C^*_{\mathcal{F}}(P)$ be the projection onto $\mathcal{F}(\mathtt{H})$ and the inclusion of $\mathcal{F}(\mathtt{H})$ in $C^*_{\mathcal{F}}(P)$ as direct summand, respectively. Note that the composition $d^n \circ d^{n-1}$ equals 0 if, and only if, the composition $\pi_{\mathtt{H}''} \circ d^n \circ d^{n-1}$ is trivial for all $\mathtt{H}'' \in P$ such that $\ell(\mathtt{H}'') = n+1$. In particular, $d^n \circ d^{n-1} \equiv 0$ if there are no $\mathtt{H}'' \in P$ with $\ell(\mathtt{H}'') = n+1$.

Every element of $C^{n-1}_{\mathcal{F}}(P)$ is a linear combination of elements of $\mathcal{F}(\mathtt{H})$ for $\mathtt{H}$ ranging in P with $\ell(\mathtt{H}) = n-1$, and d is linear. Thus, if the composition $\pi_{\mathtt{H}''} \circ d^n \circ d^{n-1} \circ \iota_{\mathtt{H}}$ equals 0 for all $\mathtt{H}, \mathtt{H}''$ such that $\ell(\mathtt{H}) + 2 = \ell(\mathtt{H}'') = n+1$, then $d^n \circ d^{n-1} \equiv 0$. We can factor the map $\pi_{\mathtt{H}''} \circ d^n \circ d^{n-1} \circ \iota_{\mathtt{H}}$ through $C^n_{\mathcal{F}}(P)$, and write

$$\pi_{\mathtt{H}''} \circ d^n \circ d^{n-1} \circ \iota_{\mathtt{H}} = \sum_{\mathtt{H} \prec \mathtt{H}' \prec \mathtt{H}''} (\pi_{\mathtt{H}''} \circ d^n \circ \iota_{\mathtt{H}'}) \circ (\pi_{\mathtt{H}'} \circ d^{n-1} \circ \iota_{\mathtt{H}}).$$

The right-hand side of the above equation vanishes if there is no $\mathtt{H}'$ such that $\mathtt{H} \prec \mathtt{H}' \prec \mathtt{H}''$. It follows that it is sufficient to check this case.

Since P is squared, if there is $\mathtt{H} \prec \mathtt{H}'_1 \prec \mathtt{H}''$, then there a unique $\mathtt{H}'_2 \ (\neq \mathtt{H}'_1)$ such that $\mathtt{H} \prec \mathtt{H}'_2 \prec \mathtt{H}''$. In other words, $\mathtt{H}, \mathtt{H}'_1, \mathtt{H}'_2, \mathtt{H}''$ form a square in P. Thus, we obtain

$$\begin{aligned}
\pi_{\mathtt{H}''} &\circ d^n \circ d^{n-1} \circ \iota_{\mathtt{H}} \\
&= (\pi_{\mathtt{H}''} \circ d^n \circ \iota_{\mathtt{H}'_1}) \circ (\pi_{\mathtt{H}'_1} \circ d^{n-1} \circ \iota_{\mathtt{H}}) + (\pi_{\mathtt{H}''} \circ d^n \circ \iota_{\mathtt{H}'_2}) \circ (\pi_{\mathtt{H}'_2} \circ d^{n-1} \circ \iota_{\mathtt{H}}) \\
&= (-1)^{\epsilon(\mathtt{H},\mathtt{H}'_1)+\epsilon(\mathtt{H}'_1,\mathtt{H}'')} \mathcal{F}(\mathtt{H}'_1 \prec \mathtt{H}'') \circ \mathcal{F}(\mathtt{H} \prec \mathtt{H}'_1) + (-1)^{\epsilon(\mathtt{H},\mathtt{H}'_2)+\epsilon(\mathtt{H}'_2,\mathtt{H}'')} \mathcal{F}(\mathtt{H}'_2 \prec \mathtt{H}'') \circ \mathcal{F}(\mathtt{H} \prec \mathtt{H}'_2) \\
&= ((-1)^{\epsilon(\mathtt{H},\mathtt{H}'_1)+\epsilon(\mathtt{H}'_1,\mathtt{H}'')} + (-1)^{\epsilon(\mathtt{H},\mathtt{H}'_2)+\epsilon(\mathtt{H}'_2,\mathtt{H}'')}) \mathcal{F}(\mathtt{H}'_2 \prec \mathtt{H}'') \circ \mathcal{F}(\mathtt{H} \prec \mathtt{H}'_2),
\end{aligned}$$

where the last equality is due to the fact the functor $\mathcal{F}$ preserves the commutative squares in P — see Remark 25. The result now follows immediately as ϵ is a sign assignment on P. □

Remark 27 The definition of the cochain complex $C_{\mathcal{F}}(P)$ relies on the structure of the input graph $\mathtt{G}$, via the associated squared and faithful poset $P \subseteq \mathrm{SG}(\mathtt{G})$, on the choice of the functor $\mathcal{F}$, and on a sign assignment ϵ on P. More in general, the same machinery can be applied without graphs but dealing only with a certain type of posets. This viewpoint was taken by Chandler, and we refer the reader to [Chandler 2019] for a more comprehensive discussion. For completeness we pursue our independently developed approach. In particular, we shall provide an independent proof of functoriality with respect to graphs, extending the generality to include also coefficients systems, in Section 5.

We conclude the section by observing that the general discussion of this section can be applied to the case of $\mathcal{G} \colon P \to A$ a contravariant functor. All proofs are straightforward adaptations of the proofs in the case of covariant functors.

3.2 Existence and uniqueness of sign assignments

The cochain complexes defined in the previous subsection may depend on the choice of the sign assignment. In this subsection, we see that this is actually not the case for a quite general class of posets, including path posets.

A sign assignment on $(S, \triangleleft)$ can be seen as a map $\epsilon \colon E(\mathtt{Hasse}(S, \triangleleft)) \to \mathbb{Z}_2$ such that (1) holds for each square $x \,\widetilde{\triangleleft}\, y$, $y' \,\widetilde{\triangleleft}\, z$ of S. Consider the Hasse graph $\mathtt{Hasse}(S, \triangleleft)$ of a poset $(S, \triangleleft)$ as a CW–complex (formally, by taking its geometric realisation).

Definition 28 Given a poset $(S, \triangleleft)$ define $\mathcal{K}(S, \triangleleft)$ as the CW–complex obtained from $(S, \triangleleft)$ by attaching to the (geometric realisation of the) Hasse graph $\mathtt{Hasse}(S, \triangleleft)$ a 2–cell $e_{x,y,y',z}$ for each square $x \,\widetilde{\triangleleft}\, y$, $y' \,\widetilde{\triangleleft}\, z$ in $(S, \triangleleft)$.

We will now show that the existence and uniqueness of a sign assignment on a poset $(S, \triangleleft)$ depends only upon the topological structure of the CW–complex $\mathcal{K}(S, \triangleleft)$. Denote by $(C^*_{\mathrm{CW}}(\mathcal{K}(S, \triangleleft); \mathbb{Z}_2), d^*_{\mathrm{CW}})$ the CW–cochain complex of $\mathcal{K}(S, \triangleleft)$, with respect to the given CW–structure, with coefficients in $\mathbb{Z}_2$. We can interpret the sign assignments as cochains in the CW–cochain complex associated to $\mathcal{K}(S, \triangleleft)$:

Lemma 29 *Let ϵ be a sign assignment on a poset $(S, \triangleleft)$, and denote by ψ the 2–cochain which associates $1 \in \mathbb{Z}_2$ to each 2–cell in $\mathcal{K}(S, \triangleleft)$. Then, ϵ defines a cochain $a(\epsilon) \in C^1_{\mathrm{CW}}(\mathcal{K}(S, \triangleleft); \mathbb{Z}_2)$ such that $d_{\mathrm{CW}}\, a(\epsilon) = \psi$. Moreover, for each $a \in C^1_{\mathrm{CW}}(\mathcal{K}(S, \triangleleft); \mathbb{Z}_2)$ such that $d_{\mathrm{CW}}\, a = \psi$ there is a unique sign assignment ϵ such that $a = a(\epsilon)$.*

Proof A 1–cocycle a with values in $\mathbb{Z}_2$ is a map from the set of 1–cells (which are the edges of the Hasse graph, in our case) to $\mathbb{Z}_2$. Since the edges of the Hasse graph correspond to the pairs in the covering relations, this is equivalent to the assignment of an element $\mathbb{Z}_2$ for each pair (x, y) such that

$x \tilde{\lhd} y$. It is left to show that the differential of a is ψ if, and only if, (1) holds for each square. Note that $d_{\mathrm{CW}} a(e) = \sum_{(x,y) \in \partial e} a(x, y)$ for 2–cells e; therefore $d_{\mathrm{CW}} a = \psi$ if, and only if,

$$a(x, y) + a(y, z) + a(x, y') + a(y', z) \equiv 1 \mod 2$$

for each square $x \tilde{\lhd} y$, $y' \tilde{\lhd} z$, concluding the proof. $\qquad\square$

It is easy to see that a poset $(S, \lhd)$ admits a sign assignment if the CW–complex $\mathcal{K}(S, \lhd)$ has trivial second homology group:

Proposition 30 *Let* $(S, \lhd)$ *be a poset. If* $\mathrm{H}^2_{\mathrm{CW}}(\mathcal{K}(S, \lhd); \mathbb{Z}_2) = 0$, *then there exists a sign assignment on* $(S, \lhd)$.

Proof Consider the cochain $\psi \colon C^{\mathrm{CW}}_2(\mathcal{K}(S, \lhd); \mathbb{Z}_2) \to \mathbb{Z}_2$ assigning $1 \in \mathbb{Z}_2$ to each 2–cell of $\mathcal{K}(S, \lhd)$. Since $\mathcal{K}(S, \lhd)$ has no 3–cells, $d_{\mathrm{CW}}(\psi) \equiv 0$ and hence ψ is a cocycle. Since, by assumption, we have $\mathrm{H}^2_{\mathrm{CW}}(\mathcal{K}(S, \lhd); \mathbb{Z}_2) = 0$, every 2–cocycle is a coboundary. Thus, there is $a \in C^1_{\mathrm{CW}}(\mathcal{K}(S, \lhd); \mathbb{Z}_2)$ such that $d_{\mathrm{CW}} a = \psi$. The statement now follows directly from the second part of Lemma 29. $\qquad\square$

The above proposition provides a condition for a poset to admit a sign assignment. We now describe when also the uniqueness is satisfied. First, we introduce the notion of isomorphisms of sign assignments.

Definition 31 Let ϵ, ϵ' be sign assignments on a poset $(S, \lhd)$. An *isomorphism of sign assignments* *between ϵ and ϵ'* is a map $\eta \colon S = V(\mathtt{Hasse}(S, \lhd)) \to \mathbb{Z}_2$ such that

(2) $$\eta(x) + \epsilon'_{x,y} = \epsilon_{x,y} + \eta(y) \mod 2$$

holds for all $x \tilde{\lhd} y$.

Roughly speaking, an isomorphism of sign assignments is a map $\eta \colon S \to \mathbb{Z}_2$ such that the elements of $\mathbb{Z}_2$ on the edges of the square

$$
\begin{array}{ccc}
x & \xrightarrow{\epsilon_{x,y}} & y \\
\eta_x \downarrow & & \downarrow \eta_y \\
x & \xrightarrow[\epsilon'_{x,y}]{} & y
\end{array}
$$
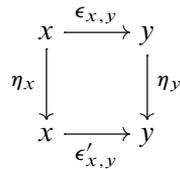
add up to $0 \in \mathbb{Z}_2$. Intuitively this condition encodes the "commutativity" of such squares. We can now provide a uniqueness result for sign assignments on posets — compare with [Putyra 2014, Lemma 5.7].

Proposition 32 *Let ϵ and ϵ' be two sign assignments on a poset $(S, \lhd)$. If* $\mathrm{H}^1_{\mathrm{CW}}(\mathcal{K}(S, \lhd); \mathbb{Z}_2) = 0$, *then there is an isomorphism η of sign assignments from ϵ to ϵ'.*

Proof Let $a(\epsilon), a(\epsilon')$ be the 1–cochains corresponding to ϵ, ϵ' as in Lemma 29. Notice that

$$d_{\mathrm{CW}}(a(\epsilon) - a(\epsilon')) = d_{\mathrm{CW}}(a(\epsilon)) - d_{\mathrm{CW}}(a(\epsilon')) = \psi - \psi = 0,$$
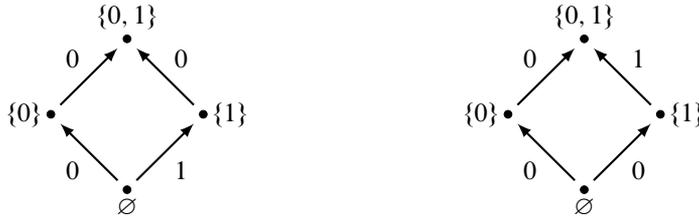
Figure 10: Two (isomorphic) sign assignments on the poset $(\wp(\{0, 1\}), \subseteq)$.

where ψ is the usual 2–cocycle assigning $1 \in \mathbb{Z}_2$ to each face of $\mathscr{K}(S, \lhd)$. Since, by assumption, $\mathrm{H}^1_{\mathrm{CW}}(\mathscr{K}(S, \lhd); \mathbb{Z}_2) = 0$, we must have $a(\epsilon) - a(\epsilon') = d_{\mathrm{CW}}(\eta)$ for some $\eta \in C^0_{\mathrm{CW}}(\mathscr{K}(S, \lhd); \mathbb{Z}_2)$. We can see η as a map

$$\eta \colon \{\text{0-cells of } \mathscr{K}(S, \lhd)\} = V(\texttt{Hasse}(S, \lhd)) \to \mathbb{Z}_2.$$

Moreover, the equality $a(\epsilon) - a(\epsilon') = d_{\mathrm{CW}}(\eta)$ applied to each edge of the Hasse graph gives precisely the condition of isomorphisms of sign assignment in (2), concluding the proof. □

Example 33 Consider the two sign assignments on the Boolean poset $(\wp(\{0, 1\}), \subseteq)$ illustrated in Figure 10. By definition $\mathscr{K}(\wp(\{0, 1\}), \subseteq)$ is a disk.

Thus, $\mathrm{H}^1_{\mathrm{CW}}(\mathscr{K}(\wp(\{0, 1\}), \subseteq); \mathbb{Z}_2) = 0$. This implies the uniqueness of the sign assignment up to isomorphism in this case. It is not difficult, in this case, to produce a concrete isomorphism:

$$\eta \colon V(\texttt{Hasse}(\wp(\{0, 1\}), \subseteq)) \to \mathbb{Z}_2, \quad v \mapsto \begin{cases} 1 & \text{if } v = \{1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Recall the definition of downward closed subposet $(S', \lhd_{|S' \times S'})$ of a poset $(S, \lhd)$ — see Definition 5. As a consequence of Proposition 32, we get the following:

Theorem 34 *Let P be a downward (or upward) closed subposet of $\mathrm{SSG}(\mathtt{G})$. Then, any two sign assignments ϵ and ϵ' on P are isomorphic.*

Proof The poset $\mathrm{SSG}(\mathtt{G})$ is a Boolean poset — see Example 4. It follows that its Hasse diagram is the 1–skeleton of an n–dimensional cube, with $n = |E(\mathtt{G})|$. This implies that $\mathscr{K}(\mathrm{SSG}(\mathtt{G}), <)$, which for $n \geq 2$ is the union of the 1– and 2–skeletons of an n–dimensional cube, has trivial homology groups in degree $i = 1$ — this being trivial for $n = 1$. Therefore, the statement follows in this case from Proposition 32. Note that there always exists a sign assignment on Boolean posets; this is true for homological reasons for $n \neq 3$, see Proposition 30, and we can even define it explicitly for all n — see for instance [Khovanov 2000, Section 3].

In the general case in which P is a downward or upward closed proper subposet of a cube, observe that it is squared (see Example 7) and it contains either the minimum or the maximum of the cube. Furthermore, the sub-CW–complex $\mathscr{K}(P, <) \subset \mathscr{K}(\mathrm{SSG}(\mathtt{G}), <)$ retracts onto the minimum (or the maximum); hence,

again by [Proposition 32](), the uniqueness of the sign assignment up to isomorphism follows. For a detailed proof of this fact we refer the interested reader to [Chandler 2019, Theorem 4.5] (to be read in conjunction with Theorems 2.9 (3) and 5.14 of that work).   □

In particular, if $P = P(\mathtt{G})$ is the path poset of a digraph $\mathtt{G}$ — see [Remark 11]() — we get:

Corollary 35 *Any two sign assignments ϵ and ϵ' on $P(\mathtt{G})$ are isomorphic.*

We conclude the section with an application to the cohomology theories defined in [Section 3.1]().

Corollary 36 *Let $\mathtt{G}$ be a digraph, $P \subseteq \mathrm{SSG}(\mathtt{G})$ be a downward (or upward) closed subposet, and $\mathcal{F} \colon P \to A$ a covariant functor to an additive category A. Then, the cochain complex $(C_{\mathcal{F}}^*(P), d^*)$ does not depend, up to isomorphism, on the choice of the sign assignment on P.*

Proof Let ϵ and ϵ' be two sign assignments on P. Denote by $(C_{\mathcal{F}}^*(P), d_{\epsilon}^*)$ and $(C_{\mathcal{F}}^*(P), d_{\epsilon'}^*)$ the associated cochain complexes defined using ϵ and ϵ', respectively. By [Theorem 34](), any two sign assignments on P are isomorphic. Let η be an isomorphism between them and define the map

$$\Phi \colon (C_{\mathcal{F}}^*(P), d_{\epsilon}^*) \to (C_{\mathcal{F}}^*(P), d_{\epsilon'}^*),$$

as $\Phi := \bigoplus_{\mathtt{H}} \Phi_{\mathtt{H}}$, where $\Phi_{\mathtt{H}} = (-1)^{\eta(\mathtt{H})} \mathrm{Id}_{\mathcal{F}(\mathtt{H})}$. Observe that this clearly gives an isomorphism of modules. Furthermore, the commutativity of Φ with the differentials is immediate by the definition of isomorphisms of sign assignments (see [(2)]()); hence it provides an isomorphism of chain complexes, concluding the proof.   □

4 Multipath cohomology

The goal of this section is to define *multipath cohomology* of directed graphs using poset homology. This will be achieved in the first subsection, whereas the second subsection is devoted to providing some computations. In particular, we will see that the multipath cohomology may be nontrivial when evaluated on trees.

4.1 Multipath cohomology

In this subsection we specialise the general construction described in [Section 3.1]() by taking as poset the path poset $P(\mathtt{G})$ ([Definition 12]()), and defining an explicit functor $\mathcal{F}_{A,M} \colon P(\mathtt{G}) \to R\text{–}\mathbf{Mod}$. In order to define $\mathcal{F}_{A,M}$ and an explicit sign assignment on $P(\mathtt{G})$, we need some auxiliary data, more precisely, an ordering on the vertices of $\mathtt{G}$. An *ordered digraph* is a digraph with a fixed well-ordering[2] of the vertices. Note that the order of the vertices induces an order of the edges of $\mathtt{G}$; this order is given by ordering the pairs source-target lexicographically. We can use the ordering on the vertices of an ordered graph to index the connected components of any subgraph $\mathtt{H} < \mathtt{G}$; the order being given according to the minimum of the vertices belonging to each component.

---

[2]Every nonempty subset has a minimal element.

Notation 37 Given a subgraph H of an ordered graph G, we will denote by $\text{index}_H(c)$ the position of a connected component c of H with respect to the aforementioned order — we start the count at 0. More precisely, if the ordered connected components of H are $c_0 < c_1 < \cdots < c_k$, then $\text{index}_H(c) = i$ if $c = c_i$. Note that the definition of index is well-posed. Whenever H is clear from the context, we will remove it from the notation of the index.

Definition 38 Consider $H \in SG(G)$ and $e \in E(G) \setminus E(H)$ such that $s(e), t(e) \in V(H)$. The *source* (resp the *target*) *index of e with respect to H* is defined as

$$s(e, H) = \text{index}_H(c) \text{ such that } s(e) \in c \quad (\text{resp. } t(e, H) = \text{index}_H(c) \text{ such that } t(e) \in c).$$

The naming is motivated by the following facts: $\text{index}(s(e)) = s(e, G_\varnothing)$ and $\text{index}(t(e)) = t(e, G_\varnothing)$, where $G_\varnothing$ denotes the spanning subgraph of G with no edges.

With this notation in place we are now ready to define a sign assignment σ_e on $P(G)$:

$$(3) \qquad \sigma_e(H, H') = \begin{cases} t(e, H) + 1 & \text{if } H' = H \cup e \text{ and } t(e, H) > s(e, H), \\ s(e, H) & \text{if } H' = H \cup e \text{ and } s(e, H) > t(e, H) \end{cases} \mod 2.$$

Lemma 39 *The function σ_e in* (3) *gives a sign assignment on $P(G)$.*

For the sake of presentation we moved the proof of the lemma to the Appendix.

Remark 40 More generally, observe that, for each faithful and squared poset $P \subseteq P(G)$, the restriction $\sigma_{e|P}$ is a sign assignment. Here we are using the faithfulness of P in $P(G)$ (and, by Proposition 8, in $SSG(G)$) to be sure that the covering relation only amounts to the addition of a single edge.

We now construct an explicit functor $\mathcal{F}_{A,M} : P(G) \to R\text{–}\mathbf{Mod}$. From now on, R will denote a commutative ring with identity, A an associative unital R–algebra and M an (A, A)–bimodule, ie M is both a left and a right A–module, and the two actions are compatible.

Let G be an ordered graph and let $v_0 \in V(G)$ be the minimum with respect to the given ordering. Given a multipath $H < G$, to each connected component of H but the one containing the vertex v_0 we associate a copy of A, and to the component containing v_0 we associate a copy of M. Then we take the ordered tensor product. More concretely, if $c_0 < \cdots < c_k$ is the set of ordered connected components of H, we define

$$(4) \qquad \mathcal{F}_{A,M}(H) := M_{c_0} \otimes_R A_{c_1} \otimes_R \cdots \otimes_R A_{c_k},$$

where all the modules are labelled by the respective component.

Assume $H' = H \cup e$. Denote by $c_0, \ldots, c_k$ the ordered components of H, denote by $c'_0, \ldots, c'_{k-1}$ the ordered components of H', and assume that the addition of e merges c_i and c_j. Then, for each $h = 0, \ldots, k-1$, there is a natural identification

$$(5) \qquad c'_h = \begin{cases} c_h & \text{if } 0 \le h < i \text{ or } i < h < j, \\ c_i \cup e \cup c_j & \text{if } h = i, \\ c_{h+1} & \text{if } j \le h < k \end{cases}$$
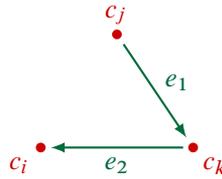
Figure 11: A schematic description of case (A) subcase (b).

for some $0 \leq i < j \leq k$. Using this identification, we define $\mu_{\mathtt{H} \prec \mathtt{H}'} : \mathcal{F}_{A,M}(\mathtt{H}) \to \mathcal{F}_{A,M}(\mathtt{H}')$ as

$$\mu_{\mathtt{H} \prec \mathtt{H}'}(a_0 \otimes \cdots \otimes a_k) = a_0 \otimes \cdots \otimes a_{s(e,\mathtt{H})-1} \otimes a_{s(e,\mathtt{H})} \cdot a_{t(e,\mathtt{H})} \otimes a_{s(e,\mathtt{H})+1} \otimes \cdots \otimes \widehat{a_{t(e,\mathtt{H})}} \otimes \cdots \otimes a_{k-1} \otimes a_k,$$

where $\widehat{a_{t(e,\mathtt{H})}}$ indicates the $a_{t(e,\mathtt{H})}$ is missing. We set

$$(6) \qquad \mathcal{F}_{A,M}(\mathtt{H} \preceq \mathtt{H}') := \begin{cases} \mu_{\mathtt{H} \prec \mathtt{H}'} & \text{if } \mathtt{H} \prec \mathtt{H}', \\ \mathrm{Id}_{\mathcal{F}_{A,M}(\mathtt{H})} & \text{if } \mathtt{H} = \mathtt{H}'. \end{cases}$$

Equations (4) and (6) describe a functor

$$(7) \qquad \mathcal{F}_{A,M} : \boldsymbol{P}(\mathtt{G}) \to R\text{–}\mathbf{Mod}$$

from the category $\boldsymbol{P}(\mathtt{G})$ associated to the path poset $P(\mathtt{G})$ to the additive category R-**Mod** of left R–modules. In fact, we have the following:

Lemma 41 *Let* $\mathtt{G}$ *be an ordered digraph. The assignment* $\mathcal{F}_{A,M}(\mathtt{H} \prec \mathtt{H}') := \mu_{\mathtt{H} \prec \mathtt{H}'}$ *in* (6) *preserves all the commutative squares in* $\boldsymbol{P}(\mathtt{G})$ *— see Remark 25.*

Proof The possible configurations of squares in $P(G)$ are described in the proof of Lemma 39, contained in the Appendix — see Figures 16 and 17. We leave the full checking to the dedicated reader and present here only one case, namely case (A), subcase (b). The remaining checks can be dealt with similarly. In the case at hand we have $\mathtt{H}' = \mathtt{H} \cup \{e_1, e_2\}$ with

$$t(e_2, \mathtt{H}) = i < j = s(e_1, \mathtt{H}) < s(e_2, \mathtt{H}) = t(e_1, \mathtt{H}) = k.$$

The schematic description of this configuration is shown in Figure 11. Now, we compute the two compositions directly, and we obtain

$$\mathcal{F}_{A,M}(\mathtt{H} \cup e_1 \prec \mathtt{H}') \circ \mathcal{F}_{A,M}(\mathtt{H} \prec \mathtt{H} \cup e_1)(a_0 \otimes \cdots \otimes a_h) = a_0 \otimes \cdots \otimes (a_j a_k) a_i \otimes \cdots \otimes \hat{a}_j \otimes \cdots \otimes \hat{a}_k \otimes \cdots \otimes a_h,$$

$$\mathcal{F}_{A,M}(\mathtt{H} \cup e_2 \prec \mathtt{H}') \circ \mathcal{F}_{A,M}(\mathtt{H} \prec \mathtt{H} \cup e_2)(a_0 \otimes \cdots \otimes a_h) = a_0 \otimes \cdots \otimes a_j (a_k a_i) \otimes \cdots \otimes \hat{a}_j \otimes \cdots \otimes \hat{a}_k \otimes \cdots \otimes a_h.$$

The statement follows in this case by the associativity of A (and the definition of (A, A)–bimodule if $i = 0$). $\qquad \square$

Proposition 42 *Let* $\mathtt{G}$ *be an ordered digraph. The assignment* $\mathcal{F}_{A,M} : \boldsymbol{P}(\mathtt{G}) \to R\text{–}\mathbf{Mod}$ *defines a covariant functor.*

Proof It is clear that $\mathcal{F}_{A,M}$ preserves the identities. Let $f_{\mathtt{H},\mathtt{H}'} \colon \mathtt{H} \to \mathtt{H}'$ be a morphism in $\boldsymbol{P}(\mathtt{G})$. The morphism $f_{\mathtt{H},\mathtt{H}'}$ can be written as the composition of the covering morphisms $f_{\mathtt{H}_i,\mathtt{H}_{i+1}}$ for any given chain $\mathtt{H} = \mathtt{H}_0 \prec \mathtt{H}_1 \prec \cdots \prec \mathtt{H}_{n-1} = \mathtt{H}'$ in $\boldsymbol{P}(\mathtt{G})$ — this is well-defined since we have a unique morphism between two related objects in $\boldsymbol{P}(\mathtt{G})$; see Remark 3. We only have to show that the composition

$$\mathcal{F}_{A,M}(\mathtt{H}_{n-2} \prec \mathtt{H}_{n-1}) \circ \cdots \circ \mathcal{F}_{A,M}(\mathtt{H}_0 \prec \mathtt{H}_1)$$

depends only on $\mathtt{H} = \mathtt{H}_0$ and $\mathtt{H}' = \mathtt{H}_{n-1}$ and not on the chosen chain.

Note that $\mathtt{H}' = \mathtt{H} \cup \{e_1, \ldots, e_{n-1}\}$, and each chain corresponds to a choice of the order in which we add the edges $e_1, \ldots, e_{n-1}$ to $\mathtt{H}$. Therefore, the proof boils down to showing that we can switch the order in which we add two edges to $\mathtt{H}$. This is equivalent to showing that $\mathcal{F}_{A,M}$ preserves the commutative squares in $\boldsymbol{P}(\mathtt{G})$. Thus, the proposition follows directly from Lemma 41. $\qquad\square$

The above proof shows that both the poset $P(\mathtt{G})$ and its squared faithful subposets are, in the language of [Chandler 2019], diamond transitive. For a more general proof of this fact in the case of downward or upward closed subposets of $\mathrm{SSG}(\mathtt{G})$, or even more in general, the reader can consult [Chandler 2019].

We conclude this section with the following theorem which is an immediate consequence of Theorem 26, Lemma 39, and Proposition 42.

Theorem 43 *Given a graph $\mathtt{G}$ the graded R–module $C_\mu^*(\mathtt{G}; A, M) := C_{\mathcal{F}_{A,M}}^*(P(\mathtt{G}))$ endowed with the map $d^* := d_{\mathcal{F}_{A,M},\sigma}^*$ is a cochain complex.*

By Corollary 36, up to isomorphism of chain complexes, $(C_\mu^*(\mathtt{G}; A, M), d^*)$ does not depend on the choice of the sign σ_e.

Assume now that M is isomorphic to A as an (A, A)–bimodule. Then, the chain complex does not depend, up to isomorphism, on the given ordering of the vertices of the graph. In other words, the isomorphism class of $(C_\mu^*(\mathtt{G}; A, A), d^*)$ depends only on the underlying graph $\mathtt{G}$ and on the algebra A:

Proposition 44 *Let $\mathtt{G}$ be an ordered digraph. Then, the cochain complex $(C_\mu^*(\mathtt{G}; A, A), d^*)$ does not depend on the choice of the ordering on $V(\mathtt{G})$.*

Proof A permutation of the ordering on the vertices of $\mathtt{G}$ induces for each $\mathtt{H} \in P(\mathtt{G})$ a permutation on the factors appearing in $\mathcal{F}(\mathtt{H})$. There is an induced natural isomorphism of modules induced by the latter, which extends to an isomorphism of chain complexes; the commutativity with the differentials is clear up to sign. The statement now follows by Corollary 36. $\qquad\square$

Remark 45 More generally, the cochain complex $(C_\mu^*(\mathtt{G}; A, M), d^*)$ does not depend, up to isomorphisms, on the choice of the order on $V(\mathtt{G})$ preserving the minimum — which can be considered as a base vertex.

We are ready to give the main definition of the paper:

Definition 46 The *multipath cohomology* $H_\mu^*(G; A, M)$ of a digraph G with (A, M)–coefficients is the homology of the cochain complex $(C_\mu^*(G; A, M), d^*)$. When $A = M$ we simply write $H_\mu^*(G; A)$.

Consider the category **Digraph**$_*$ of pointed digraphs, ie digraphs with the choice of a base vertex, and morphisms of pointed digraphs, ie morphisms of digraphs that preserve the base vertex. Then, we can define multipath cohomology of a pointed digraph (G, v_0) with (A, M)–coefficients as the homology of the cochain complex $(C_\mu^*(G; A, M), d^*)$. Note that in the case $M \neq A$ we need to keep track of the base vertex because the associated cohomology groups $H_\mu^*(G; A, M)$ may depend upon this choice — see Remark 51.

We conclude this subsection by observing that the sign assignment on $SSG(G)$, given in Example 23, induces, by restriction, a sign assignment on the path poset $P(G)$. The cochain complex obtained from this sign assignment and the one obtained from σ_e are isomorphic. However, this is not true for more general subposets of $P(G)$ (as it depends on their topology) and the two constructions may lead to nonisomorphic cohomology theories of digraphs.

4.2 Computations and examples

In this section we provide some computations of multipath cohomology — see Table 1. Further calculations, new computational tools and more general results concerning the structure of multipath homology, are developed in [Caputi et al. 2023; 2024b].

For the whole section, unless otherwise specified, we will always implicitly assume both M and A to be the ground ring R, and $R = \mathbb{K}$ to be a field. Tensor products $\otimes$ will always be tensor products over $\mathbb{K}$.

We remark here that, from our computations, see Table 1, it follows that there exist trees with nontrivial multipath cohomology. Most digraph homology theories known to the authors — as path homology, clique homology and Hochschild homology of digraphs, or Turner–Wagner homology with constant coefficients — vanish on trees.

Our first example is the noncoherently oriented linear digraph on three vertices. In this case we provide the explicit computation of the multipath cohomology both with constant coefficients, that is, $M = A = \mathbb{K}$, and in a nonconstant setting, namely $M = A = \mathbb{K}[x]/(x^2)$. As we will see, these two coefficients provide different cohomologies, showing that the multipath cohomology actually depends on the choices of A and M. Note also that, being the base ring a field, this example additionally shows that the classical universal coefficients theorem is not sufficient to recover the cohomology computed using A from the cohomology computed using R.

Example 47 Let G be the noncoherent linear digraph on three vertices v_0, v_1, v_2 — see Figure 8. Application of the functor $\mathcal{F}_{A,A}$ on (the category associated to) its path poset $P(G)$ gives the following diagram of $\mathbb{K}$–modules:

$$A_{v_0} \otimes_\mathbb{K} A_{v_1} \otimes_\mathbb{K} A_{v_2} \xrightarrow{(m \otimes \mathrm{Id}_A) \oplus (\mathrm{Id}_A \otimes m)} A_{(v_0, v_1)} \otimes_\mathbb{K} A_{v_2} \oplus A_{v_0} \otimes_\mathbb{K} A_{(v_2, v_1)},$$

where we have decorated the modules with the components of the corresponding multipaths, and the arrows with the induced signs σ_e as by (3). The map on the left sends the elementary tensor product $a_0 \otimes a_1 \otimes a_2 \in A_{v_0} \otimes_{\mathbb{K}} A_{v_1} \otimes_{\mathbb{K}} A_{v_2}$ to the element $(a_0 \cdot a_1) \otimes a_2 \in A_{(v_0,v_1)} \otimes_{\mathbb{K}} A_{v_2}$, whereas the map on the right sends the same element to $a_0 \otimes (a_2 \cdot a_1) \in A_{v_0} \otimes_{\mathbb{K}} A_{(v_2,v_1)}$. If $A = \mathbb{K}$, using the identification $\mathbb{K} \otimes_{\mathbb{K}} \mathbb{K} \cong \mathbb{K}$ and the commutativity of $\mathbb{K}$, we get the cochain complex

$$0 \to \mathbb{K} \xrightarrow{d^0 = (\mathrm{Id}_{\mathbb{K}}, \mathrm{Id}_{\mathbb{K}})} \mathbb{K}^2 \xrightarrow{d^1 = 0} 0.$$

It is now straightforward that the homology of such a cochain complex is concentrated in degree 1 and is of dimension 1.

Now, take A to be the $\mathbb{K}$–algebra $\mathbb{K}[x]/(x^2)$. Fix the basis $e_0 = 1$ and $e_1 = x$ for A as a $\mathbb{K}$–vector space. The basis for a tensor product of copies of A will be given by elementary tensors of e_0 and e_1 ordered lexicographically. We can now write explicitly the matrix associated to the differential d^0 with respect to these bases, which yields a matrix M_{d^0} of rank 6 (over any field). Therefore, we have that $\dim(H^0_\mu(\mathbb{G}; A, A)) = \dim(\mathrm{Ker}(d^0)) = 2$, and that $\dim(H^1_\mu(\mathbb{G}; A, A)) = 8 - \dim(\mathrm{Img}(d^1)) = 2$, concluding our computations.

To facilitate the calculations in the remaining examples, we will use some basic notions of algebraic Morse theory; a general reference is [Kozlov 2008, Chapter 11, Section 3]. Roughly speaking, algebraic Morse theory gives a way to reduce a (co)chain complex by eliminating acyclic summands via changes of bases.

The theory works as follows: Consider a finitely generated complex of $\mathbb{K}$–vector spaces, say (C^*, d^*), and a basis $B_i = \{b^i_j\}_j$ of C^i as a $\mathbb{K}$–vector space for each i. With respect to these bases, the differential can be expressed as

$$d(b^i_j) = \sum_h c^{i+1}_{j,h} b^{i+1}_h$$

for some $c^{i+1}_{j,h} \in \mathbb{K}$. One can now construct a digraph $\mathbb{C}$ by taking $V(\mathbb{C}) = \bigcup_i B_i$, and $(b^i_k, b^j_h) \in E(\mathbb{C})$ if, and only if, $i = j - 1$ and the coefficient $c^{i+1}_{k,h}$ is nontrivial.

An *acyclic matching M* on a graph $\mathbb{C}$ is a subset of pairwise disjoint[3] edges of $\mathbb{C}$ such that the graph obtained from $\mathbb{C}$ by changing the orientations of the edges in M has no cycles, ie there are no embedded copies of P_n in $\mathbb{C}$.

The main result in algebraic Morse theory (see [Kozlov 2008, Theorem 11.24]) is that, given an acyclic matching M on $\mathbb{C}$, the complex (C^*, d^*) is quasi-isomorphic to a complex (C^*_M, d^*_M), where C^i_M is generated by all the b^i_j's that are not incident to the edges in M.

Remark 48 If M is an acyclic matching and $\{v \in V(\mathbb{C}) \mid v = s(e) \text{ or } v = t(e), \ e \in M\} = V(\mathbb{C})$, then the complex (C^*, d^*) has trivial homology.

---

[3]Two edges e and e' are said to be disjoint if the sets $\{s(e), t(e)\}$ and $\{s(e'), t(e')\}$ are disjoint.

Remark 49 If M is an acyclic matching and

$$V(\mathtt{C}) \setminus \{v \in V(\mathtt{C}) \mid v = s(e) \text{ or } v = t(e), \, e \in M\} \subseteq B_i$$

for a fixed i, then (C_M^*, d_M^*) is concentrated in degree i. Therefore, (C_M^*, d_M^*) has a trivial differential. Hence the homology of (C^*, d^*) is concentrated in degree i, and it is isomorphic to C_M^i.

In the following examples, for each digraph $\mathtt{G}$, we can take the graph $\mathtt{C}$ to be the Hasse graph of the path poset $P(\mathtt{G})$. This is due to the following two facts:

• All tensor products are taken over $\mathbb{K}$ and $A = M = \mathbb{K}$; hence $\mathcal{F}_{A,M}(\mathtt{H}) \cong \mathbb{K}$ has a single generator for each multipath $\mathtt{H}$ in the path poset.

• For each pair of multipaths $\mathtt{H}, \mathtt{H}'$ such that $\mathtt{H} \prec \mathtt{H}'$, the map $\mathcal{F}_{A,M}(\mathtt{H} \prec \mathtt{H}')$, under the identifications $\mathcal{F}_{A,M}(\mathtt{H}) \cong \mathbb{K}$ and $\mathcal{F}_{A,M}(\mathtt{H}') \cong \mathbb{K}$, can be taken to be the identity up to a sign.

We can now proceed with the computation of the multipath cohomology of the n–step graph $\mathtt{I}_n$.

Example 50 Let $\mathtt{I}_n$ be the n–step graph in Figure 1. We claim that $\mathrm{H}_\mu^*(\mathtt{I}_n; \mathbb{K}) = 0$ for all $n > 0$.

If $n = 0$, we have the degenerate case where $\mathtt{I}_n$ is just a vertex with no edges. By definition, the cochain complex $(C_\mu^*(\mathtt{I}_0; \mathbb{K}), d^*)$ is just a copy of $\mathbb{K}$ in degree 0 and has trivial differential. Hence, we have $\mathrm{H}_\mu^*(\mathtt{I}_0; \mathbb{K}) = \mathrm{H}_\mu^0(\mathtt{I}_0; \mathbb{K}) = \mathbb{K}$.

Let us turn back to the general computation. Notice that the path poset $P(\mathtt{I}_n)$ is a Boolean poset — see Example 13. Since $\mathcal{F}_{A,M}(\mathtt{H}) \cong \mathbb{K}$ for each multipath $\mathtt{H} \in P(\mathtt{I}_n)$, it follows that

$$C_\mu^k(\mathtt{I}_n; \mathbb{K}, \mathbb{K}) = \bigoplus_{\substack{\mathtt{H} \in P \\ \ell(\mathtt{H}) = k}} \mathcal{F}_{A,M}(\mathtt{H}) \cong \mathbb{K}^{\binom{n}{k}}$$

for each $k = 0, \ldots, n$. In other words, the resulting cochain complex $C_\mu^*(\mathtt{I}_n; \mathbb{K}, \mathbb{K})$ is of the form

$$0 \to \mathbb{K} \xrightarrow{d^0} \mathbb{K}^n \xrightarrow{d^1} \cdots \to \mathbb{K}^{\binom{n}{k}} \xrightarrow{d^{\binom{n}{k}}} \mathbb{K}^{\binom{n}{k+1}} \to \cdots \to \mathbb{K}^n \xrightarrow{d^n} \mathbb{K} \xrightarrow{d^{n+1}} 0.$$

An acyclic matching (the check of the nonexistence of cycles is left to the reader) on the Hasse graph of $P(\mathtt{I}_n) \cong \wp(\{0, \ldots, n-1\})$ is given by all edges $(s, s \cup \{0\})$ with $s \in \wp(\{1, \ldots, n-1\})$. Since each $s \in \wp(\{0, \ldots, n-1\})$ either contains 0 or does not, this matching touches all vertices of $\mathtt{Hasse}(P(\mathtt{I}_n))$, and our claim follows from Remark 48.

In the general case $A \neq \mathbb{K}$, the computation that $\mathrm{H}_\mu^*(\mathtt{I}_n; A) = 0$ is more convoluted. In Corollary 81 we will prove the claim for every unital algebra A and positive degrees. In [Caputi et al. 2023] we prove a more general result on the vanishing of multipath cohomology for $A = \mathbb{K}$.

In degree 0, the multipath cohomology is possibly not trivial — eg $\mathrm{H}_\mu^*(\mathtt{I}_n; A) \neq 0$; see Corollary 81. In the next remark we see that $\mathrm{H}_\mu^*(-; A, M)$, when $M \neq A$, depends on the choice of the base vertex.
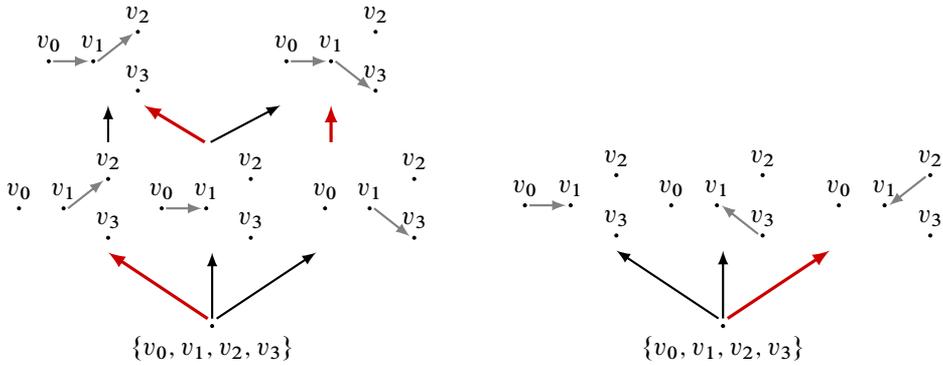
Figure 12: Acyclic matchings (in red and thicker) in the path posets of the graphs Y_1(left) and Y_2 (right) depicted in Figure 4(a) and (c).

Remark 51 When the bimodule M is not the R–algebra A itself, the multipath cohomology of a digraph G may depend upon the choice of the base vertex. As an example, let I_2 be the 2–step graph on vertices v and w and the only directed edge (v, w). Choose first v to be the base vertex; then, the associated cochain complex $\mathrm{C}_\mu(\mathrm{G}; A, M)$ is

$$0 \to M \otimes A \xrightarrow{d^0} M \to 0,$$

where d_0 is induced by the left action: $(m \otimes a) \mapsto m \cdot a$. If we choose the base vertex to be w, we get

$$0 \to A \otimes M \xrightarrow{d^0} M \to 0,$$

where now d_0 is induced by the right action: $(a \otimes m) \mapsto a \cdot m$. Therefore, if left and right action do not agree, then the homology groups may differ, in this case.
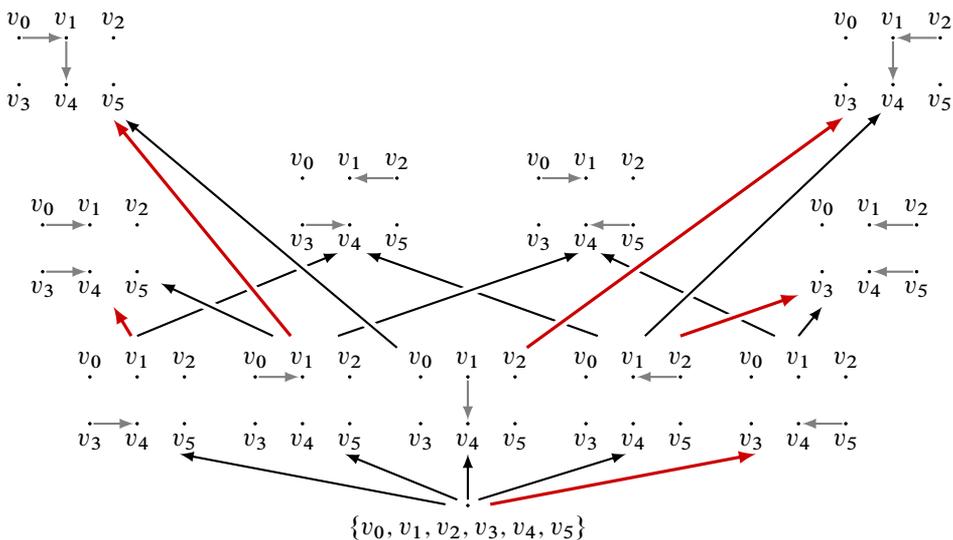
Figure 13: An acyclic matching (in red and thicker) in the path poset of the H–shaped digraph in Figure 6.

We proceed with the computation of the multipath cohomology groups of the examples in Table 1.

Example 52 Consider the graphs Y_1 and Y_2 depicted in Figure 4(a) and (c), respectively. Let us start with Y_1. In this case we have an acyclic matching on $\mathrm{Hasse}(P(Y_1))$ which touches all vertices (see Figure 12). It follows from Remark 48 that $H^*_\mu(Y_1; \mathbb{K}) = 0$.

Moving on to the graph Y_2, all (nonempty) acyclic matchings on $\mathrm{Hasse}(P(Y_2))$ consist of a single edge going from $(Y_2)_\varnothing$ (ie the multipath with no edges) to a multipath with a single edge (eg see Figure 12). This leaves only two vertices unmatched (ie not incident to the edges in the matching), both corresponding to multipaths with a single edge (and thus representing two generators in cohomological degree 1). It follows from Remark 49 that $H^*_\mu(Y_2; \mathbb{K}) = H^1_\mu(Y_2; \mathbb{K}) \cong \mathbb{K}^2$.

Example 53 Let G be the digraph illustrated in Figure 6. An acyclic matching M on the Hasse graph associated to $P(G)$ is shown in Figure 13. There are only two multipaths not incident to the edges in M, both with two edges. It follows from Remark 49 that $H^*_\mu(G; \mathbb{K}) = H^2_\mu(G; \mathbb{K}) \cong \mathbb{K}^2$.

5 Functorial properties and exact sequences

The aim of the following section is to better understand the functorial properties of multipath cohomology. The machinery developed here can be adapted also to other contexts and to the general framework described in Section 3. As an application, in Section 7.2 we will clarify the relationship between multipath cohomology and chromatic homology.

Let **Digraph**(n) be the subcategory of the category **Digraph** consisting of digraphs with precisely n vertices, and morphisms of digraphs. In this section, among others, we prove the following functoriality result, which is one of the main results of the paper.

Theorem 54 (Theorem 1) *Let R–**Alg** be the category of unital R–algebras, **Digraph**$^{\mathrm{op}}(n)$ the opposite category of **Digraph**(n), and R–**Mod**$^{\mathrm{gr}}$ the category of graded R–modules. Then, multipath cohomology*

$$H_\mu \colon \mathbf{Digraph}^{\mathrm{op}}(n) \times R\text{--}\mathbf{Alg} \to R\text{--}\mathbf{Mod}^{\mathrm{gr}}$$

is a bifunctor for all $n \in \mathbb{N}$.

We start by discussing the functoriality of multipath cohomology with respect to the algebras.

Proposition 55 *Let G be a graph, and let P be a squared and faithful subposet of $\mathrm{SSG}(G)$ with a fixed sign assignment. Then*

$$H^*_{\mathcal{F}_{-,-}}(P) \colon R\text{--}\mathbf{Alg} \to R\text{--}\mathbf{Mod}^{\mathrm{gr}},$$

which associates to A the graded R–module $H^*_{\mathcal{F}_{A,A}}(P)$, is a covariant functor. In particular, the multipath cohomology of a fixed graph is covariant with respect to morphisms of R–algebras.

Proof Let A be an R–algebra, and let $f: A \to B$ be a homomorphism of R–algebras. Recall the definition of the functor $\mathcal{F}_{A,A}$ from Section 4.1; we have $\mathcal{F}_{A,A}(\mathrm{H}) := A_{c_1} \otimes_R \cdots \otimes_R A_{c_k}$ for each $\mathrm{H} \in P$, and $\mathcal{F}_{A,A}(\mathrm{H} \prec \mathrm{H}')$ is induced by the multiplication. Since $f: A \to B$ is an R–algebra homomorphism, it induces maps between the tensor powers

$$f \otimes \cdots \otimes f: A_{c_1} \otimes_R \cdots \otimes_R A_{c_k} = \mathcal{F}_{A,A}(\mathrm{H}) \to \mathcal{F}_{B,B}(\mathrm{H}) = B_{c_1} \otimes_R \cdots \otimes_R B_{c_k}.$$

For each $n \in \mathbb{N}$, these extend to a map

$$C^n_{\mathcal{F}_{A,A}}(P) = \bigoplus_{\substack{\mathrm{H} \in P \\ \ell(\mathrm{H})=n}} \mathcal{F}_{A,A}(\mathrm{H}) \to \bigoplus_{\substack{\mathrm{H} \in P \\ \ell(\mathrm{H})=n}} \mathcal{F}_{B,B}(\mathrm{H}) = C^n_{\mathcal{F}_{B,B}}(P)$$

because f extends linearly to directed sums. Note that the sign assignment is the same on both complexes. Since f commutes with the multiplication, the induced map commutes with the differentials. Thus the map induced by f is a map of cochain complexes. The fact that this construction respects compositions is straightforward since $f^{\otimes k} \circ g^{\otimes k} = (f \circ g)^{\otimes k}$ for any composable morphisms of R–algebras f and g. $\square$

We remark here that an R–algebra homomorphism $f: A \to B$ provides a natural transformation between the two functors $\mathcal{F}_{A,A}: \boldsymbol{P} \to R\text{–}\mathbf{Mod}$ and $\mathcal{F}_{B,B}: \boldsymbol{P} \to R\text{–}\mathbf{Mod}$; this follows since f extends to tensor powers and directed sums. A natural transformation $\eta: \mathcal{F} \to \mathcal{G}$ between two functors $\mathcal{F}, \mathcal{G}: \boldsymbol{P} \to \boldsymbol{A}$, which preserves the biproducts in $\boldsymbol{A}$, induces a morphism of cochain complexes $\eta^*: C_{\mathcal{F}}(P) \to C_{\mathcal{G}}(P)$.

Before turning back to multipath cohomology, we consider the behaviour of the cohomology $\mathrm{H}_{\mathcal{F}}$ under a change of graph. First, we need a "coherent" way to choose, for each graph, a squared subposet of $\mathrm{SSG}(\mathrm{G})$. Recall that for a poset P we denote by $\boldsymbol{P}$ the associated category — see Remark 3.

Definition 56 Let $S: \mathbf{Digraph} \to \mathbf{Poset}$ be a covariant functor. The functor S is called *path-like* if the following properties hold for each regular morphism of digraphs $\phi: \mathrm{G}' \to \mathrm{G}$:

- $S(\mathrm{G}) \subseteq \mathrm{SSG}(\mathrm{G})$ is a faithful subposet.
- $S(\phi)(S(\mathrm{G}'))$ is a downward closed subposet of $S(\mathrm{G})$.
- $S(\phi)$, seen as a functor between the associated categories $\boldsymbol{S}(\mathrm{G}')$ and $\boldsymbol{S}(\mathrm{G})$, is faithful[4] as a functor.

Example 57 The functors $\mathrm{SSG}: \mathbf{Digraph} \to \mathbf{Poset}$ and $P: \mathbf{Digraph} \to \mathbf{Poset}$ associating to a digraph G the poset of spanning subgraphs and the path poset, respectively, are path-like functors. This follows from Remark 17 in the case of the functor P; in a similar way, this is also true for the functor SSG.

Observe that, if S is a path-like functor, then $S(\mathrm{G}) = S(\mathrm{Id}_{\mathrm{G}})(S(\mathrm{G}))$ is squared.

The second ingredient needed is a way to fix $\mathcal{F}$ for each graph. Let $S: \mathbf{Digraph} \to \mathbf{Poset}$ be a covariant functor and $\boldsymbol{A}$ an Abelian category (eg R–$\mathbf{Mod}$).

---

[4]A functor is called *faithful* if, for each pair of objects, it is injective on the sets of morphisms between them.

Definition 58 A *coefficients system* for S is family of functors $\{\mathcal{F}_{S,\mathtt{G}}\colon \boldsymbol{S}(\mathtt{G}) \to \boldsymbol{A}\}_{\mathtt{G}}$ such that, given a regular morphism of digraphs $\phi\colon \mathtt{G}' \to \mathtt{G}$, the associated functor $S(\phi)\colon \boldsymbol{S}(\mathtt{G}') \to \boldsymbol{S}(\mathtt{G})$ makes the following diagram commute:

$$
\begin{array}{ccc}
\boldsymbol{S}(\mathtt{G}') & \xrightarrow{\;\;S(\phi)\;\;} & \boldsymbol{S}(\mathtt{G}) \\
& \searrow{\scriptstyle \mathcal{F}_{S,\mathtt{G}'}} \quad {\scriptstyle \mathcal{F}_{S,\mathtt{G}}}\swarrow & \\
& \boldsymbol{A} &
\end{array}
$$

Remark 59 The functor $\mathcal{F}_{A,A}$ is not a coefficients system for the functor path poset P unless either we restrict to **Digraph**$(n) \subset$ **Digraph**, or we work with constant coefficients — ie $A = R$.

Notation 60 For $\phi\colon \mathtt{G}' \to \mathtt{G}$ a regular map of digraphs and $S\colon \textbf{Digraph} \to \textbf{Poset}$ a functor, we denote by $S_\phi(\mathtt{G}', \mathtt{G})$ the poset

$$
S_\phi(\mathtt{G}', \mathtt{G}) := S(\mathtt{G}) \setminus S(\phi)(S(\mathtt{G}')).
$$

We are ready to compare, under mild hypotheses, the cochain complexes associated to two graphs.

Remark 61 Recall that the complex $C_{\mathcal{F}}^*(P)$ depends also on a sign assignment ϵ on P and should have been denoted by $C_{\mathcal{F}}^*(P, \epsilon)$. By [Theorem 34](), if $P \subseteq \mathrm{SSG}(\mathtt{G})$ is upward or downward closed, then $C_{\mathcal{F}}^*(P, \epsilon) \cong C_{\mathcal{F}}^*(P, \epsilon')$ for any two sign assignments ϵ, ϵ' on P. This fact motivated the removal of the sign assignment from the notation.

When comparing complexes associated to different graphs, their subcomplexes, or their quotient complexes, we need to be more careful; it is often the case that we have a chain map

$$
f_0\colon C_{\mathcal{F}}^*(P, \epsilon_0) \to C_{\mathcal{F}'}^*(P', \epsilon_0'),
$$

while we might need a chain map

$$
f_1\colon C_{\mathcal{F}}^*(P, \epsilon_1) \to C_{\mathcal{F}'}^*(P', \epsilon_1').
$$

In our case, P and P' will be either upward or downward closed. Hence, to obtain f_1 it is sufficient to compose f_0 with isomorphisms associated to the change of sign assignments, say η and η', such that the diagram

$$
\begin{array}{ccc}
C_{\mathcal{F}}^*(P, \epsilon_1) & \xrightarrow{\;\;f_1\;\;} & C_{\mathcal{F}'}^*(P', \epsilon_1') \\
{\scriptstyle \eta}\downarrow & & \uparrow{\scriptstyle \eta'} \\
C_{\mathcal{F}}^*(P, \epsilon_0) & \xrightarrow{\;\;f_0\;\;} & C_{\mathcal{F}'}^*(P', \epsilon_0')
\end{array}
$$

is commutative. Formally, in order to prove functoriality, one needs to find a coherent way to fix the isomorphisms η, η' once and for all. One approach would be to extend the category of posets to pairs of poset-sign assignments, and expand the notion of coefficient systems to this setting. This can be

done formally — compare with [Chandler 2019, Sections 6 and 7], where a similar approach is pursued. Nonetheless, for the sake of simplicity and to ease the notation, signs and the induced isomorphisms will be treated naively in this section. We do not fix them nor require compatibility; instead we just make use of the existence of such isomorphisms.

From now on, given any $P \subseteq \mathrm{SSG}(\mathsf{G})$, for some digraph G, we fix a sign assignment on P as a restriction of a (fixed) sign assignment on $\mathrm{SSG}(\mathsf{G})$. This choice is immaterial, up to isomorphism of the complex $C_{\mathcal{F}}^*(P)$, when assuming P to be a downward (or upward) closed subposet of $\mathrm{SSG}(\mathsf{G})$ by Theorem 34.

Recall that ℓ_P denotes the level in a faithful subposet $P \subset \mathrm{SG}(\mathsf{G})$; see Definition 24.

Proposition 62 *Let* $S : \textbf{Digraph} \to \textbf{Poset}$ *be a path-like functor, and* $\mathcal{F}_{S,-}$ *be a coefficient system for S. Then, we have the following short exact sequence of cochain complexes:*

$$0 \to C_{\mathcal{F}_{S,\mathsf{G}}}^*(S_\phi(\mathsf{G}', \mathsf{G}))\big[-\min_{x \in S_\phi(\mathsf{G}',\mathsf{G})}\{\ell_{S(\mathsf{G})}(x)\}\big] \to C_{\mathcal{F}_{S,\mathsf{G}}}^*(S(\mathsf{G})) \to C_{\mathcal{F}_{S,\mathsf{G}'}}^*(S(\mathsf{G}')) \to 0.$$

Proof By definition, we have

$$C_{\mathcal{F}_{S,\mathsf{G}}}^n(S(\mathsf{G})) = \bigoplus_{\substack{\mathsf{H} \in S(\mathsf{G}) \\ \ell_{S(\mathsf{G})}(\mathsf{H})=n}} \mathcal{F}_{S,\mathsf{G}}(\mathsf{H}),$$

$$C_{\mathcal{F}_{S,\mathsf{G}}}^n(S_\phi(\mathsf{G}', \mathsf{G}))\big[-\min_{x \in S_\phi(\mathsf{G}',\mathsf{G})}\{\ell_{\mathsf{G}}(x)\}\big] = \bigoplus_{\substack{\mathsf{H} \in S_\phi(\mathsf{G}',\mathsf{G}) \\ \ell_{S(\mathsf{G})}(\mathsf{H})=n}} \mathcal{F}_{S,\mathsf{G}}(\mathsf{H}),$$

where we used $\ell_{S(\mathsf{G})}(\mathsf{H}) = \ell_{S_\phi(\mathsf{G}',\mathsf{G})}(\mathsf{H}) + \min\{\ell_{\mathsf{G}}(x) \mid x \in S_\phi(\mathsf{G}', \mathsf{G})\}$ for $\mathsf{H} \in S_\phi(\mathsf{G}', \mathsf{G})$. As a consequence, we get a natural inclusion of cochain complexes. Note that the inclusion commutes with the differential due to the fact that the poset $S_\phi(\mathsf{G}', \mathsf{G})$ is upward closed, and the sign assignment on $S_\phi(\mathsf{G}', \mathsf{G})$ is induced by $\mathrm{SSG}(\mathsf{G})$.

We need to identify the quotient, with respect to this inclusion, with the cochain complex associated to $S(\mathsf{G}')$. At the level of modules, we have

$$\frac{C_{\mathcal{F}_{S,\mathsf{G}}}^n(S(\mathsf{G}))}{C_{\mathcal{F}_{S,\mathsf{G}}}^n(S_\phi(\mathsf{G}', \mathsf{G}))\big[-\min_{x \in S_\phi(\mathsf{G}',\mathsf{G})}\{\ell_{\mathsf{G}}(x)\}\big]} = \bigoplus_{\substack{\mathsf{H} \in S(\mathsf{G}) \setminus S_\phi(\mathsf{G}',\mathsf{G}) \\ \ell_{S(\mathsf{G})}(\mathsf{H})=n}} \mathcal{F}_{S,\mathsf{G}}(\mathsf{H}).$$

Since $S(\phi)(S(\mathsf{G}'))$ and $S_\phi(\mathsf{G}', \mathsf{G})$ are, by definition, complementary in $S(\mathsf{G})$, we can identify the above quotient with $C_{\mathcal{F}_{S,\mathsf{G}}}^*(S(\phi)(S(\mathsf{G}')))$. Now, the components of the differentials corresponding to the coverings $\mathsf{H}' \prec \mathsf{H}$ with $\mathsf{H} \notin S(\phi)(S(\mathsf{G}'))$ are set to 0 in the quotient. Thus, the above identification commutes with the differentials, hence inducing an isomorphism of cochain complexes, since the sign assignment on the poset $S(\phi)(S(\mathsf{G}'))$ is induced by $\mathrm{SSG}(\mathsf{G})$.

To conclude the proof, we need to identify $C^*_{\mathcal{F}_{S,\mathsf{G}}}(S(\phi)(S(\mathsf{G}')))$ with $C^*_{\mathcal{F}_{S,\mathsf{G}'}}(S(\mathsf{G}'))$. The functor S is path-like. Therefore, by definition, we have that

$$\mathcal{F}_{S,\mathsf{G}'}(\mathtt{H}) = \mathcal{F}_{S,\mathsf{G}}(S(\phi)(\mathtt{H})),$$

and similarly for the maps associated to the covering relations. This gives us an identification of $C^*_{\mathcal{F}_{S,\mathsf{G}}}(S(\phi)(S(\mathsf{G}')))$ with $C^*_{\mathcal{F}_{S,\mathsf{G}'}}(S(\mathsf{G}'))$ as graded R–modules. Observe that there is no shift in the identification because $S(\phi)(S(\mathsf{G}'))$ is downward closed. This identification commutes with the differentials up to an isomorphism induced by a change of sign assignment in one of the complexes. Composing the quotient map with such isomorphism gives us the desired short exact sequence. $\qquad\square$

We now consider the functor S to be either the path poset functor P or SSG, and the functor $\mathcal{F}$ to be the functor $\mathcal{F}_{A,A}$ for A an R–algebra.

Proposition 63 *Let $\phi\colon \mathsf{G}' \to \mathsf{G}$ be a regular morphism of digraphs. The inclusion of $S(\mathsf{G}')$ in $S(\mathsf{G})$ induces the following short exact sequence of complexes*:

$$0 \to C^*_{\mathcal{F}_{A,A}}(S_\phi(\mathsf{G}',\mathsf{G}))\Big[-\min_{x \in S_\phi(\mathsf{G}',\mathsf{G})}\{\ell(x)\}\Big] \to C^*_{\mathcal{F}_{A,A}}(S(\mathsf{G})) \to C^*_{\mathcal{F}_{A,A}}(S(\mathsf{G}')) \otimes A^{\otimes\#(V(\mathsf{G})\setminus V(\mathsf{G}'))} \to 0.$$

In particular, if G' is a spanning subgraph of G, we have the short exact sequence

$$(8) \qquad 0 \to C^*_{\mathcal{F}_{A,A}}(S_\phi(\mathsf{G}',\mathsf{G}))\Big[-\min_{x \in S_\phi(\mathsf{G}',\mathsf{G})}\{\ell(x)\}\Big] \to C^*_{\mathcal{F}_{A,A}}(S(\mathsf{G})) \xrightarrow{\pi_{\mathsf{G},\mathsf{G}'}} C^*_{\mathcal{F}_{A,A}}(S(\mathsf{G}')) \to 0.$$

Proof The proof proceeds exactly as the proof of Proposition 62, until the identification of the complexes $C^*_{\mathcal{F}_{A,A}}(S(\mathsf{G}'))$ and $C^*_{\mathcal{F}_{A,A}}(S(\phi)(S(\mathsf{G}')))$. At this point, we need to use that the family of functors $\mathcal{F}_{S,-} = \mathcal{F}_{A,A}$ is a coefficient system; however, this is not true — see Remark 59. Nonetheless, we have

$$\mathcal{F}_{S,\mathsf{G}'}(\mathtt{H}) = \mathcal{F}_{S,\mathsf{G}}(S(\phi)(\mathtt{H})) \otimes A^{\otimes\#(V(\mathsf{G})\setminus V(\mathsf{G}'))},$$

and the identification extends to the maps associated to the covering relations by tensoring with the opportune tensor power of Id_A. The proof now continues exactly as in Proposition 62. We conclude the proof by observing that if $\mathsf{G}' \in \mathrm{SSG}(\mathsf{G})$, we have $A^{\otimes\#(V(\mathsf{G})\setminus V(\mathsf{G}'))} = R$, and the statement follows. $\qquad\square$

With the same notation, we can now consider compositions of morphisms of digraphs:

Lemma 64 *If $\mathsf{G}'' \in \mathrm{SSG}(\mathsf{G})$ and $\mathsf{G}'' \subseteq \mathsf{G}' \subseteq \mathsf{G}$, then $\pi_{\mathsf{G},\mathsf{G}'} \circ \pi_{\mathsf{G}',\mathsf{G}''} = \pi_{\mathsf{G},\mathsf{G}''}$, where $\pi_{\mathsf{G},\mathsf{G}'}$ is the induced morphism in* (8).

Proof We can explicitly write the maps:

$$
\begin{array}{ccc}
C^n_{\mathcal{F}_{A,A}}(S(\mathsf{G})) = \bigoplus_{\substack{\mathtt{H} \in S(\mathsf{G}) \\ \ell_\mathsf{G}(\mathtt{H})=n}} \mathcal{F}_{A,A}(\mathtt{H}) & \xrightarrow{\pi_{\mathsf{G},\mathsf{G}''}} & \bigoplus_{\substack{\mathtt{H} \in S(\mathsf{G}'') \\ \ell_\mathsf{G}(\mathtt{H})=n}} \mathcal{F}_{A,A}(\mathtt{H}) = C^n_{\mathcal{F}_{A,A}}(S(\mathsf{G}'')) \\
{\scriptstyle \pi_{\mathsf{G},\mathsf{G}'}}\Big\downarrow & {\scriptstyle \pi_{\mathsf{G}',\mathsf{G}''}} \nearrow & \\
C^n_{\mathcal{F}_{A,A}}(S(\mathsf{G}')) = \bigoplus_{\substack{\mathtt{H} \in S(\mathsf{G}') \\ \ell_{\mathsf{G}'}(\mathtt{H})=n}} \mathcal{F}_{A,A}(\mathtt{H}) & &
\end{array}
$$

Since each of the maps above restricts to the identity for H appearing in the summands, and is zero otherwise, we get a commutative diagram of cochain complexes. Note that we are implicitly using the fact that the (family of) functor(s) $\mathcal{F}_{A,A}$ is a coefficients system (since $\mathtt{G}'$, $\mathtt{G}''$ are spanning subgraphs of $\mathtt{G}$) for $S = \mathrm{SSG}$ or $S = P$, and Remark 61. $\square$

We are ready to conclude the proof of the functoriality.

Proof of Theorem 1 The statement follows from Lemma 64, giving the functoriality with respect to maps of digraphs, and Proposition 55, giving the functoriality with respect to maps of R–algebras. $\square$

We conclude the section with the result of functoriality with respect to change of base rings:

Theorem 65 *Let* **Ring** *be the category of unital rings and* $\mathbf{Ab}^{\mathrm{gr}}$ *be the category of graded Abelian groups. Then, the multipath cohomology*

$$\mathrm{H}_\mu(-;-)\colon \mathbf{Digraph}^{\mathrm{op}} \times \mathbf{Ring} \to \mathbf{Ab}^{\mathrm{gr}}$$

is a bifunctor.

Proof For a homomorphism $f\colon R \to S$ of rings, there is an extension-of-scalars functor along f defined as $S \otimes_R (-)\colon R\text{–}\mathbf{Mod} \to S\text{–}\mathbf{Mod}$, where the tensor product in S is regarded as an R–module via the map f. In this way, we get natural isomorphisms $S \otimes_R R \cong S$ (more generally, it is true that if R is commutative and M an R–module, then $M \otimes_R R \cong M$), and, for each product $R \otimes_R \cdots \otimes_R R$, isomorphisms $S \otimes_R R \otimes_R \cdots \otimes_R R \cong S \otimes_R R \cong S$. Reasoning as in Lemma 64 and Theorem 1 gives the functoriality with respect to all regular maps of digraphs (with any finite number of vertices). $\square$

6 Other poset (co)homologies and Turner and Wagner's approach

The definition of multipath cohomology given in Section 4 uses a certain homology of posets which we referred to as poset homology. After application of the path poset functor $P\colon \mathbf{Digraph} \to \mathbf{Poset}$ — see Remark 17 — other (co)homology theories of posets can also be used to get similar graph (co)homology theories, for example, the general *functor homology* (of categories) — see eg [Gabriel and Zisman 1967; Mac Lane 1971] — or the *cellular cohomology* (of posets) introduced in [Everitt and Turner 2015]. In this section, we provide a brief review of these (co)homology theories, and compare them with poset homology on (suitable modifications of) path posets. In particular, we argue that, after mild modifications, we can interpret multipath cohomology groups as (cellular and hence) functor cohomology groups, shedding light on the nature of multipath cohomology.

6.1 Functor homology (of posets)

For a poset P, recall that $\boldsymbol{P}$ denotes its associated category — see Remark 3. Given a functor $\mathcal{F}\colon \boldsymbol{P} \to \boldsymbol{A}$, where $\boldsymbol{A}$ is a complete and cocomplete Abelian category, we can define the *functor homology* (resp *cohomology*) groups $\mathrm{H}_*(\boldsymbol{P}; \mathcal{F})$ (resp $\mathrm{H}^*(\boldsymbol{P}; \mathcal{F})$) as the associated higher colimits (resp higher limits). For the

sake of completeness, we spell out the definition. Denote by $\mathbf{1}$ the category with a single object and a single morphism. Then, there is a unique functor $\mathcal{T}\colon \boldsymbol{P} \to \mathbf{1}$. Since A is complete and cocomplete, both left and right Kan extensions of $\mathcal{F}$ exist. In particular, the left Kan extension $\mathrm{Lan}_{\mathcal{T}} \mathcal{F}$ of $\mathcal{F}$ along $\mathcal{T}$ exists, and it yields the colimit functor of $\mathcal{F}$.

Definition 66 [Mac Lane 1971] The *functor homology* $\mathrm{H}_n(\boldsymbol{P}; \mathcal{F})$ *of* $\boldsymbol{P}$ *with coefficients in* $\mathcal{F}$ is the nth left derived functor of $\mathrm{Lan}_{\mathcal{T}} \mathcal{F}$.

Analogously, the right Kan extension along $\mathcal{T}$ yields the limit of $\mathcal{F}$; thus, $\mathrm{H}^n(\boldsymbol{P}; \mathcal{F})$ is given by the nth derived functor of $\lim \mathcal{F}$. Definition 66 is rather abstract; more concretely, $\mathrm{H}_*(\boldsymbol{P}; \mathcal{F})$ can be computed (see [Gabriel and Zisman 1967]) as the homology groups of the chain complex

$$\cdots \xrightarrow{\partial_n} \bigoplus_{c_0 \to \cdots \to c_n} \mathcal{F}(c_0) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} \bigoplus_{c_0 \to c_1 \to c_2} \mathcal{F}(c_0) \xrightarrow{\partial_1} \bigoplus_{c_0 \to c_1} \mathcal{F}(c_0) \xrightarrow{\partial_0} \bigoplus_{c_0 \in \boldsymbol{P}} \mathcal{F}(c_0) \to 0$$

with differential

$$\partial_n(f(c_0 \to \cdots \to c_{n+1})) = \mathcal{F}(c_0 \to c_1) f(c_1 \to \cdots \to c_{n+1}) + \sum_{i=1}^{n+1} (-1)^i f(c_0 \to \cdots \to \hat{c}_i \to \cdots \to c_{n+1}),$$

where $\hat{c}_i$ means that c_i is missing, and $(c_0 \to \cdots \to c_n)$ denotes the inclusion of $f \in \mathcal{F}(c_0)$ into the summand corresponding to the sequence $c_0 \to \cdots \to c_{n+1}$. Dually, the Roos complex [1961] computes the functor cohomology groups $\mathrm{H}^n(\boldsymbol{P}; \mathcal{F})$. More precisely, $\mathrm{H}^*(\boldsymbol{P}; \mathcal{F})$ is the cohomology of the cochain complex

$$0 \to \prod_{c_0 \in \boldsymbol{P}} \mathcal{F}(c_0) \xrightarrow{d^0} \prod_{c_0 \to c_1} \mathcal{F}(c_1) \xrightarrow{d^1} \prod_{c_0 \to c_1 \to c_2} \mathcal{F}(c_2) \xrightarrow{d^2} \cdots \xrightarrow{d^{n-1}} \prod_{c_0 \to \cdots \to c_n} \mathcal{F}(c_n) \xrightarrow{d^n} \cdots$$

endowed with differential d^n, whose evaluation on $f \in \prod_{c_0 \to \cdots \to c_n} \mathcal{F}(c_n)$, is given by

$$d^n(f)(c_0 \to \cdots \to c_{n+1})$$
$$= (-1)^{n+1} \mathcal{F}(c_n \to c_{n+1}) f(c_0 \to \cdots \to c_n) + \sum_{i=0}^{n} (-1)^i f(c_0 \to \cdots \to \hat{c}_i \to \cdots \to c_{n+1}).$$

Note that here $(c_0 \to \cdots \to c_n)$ denotes the projection onto the factor corresponding to the sequence $c_0 \to \cdots \to c_{n+1}$. In other words, functor (co)homology groups are defined as the (co)homology groups of a suitable (co)simplicial replacement. We also point out that similar constructions can be performed using contravariant functors instead of covariant.

The homology of a category with coefficients in a functor has been extensively studied and the literature on it is very rich. When restricting to constant functors, the functor (co)homology groups depend only on the geometric realisation of the source category — see [Quillen 1973]. In particular, by Corollary 2 of that work, every poset with an initial element has with respect to the constant functor the homology of a point. We now provide an example.

Example 67 Consider the path poset associated to the digon digraph — see Figure 9. Its associated category is the pushout category $1 \leftarrow 0 \rightarrow 2$, where the initial object 0 corresponds to the empty multipath. For an Abelian category A and functor $\mathcal{F}$, set $f := \mathcal{F}(0 \rightarrow 1)$ and $g := \mathcal{F}(0 \rightarrow 2)$. The corresponding functor homology groups are the homology groups of the chain complex

$$0 \rightarrow A_0 \oplus A_0 \rightarrow A_0 \oplus A_1 \oplus A_2 \rightarrow 0,$$

where A_0, A_1, A_2 are objects of A with $\mathcal{F}(i) = A_i$, and the only nontrivial map is given by

$$(a, b) \mapsto (-(a + b), f(a), g(b)).$$

The homology groups of the complex are therefore $H_0(P; \mathcal{F}) = \operatorname{colim} \mathcal{F}$, $H_1(P; \mathcal{F}) \simeq \ker(f) \cap \ker(g)$, and they are 0 in higher degrees. Note that the functor cohomology groups are trivial in all degrees but the 0th (in which it agrees with A_0), because the category has an initial object. Note also that the poset homology groups as defined in Section 3 would be given by the kernel and image of $f - g$.

Assume now that $\mathcal{F}$ takes values in $A = \mathbf{Ab}$, the category of Abelian groups, and assume that $\mathcal{F}$ sends every morphism in P, ie every $x \leq y$ in P, to an isomorphism of A. Then, $\mathcal{F}$ induces a local coefficient system on the classifying space[5] BP of P, ie on the order complex of P. Quillen [1973] has shown that there is an isomorphism

$$H_*(P, \mathcal{F}) \cong H_*(BP, \mathcal{F})$$

between the homology groups of the category P and the classical homology groups of the space BP, with local coefficients (here for simplicity denoted with the same symbol $\mathcal{F}$). In order to show it, one considers the skeleton filtration

$$BP^{(0)} \subseteq BP^{(1)} \subseteq \cdots$$

and the associated spectral sequence with E^1–term $E^1_{p,q} = H_{p+q}(BP^{(p)}, BP^{(p-1)}, \mathcal{F})$. When $q = 0$, the E^1–term yields the homology groups $H_p(P, \mathcal{F})$. The spectral sequence converges to $H_p(BP, \mathcal{F})$, providing the isomorphism. In a similar fashion, Turner and Everitt have defined the so-called cellular cohomology groups of posets, as we shall recall in the next subsection.

6.2 Cellular poset cohomology

Cellular poset (co)homology is a rather general (co)homology theory of posets introduced in [Everitt and Turner 2015]. The cellular poset (co)chain groups are defined using a relative version of functor (co)homology and, for a rather large class of posets, it agrees with functor (co)homology, providing a tool to the computation of the higher (co)limits of functors on posets. We now proceed by reviewing its definition in the cohomological case (the homological case is analogous).

[5]The geometric realisation of the nerve.
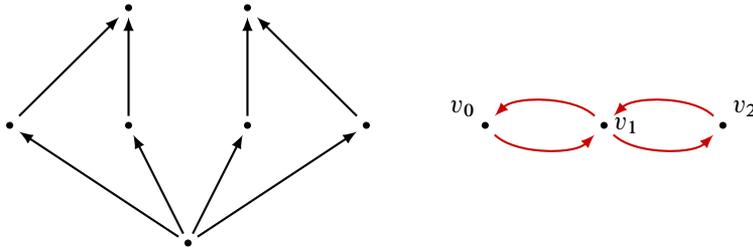
Figure 14: The poset P (left) and a graph realising P as its path poset.

In what follows, we assume that P is a finite and ranked poset, with rank function $\mathrm{rk} \colon P \to \mathbb{N}$. Let $r :=$ $\max\{\mathrm{rk}\, x \mid x \in P\}$ be the maximum rank; then one can filter P with subposets

$$P^k := \{x \in P \mid \mathrm{rk}(x) \geq r - k\},$$

yielding a filtration $P^0 \subseteq P^1 \subseteq \cdots \subseteq P^r = P$. Let $\mathcal{F}$ be a *contravariant* functor (a presheaf) on (the associated category of) P.

Definition 68 [Everitt and Turner 2015, Definition 2.1] The cellular cochain complex has cochain groups

$$\mathrm{C}^i_{\mathrm{cell}}(P; \mathcal{F}) := \mathrm{H}^n(P^i, P^{i-1}, \mathcal{F}),$$

where $\mathrm{H}^n(P^i, P^{i-1}, \mathcal{F})$ are the relative functor cohomology groups.

The differentials are also induced from functor cohomology — see [Everitt and Turner 2015] for a description of the differential. Observe that, as taking the classifying space of a category is natural, the relative cohomology groups appearing in the definition can be interpreted as the usual relative cohomology groups (of the associated classifying spaces). One can compute explicitly this complex via the formulae

$$\mathrm{C}^i_{\mathrm{cell}}(P; \mathcal{F}) = \begin{cases} \bigoplus_{\mathrm{rk}(x)=n} \mathcal{F}(x), & i = 0, \\ \bigoplus_{\mathrm{rk}(x)=n-i} \widetilde{\mathrm{H}}^{i-1}(|NP_{>x}|, \mathcal{F}(x)), & i > 0, \end{cases}$$

where N denotes the nerve, $|\cdot|$ denotes the geometric realisation, and $\widetilde{\mathrm{H}}^*$ denotes the usual reduced singular cohomology — see [Everitt and Turner 2015, Propositions 3, 4 and 5].

Note that the functors appearing in the definition of cellular cohomology are contravariant, and hence defined on P^{op}. The constant functor can be seen both as a covariant and as a contravariant functor; hence computations can be carried on in both cases. We now proceed with an example of calculation, computing the cellular cohomology of a path poset, with respect to the contravariant constant functor.

Example 69 Consider the poset P and the graph G represented in Figure 14. The path poset $P(\mathrm{G})$ is isomorphic to P. For a fixed field $\mathbb{K}$, consider the constant functor $\mathbb{K}$ on the category associated to the poset P.

We now compute the cellular cochain groups of the poset P. First, observe that the poset P is ranked with rank function rk given by the distance from the minimum; this function is bounded with maximum value

$r = 2$, which is achieved by the maximal elements. The corank function is defined to be $|x| := 2 - \mathrm{rk}(x)$. In degree 0, the cellular cochain complex is generated by the (evaluation of the constant) functor $\mathbb{K}$ at the maxima, obtaining

$$C^0_{\mathrm{cell}}(P; \mathbb{K}) \cong \mathbb{K}^2.$$

In order to analyse the higher degrees, we use [Everitt and Turner 2015, Proposition 3]

$$C^n_{\mathrm{cell}}(P; \mathbb{K}) \cong \prod_{|x|=n} \mathrm{H}^n(P_{\geq x}, P_{>x}; \mathbb{K}),$$

with the convention that, $\mathrm{H}^n(P_{\geq x}, \varnothing; k) = \mathrm{H}^n(P_{\geq x}; k)$ — see [loc cit, page 140] — and where H^* denotes functor cohomology. Then, for the elements x in P of corank 1, we get

$$\mathrm{H}^1(P_{\geq x}, P_{>x}; \mathbb{K}) \cong \widetilde{\mathrm{H}}^0(P_{>x}; \mathbb{K})$$

by [loc cit, Proposition 4]. As $P_{>x}$ consists of a single point, we get $\widetilde{\mathrm{H}}^0(P_{>x}; \mathbb{K}) \cong \widetilde{\mathrm{H}}^0(\{*\}; \mathbb{K}) \cong 0$ (see [loc cit, page 140]). Therefore, we have

$$C^1(P; \mathbb{K}) \cong 0.$$

We conclude the computation of the cellular cohomology groups by analysing $C^2(P; \mathbb{K})$, as there are no elements of corank ≥ 3. There is only a single element m of corank 2, given by the minimum of P, and the geometric realisation of $P_{>m}$ consists of two intervals. By [loc cit, Propositions 3 and 4],

$$C^2_{\mathrm{cell}}(P; \mathbb{K}) \cong \widetilde{\mathrm{H}}^1(P_{>m}; \mathbb{K}) \cong 0.$$

Therefore, it follows that the cellular cohomology, in this case, is concentrated in degree 0, where its dimension is 2.

Arguing as in Example 69, we have the analogue of Example 67:

Example 70 Consider the path poset associated to the digon digraph — see Figure 9. Then, P has a unique element m of rank 0 and two elements of rank 1. Then,

$$C^0_{\mathrm{cell}}(P; \mathbb{K}) \cong \mathbb{K}^2$$

generated by the elements of rank 1. The group $C^1_{\mathrm{cell}}(P; \mathbb{K})$, instead, is isomorphic to $\widetilde{\mathrm{H}}^1(P_{>m}; \mathbb{K}) \cong \mathbb{K}$. The differential acts by $(x, y) \mapsto x - y$, giving $\mathrm{H}^1_{\mathrm{cell}}(P; \mathbb{K}) \cong \mathbb{K}$ and 0 in other degrees. When passing to arbitrary coefficients, as in Example 67, let $A_i := \mathcal{F}(i)$ and set $f^* := \mathcal{F}(0 \to 1)$, $g^* := \mathcal{F}(0 \to 2)$. Then, the cellular cochain complex becomes

$$0 \to A_1 \oplus A_2 \to A_0 \to 0,$$

with unique differential $(a, b) \mapsto f^*(a) - g^*(b)$.

Using a spectral sequences argument, one can prove that, for certain ranked and finite posets, cellular cohomology groups compute the higher limits of (a contravariant functor) $\mathcal{F}$. We first recall — see [Everitt and Turner 2015, Definition 3.1] — that a ranked poset is *cellular* if, and only if, for every contravariant

functor $\mathcal{F}$ on $\boldsymbol{P}$, the relative functor cohomology groups $\mathrm{H}^i(\boldsymbol{P}^n, \boldsymbol{P}^{n-1}, \mathcal{F})$ are 0 for all $i \neq n$. For example, for X a regular CW–complex, the face poset $P(X)^{\mathrm{op}}$ with reversed inclusion (hence, $x \leq y$ if, and only if, $y \subseteq x$) is cellular — see [Everitt and Turner 2015, Section 4.1]. By [loc cit, Theorem 1], when P is a cellular poset and $\mathcal{F} \colon \boldsymbol{P} \to \mathbf{Ab}$ is a contravariant functor, there are isomorphisms $\mathrm{H}^*_{\mathrm{cell}}(\boldsymbol{P}, \mathcal{F}) \cong \mathrm{H}^*(\boldsymbol{P}, \mathcal{F})$ between cellular cohomology groups and the functor cohomology groups, showing that for a large class of posets cellular (co)chain groups compute the higher (co)limits.

6.3 Comparisons on path posets

In this subsection we restrict to posets arising as path posets of digraphs. The idea of defining graph homologies using the path poset is, to the best of the authors' knowledge, due to Turner and Wagner, and inspired this work. In [Turner and Wagner 2012], they make use of functor homology to define a graph homology, as the functor homology groups of the (category associated to the) path poset. In the special case $\mathcal{F} = \mathcal{F}_{A,M}$, that is, the functor defined in (7) (or, better, a symmetrised version of it, see [loc. cit.]), we get what we call the *Turner–Wagner homology* TW of $\mathtt{G}$:

$$\mathrm{TW}_*(\mathtt{G}; A, M) := \mathrm{H}_*(\boldsymbol{P}(\mathtt{G}); \mathcal{F}_{A,M}).$$

Here we point out a small technical issue; if the module M is different from A, we have to fix a base vertex, and the theory provides a homology for *based digraphs*, ie graphs with a base vertex, exactly as in our case — see Remark 45. As every category with an initial element, with respect to the constant functor, has the homology of a point, we obtain the following:

Remark 71 We have $\mathrm{TW}_0(\mathtt{G}; R, R) \cong R$ and $\mathrm{TW}_i(\mathtt{G}; R, R) = 0$ for $i > 0$.

An immediate consequence of the previous remark and of the examples in Section 4.2, along with Example 67, is the following result.

Remark 72 The (co)homologies TW and H_μ are not isomorphic nor dual to each other.

In order to understand the precise relation between the multipath cohomology of a graph and the Turner–Wagner homology, we use cellular cohomology as an intermediate theory. In the following, we aim to show that, after some mild modifications of the path poset, all these theories agree. However, despite the similarities, it is easy to see that these are different "on the nose":

Example 73 Consider the path poset P_1 of the digon graph — see Figure 3. Note that the associated category is the pushout category $1 \leftarrow 0 \rightarrow 2$. As shown in Example 67, for an algebra A and the functor $\mathcal{F}_{A,A} \colon \boldsymbol{P} \to A$ described in (7), we have $\mathcal{F}_{A,A}(0) = A \otimes A$ and $\mathcal{F}_{A,A}(1) = \mathcal{F}_{A,A}(2) = A$. The functor homology groups are the homology groups of the complex

$$0 \to A \otimes A \oplus A \otimes A \to A \otimes A \oplus A \oplus A \to 0$$

whose differential is given by

$$(a_0 \otimes b_0, a_1 \otimes b_1) \mapsto (a_0 \otimes b_0 + a_1 \otimes b_1, -a_0 b_0, -a_1 b_1).$$

Note that, as the path poset has a minimum, the functor *cohomology* groups are all trivial in higher degree, and isomorphic to $A \otimes A$ in degree 0. The functor $\mathcal{F}_{A,A}$ is not directly defined on $\boldsymbol{P}^{\mathrm{op}}$, so we cannot directly compute the associated cellular cohomology groups. However, $\mathcal{F}_{A,A}$ can be seen as a contravariant functor on $\boldsymbol{P}^{\mathrm{op}}$, in which case the cellular cohomology groups can be computed. Observe that the only nontrivial cellular cochain group in this case is $C^0_{\mathrm{cell}}(\boldsymbol{P}^{\mathrm{op}}, \mathcal{F}_{A,A}) \cong A \otimes A$. Note also that the cellular *homology* groups would be trivial because of the analogue of [Everitt and Turner 2015, Theorem 1] in this context. When considering the multipath cohomology cochain complex, we get

$$0 \to A \otimes A \to A \oplus A \to 0,$$

with unique differential

$$a \otimes b \mapsto (ab, -ba).$$

To be concrete, when $A = \mathbb{K}$ we get that functor homology and cellular cohomology are both of dimension 1 concentrated in degree 0, whereas multipath cohomology is of dimension 0 concentrated in degree 1.

The previous example shows that the poset homology theories described in this section, when evaluated at the path poset, are not the same on the nose. However, they become all equivalent after some mild modification of the path poset, as we now shall explain.

Let $\mathtt{G}$ be a digraph and let $\boldsymbol{P}(\mathtt{G})^{\mathrm{op}}$ be the opposite category (with the same objects as $\boldsymbol{P}(\mathtt{G})$ but reversed arrows) of $\boldsymbol{P}(\mathtt{G})$. Consider the category $\boldsymbol{Q}(\mathtt{G}) := \boldsymbol{P}(\mathtt{G})^{\mathrm{op}} \setminus \{\varnothing\}$ obtained from $\boldsymbol{P}(\mathtt{G})^{\mathrm{op}}$ by removing the empty multipath—ie the terminal object in $\boldsymbol{P}(\mathtt{G})^{\mathrm{op}}$. Note that $\mathcal{F}_{A,M}$ is a functor on $\boldsymbol{P}(\mathtt{G})$, and hence a presheaf on $\boldsymbol{P}(\mathtt{G})^{\mathrm{op}}$. Then, the cellular cochain groups $C^i_{\mathrm{cell}}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M})$ and $C^{i+1}_\mu(\mathtt{G}; \mathcal{F}_{A,M})$ are isomorphic for all $i \geq 0$. Furthermore, this isomorphism is an isomorphism of chain complexes $C^*_{\mathrm{cell}}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M}) \cong C^{*\geq 1}_\mu(\mathtt{G}; \mathcal{F}_{A,M})$. Therefore we obtain the following remark.

Remark 74 Although $P(\mathtt{G})$ is not cellular in the sense of [Everitt and Turner 2015], $Q(\mathtt{G})$ is — see [loc cit, Section 4.1]; thus the previous isomorphism of cochain complexes, together with [loc cit, Theorem 1], provides isomorphisms

$$\mathrm{H}^i_\mu(\mathtt{G}; A, M) \cong \mathrm{H}^{i-1}_{\mathrm{cell}}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M}) \cong \mathrm{H}^{i-1}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M})$$

of cohomology groups between $\mathrm{H}^i_\mu(\mathtt{G}; A, M)$, the cellular cohomology $\mathrm{H}^{i-1}_{\mathrm{cell}}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M})$ and the functor cohomology groups $\mathrm{H}^{i-1}(\boldsymbol{Q}(\mathtt{G}); \mathcal{F}_{A,M})$ for all $i > 1$.

In light of Remark 74, one can wonder if the graded module obtained by removing the minimum from the path poset in the Turner–Wagner construction and multipath cohomology become related. However, this is not generally the case, as shown by the next example. Before that, recall that the face poset of a simplicial complex X is the poset on the set of simplices of X, ordered by containment. The augmented face poset of X is its face poset together with a minimum element $\varnothing$ corresponding to the empty simplex.

Example 75 The path poset $P(\mathtt{G})$ is the augmented face poset of a topological space $X = X(\mathtt{G})$ — see [Caputi et al. 2023, Section 6]. Note that the geometric realisation of an augmented face poset is always contractible (since it is the cone on the geometric realisation of the face poset). In particular, the geometric realisation $B(P(\mathtt{G}))$ is the cone over $B(P(\mathtt{G}) \setminus \{\varnothing\})$. For $A = R$ the base ring, the functor homology of (the category associated to) $P(\mathtt{G}) \setminus \{\varnothing\}$ with coefficients in $\mathcal{F}_{R,R}$ agrees with the simplicial homology of $X \simeq B(P(\mathtt{G}) \setminus \{\varnothing\})$ with coefficients in R. On the other hand, it is not difficult to see — see [loc cit, Theorem 6.8] — that multipath cohomology is simplicial, ie

$$\widetilde{\mathrm{H}}^n(X; R) \cong \mathrm{H}_\mu^{n+1}(\mathtt{G}; R),$$

where $\widetilde{\mathrm{H}}^*$ denotes the reduced simplicial cohomology. Therefore, although the Turner–Wagner homology

$$\mathrm{TW}_*(\mathtt{G}; \mathcal{F}_{R,R}) = \mathrm{H}_*(P(\mathtt{G}); \mathcal{F}_{R,R})$$

is always trivial (as $P(\mathtt{G})$ is an augmented face poset), after removing the minimum element, the associated functor homology $\mathrm{H}_i(P(\mathtt{G}) \setminus \{\varnothing\}; \mathcal{F}_{R,R})$ is not. In fact, the functor homology $\mathrm{H}_i(P(\mathtt{G}) \setminus \{\varnothing\}; \mathcal{F}_{R,R})$ and the multipath cohomology groups $\mathrm{H}_\mu^{i+1}(\mathtt{G}; R)$ are related, for $i \geq 2$, by the standard universal coefficients theorem. The induced short exact sequence

$$0 \to \mathrm{Ext}_R^1(\mathrm{H}_{i-1}(P(\mathtt{G}) \setminus \{\varnothing\}; \mathcal{F}_{R,R})) \to \mathrm{H}_\mu^{i+1}(\mathtt{G}; R) \to \mathrm{Hom}_R(\mathrm{H}_i(P(\mathtt{G}) \setminus \{\varnothing\}; \mathcal{F}_{R,R})) \to 0$$

features an Ext functor, which is nontrivial in general. For instance, taking $A = R = \mathbb{Z}$, the multipath cohomology of the bipartite complete graph $\mathtt{K}_{5,5}$ has 3–torsion [Caputi et al. 2024a, Proposition 4.5].

The connection shown in the previous example between functor homology and multipath cohomology is given by two facts; the first, that functor (co)homology of a category, for nice functors, agrees with the usual (co)homology of the classifying space (with local coefficients as in [Quillen 1978, Section 7]), and the second, that the classifying space of the opposite category is naturally homeomorphic to the classifying space of the category itself. As shown in Remark 74, multipath cohomology and functor cohomology agree (in degree $i \geq 2$) when we pass to the opposite category associated to the path poset. Then, in the Turner–Wagner approach, which uses functor homology, one computes the higher colimits of $\mathcal{F}$, whereas multipath cohomology provides a way to compute the higher limits of $\mathcal{F} \circ \mathrm{op}$; when $\mathcal{F}$ is a coefficient system in the sense of Quillen, as in the case of constant functors, higher limits and colimits are computed as usual cohomology on the classifying spaces; then, as the op functor does not change the homotopy type of the classifying spaces, the assertion follows. Note that this does not provide a precise relation for nonlocal coefficients (eg $\mathcal{F}_{A,M}$, $A \neq \mathbb{K}$). In particular, this reasoning does not provide a precise relationship between the Turner–Wagner and multipath cohomologies.

Remark 76 All said above provides an alternative way to define multipath cohomology; ie after passing to the path poset and removing the minimum, one can take the opposite associated category and compute (equivalently) either the higher limits of $\mathcal{F}_{A,M}$ or the associated cellular cohomology groups (as the obtained poset is now cellular). However, functor and cellular cohomologies are not directly computable from the definitions, whereas poset homology happens to be quite computable, also algorithmically. The

approach has shown to be fruitful in computing the multipath cohomology of all linear graphs — see [Caputi et al. 2023].

To conclude the comparisons, we point out that, in special cases, like the linear graph $\mathtt{I}_n$ and the polygonal graph $\mathtt{P}_n$, the difference between the multipath and Turner–Wagner homologies is controlled. This is also due to the fact that both homologies provide roughly the same amount of information as the chromatic homology — see [Everitt and Turner 2009; Turner and Wagner 2012] for relations between TW and the chromatic homology. In the next subsection we recall the definition of the latter homology, and prove a comparison result for the graphs $\mathtt{I}_n$ and $\mathtt{P}_n$.

7 Comparison with chromatic homology

We now compare multipath cohomology with chromatic homology of unoriented graphs [Helme-Guizon and Rong 2005; Przytycki 2010]. The latter can be seen as a special case of the construction in Section 3; in light of this observation, we can interpret multipath cohomology as an extension of chromatic homology to the directed setting.

In the first subsection, we briefly revise the construction of the chromatic homology (both in its original version [Helme-Guizon and Rong 2005] and in Przytycki's variant [2010]). We argue that the multipath cohomology of a graph differs from either of these theories computed for the underlying unoriented graph. This uses the fact that multipath cohomology is sensible to orientations. Nonetheless, in the special case of coherently oriented polygonal graphs and linear graphs, we prove that the two (co)homology theories contain the same amount of information — see Theorem 80. As a consequence, see Corollary 82, we obtain that the multipath cohomology of the coherently oriented polygon recovers (a truncated version of) the Hochschild homology of its coefficients. As an application of the functoriality, in the second subsection we clarify the relationship between multipath cohomology and chromatic homology, providing the long exact sequence relating multipath and chromatic cohomologies.

7.1 Chromatic homologies

In this subsection we review the construction of two graph homology theories. The first of these homologies goes under the name of chromatic homology and was introduced in [Helme-Guizon and Rong 2005]. The second homology is a variation of the chromatic homology, and it is due to Przytycki [2010].

Let $\mathtt{G}$ denote a *unoriented* graph with ordered edges and a base vertex v_0. Let A be a *commutative* unital R–algebra, and M be an (A, A)–bimodule. Assume that the A–action on M is symmetric — that is, $a \cdot m = m \cdot a$ for all $m \in M$ and $a \in A$. To each spanning subgraph $\mathtt{H} \in \mathrm{SSG}(\mathtt{G})$ we associate the module

$$M(\mathtt{H}) = M \otimes \bigotimes_{c \not\ni v_0} A_c,$$

where c ranges among the connected components of $\mathtt{H}$ — ordered arbitrarily. If $\mathtt{H} \prec \mathtt{H}'$ then $\mathtt{H} \cup e = \mathtt{H}'$ for some edge e. We can define a map $\mathtt{d}_{\mathtt{H} \prec \mathtt{H}'} : M(\mathtt{H}) \to M(\mathtt{H}')$ (see [Helme-Guizon and Rong 2005]). There

are two cases to consider depending on the number of components merged by e:

(i) The edge e is incident to two distinct components of H. We have a natural identification of the components of H and H′ that do not share vertices with e. Furthermore, precisely two distinct components, say c_1 and c_2, of H are merged into a single component of H′, say c'. The map $d_{H \prec H'} : M(H) \to M(H')$ is defined as the identity on all factors but those corresponding to c_1 and c_2, where it behaves as follows:

$$A_{c_1} \otimes A_{c_2} \to A_{c'}, \quad a \otimes b \mapsto ab = ba,$$

or if c_{3-i} for $i \in \{1, 2\}$ contains the marked vertex,

$$M \otimes A_{c_i} \to A_{c'}, \quad a \otimes b \mapsto a \cdot b;$$

(ii) The edge e is incident to a single component of H. There is a natural identification of *all* components of H and H′, and the map $d_{H \prec H'} : M(H) \to M(H')$ is taken to be the corresponding identification of the associated modules.

Similarly, Przytycki [2010] defines the map $\hat{d}_{H \prec H'} : M(H) \to M(H')$ as above, but setting it to be the zero map instead of the identity in case (ii). The cochain complexes

$$(C^*_{\text{Chrom}}(G; A, M), d^*) \quad \text{and} \quad (\widehat{C}^*_{\text{Chrom}}(G; A, M), \hat{d}^*)$$

are defined as

$$C^i_{\text{Chrom}}(G; A, M) = \widehat{C}^i_{\text{Chrom}}(G; A, M) = \bigoplus_{\substack{H \subset G \\ \#E(H) = i}} M(H),$$

and, for $x \in M(H)$,

$$d(x) = \sum_{H \prec H'} (-1)^{\zeta(H \prec H')} d_{H \prec H'}(x) \quad \text{and} \quad \hat{d}(x) = \sum_{H \prec H'} (-1)^{\zeta(H \prec H')} \hat{d}_{H \prec H'}(x),$$

where ζ is defined as

(9) $$\zeta(H \prec H \cup e) = \begin{cases} 0 & \text{if an even number of edges preceding } e \text{ belong to H,} \\ 1 & \text{otherwise.} \end{cases}$$

Remark 77 The chain complexes $(C^*_{\text{Chrom}}(G; A, M), d^*)$ and $(\widehat{C}^*_{\text{Chrom}}(G; A, M), \hat{d}^*)$ do not depend on the ordering of the edges up to isomorphism — see [Helme-Guizon and Rong 2005; Przytycki 2010].

Recall that I_n denotes the n–step graph in Figure 1, and P_n denotes the polygonal graph in Figure 2.

Remark 78 In the special case of the coherently oriented line graph I_n and of the polygon P_n, the cochain complexes $(C^*_{\text{Chrom}}(G; A, M), d^*)$ and $(\widehat{C}^*_{\text{Chrom}}(G; A, M), \hat{d}^*)$ can be extended, using the orientation of I_n and of P_n, to the noncommutative context — see [Przytycki 2010, Remark 2.3 (ii)] — and this extension is perfectly identical to our definition of μ — see Section 4.1.

Observe that the chromatic homology theories can be recovered from the framework in Section 3.1.

Remark 79 Consider an unoriented graph G and the poset $P = \mathrm{SSG}(G) \supseteq P(G)$. Recall that $\boldsymbol{P}$ denotes the category associated to P. Consider the covariant functor $\mathcal{F} \colon \boldsymbol{P} \to R\text{–}\mathbf{Mod}$ defined by extending the functor $\mathcal{F}_{A,M} \colon \boldsymbol{P}(G) \to R\text{–}\mathbf{Mod}$, see (7), to the whole $\mathrm{SSG}(G)$. This extension is defined as follows: when the covering relation $H \prec H'$ is as in case (ii) above, $\mathcal{F}$ is either the identity or the 0–map, depending whether we want to recover $(C^*_{\mathrm{Chrom}}(G; A, M), d^*)$ or $(\hat{C}^*_{\mathrm{Chrom}}(G; A, M), \hat{d}^*)$. These constructions do not depend on signs by Corollary 36.

The following theorem establishes a first relation between multipath and chromatic (co)homologies.

Theorem 80 *Let A be a unital R–algebra, and M an (A, A)–bimodule. Then, we have the following isomorphisms of chain complexes (of R–modules):*

$$(10) \qquad (C^*_\mu(I_n; A, M), d) \cong (\hat{C}^*_{\mathrm{Chrom}}(I_n; A, M), \hat{d}) \cong (C^*_{\mathrm{Chrom}}(I_n; A, M), d),$$

$$(11) \qquad (C^*_\mu(P_n; A, M), d) \oplus (M[n+1], 0) \cong (\hat{C}^*_{\mathrm{Chrom}}(P_n; A, M), \hat{d}),$$

where $(M[n+1], 0)$ is the cochain complex consisting of a copy of M in degree $n + 1$.

Proof By Remark 78, the cochain complexes $(C^*_{\mathrm{Chrom}}(G; A, M), d^*)$ and $(\hat{C}^*_{\mathrm{Chrom}}(G; A, M), \hat{d}^*)$ can be defined for arbitrary unital R–algebras, using the orientation of the coherently oriented n–step graph I_n or the polygon P_n. The proof follows directly from Remark 79 by noticing that $\mathrm{SSG}(I_n) = P(I_n)$ and $\mathrm{SSG}(P_n)$ is the poset $P(P_n) \cup \{P_n\}$ obtained from the path poset $P(P_n)$ by adding the P_n as the maximum. $\qquad\square$

Corollary 81 *Let A be a unital R–algebra and R a principal ideal domain. Then, for all $n \in \mathbb{N}$, we have $\mathrm{H}^i_\mu(I_n; A) = 0$ for all $i \in \mathbb{N} \setminus \{0\}$, and*

$$\mathrm{rank}_R(\mathrm{H}^0_\mu(I_n; A)) = \begin{cases} \mathrm{rank}_R(A)(\mathrm{rank}_R(A) - 1)^n, & n > 0, \\ \mathrm{rank}_R(A), & n = 0. \end{cases}$$

Proof By (10), the statement follows directly from [Przytycki 2010, Lemma 3.3]. $\qquad\square$

For a unital R–algebra A and an (A, A)–bimodule M, denote by $\mathrm{HH}_*(A, M)$ the *Hochschild homology* of A with coefficients in the bimodule M — see, for instance, [Loday 1992, Section 1.1.3] for the definition. Let $\hat{\mathrm{H}}^*_{\mathrm{Chrom}}(G; A, M)$ denote the homology of the complex $(\hat{C}^*_{\mathrm{Chrom}}(G; A, M), \hat{d})$. We conclude the section showing that the multipath cohomology groups of the polygon agree with the Hochschild homology of A with coefficients in the bimodule M:

Corollary 82 *Let A be a flat unital R–algebra, and M an (A, A)–bimodule and let P_n be the polygon (see Figure 2). Then, we have the following chain of isomorphisms of homology groups:*

$$\mathrm{H}^i_\mu(P_n; A, M) \cong \hat{\mathrm{H}}^i_{\mathrm{Chrom}}(P_n; A, M) \cong \mathrm{HH}_{n-i}(A, M) \quad \textit{for } i = 1, \dots, n.$$

Proof The result follows directly from [Przytycki 2010, Theorem 3.1] and (11). $\qquad\square$

Note that, by [Turner and Wagner 2012, Theorem 1], we have $\mathrm{TW}_i(\mathtt{P}_n; A, M) \cong \hat{\mathrm{H}}_{\mathrm{Chrom}}^{n-i}(\mathtt{P}_n; A, M)$ for i in the set $\{1, \ldots, n\}$. From which follows the isomorphism with the multipath cohomology in this case.

We conclude this section by remarking that in general the chromatic and the multipath homologies are distinct also in the case where A is commutative.

Proposition 83 *The cohomologies* $\mathrm{H}_{\mathrm{Chrom}}$ *and* H_μ *are not isomorphic.*

Proof The multipath homology of the noncoherent 3–step graph is different from the homology of $\mathtt{I}_3$. Since the chromatic homology does not distinguish orientations, the statement follows. □

7.2 Short exact sequences and chromatic homology

Here we apply the machinery developed in Section 5 to obtain a long exact sequence featuring both multipath and chromatic homologies. This clarifies the relationship between the two homology theories. As an application we recover the isomorphisms in the case of the linear graph and polygonal graph, when A is a commutative R–algebra.

For an oriented graph $\mathtt{G}$, let $\hat{C}_{\mathrm{Chrom}}^*(\mathtt{G}; A)$ be the chromatic cochain complex of the underlying unoriented graph. From the results in the previous section, it follows immediately:

Proposition 84 *Let $\mathtt{G}$ be an oriented graph, and let A be a commutative R–algebra. Then, we have the following short exact sequence of complexes*

$$0 \to \widetilde{C}_\mu(\mathtt{G}; A) \to \hat{C}_{\mathrm{Chrom}}(\mathtt{G}; A) \to C_\mu(\mathtt{G}; A) \to 0,$$

where we set

$$\widetilde{C}_\mu(\mathtt{G}; A) := C_{\mathcal{F}_{A,A}}(\mathrm{SSG}(\mathtt{G}) \setminus P(\mathtt{G}))\Big[-\min_{x \in \mathrm{SSG}(\mathtt{G}) \setminus P(\mathtt{G})}\{\ell(x)\}\Big]$$

and we extended $\mathcal{F}_{A,A}(\mathtt{H} \prec \mathtt{H} \cup e)$ to be zero if the number of components of $\mathtt{H}$ and $\mathtt{H} \cup e$ is the same.

Proof Fix a graph $\mathtt{G}$ and consider $\mathcal{F}: \mathbf{SSG}(\mathtt{G}) \to A$. Following the proof of Proposition 62 almost verbatim, we obtain that if P is a downward closed subposet of $\mathrm{SSG}(\mathtt{G})$, then we have the following short exact sequence of chain complexes:

$$0 \to C_{\mathcal{F}|\mathrm{SSG}(\mathtt{G}) \setminus P}(\mathrm{SSG}(\mathtt{G}) \setminus P)\Big[-\min_{x \in \mathrm{SSG}(\mathtt{G}) \setminus P}\{\ell(x)\}\Big] \to C_{\mathcal{F}}(\mathrm{SSG}(\mathtt{G})) \to C_{\mathcal{F}|P}(P) \to 0,$$

where the sign assignments are induced by any sign assignment on $\mathrm{SSG}(\mathtt{G})$. The statement now follows by taking $P = P(\mathtt{G})$ and $\mathcal{F} = \mathcal{F}_{A,A}$. □

As a consequence we (partially) recover one of the main results of this paper:

Corollary 85 *Let A be a commutative unital R–algebra. Then, we have the following isomorphisms of chain complexes (of R–modules):*

$$(12) \qquad\qquad (C_\mu^*(\mathtt{I}_n; A), d) \cong (\hat{C}_{\mathrm{Chrom}}^*(\mathtt{I}_n; A), \hat{\mathtt{d}}) \cong (C_{\mathrm{Chrom}}^*(\mathtt{I}_n; A), \mathtt{d}),$$

$$(13) \qquad\qquad (C_\mu^*(\mathtt{P}_n; A), d) \oplus (A[n+1], 0) \cong (\hat{C}_{\mathrm{Chrom}}^*(\mathtt{P}_n; A), \hat{\mathtt{d}}),$$

where $(A[n+1], 0)$ indicates the cochain complex consisting of a copy of A in degree $n+1$.

Proof It is sufficient to notice that the poset $\mathrm{SSG}(\mathtt{G}) \setminus P(\mathtt{G})$ is either empty (if $\mathtt{G} = \mathtt{I}_n$) or a single point (if $\mathtt{G} = \mathtt{P}_n$). The corollary is an immediate consequence of Proposition 84.　　　　　$\square$

8 Open questions

In this section we gather some open questions.

Question 86 (full functoriality) We have shown in Section 5 that multipath cohomology is a bifunctor when restricting either to the category of rings or to the category of graphs with same number of vertices. Is it possible to lift this result simultaneously to the full categories **Digraph** of directed graphs and R-**Alg** of R–algebras? If not, what are the obstructions to this extension?

Question 87 (cyclic homology theories and extensions) One of the main properties of $\mathrm{H}_\mu(-; A)$ (for a fixed A) is that it recovers (a truncation of) the Hochschild homology of A. To the best of the authors' knowledge, it is still an open question by [Przytycki 2010] whether or not it is possible to recover, in a similar fashion, also the cyclic homology groups of A — see [Loday 1992] for the definition. Moreover, the construction in Section 3.1 can be generalised, by application of the nerve functor and a suitable adaptation, to the realm of ∞–categories — see [Lurie 2009]. In particular, this generalisation should hold for functors in the module categories over commutative ring spectra. A topological enhancement of the cyclic homology theories is given by the so-called *topological Hochschild homology* (or *topological cyclic homology*) — see [Nikolaus and Scholze 2018]. Do we have for topological Hochschild homology, cyclic homology, negative homology, or periodic homology, a result similar to Corollary 82?

Question 88 (categorification of graph invariants) The chromatic homology is named after the chromatic polynomial, which can be obtained as the graded Euler characteristic of the chromatic homology. In other terms, we can say that the chromatic homology is a categorification of the chromatic polynomial. This holds, of course, for a specific choice of the (commutative) algebra A (eg it must be graded or filtered, and its graded dimension should be the chromatic polynomial of a vertex). The first question is: are there natural choices of the algebra A such that the appropriate Euler characteristic of $C_\mu^*(\mathtt{G}; A)$ is a known invariant of the graph $\mathtt{G}$? In general, what are the combinatorial properties of the graded Euler characteristic of the multipath cohomology of a graph with coefficients in a graded algebra?

Question 89 (relationship with Turner–Wagner theory)   We showed that the chromatic homology and the multipath cohomology, when both are defined (ie A commutative), fit into a long exact sequence. Does there exist a long exact sequence, or a spectral sequence, featuring both the multipath and Turner–Wagner homologies with general coefficients?

Question 90 (spectral sequences and applications)   A very interesting and deep feature of Khovanov homology is that it admits a spectral sequence which abuts to a very simple homology called Lee homology [2005]. From this and similar spectral sequences one can extract numerical invariants with interesting applications to low-dimensional topology and knot theory. More importantly, these spectral sequences provide structural information on Khovanov homology. Chromatic homology mimics Khovanov homology. Hence, it is not surprising to find similar spectral sequences and invariants in the context of chromatic homology [Chmutov et al. 2008]. This spectral sequence has been used to prove structural properties of chromatic homology. Are there similar spectral sequences for multipath cohomology? If yes, which kind of information can be extracted from them?

Question 91 (persistent multipath cohomology)   Persistent homology [Edelsbrunner et al. 2002; Zomorodian and Carlsson 2005] is nowadays one of the main tools adopted in topological data analysis, with applications in several domains. One usually starts with a fixed number of data points, joined by (weighted) edges representing the connections between them. These edges are typically added gradually; that is, we have a filtration of the resulting (unoriented) graph G. This filtration is a sequence $G_0 \subset \cdots \subset G_n$ of spanning subgraphs of G. Then, one uses the functorial properties of the classical homology to obtain information in the form of persistent homology groups. Within this framework, one usually works with unoriented graphs, but in concrete applications, graphs are often directed; it is also interesting to compare the undirected versus the directed information (see [Caputi et al. 2021]). Multipath cohomology is a cohomology theory of directed graphs and it is functorial with respect to morphisms of digraphs with the same number of vertices. It is hence natural to define a persistent multipath cohomology for filtrations of digraphs. Which information of the input data can multipath cohomology capture? How does it compare with the analysis using unoriented graphs?

Appendix   Proof of Lemma 39

Lemma 92   *The function σ_e in* (3) *gives a sign assignment on* $P(G)$.

Proof   Consider a square $H \prec H'_1, H'_2 \prec H''$ in $P(G)$. Then, there exist two edges e_1 and e_2 of G such that $H'_1 = H \cup e_1$, $H'_2 = H \cup e_2$, and $H'' = H'_2 \cup e_1 = H'_1 \cup e_2$ (see Example 4 and Figure 15).

The proof is split in cases, according to the number of components of H which are merged by adding the edges e_1 and e_2. First, adding both e_1 and e_2 to H decreases the number of connected components by at most 2. Second, the result of the addition of e_1 and e_2 must still be a multipath — submultipath of H'' to be precise. In particular, observe that cycles are not allowed. It follows that there are two cases:
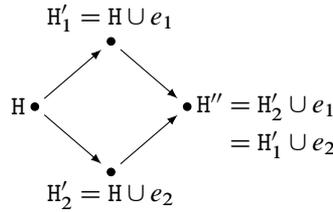
Figure 15: A square in $P(\mathtt{G})$: four multipaths such that $\mathtt{H} \prec \mathtt{H}'_1, \mathtt{H}'_2 \prec \mathtt{H}''$.
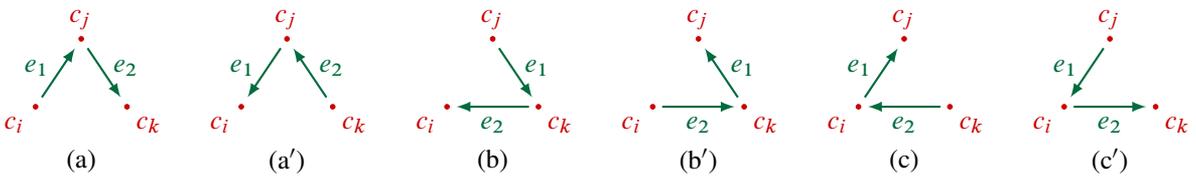

Figure 16: A schematic description of the subcases of case (A). Note that, since the merge of the components c_i, c_j, and c_k must be a path, all possible orientations of e_1 and e_2 are precisely those illustrated.

(A) Three connected components of $\mathtt{H}$ merge into a single connected component of $\mathtt{H}''$.

(B) Four connected components of $\mathtt{H}$ merge into two connected components of $\mathtt{H}''$.

All cases are divided into subcases depending on the indices of the components involved (to be more precise, on the relative order of said indices), and on the orientations of the edges e_1 and e_2 — see Figures 16 and 17. We now proceed with the core of the proof.

(A) Three connected components, c_i, c_j, and c_k, of the multipath $\mathtt{H}$ are merged into a single component of $\mathtt{H}''$. Without loss of generality, up to a permutation of the labels of the components, we may assume that $i < j < k$. Note that e_1 and e_2 cannot be incident to the same pair of components; otherwise $\mathtt{H}''$ would contain a loop. We have six subcases in total — see Figure 16. Since the result of merging the components c_i, c_j, and c_k must be a unique simple path, the orientations of e_1 and e_2 must be coherent; that is, the source of an edge has to be the target of the previous one in the resulting path, and the edges e_1 and e_2 cannot have same sources or targets — eg if the source of e_1 lies in c_k, then the source of e_2 cannot lie in c_k. We report in Table 2 the result of the computation of the signs of σ_e in this case.

(B) Four connected components of $\mathtt{H}$, say c_i, c_j, c_k and c_h, are pairwise merged to obtain exactly two connected components of $\mathtt{H}''$. Without loss of generality we may assume $i < j < k < h$. We have twelve relevant cases, but we can reduce them to six; in fact, a change in the orientation of the edges induces a change of the parity of the index. As a consequence, a simultaneous change in the orientations of both e_1 and e_2 affect our computation by a global sign. All six cases are shown in Figure 17 and the results are summarised in Table 3.

It follows from (A) and (B) that σ_e is a sign assignment on $P(\mathtt{G})$, which concludes the proof. $\qquad\square$

| subcase	$\sigma_e(H, H_1')$	$\sigma_e(H_1', H'')$	$\sigma_e(H, H_2')$	$\sigma_e(H_2', H'')$
(a)	$j+1$	k	$k+1$	$j+1$
(a$'$)	j	$k-1$	k	j
(b)	$k+1$	j	k	j
(b$'$)	k	$j+1$	$k+1$	$j+1$
(c)	$j+1$	$k-1$	k	$j+1$
(c$'$)	j	k	$k+1$	j

Table 2: Computations for all subcases of case (A).
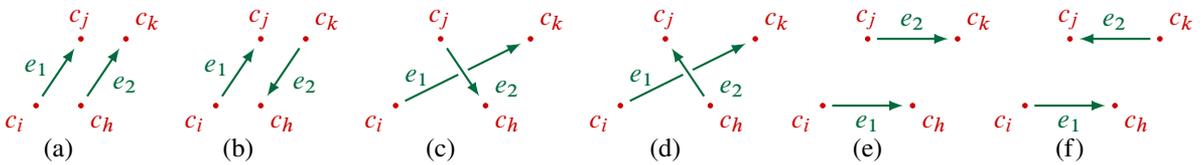
Figure 17: A schematic description of the subcases of case (B) up to a global change in the orientations of e_1 and e_2.

subcase	$\sigma_e(H, H_1')$	$\sigma_e(H_1', H'')$	$\sigma_e(H, H_2')$	$\sigma_e(H_2', H'')$
(a)	$j+1$	k	$k+1$	$j+1$
(b)	$j+1$	$k-1$	k	$j+1$
(c)	$k+1$	h	$h+1$	$k+1$
(d)	$k+1$	$h-1$	h	$k+1$
(e)	$h+1$	$k+1$	$k+1$	h
(f)	$h+1$	k	k	h

Table 3: Computations for all relevant subcases of case (B).

References

[Aharoni et al. 2005] **R Aharoni**, **E Berger**, **R Meshulam**, *Eigenvalues and homology of flag complexes and vector representations of graphs*, Geom. Funct. Anal. 15 (2005) 555–566 MR Zbl

[Björner 1984] **A Björner**, *Posets, regular CW complexes and Bruhat order*, European J. Combin. 5 (1984) 7–16 MR Zbl

[Caputi and Riihimäki 2024] **L Caputi**, **H Riihimäki**, *Hochschild homology, and a persistent approach via connectivity digraphs*, J. Appl. Comput. Topol. 8 (2024) 1121–1170 MR

[Caputi et al. 2021] **L Caputi**, **A Pidnebesna**, **J Hlinka**, *Promises and pitfalls of topological data analysis for brain connectivity analysis*, NeuroImage 238 (2021) art. id. 118245

[Caputi et al. 2023] **L Caputi**, **C Collari**, **S Di Trani**, *Combinatorial and topological aspects of path posets, and multipath cohomology*, J. Algebraic Combin. 57 (2023) 617–658 MR Zbl

[Caputi et al. 2024a] **L Caputi**, **D Celoria**, **C Collari**, *Monotone cohomologies and oriented matchings*, Homology Homotopy Appl. 26 (2024) 137–161 Zbl

[Caputi et al. 2024b]   **L Caputi**, **C Collari**, **S Di Trani**, **J P Smith**, *On the homotopy type of multipath complexes*, Mathematika 70 (2024) art. id. e12235   MR Zbl

[Chandler 2019]   **A Chandler**, *Thin posets*, *CW posets*, *and categorification*, preprint (2019)   arXiv 1911.05600

[Chen et al. 2001]   **B Chen**, **S-T Yau**, **Y-N Yeh**, *Graph homotopy and Graham homotopy*, Discrete Math. 241 (2001) 153–170   MR Zbl

[Chmutov et al. 2008]   **M Chmutov**, **S Chmutov**, **Y Rong**, *Knight move in chromatic cohomology*, European J. Combin. 29 (2008) 311–321   MR Zbl

[Edelsbrunner et al. 2002]   **H Edelsbrunner**, **D Letscher**, **A Zomorodian**, *Topological persistence and simplification*, Discrete Comput. Geom. 28 (2002) 511–533   MR Zbl

[Everitt and Turner 2009]   **B Everitt**, **P Turner**, *Homology of coloured posets: a generalisation of Khovanov's cube construction*, J. Algebra 322 (2009) 429–448   MR Zbl

[Everitt and Turner 2015]   **B Everitt**, **P Turner**, *Cellular cohomology of posets with local coefficients*, J. Algebra 439 (2015) 134–158   MR Zbl

[Gabriel and Zisman 1967]   **P Gabriel**, **M Zisman**, *Calculus of fractions and homotopy theory*, Ergebnisse der Math. 35, Springer (1967)   MR Zbl

[Govc et al. 2021]   **D Govc**, **R Levi**, **J P Smith**, *Complexes of tournaments, directionality filtrations and persistent homology*, J. Appl. Comput. Topol. 5 (2021) 313–337   MR Zbl

[Grigoryan et al. 2016]   **A A Grigoryan**, **Y Lin**, **Y V Muranov**, **S Yau**, *Path complexes and their homologies*, Fundam. Prikl. Mat. 21 (2016) 79–128   MR Zbl   In Russian; translated in J. Math. Sci. 248 (2020) 564–599

[Happel 1989]   **D Happel**, *Hochschild cohomology of finite-dimensional algebras*, from "Séminaire d'Algèbre Paul Dubreil et Marie-Paul Malliavin", Lecture Notes in Math. 1404, Springer (1989) 108–126   MR Zbl

[Helme-Guizon and Rong 2005]   **L Helme-Guizon**, **Y Rong**, *A categorification for the chromatic polynomial*, Algebr. Geom. Topol. 5 (2005) 1365–1388   MR Zbl

[Ivashchenko 1994]   **A V Ivashchenko**, *Contractible transformations do not change the homology groups of graphs*, Discrete Math. 126 (1994) 159–170   MR Zbl

[Jonsson 2008]   **J Jonsson**, *Simplicial complexes of graphs*, Lecture Notes in Math. 1928, Springer (2008)   MR Zbl

[Khovanov 2000]   **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426   MR Zbl

[Kozlov 2008]   **D Kozlov**, *Combinatorial algebraic topology*, Algor. Computat. Math. 21, Springer (2008)   MR Zbl

[Lee 2005]   **E S Lee**, *An endomorphism of the Khovanov invariant*, Adv. Math. 197 (2005) 554–586   MR Zbl

[Loday 1992]   **J-L Loday**, *Cyclic homology*, Grundl. Math. Wissen. 301, Springer (1992)   MR Zbl

[Lurie 2009]   **J Lurie**, *Higher topos theory*, Ann. of Math. Stud. 170, Princeton Univ. Press (2009)   MR Zbl

[Mac Lane 1971]   **S Mac Lane**, *Categories for the working mathematician*, Graduate Texts in Math. 5, Springer (1971)   MR Zbl

[Nikolaus and Scholze 2018]   **T Nikolaus**, **P Scholze**, *On topological cyclic homology*, Acta Math. 221 (2018) 203–409   MR Zbl

[Przytycki 2010]   **J H Przytycki**, *When the theories meet: Khovanov homology as Hochschild homology of links*, Quantum Topol. 1 (2010) 93–109   MR Zbl

[Putyra 2014]   **K K Putyra**, *A 2–category of chronological cobordisms and odd Khovanov homology*, from "Knots in Poland, III: Part 3", Banach Center Publ. 103, Polish Acad. Sci. Inst. Math., Warsaw (2014) 291–355   MR Zbl

[Quillen 1973]   **D Quillen**, *Higher algebraic K–theory, I*, from "Algebraic K–theory, I: Higher K–theories" (H Bass, editor), Lecture Notes in Math. 341, Springer (1973) 85–147   MR Zbl

[Quillen 1978]   **D Quillen**, *Homotopy properties of the poset of nontrivial p–subgroups of a group*, Adv. Math. 28 (1978) 101–128   MR Zbl

[Roos 1961]   **J-E Roos**, *Sur les foncteurs dérivés de* $\varprojlim$: *applications*, C. R. Acad. Sci. Paris 252 (1961) 3702–3704   MR Zbl

[Turner and Wagner 2012]   **P Turner**, **E Wagner**, *The homology of digraphs as a generalization of Hochschild homology*, J. Algebra Appl. 11 (2012) art. id. 1250031   MR Zbl

[Wachs 2003]   **M L Wachs**, *Topology of matching, chessboard, and general bounded degree graph complexes*, Algebra Universalis 49 (2003) 345–385   MR Zbl

[West 1996]   **D B West**, *Introduction to graph theory*, Prentice Hall, Upper Saddle River, NJ (1996)   MR Zbl

[Zomorodian and Carlsson 2005]   **A Zomorodian**, **G Carlsson**, *Computing persistent homology*, Discrete Comput. Geom. 33 (2005) 249–274   MR Zbl

Institute of Mathematics, University of Aberdeen
Aberdeen, United Kingdom

New York University Abu Dhabi
Abu Dhabi, United Arab Emirates

Dipartimento di Matematica "G Castelnuovo", Sapienza Università di Roma
Rome, Italy

luigi.caputi@unibo.it,   carlo.collari@dm.unipi.it,   sabino.ditrani@uniroma1.it

Strong topological rigidity of noncompact orientable surfaces

SUMANTA DAS

We show that every orientable infinite-type surface is properly rigid as a consequence of a more general result. Namely, we prove that if a homotopy equivalence between any two noncompact orientable surfaces is a proper map, then it is properly homotopic to a homeomorphism, provided the surfaces are neither the plane nor the punctured plane. Thus all noncompact orientable surfaces, except the plane and the punctured plane, are *topologically rigid in a strong sense*.

1 Introduction

All manifolds will be assumed to be second countable and Hausdorff. A surface is a 2–dimensional manifold with an empty boundary. All surfaces will be considered connected and orientable. We say a surface is of *infinite type* if its fundamental group is not finitely generated; otherwise, we say it is of *finite type*.

A fundamental question in topology is whether two closed n–manifolds that are homotopy equivalent to each other are homeomorphic. This has a positive answer in dimension 2, as two closed surfaces with isomorphic fundamental groups are homeomorphic. But the same doesn't happen in other dimensions; for example, there are homotopy equivalent lens spaces (a particular type of spherical 3–manifolds) that are not homeomorphic. A closed topological n–manifold M is said to be *topologically rigid* if any homotopy equivalence $N \to M$ with a closed topological n–manifold N as the source is homotopic to a homeomorphism. The *Borel conjecture* (see Rosenberg [34, Conjecture (A Borel)]) asserts that every closed aspherical (ie $\pi_k = 0$ if $k \neq 1$) manifold is topologically rigid. In dimension 2, every closed surface is topologically rigid. This is known as the *Dehn–Nielsen–Baer theorem*; see Dehn [12, Appendix]. The Borel conjecture is known to be true in other dimensions under some additional hypotheses; for example, see Waldhausen [39, Theorem 6.1] and Gabai, Meyerhoff, and Thurston [21, Theorem 0.1(i)] for dimension 3, and for high dimensions see Farrell and Jones [18, proof of Theorem 3.2].

Though noncompact manifolds are not rigid in the above sense, for example in McMillan [29, Theorem 2], the author has constructed (generalizing a construction given by J H C Whitehead) uncountably many contractible open subsets of $\mathbb{R}^3$ such that any two of them are not homeomorphic. Similarly, for noncompact surfaces, we have several examples. In the case of finite-type surfaces we may consider the once-punctured torus and thrice-punctured sphere, which are homotopy equivalent but nonhomeomorphic,

as any homomorphism preserves the cardinality of the puncture set as well as the genus. On the other hand, up to homotopy equivalence, there is precisely one infinite-type surface, but up to homeomorphism there are $2^{\aleph_0}$–many infinite-type surfaces (see Proposition 3.1.11). We consider only noncompact surfaces and discuss their topological rigidity in the *proper category*. Here proper category means the category of spaces with proper maps (recall that a map from a space X to a space Y is called a *proper map* if the inverse image of each compact subset of Y is a compact subset of X). We first define the analogs of homotopy, homotopy equivalence, etc in the proper category.

If a homotopy $\mathcal{H}: X \times [0, 1] \to Y$ is a proper map, then we call $\mathcal{H}$ a *proper homotopy*. Two proper maps from X to Y are said to be *properly homotopic* if there is a proper homotopy between them. We say that a proper map $f: X \to Y$ is a *proper homotopy equivalence* if there exists a proper map $g: Y \to X$ such that both $g \circ f$ and $f \circ g$ are properly homotopic to the identity maps (when such a g exists, g is a *proper homotopy inverse* of f). Two spaces X and Y are said to have the same *proper homotopy type* if there is a proper homotopy equivalence between them. It is worth noting that homotopy through proper maps is a weaker notion than proper homotopy. For example, consider $H: \mathbb{C} \times [0, 1] \to \mathbb{C}$ given by $H(z, t) := tz^2 - z$. Being a polynomial, each $H(-, t)$ is proper. But H itself is not proper as $H(n, 1/n) = 0$ for all integers $n \geq 1$.

The analog of topological rigidity in the proper category is defined as follows: a noncompact topological manifold M without boundary is said to be *properly rigid* if, whenever N is another boundaryless topological manifold of the same dimension and $h: N \to M$ is a proper homotopy equivalence, h is properly homotopic to a homeomorphism. The analog of the Borel conjecture in the proper category, often called the *proper Borel conjecture* (see Chang and Weinberger [11, Conjecure 3.1]), asserts that every noncompact aspherical topological manifold without boundary is properly rigid.

It is known that noncompact finite-type surfaces are properly rigid. Further, using the algebraic tools of classification of noncompact surfaces [23, Theorem 4.1], Goldman showed that two noncompact surfaces of the same proper homotopy type are homeomorphic; see [22, Corollary 11.1]. We show that infinite-type surfaces are also properly rigid. In fact, we show the rigidity of all noncompact surfaces, except for the plane and the punctured plane, under a weaker assumption, namely only assuming the existence of a homotopy inverse, which a priori may or may not be proper. For brevity, define a weaker version of proper homotopy equivalence:

Definition   A homotopy equivalence is said to be a *pseudoproper homotopy equivalence* if it is proper.

Indeed, a proper homotopy equivalence is a pseudoproper homotopy equivalence, though not conversely: a pseudoproper homotopy equivalence has an "ordinary" homotopy inverse but may not have a proper homotopy inverse. For example, consider the φ and ψ below. Our main theorem is the following:

Theorem   Let $f: \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact surfaces. Then Σ' is homeomorphic to Σ. If we further assume that Σ is homeomorphic to neither the plane nor the punctured plane, then f is a proper homotopy equivalence, and there exists a homeomorphism $g_{\text{homeo}}: \Sigma \to \Sigma'$ as a proper homotopy inverse of f.

The reason for the exclusion of the plane and the punctured plane from the hypothesis is almost immediate. Consider $\varphi \colon \mathbb{C} \ni z \mapsto z^2 \in \mathbb{C}$ and $\psi \colon \mathbb{S}^1 \times \mathbb{R} \ni (z, x) \mapsto (z, |x|) \in \mathbb{S}^1 \times \mathbb{R}$. Both of these proper maps are homotopy equivalences, but neither is a proper homotopy equivalence as the degree of a proper homotopy equivalence is ± 1 (see Section 2.6), though $\deg(\varphi) = \pm 2$ (as φ is a twofold branched covering) and $\deg(\psi) = 0$ (as ψ is not surjective; see Lemma 3.6.4.1).

In general, additional assumptions must be imposed on a pseudoproper homotopy equivalence to become a proper homotopy equivalence. For example, using the binary symmetry of the Cantor tree $\mathcal{T}_{\text{Cantor}}$, we have a twofold branched covering $f_{\text{Cantor}} \colon \mathcal{T}_{\text{Cantor}} \to \mathcal{T}_{\text{Cantor}}$ which is undoubtedly a pseudoproper homotopy equivalence (trees are contractible) but not a proper homotopy equivalence (the induced map on $\text{Ends}(\mathcal{T}_{\text{Cantor}})$ by f_{Cantor} is noninjective; see parts (1) and (3) of Proposition 2.3.1). Here is another example: Let M be a connected noncompact contractible boundaryless manifold of dimension $n \geq 2$, and let $f \colon M \to M$ be the composition of a proper map $M \to [0, \infty)$ (using partition of unity) and a nonsurjective proper map $[0, \infty) \to M$ corresponding to an end of M (using compact exhaustion by connected codimension 0–submanifolds; see Guilbault [24, Exercise 3.3.18]). Then f is a pseudoproper homotopy equivalence (M is contractible) but not a proper homotopy equivalence (a proper homotopy equivalence is a surjective map as its degree is ± 1; see Lemma 3.6.4.1).

Brown showed that a pseudoproper homotopy equivalence between two connected finite-dimensional locally finite simplicial complexes is a proper homotopy equivalence if and only if it induces a homeomorphism on the spaces of ends and isomorphisms on all proper homotopy groups [7, Whitehead theorem]. Farrell, Taylor, and Wagoner [19, Corollary 4.10] showed that if $f \colon M \to N$ is a pseudoproper homotopy equivalence between two simply connected noncompact boundaryless n–dimensional smooth manifolds, where both M and N both are simply connected at infinity, then f is a proper homotopy equivalence if and only if $\deg(f) = \pm 1$. Another interesting statement in this context is that a proper map $f \colon X \to Y$ between two locally finite infinite connected 1–dimensional CW–complexes is a proper homotopy equivalence if $\text{Ends}(f)$ is a homeomorphism and f is an extension of a proper homotopy equivalence $X_g \to Y_g$ (where X_g (resp. Y_g) denotes the smallest connected subcomplex of X (resp. Y) that contains all immersed loops of X (resp. Y)); see Algom-Kfir and Bestvina [1, Corollary 3.7].

We conclude this section by citing a few more related results of two different flavors: when does a proper homotopy equivalence exist, and if it does exist, does it determine the space up to homeomorphism. Similar to Kerékjártó's classification theorem (see Theorem 2.4.1), there exists a classification of graphs up to proper homotopy type: two locally finite infinite connected 1–dimensional CW–complexes X and Y have the same proper homotopy type if and only if $\text{rank}(\pi_1(X)) = \text{rank}(\pi_1(Y))$ and there exists a homeomorphism $\varphi \colon \text{Ends}(X) \to \text{Ends}(Y)$ with $\varphi(\text{Ends}(X_g)) = \text{Ends}(Y_g)$; see Ayala, Dominguez, Márquez, and Quintero [3, Theorem 2.7].

As stated earlier, any two noncompact surfaces of the same proper homotopy type are homeomorphic. Sometimes this also happens in other dimensions; for instance, a boundaryless topological manifold

of dimension $n \geq 3$ with the same proper homotopy type as $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^n$; see Edwards [14, Theorem 1] for $n = 3$, Freedman [20, Corollary 1.2] for $n = 4$, and Siebenmann [36, Corollary 1.4] for $n \geq 5$. In contrast, there are exotic pairs: two noncompact connected boundaryless manifolds N and M of dimension $n \geq 5$ exist, where N is smoothable and M is a nonuniform arithmetic manifold, such that M and N have the same proper homotopy type but M is not homeomorphic to N; see Chang and Weinberger [10, Theorem 2.6; 11, pages 137 and 138].

1.1 Main results

The analog of Farb and Margalit's [17, first proof of Theorem 8.9] in the proper category is Theorem 2, which follows almost directly from our main result Theorem 1. Indeed, Theorem 1 is more general.

Theorem 1 (strong topological rigidity) *Let* $f : \Sigma' \to \Sigma$ *be a pseudoproper homotopy equivalence between two noncompact surfaces. Suppose* Σ *is homeomorphic to neither* $\mathbb{R}^2$ *nor* $\mathbb{S}^1 \times \mathbb{R}$. *Then* Σ' *is homeomorphic to* Σ *and* f *is properly homotopic to a homeomorphism.*

Theorem 2 (proper rigidity) *If* $f : \Sigma' \to \Sigma$ *is a proper homotopy equivalence between two noncompact surfaces, then* Σ' *is homeomorphic to* Σ *and* f *is properly homotopic to a homeomorphism.*

A theorem of Edmonds [13, Theorem 3.1] says that any π_1–injective map of degree 1 between two closed surfaces is homotopic to a homeomorphism. The analogous fact for noncompact surfaces is Theorem 3, which classifies all π_1–injective degree 1 maps between two noncompact surfaces and also follows almost directly from Theorem 1.

Theorem 3 (classification of π_1–injective degree 1 maps) *Let* Σ *and* Σ' *be any two noncompact oriented surfaces. Suppose there exists a* π_1–*injective proper map* $f : \Sigma' \to \Sigma$ *of degree* ± 1. *Then* Σ *is homeomorphic to* Σ' *and* f *is properly homotopic to a homeomorphism.*

Proofs of Theorems 1, 2, and 3 can be found in Section 4. A statement equivalent to Theorem 1 is claimed by Brittenham [6, Proposition 2.1(b)] referencing his unpublished work [5], and where the proof of [5] is claimed to be in the spirit of a result of Brown and Tucker [9].

1.2 Outline of the proof of Theorem 1

Let $f : \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces. Suppose Σ is homeomorphic to neither $\mathbb{R}^2$ nor $\mathbb{S}^1 \times \mathbb{R}$.

1.2.1 Decomposition and transversality Let $\mathscr{C}$ be a locally finite pairwise-disjoint collection of smoothly embedded circles on Σ such that $\mathscr{C}$ decomposes Σ into bordered subsurfaces, and a complementary component of this decomposition is homeomorphic to the one-holed torus, the pair of pants, or the punctured disk (see Theorem 3.1.5).

Properly homotope f to make it smooth as well as transverse to $\mathscr{C}$. Thus $f^{-1}(\mathcal{C})$ is either *empty* or a pairwise-disjoint finite collection of smoothly embedded circles on Σ' for each component $\mathcal{C}$ of $\mathscr{C}$; see Theorem 3.2.3.

1.2.2 Removing unnecessary circles Following the steps below, we properly homotope f further so that for each component of $\mathcal{C}$ of $\mathscr{C}$, either $f^{-1}(\mathcal{C})$ is empty or $f|f^{-1}(\mathcal{C}) \to \mathcal{C}$ is a homeomorphism.

(1) Notice that $f^{-1}(\mathscr{C})$ may have infinitely many disk-bounding components. But, in such a case, an arbitrarily large disk in Σ' bounded by a component of the locally finite collection $f^{-1}(\mathscr{C})$ is not possible as $\Sigma' \not\cong \mathbb{R}^2$ (see Lemma 3.3.1), ie there always exists an "outermost disk" bounded by some component of $f^{-1}(\mathscr{C})$. Now, properly homotope f to remove all disk bounding components of $f^{-1}(\mathscr{C})$ upon considering all these outermost disks simultaneously; see Theorem 3.3.5.

(2) Thereafter, using π_1–bijectivity of f, properly homotope f to map each (primitive) component of $f^{-1}(\mathscr{C})$ onto a component of $\mathscr{C}$ homeomorphically; see Theorem 3.4.3.

(3) Since f has homotopy left inverse, any two components of $f^{-1}(\mathscr{C})$ cobound an annulus in Σ' if and only if their f–images are the same, ie an arbitrarily large annulus in Σ' cobounded by two components of $f^{-1}(\mathscr{C})$ is impossible. So, considering all these "outermost annuli" simultaneously, we complete the goal, as stated in the beginning; see Theorem 3.5.3.

1.2.3 Showing f is a degree ± 1 map (see Theorem 3.6.3.1) To rule out the possibility that $f^{-1}(\mathcal{C})$ is empty, where $\mathcal{C}$ is a component of $\mathscr{C}$, we prove $\deg(f) = \pm 1$. This is because $\deg(f)$ remains the same after any proper homotopy of f, and a map of nonzero degree is surjective; see Lemmas 3.6.4.1 and 3.6.4.3. Our aim is to properly homotope f to obtain a closed disk $\mathcal{D} \subseteq \Sigma$ such that $f|f^{-1}(\mathcal{D}) \to \mathcal{D}$ becomes a homeomorphism, and thus we show $\deg(f) = \pm 1$; see Theorem 2.6.1. The argument is based on finding a smoothly embedded finite-type bordered surface S in Σ such that, for each component c of ∂S, we have $f^{-1}(c) \neq \varnothing$, even after any proper homotopy of f. Depending on the nature of S, we consider two cases:

(1) If Σ is either an infinite-type surface or any $S_{g,0,p}$ with high complexity ($g + p \geq 4$ or $p \geq 6$), then using π_1–surjectivity of f, we can choose S as a smoothly embedded pair of pants in Σ such that $\Sigma \setminus S$ has at least two components and every component of $\Sigma \setminus S$ has a nonabelian fundamental group; see Lemmas 3.6.1.2 and 3.6.1.4. Properly homotope f so that it becomes transverse to ∂S. Then remove unnecessary components from the transverse preimage $f^{-1}(\partial S)$. Thus after a proper homotopy, we may assume $f|f^{-1}(c) \to c$ is a homeomorphism for each component c of ∂S. Now, since f is π_1–injective, by the rigidity of the pair of pants (see Theorem 3.6.1.9), after a proper homotopy one can show that $f|f^{-1}(S) \to S$ is a homeomorphism; see Lemma 3.6.1.10. Therefore the required $\mathcal{D}$ can be any disk in $\text{int}(S)$.

(2) If Σ is a finite-type surface, then we choose a smoothly embedded punctured disk S in Σ so that the puncture of S is an end $e \in \text{im}(\text{Ends}(f)) \subseteq \text{Ends}(\Sigma)$. By Theorem 3.6.2.1, this means every deleted neighborhood of e in Σ intersects $\text{im}(f)$, even after any proper homotopy of f. Now, properly

homotope f so that it becomes transverse to ∂S. Then remove unnecessary components from the transverse preimage $f^{-1}(\partial S)$, ie after a proper homotopy we may assume $f \,|\, f^{-1}(\partial S) \to \partial S$ is a homeomorphism (as $\Sigma \not\cong \mathbb{S}^1 \times \mathbb{R}$, the fundamental group of $\Sigma \setminus S$ is nonabelian, and so π_1–surjectivity of f says $f^{-1}(\partial S) \neq \varnothing$, even after any proper homotopy of f). Since f is π_1–injective, by the proper rigidity of the punctured disk (see Theorem 3.6.2.4), after a proper homotopy, one can show that $f \,|\, f^{-1}(S) \to S$ is a homeomorphism; see Lemma 3.6.2.3. Therefore the required $\mathcal{D}$ can be any disk in int(S).

1.2.4 Inverse decomposition By the last three parts, after a proper homotopy, removing unnecessary components from the transverse preimage $f^{-1}(\mathscr{C})$, we may assume that $f \,|\, f^{-1}(C) \to C$ is a homeomorphism for each component C of $\mathscr{C}$. Thus $\mathscr{C}$ and $f^{-1}(\mathscr{C})$ decompose Σ and Σ', respectively, and there is a shape-preserving bijective correspondence between these two collections of complementary components (see Lemmas 3.6.1.10 and 3.6.2.3). On each complementary component, apply either the rigidity of compact bordered surfaces (see Theorem 3.6.1.9) or the proper rigidity of the punctured disk (see Theorem 3.6.2.4). Thus, we have a collection of boundary-relative proper homotopies such that by pasting them, a proper homotopy from f to a homeomorphism $\Sigma' \to \Sigma$ can be constructed; see the proof of Theorem 1 in Section 4.

2 Background

2.1 Conventions

A *bordered surface* (resp. *surface*) is a connected orientable 2–dimensional manifold with a nonempty (resp. an empty) boundary. For integers $g \geq 0$, $b \geq 0$, and $p \geq 0$, denote the connected orientable 2–manifold of genus g with b boundary components by $S_{g,b}$, and let $S_{g,b,p}$ be the 2–manifold after removing p points from int($S_{g,b}$). Note that for a manifold M, we use int(M) to denote the interior of M. Sometimes $S_{0,1}$, $S_{0,2}$, $S_{0,3}$, $S_{1,2}$, and $S_{0,1,1}$ will be called a disk, an annulus, a pair of pants, a two-holed torus, and a punctured disk, respectively.

We say a connected 2–manifold with or without boundary is of *infinite type* if its fundamental group is not finitely generated; otherwise, we say it is of *finite type*.

2.2 Simple closed curves on 2–manifolds

Definition 2.2.1 Let S be a connected orientable 2–dimensional manifold with or without boundary. A *circle* (resp. *smoothly embedded circle*) on S is the image of an embedding (resp. a smooth embedding) of $\mathbb{S}^1$ into S. We say a circle $\mathcal{C}$ on S is a *trivial circle* if there is an embedded disk $\mathcal{D}$ in S such that $\partial \mathcal{D} = \mathcal{C}$, and we say a circle $\mathcal{C}$ on S is a *primitive circle* if it is not a trivial circle.

The following theorem justifies naming a nondisk bounding circle a primitive circle: a primitive circle represents a primitive element of the fundamental group. Recall that an element g of a group G is *primitive* if there does not exist any $h \in G$ such that $g = h^k$, where $|k| > 1$.

Theorem 2.2.2 [15, Theorems 1.7. and 4.2] *Let S be a connected orientable 2–dimensional manifold with or without boundary. Let C be a primitive circle on S, and let $f : \mathbb{S}^1 \hookrightarrow S$ be an embedding with $f(\mathbb{S}^1) = C$. Then $[f] \in \pi_1(S)$ is a primitive element. In particular, $[f]$ is a nontrivial element of $\pi_1(S)$.*

Recall that for a path-connected space X, there is a bijective correspondence between the set of all conjugacy classes of $\pi_1(X, *)$ and the set of all free homotopy classes of maps $\mathbb{S}^1 \to X$. The next theorem says that two pairwise-disjoint freely homotopic primitive circles on a 2–manifold cobound an annulus.

Theorem 2.2.3 [15, Lemma 2.4] *Let S be a connected orientable 2–dimensional manifold with or without boundary. Let $\ell_0, \ell_1 : \mathbb{S}^1 \hookrightarrow S$ be two embeddings such that $\ell_0(\mathbb{S}^1)$ is a smoothly embedded submanifold of S and $\ell_0(\mathbb{S}^1) \cap \ell_1(\mathbb{S}^1) = \varnothing$. If ℓ_0 and ℓ_1 represent the same nontrivial conjugacy class in $\pi_1(S, *)$, then there is an embedding $\mathcal{L} : \mathbb{S}^1 \times [0, 1] \hookrightarrow S$ such that $\mathcal{L}(-, 0) = \ell_0$ and $\mathcal{L}(-, 1) = \ell_1$.*

2.3 Ends of spaces

Let X be a connected separable locally compact locally connected Hausdorff ANR (absolute neighborhood retract) space. For example, X can be any connected topological manifold. We say X admits an *efficient exhaustion by compacta* if there is a nested sequence $K_1 \subseteq K_2 \subseteq \cdots$ of compact connected subsets of X such that $\bigcup_i K_i = X$, $K_i \subseteq \operatorname{int}(K_{i+1})$, $\bigcap_i (X \setminus K_i) = \varnothing$, and the closure of each component of any $X \setminus K_i$ is noncompact. For the existence of efficient exhaustion of X by compacta, see [24, Exercise 3.3.4].

Let $\operatorname{Ends}(X)$ be the set of all sequences $(V_1, V_2, \dots)$, where V_i is a component of $X \setminus K_i$ and $V_1 \supseteq V_2 \supseteq \cdots$. Set $X^\dagger := X \cup \operatorname{Ends}(X)$ with the topology generated by the basis consisting of all open subsets of X, and all sets $V_i^\dagger$, where

$$V_i^\dagger := V_i \cup \{(V_1', V_2', \dots) \in \operatorname{Ends}(X) : V_i' = V_i\}.$$

Then $X^\dagger$ is separable, compact, and metrizable, so X is an open dense subset of $X^\dagger$; it is known as the *Freudenthal compactification* of X (recall that we say a space X_c is a *compactification* of X if X_c is compact Hausdorff space and X is a dense subset of X_c). The subspace $\operatorname{Ends}(X)$ of $X^\dagger$ is a totally disconnected space; hence $\operatorname{Ends}(X)$ is a closed subset of the Cantor set.

The Freudenthal compactification *dominates* any other compactification: If $\widetilde{X}$ is a compactification of X such that $\widetilde{X} \setminus X$ is totally disconnected, then there exists a map $f : X^\dagger \to \widetilde{X}$ extending Id_X.

Also, the Freudenthal compactification is *unique* in the following sense: If $X^{\dagger\dagger}$ is a compactification of X such that $X^{\dagger\dagger} \setminus X$ is totally disconnected and $X^{\dagger\dagger}$ dominates any other compactification, then there exists a homeomorphism $X^{\dagger\dagger} \to X^\dagger$ extending Id_X; see [22, Theorem 3.1]. Thus the definition of $\operatorname{Ends}(X)$ is independent of the choice of efficient exhaustion of X by compacta.

Now we consider a relationship between Ends and proper maps:

Proposition 2.3.1 [24, Proposition 3.3.12] *Let X and Y be two connected separable locally compact locally connected Hausdorff ANRs. Then we have the following*:

(1) *Every proper map $f : X \to Y$ induces a map $\mathrm{Ends}(f) : \mathrm{Ends}(X) \to \mathrm{Ends}(Y)$ that can be used to extend $f : X \to Y$ to a map $f^\dagger : X^\dagger \to Y^\dagger$ between the Freudenthal compactifications.*

(2) *If two proper maps $f_0, f_1 : X \to Y$ are properly homotopic, then $\mathrm{Ends}(f_0) = \mathrm{Ends}(f_1)$.*

(3) *If $f : X \to Y$ is a proper homotopy equivalence, then $\mathrm{Ends}(f) : \mathrm{Ends}(X) \to \mathrm{Ends}(Y)$ is a homeomorphism.*

More about the ends of spaces and proper homotopy can be found in [31; 26].

2.4 Kerékjártó's classification theorem and Ian Richards' representation theorem

Let Σ be a noncompact surface with an efficient exhaustion $\{K_i\}_1^\infty$. Let $e := (V_1, V_2, \dots) \in \mathrm{Ends}(\Sigma)$ be an end, where V_i is a component of $X \setminus K_i$. We say e is a *planar end* if V_i is embeddable in $\mathbb{R}^2$ for some positive integer i. An end is said to be *nonplanar* if it is not planar. Denote the subspace of $\mathrm{Ends}(\Sigma)$ consisting of all planar (resp. nonplanar) ends by $\mathrm{Ends}_p(\Sigma)$ (resp. $\mathrm{Ends}_{np}(\Sigma)$). Note that $\mathrm{Ends}_p(\Sigma)$ is an open subset of $\mathrm{Ends}(\Sigma)$. Define the *genus* of Σ as $g(\Sigma) := \sup g(\mathcal{S})$, where $\mathcal{S}$ is a compact bordered subsurface of Σ. Therefore, the genus counts the number of handles of a surface, ie the number of embedded copies of $S_{1,1}$ in a surface, which may be any nonnegative integer or countably infinite.

Theorem 2.4.1 (Kerékjártó's classification of noncompact surfaces [33, Theorem 1]) *Let Σ and Σ' be noncompact surfaces of genus g and g', respectively. Then Σ is homeomorphic to Σ' if and only if $g = g'$ and there is a homeomorphism $\varphi : \mathrm{Ends}(\Sigma) \to \mathrm{Ends}(\Sigma')$ with $\varphi(\mathrm{Ends}_{np}(\Sigma)) = \mathrm{Ends}_{np}(\Sigma')$.*

Theorem 2.4.2 (realization of ends and representation of a noncompact surface [33, Theorems 2 and 3]) *Let $\mathscr{E}_{np} \subseteq \mathscr{E}$ be two closed totally disconnected subsets of $\mathbb{S}^1$, and let $\mathscr{G}$ be an at most countable set such that $\mathscr{E} \neq \varnothing$, and $\mathscr{E}_{np} \neq \varnothing$ if and only if $\mathscr{G}$ is infinite. Define $\mathbb{D} := \{z \in \mathbb{C} : 0 \leq |z| \leq 1\}$. Then there exists a pairwise-disjoint collection $\{\mathcal{D}_i : i \in \mathscr{G}\}$ of disks in $\mathrm{int}(\mathbb{D})$ such that a point $p \in \mathbb{D}$ is an element of $\mathscr{E}_{np}$ if and only if every neighborhood of p in $\mathbb{D}$ contains infinitely many elements of $\{\mathcal{D}_i : i \in \mathscr{G}\}$. Moreover, $S := (\mathbb{D} \setminus \mathscr{E}) \setminus \bigcup_{i \in \mathscr{G}} \mathrm{int}(\mathcal{D}_i)$ is a noncompact bordered surface, and*

$$DS := \frac{(S \times 0) \sqcup (S \times 1)}{(p, 0) \sim (p, 1)} \quad \text{for } p \in \partial S$$

is a genus-$|\mathscr{G}|$ noncompact surface with $\mathrm{Ends}(DS) \cong \mathscr{E}$ and $\mathrm{Ends}_{np}(DS) \cong \mathscr{E}_{np}$.

Thus, given any noncompact surface Σ, in this procedure, if we assume $\mathscr{E}_{np} \subseteq \mathscr{E}$ is homeomorphic to the pair $\mathrm{Ends}_{np}(\Sigma) \subseteq \mathrm{Ends}(\Sigma)$, and $|\mathscr{G}|$ is equal to $g(\Sigma)$, then $DS \cong \Sigma$ by Theorem 2.4.1.

Remark 2.4.3 The classification of noncompact bordered surfaces is also possible: When the boundary is compact, it follows from Theorem 2.4.1 with [38, Proposition A.3]. When each boundary component
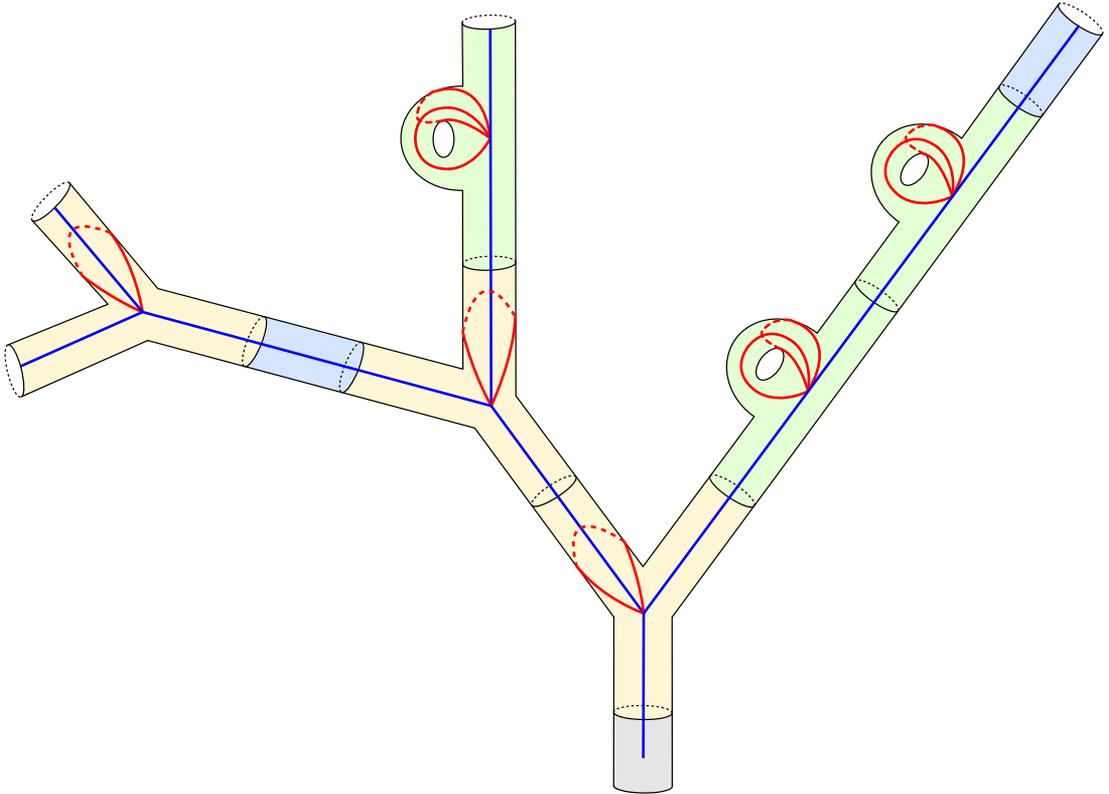
Figure 1: The inductive construction of any noncompact surface and its spine uses four compact bordered surfaces: the disk, annulus, pair of pants, and torus with two holes.

is compact, this follows from [4] (based on the classification of their interiors) or [38, Theorem A.7] (based on the classification of noncompact surfaces obtained from gluing a disk along each boundary component). For arbitrary boundary, see [8, Theorem 2.2].

2.5 Goldman's inductive procedure of constructing all noncompact surfaces

A noncompact surface Σ_{std} is said to be in *standard form* if it is built up from four building blocks, $S_{0,1}$, $S_{0,2}$, $S_{0,3}$, and $S_{1,2}$, in the following inductive manner: Start with $S_{0,1}$. Suppose the ith step of the induction has already been done. Let K_i be the compact bordered subsurface of Σ_{std} after the ith step of induction. In particular, $K_1 \cong S_{0,1}$. Now, to obtain K_{i+1} from K_i, consider one of the last three building blocks, say $\mathcal{S}$ (homeomorphic to $S_{0,2}$, $S_{0,3}$, or $S_{1,2}$); finally, suitably identify one boundary circle of $\mathcal{S}$ with a boundary circle of K_i; see Figure 1.

Theorem 2.5.1 [23, Section 2.6; 27, page 173] *Let Σ be a noncompact surface. Then Σ is homeomorphic to a noncompact surface Σ_{std} in standard form. Thus every noncompact surface is homeomorphic to a noncompact surface constructed using an inductive procedure as above, though two noncompact surfaces obtained from two different inductive procedures may be homeomorphic.*

Theorem 2.5.2 [23, Section 2.6. and Section 7.3] *The graph in Figure 1 consisting of blue straight line segments and red circles is a deformation retract of the surface Σ. Thus Σ is homotopy equivalent to the wedge of at most countably many circles. In particular, $\pi_1(\Sigma)$ is free.*

Remark 2.5.3 An alternative proof of the last two sentences of Theorem 2.5.2 is [35, Lemma 3.2.2].

2.6 The degree of a proper map

We use singular cohomology with compact support to define the notion of the degree of a proper map. Recall that for a topological manifold X, the rth singular cohomology with compact support $H_c^r(X, \partial X; \mathbb{Z})$ is equal to the direct limit $\varinjlim H^r(X, \partial X \cup (X \setminus K); \mathbb{Z})$, where K is a compact subset of X and various maps to define this direct system are inclusion induced maps. Hence, for a compact subset K of X, the definition of the direct limit yields an *obvious map* $H^r(X, \partial X \cup (X \setminus K); \mathbb{Z}) \to H_c^r(X, \partial X; \mathbb{Z})$. It is worth noting that when X is a compact topological manifold, $H_c^r(X, \partial X; \mathbb{Z}) = H^r(X, \partial X; \mathbb{Z})$ for all r.

Let X and Y be two topological manifolds. If $f: X \to Y$ is a proper map with $f(\partial X) \subseteq \partial Y$, then for each r, f induces a map $H_c^r(f): H_c^r(Y, \partial Y; \mathbb{Z}) \to H_c^r(X, \partial X; \mathbb{Z})$ such that H_c^r becomes a functor in the following sense: the induced map of the identity is the identity, and the induced map of a (well-defined) composition of two proper maps (each of which sends boundary into boundary) is the composition of their induced maps. Moreover, if $\mathcal{H}: X \times [0, 1] \to Y$ is a proper homotopy such that $\mathcal{H}(\partial X, t) \subseteq \partial Y$ for each $t \in [0, 1]$, then $H_c^r(\mathcal{H}(-, 0)) = H_c^r(\mathcal{H}(-, 1))$ for all r. For more details see [37, pages 320, 322, 323, 339, and 341].

Let M be a connected orientable topological n–manifold. Then $H_c^n(M, \partial M; \mathbb{Z})$ is an infinite cyclic group; see [37, page 342]. If we choose an orientation of M (ie M is oriented), then there exists a unique element $[M] \in H_c^n(M, \partial M; \mathbb{Z})$ such that $[M]$ generates $H_c^n(M, \partial M; \mathbb{Z})$, and for each $x \in M \setminus \partial M$, the unique generator of $H^n(M, M \setminus x; \mathbb{Z})$, which comes from the chosen orientation of M, is sent to $[M]$ by the obvious isomorphism $H^n(M, M \setminus x; \mathbb{Z}) \to H_c^n(M, \partial M; \mathbb{Z})$; see [16, proof of Lemma 2.1]. Thus if $f: M \to N$ is a proper map between two connected oriented topological n–manifolds with $f(\partial M) \subseteq \partial N$, the (*compactly supported cohomological*) *degree* of f is the unique integer $\deg(f)$ defined by $H_c^n(f)([N]) = \deg(f)[M]$.

By the previous two paragraphs, we have the following:

(i) When manifolds are compact, the notion of compactly supported cohomological degree agrees with the notion of the usual degree defined by singular cohomology.

(ii) The degree is proper homotopy invariant: if $f, g: M \to N$ are proper maps between two connected oriented topological n–manifolds with $f(\partial M) \cup g(\partial M) \subseteq \partial N$ such that there is a proper homotopy $\mathcal{H}: M \times [0, 1] \to N$ with $\mathcal{H}(\partial M \times [0, 1]) \subseteq \partial N$ from f to g, then $\deg(f) = \deg(g)$.

(iii) The degree is multiplicative: the degree of the (well-defined) composition of two proper maps (each of which sends boundary into boundary) is the product of their degrees.

Therefore the degree of a proper homotopy equivalence between two oriented connected boundaryless n–manifolds is ± 1 due to (ii) and (iii) above. We use the following well-known characterizations of a map of degree ± 1. In the below two theorems, "D is a disk in a smooth n–manifold X" means D is the image of $\{z \in \mathbb{R}^n : |z| \leq 1\}$ under a smooth embedding $\{z \in \mathbb{R}^n : |z| \leq 2\} \hookrightarrow X$.

Theorem 2.6.1 [16, Lemma 2.1b] *Let $f : M \to N$ be a proper map between two connected oriented smooth manifolds of the same dimension such that $f^{-1}(\partial N) = \partial M$. Suppose for a disk D in $\mathrm{int}(N)$, $f^{-1}(D)$ is a disk in $\mathrm{int}(M)$ such that f maps $f^{-1}(D)$ homeomorphically onto D. Then $\deg(f) = +1$ or -1 according to whether $f | f^{-1}(D) \to D$ is orientation preserving or orientation reversing.*

The following theorem is due to Hopf, and says that for a degree 1 map, we can achieve such a disk with nice properties, as mentioned in Theorem 2.6.1, after a proper homotopy:

Theorem 2.6.2 [16, Theorems 3.1 and 4.1] *Let $f : M \to N$ be a proper map between two connected oriented smooth manifolds of the same dimension such that $f^{-1}(\partial N) \subseteq \partial M$. Suppose $\deg(f) = \pm 1$. Then there is a proper map $g : M \to N$ with $g(\partial M) \subseteq \partial N$ and a homotopy $\mathcal{H} : M \times [0, 1] \to N$ from f to g with the following properties:*

- *There exists a compact subset $K \subseteq \mathrm{int}(M)$ such that $\mathcal{H}(x, t) = f(x)$ for all $(x, t) \in (M \setminus K) \times [0, 1]$. In particular, H is a proper homotopy and $\mathcal{H}(\partial M, t) \subseteq \partial N$ for all $t \in [0, 1]$.*
- *There exists a disk $D \subseteq \mathrm{int}(N)$ such that $g^{-1}(D)$ is a disk in $\mathrm{int}(M)$ and $g | g^{-1}(D) \to D$ is a homeomorphism.*

The theorem below is due to Olum, and roughly says that when there is a degree 1 map, the domain is more massive than the codomain.

Theorem 2.6.3 [16, Corollary 3.4] *Let $f : M \to N$ be a proper map between two connected oriented topological manifolds of the same dimension such that $f(\partial M) \subseteq \partial N$. If $\deg(f) = \pm 1$, then $\pi_1(f) : \pi_1(M) \to \pi_1(N)$ is surjective.*

3 Ingredients for proving Theorem 1

3.1 Decomposition of a noncompact surface into pairs of pants and punctured disks

Every compact surface of genus $g \geq 2$ is the union (with pairwise-disjoint interiors) of $(2g-2)$–many copies of the pair of pants, but the same thing doesn't happen for noncompact surfaces. For example, the thrice punctured sphere is not a union (with pairwise-disjoint interiors) of copies of the pair of pants; we need copies of the punctured disk. The main aim of this section is to prove that every noncompact surface, except the plane and the once punctured torus, decomposes into copies of the pair of pants and copies of the punctured disk when we cut it along a collection of circles, where each circle of this collection has an open neighborhood that does not intersect with any other circles of this collection.

First, we define some terminology:

Definition 3.1.1 Let X be a space, and let $\{X_\alpha : \alpha \in \mathscr{I}\}$ be a collection of subsets of X. We say $\{X_\alpha : \alpha \in \mathscr{I}\}$ is a *locally finite collection* and write $X_\alpha \to \infty$ if, for each compact subset K of X, $X_\alpha \cap K = \varnothing$ for all but finitely many $\alpha \in \mathscr{I}$.

Definition 3.1.2 Let $\mathscr{A}$ be a pairwise-disjoint collection of smoothly embedded circles on a surface Σ. We say $\mathscr{A}$ is a *locally finite curve system* (in short, LFCS) on Σ if $\mathscr{A}$ is a locally finite collection.

Remark 3.1.3 Let $\mathscr{A}$ be an LFCS on a surface Σ. Notice that $\bigcup \mathscr{A}$ (ie the union of all elements of $\mathscr{A}$) is a closed subset of Σ as well as a smoothly embedded submanifold of Σ such that the set of all components of $\bigcup \mathscr{A}$ is $\mathscr{A}$. But to simplify notation, whenever needed we will think of $\mathscr{A}$ and $\bigcup \mathscr{A}$ as the same without any harm.

Definition 3.1.4 Let $\mathscr{A}$ be an LFCS on a surface Σ. Suppose there exists an at most countable collection $\{\Sigma_n\}$ of bordered subsurfaces of Σ such that

(1) each Σ_n is a closed subset of Σ,

(2) $\text{int}(\Sigma_n) \cap \text{int}(\Sigma_m) = \varnothing$ if $n \neq m$,

(3) $\bigcup_n \Sigma_n = \Sigma$, and

(4) $\bigcup_n \partial \Sigma_n = \bigcup \mathscr{A}$.

In this case, we say $\mathscr{A}$ *decomposes Σ into bordered subsurfaces*, where *complementary components* are $\{\Sigma_n\}$. Also, we call each component of $\mathscr{A}$ a *decomposition circle*.

The following theorem asserts that any noncompact surface other than the plane has a decomposition, where each complementary part is either a pair of pants, a one-holed torus, or a punctured disk. This decomposition of the codomain of a pseudoproper homotopy equivalence will be used in all cases.

Theorem 3.1.5 *Let Σ be a noncompact surface not homeomorphic to $\mathbb{R}^2$. Then there is an LFCS $\mathscr{C}$ on Σ such that $\mathscr{C}$ decomposes Σ into bordered subsurfaces, and a complementary component of this decomposition is homeomorphic to either $S_{1,1}$ (used at most once), $S_{0,3}$, or $S_{0,1,1}$.*

Proof It is enough to find a collection $\{\Sigma_n\}$ of bordered subsurfaces of Σ with the four properties, as mentioned in Definition 3.1.4, so that each Σ_n is homeomorphic to either $S_{0,3}$, $S_{1,1}$, or $S_{0,1,1}$. For that, consider an inductive construction of Σ; see Theorem 2.5.1. Now, a finite sequence of annuli, when added to the compact bordered surface used just before it, can be ignored. Thus we may assume $S_{0,3}$ or $S_{1,2}$ is used after $S_{0,1}$ without loss of generality because $\Sigma \not\cong \mathbb{R}^2$, and hence, pushing $S_{0,1}$ into $\text{int}(S_{0,3})$ or $\text{int}(S_{1,2})$, we end up with $S_{0,2}$ (which can be ignored) or $S_{1,1}$. We complete the proof by observing that $S_{1,2}$ can be decomposed into two copies of $S_{0,3}$, and $S_{0,1,1}$ is the union (with pairwise-disjoint interiors) of countably many copies of $S_{0,2}$. $\square$

Remark 3.1.6 A statement closely related to Theorem 3.1.5 is [2, Theorem 1.1], which says that *"every surface except for the sphere, the plane, and the torus is the union (with pairwise-disjoint interiors) of copies of the pair of pants and copies of the punctured disk"*. But due to Definition 3.1.4(4), if we want
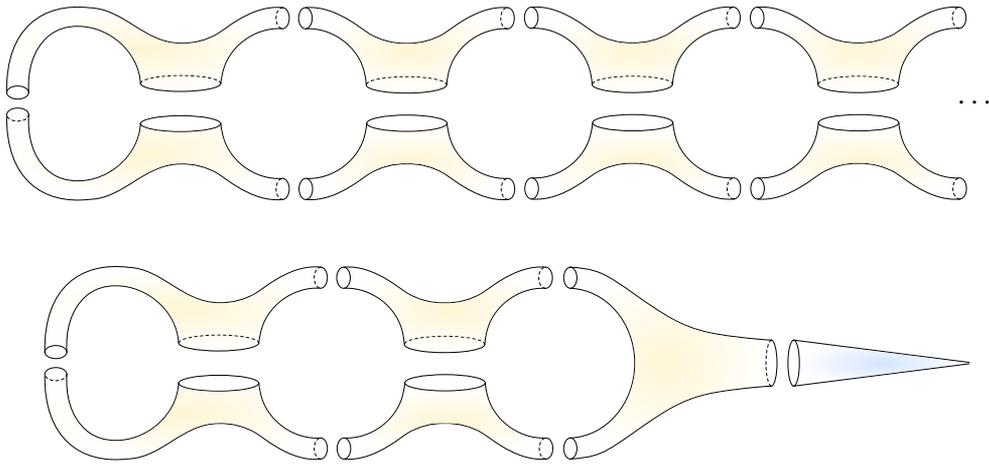
Figure 2: Top: decomposition of the Loch Ness monster into countably infinitely many copies of the pair of pants. Bottom: decomposition of $S_{3,0,1}$ into five copies of the pair of pants and a copy of the punctured disk.

that any complementary component is homeomorphic to only either $S_{0,3}$ or $S_{0,1,1}$, then we also need to assume that $\Sigma \ncong S_{1,0,1}$; see Figure 2 and Theorem 3.1.7.

Theorem 3.1.7 *Let Σ be a noncompact surface that is not homeomorphic to either $\mathbb{R}^2$ or $S_{1,0,1}$. Then there is an LFCS $\mathscr{C}'$ on Σ such that $\mathscr{C}'$ decomposes Σ into bordered subsurfaces, and a complementary component of this decomposition is homeomorphic to either $S_{0,3}$ or $S_{0,1,1}$.*

Proof It is enough to find a collection $\{\Sigma_n\}$ of bordered subsurfaces of Σ with the four properties of Definition 3.1.4 such that each Σ_n is homeomorphic to either $S_{0,3}$ or $S_{0,1,1}$. For that, consider an inductive construction of Σ; see Theorem 2.5.1. We will divide the whole proof into two cases, depending on whether Σ has at least two ends.

At first, suppose the number of ends of Σ is at least two. Now, the definition of the space of ends tells us that we need to use at least one pair of pants in the inductive construction of Σ. By Lemma 3.1.8, we may assume that in this inductive construction, a pair of pants is used just after the disk. An argument similar to before (see the proof of Theorem 3.1.5) concludes this case.

Next, consider the case when the number of ends of Σ is precisely one. That is, Σ can be either the Loch Ness monster (the infinite-genus surface with one end) or $S_{g,0,1}$ with $g \geq 2$. The Loch Ness monster decomposes into countably infinitely many copies of the pair of pants, and $S_{g,0,1}$ with $g \geq 2$ decomposes into $(2g-1)$–many copies of the pair of pants and a copy of the punctured disk; see Figure 2. □

To prove Theorem 3.1.7 we used Lemma 3.1.8, which says that in an inductive construction of a noncompact surface, interchanging the positions of the compact bordered surfaces used in the first few inductive steps doesn't change the homeomorphism type; its proof is based on the observation that the portions outside compact subsets determine the space of ends.
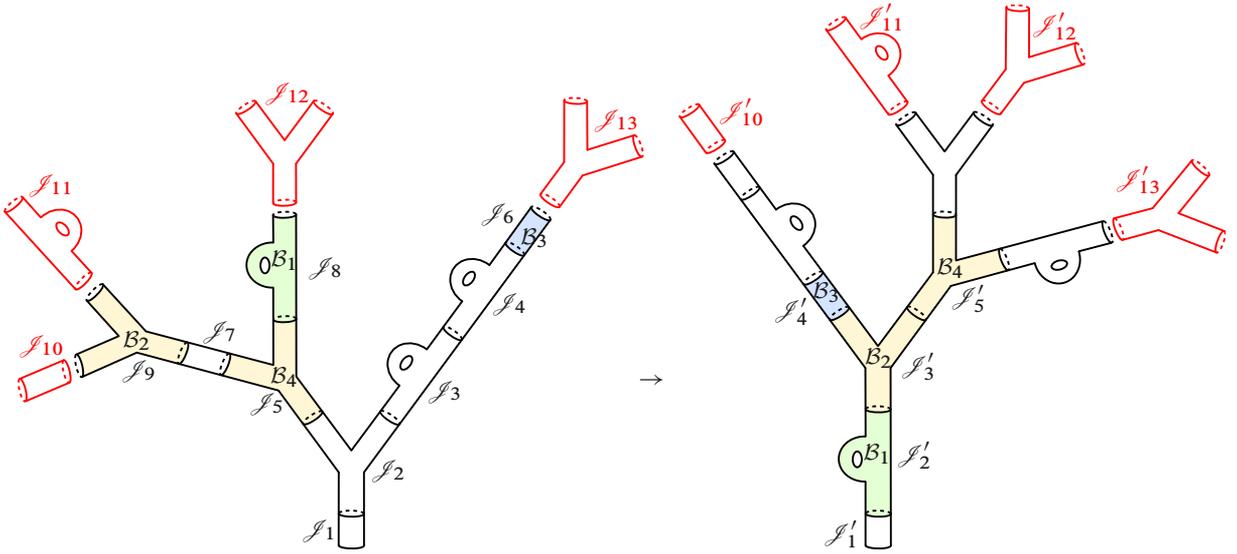
Figure 3: $\mathcal{I}_r$ (resp. $\mathcal{I}'_r$) denotes the rth step of $\mathcal{I}$ (resp. $\mathcal{I}'$). Here $n_0 = 9$ and $n = 4$. Also, the red-colored compact bordered surfaces are the portions of S, and the inductive construction of S given by $\mathcal{I}'$ is inherited from the inductive construction of S given by $\mathcal{I}$.

Lemma 3.1.8 *Let Σ be a noncompact surface with some inductive construction $\mathcal{I}$. Denote the compact bordered subsurface of Σ after the ith step of $\mathcal{I}$ by K_i. Suppose*

$$\{\mathcal{B}_1, \ldots, \mathcal{B}_n : \text{ each } \mathcal{B}_l \text{ is homeomorphic to } S_{0,2}, S_{0,3}, \text{ or } S_{1,2}\}$$

is a finite collection of compact bordered surfaces such that $\mathcal{B}_l$ is used to construct K_{i_l+1} from K_{i_l} for each $l = 1, \ldots, n$. Then there exists a noncompact surface Σ' with an inductive construction $\mathcal{I}'$ such that $\Sigma' \cong \Sigma$ and $\mathcal{B}_l$ is used to construct K'_{l+1} from K'_l for each $l = 1, \ldots, n$, where K'_i denotes the compact bordered subsurface of Σ after the ith step of $\mathcal{I}'$.

Proof Let n_0 be a positive integer such that K_{n_0} contains each $\mathcal{B}_l$. Define $S := \Sigma \setminus \text{int}(K_{n_0})$. Thus S is a bordered subsurface of Σ with $\partial S = \partial K_{n_0}$. Now consider all copies of different building blocks used up to the n_0^{th} step of $\mathcal{I}$, and inside K_{n_0} interchange them so that $\mathcal{B}_1, \ldots, \mathcal{B}_n$ comes just after the initial disk K_1 one by one following the increasing order of their indices. Denote the result of this interchange process by K'_{n_0}. So $K_{n_0} \cong K'_{n_0}$ as $g(K_{n_0}) = g(K'_{n_0})$ and $\partial K_{n_0} \cong \partial K'_{n_0}$. Define a noncompact surface Σ' as $\Sigma' := K'_{n_0} \cup_{\partial K'_{n_0} \equiv \partial S} S$. Therefore $\Sigma \setminus K_{n_0} = \text{int}(S) = \Sigma' \setminus K'_{n_0}$ (notice that we are thinking of S as a subset of Σ' using the obvious embedding $S \hookrightarrow \Sigma'$).

Choose an inductive construction $\mathcal{I}'_{\leq n_0}$ of K'_{n_0} such that the ith element of the ordered sequence $K_1, \mathcal{B}_1, \ldots, \mathcal{B}_l$ is used in the ith step of $\mathcal{I}'_{\leq n_0}$. Then $\mathcal{I}$ gives a truncated inductive construction $\mathcal{I}|S$ on S starting from the $(n_0+1)^{\text{th}}$ step. Now $\mathcal{I}'_{\leq n_0}$ followed by $\mathcal{I}|S$ together gives an inductive construction $\mathcal{I}'$ of Σ'. Roughly it means $\mathcal{I}'$ is the same as the inductive construction of Σ, except for the first few steps.

Denote the compact bordered subsurface of Σ' after the ith step of $\mathscr{I}'$ by K_i'. To complete the proof, we show $\Sigma' \cong \Sigma$ using Theorem 2.4.1.

Consider the efficient exhaustion $\{K_i\}$ (resp. $\{K_i'\}$) of Σ (resp. Σ') by compacta to define $\mathrm{Ends}(\Sigma)$ (resp. $\mathrm{Ends}(\Sigma')$). Recall that the space of ends remains the same up to homeomorphism even if we choose a different efficient exhaustion by compacta; see Section 2.3. By $\Sigma \setminus K_{n_0} = \mathrm{int}(S) = \Sigma' \setminus K_{n_0}'$, for every sequence $(V_1, V_2, \dots) \in \mathrm{Ends}(\Sigma)$, there exists a unique sequence $(V_1', V_2', \dots) \in \mathrm{Ends}(\Sigma')$ such that $V_i = V_i'$ for all integers $i \geq n_0$, and conversely. Thus there exists a homeomorphism $\varphi \colon \mathrm{Ends}(\Sigma) \to \mathrm{Ends}(\Sigma')$ with $\varphi(\mathrm{Ends}_{\mathrm{np}}(\Sigma)) = \mathrm{Ends}_{\mathrm{np}}(\Sigma')$. Also, $\Sigma \setminus K_{n_0} = \mathrm{int}(S) = \Sigma' \setminus K_{n_0}'$ and $K_{n_0} \cong K_{n_0}'$ together imply $g(\Sigma) = g(\Sigma')$. Therefore $\Sigma' \cong \Sigma$ by Theorem 2.4.1. $\qquad\square$

The spine construction of Goldman's inductive procedure shows that every noncompact surface Σ (possibly of infinite type) is the interior of a bordered surface: consider the graph $\mathcal{G}$ consisting of blue straight line segments and red circles, as given in Figure 1. Any thickening [23, Definition 7.2] of $\mathcal{G}$ in Σ is the interior of a bordered subsurface S of Σ. Now [23, Corollary 7.2. and Section 7.3] says that $\mathrm{int}(S) \cong \Sigma$. When Σ is of finite type, we prove the same thing differently in the following theorem:

Theorem 3.1.9 *A noncompact finite-type surface is the interior of a compact bordered surface. In particular, if a noncompact surface has infinite cyclic (resp. trivial) fundamental group, then it is homeomorphic to $\mathbb{S}^1 \times \mathbb{R}$ (resp. $\mathbb{R}^2$).*

Proof Let Σ be a finite-type noncompact surface. Consider an inductive construction of Σ; see Theorem 2.5.1. Since $\pi_1(\Sigma)$ is finitely generated, Theorem 2.5.2 says that Σ is homotopy equivalent to $\bigvee^{2r+s} \mathbb{S}^1$, where in this inductive construction $r \in \mathbb{N}$ is the total number of copies of $S_{1,2}$, and $s \in \mathbb{N}$ is the total number of copies of $S_{0,3}$; see Figure 1. Thus there is an integer n such that $\overline{\Sigma \setminus K_n}$ (where K_n is the compact bordered subsurface of Σ after n^{th} inductive step) is a finite collection of punctured disks. Now $g(\Sigma) = g(\mathrm{int}(K_n))$. Also, each end of Σ (resp. $\mathrm{int}(K_n)$) is planar, and the total number of ends of Σ (resp. $\mathrm{int}(K_n)$) is the same as the number of components of ∂K_n. By Theorem 2.4.1, $\Sigma \cong \mathrm{int}(K_n)$.

If Σ is a noncompact surface with infinite-cyclic fundamental group, then any inductive construction of Σ contains no copy of $S_{1,2}$ but precisely one copy of $S_{0,3}$, ie $\Sigma \cong \mathbb{S}^1 \times \mathbb{R}$. Similarly, if Σ is a noncompact surface with trivial fundamental group, then any inductive construction of Σ has no copy of $S_{0,3}$ as well as no copy of $S_{1,2}$, ie $\Sigma \cong \mathbb{R}^2$. $\qquad\square$

The proposition below follows directly from Goldman's inductive construction, so we quote it without proof. It says that an infinite-type surface has a finite genus only if it has infinitely many ends. On the other hand, Theorem 2.4.2 guarantees the existence of an infinite-type surface of infinite genus with infinitely many ends.

Proposition 3.1.10 *A noncompact surface is of a finite genus if and only if the total number of copies of $S_{1,2}$ used in any inductive construction of Σ is finite. Thus if an infinite-type surface has a finite genus, it must have infinitely many ends.*

This section's final fact (as promised in the introduction) says that the fundamental group alone can't detect the homeomorphism type of an infinite-type surface.

Proposition 3.1.11 *Up to homotopy equivalence, there is exactly one infinite-type surface, but up to homeomorphism, there are $2^{\aleph_0}$–many infinite-type surfaces.*

Proof Any infinite-type surface is homotopy equivalent to the wedge of countably infinitely many circles; see Theorem 2.5.2. Thus, any two infinite-type surfaces are homotopy equivalent.

Now, we prove that up to homeomorphism, there are $2^{\aleph_0}$–many infinite-type surfaces. Notice that except for the first step, in each step of Goldman's inductive procedure, we use $S_{0,3}$, or $S_{0,2}$, or $S_{1,2}$. Thus we have at most $3^{\aleph_0} = 2^{\aleph_0}$–many noncompact surfaces, up to homeomorphism. Therefore it is enough to show that this upper bound is reachable. Let τ be a nonempty closed subset of the Cantor set. By Theorem 2.4.2, there exists an infinite-genus surface Σ_τ such that $\mathrm{Ends}(\Sigma_\tau) = \mathrm{Ends}_{np}(\Sigma_\tau) \cong \tau$. Therefore, if τ_1 and τ_2 are two nonhomeomorphic nonempty closed subsets of the Cantor set, then Σ_{τ_1} is not homeomorphic to Σ_{τ_2} by Theorem 2.4.1. Now [32, Theorem 2] says that up to homeomorphism, there are $2^{\aleph_0}$–many closed subsets of the Cantor set.   □

3.2 Transversality of a proper map with respect to all decomposition circles

In the previous section, the codomain of a pseudoproper homotopy equivalence has been decomposed into finite-type bordered surfaces by a locally finite pairwise-disjoint collection of circles. This section aims to properly homotope the pseudoproper homotopy equivalence to make it transverse to each decomposition circle.

The theorem below follows from the theory developed in the appendix. We aim to use it to impose a 1–dimensional submanifold structure on the inverse image of each decomposition circle.

Theorem 3.2.1 *Let $f : \Sigma' \to \Sigma$ be a proper map between noncompact surfaces, and let $\mathscr{A}$ be an LFCS on Σ. Then f can be properly homotoped to make it smooth as well as transverse to the manifold $\mathscr{A}$.*

Proof Using Theorem A.1.1, after a proper homotopy, we may assume that f is a smooth proper map. After that, using Theorem A.1.2, properly homotope f so that it becomes transverse to $\mathscr{A}$.   □

Remark 3.2.2 Note that in Theorem 3.2.1, we have no control over those proper homotopies, which make the proper map f smooth as well as transverse to $\mathscr{A}$, ie after these proper homotopies, $f^{-1}(\mathscr{A})$ can be empty, even if these proper homotopies start with a surjective proper map. A remedy for this is to assume $\deg(f) \neq 0$. This is because the degree is invariant under proper homotopy, and a map of nonzero degree is surjective; see Lemmas 3.6.4.1 and 3.6.4.3. If f is a proper homotopy equivalence, then f has a proper homotopy inverse; hence $\deg(f) \neq 0$ (see Section 2.6). But if f is a pseudoproper

homotopy equivalence, then we don't know (at least till this stage) whether f has a proper homotopy inverse or not (though it has a homotopy inverse). Later in Section 3.6, using π_1–bijectivity, we will show that most pseudoproper homotopy equivalence is a map of degree ± 1.

The following theorem says that the transverse preimage of an LFCS under a proper map is an LFCS.

Theorem 3.2.3 *Let $f : \Sigma' \to \Sigma$ be a smooth proper map between two noncompact surfaces, and let $\mathscr{A}$ be an LFCS on Σ such that $f \pitchfork \mathscr{A}$. Then for each component $\mathcal{C}$ of $\mathscr{A}$, either $f^{-1}(\mathcal{C})$ is empty or a pairwise-disjoint finite collection of smoothly embedded circles on Σ'. Therefore $f^{-1}(\mathscr{A})$ is an LFCS on Σ'.*

Proof By the definition of transversality, $f \pitchfork \mathscr{A}$ implies $f \pitchfork \mathcal{C}$ for each component $\mathcal{C}$ of $\mathscr{A}$. Thus $f^{-1}(\mathcal{C})$ is either empty or is a compact (since f is proper) 1–dimensional boundaryless smoothly embedded submanifold of Σ'. By the classification of closed 1–dimensional manifolds, this completes the first part.

Next, if possible, let K' be a compact subset of Σ' such that K' intersects infinitely many components of $f^{-1}(\mathscr{A})$. By the first part, this means the compact set $f(K')$ intersects infinitely many components of $\mathscr{A}$, which contradicts the fact that $\mathscr{A}$ is a locally finite collection. $\square$

3.3 Disk removal

As previously observed, after a proper homotopy, the number of components in the collection of transverse preimages of all decomposition circles can be infinite, and many components (possibly infinitely many) of this collection, may be trivial circles. Here our goal is to group all these trivial circles in terms of the size of the disk bounded by a trivial circle and then remove all groups of trivial circles simultaneously by a proper homotopy.

Our intended grouping requires a technical lemma, which asserts that on a nonsimply connected surface, an LFCS consisting of concentric trivial circles doesn't exist. Roughly, it means, on a nonsimply connected surface, arbitrarily large disks bounded by components of an LFCS don't exist.

Lemma 3.3.1 *Let Σ be a surface, and let $\mathscr{A} := \{\mathcal{C}_i : i \in \mathbb{N}\}$ be an LFCS on Σ such that for each i the circle $\mathcal{C}_i$ bounds a disk $\mathcal{D}_i \subset \Sigma$ with $\mathcal{C}_i \subset \operatorname{int}(\mathcal{D}_{i+1})$. Then Σ is homeomorphic to $\mathbb{R}^2$.*

Proof At first, notice that Σ must be noncompact as $\mathscr{A}$ is a locally finite pairwise-disjoint infinite collection of circles. Using inductive construction (see Theorem 2.5.1), we have a sequence $\{S_j : j \in \mathbb{N}\}$ of compact bordered subsurfaces of Σ such that $\bigcup_j S_j = \Sigma$ and $S_j \subset \operatorname{int}(S_{j+1})$ for each $j \in \mathbb{N}$. Consider any $p \in \Sigma$. A $j_0 \in \mathbb{N}$ exists such that $p \in S_{j_0}$ and $S_{j_0} \cap (\cup_i \mathcal{C}_i) \neq \varnothing$. Since $\mathscr{A}$ is a locally finite collection, only finitely many components of $\mathscr{A}$ intersect the compact set S_{j_0}. Let $\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_l}$ be the only components of $\mathscr{A}$ intersecting S_{j_0}, where $i_1 < \cdots < i_l$. Pick an integer $i_0 > i_l$. Then $\mathcal{C}_{i_0} \cap S_{j_0} = \varnothing$. Now, since $\mathcal{C}_{i_l} \subset \operatorname{int}(\mathcal{D}_{i_0})$, S_{j_0} is connected, and Σ locally Euclidean, we can say that $S_{j_0} \subseteq \operatorname{int}(\mathcal{D}_{i_0})$. Thus every point $x \in \Sigma$ has an open neighborhood $\mathcal{U}_x$ in Σ such that $\mathcal{U}_x \subseteq \mathcal{D}_i$ for some $i \in \mathbb{N}$. Therefore every loop on Σ is contained in a disk $\mathcal{D}_i$ for some large $i \in \mathbb{N}$, ie Σ is simply connected. By Theorem 3.1.9, $\Sigma \cong \mathbb{R}^2$. $\square$

The following lemma is the primary tool for showing that a homotopy is proper. It tells how a proper map can be properly homotoped so that it changes on infinitely many pairwise-disjoint compact sets.

Lemma 3.3.2 *Let* $f : \Sigma' \to \Sigma$ *be a proper map between two noncompact surfaces, and let* $\{\Sigma'_n : n \in \mathbb{N}\}$ *be a pairwise-disjoint collection of compact bordered subsurfaces of* Σ'. *For each* $n \in \mathbb{N}$, *suppose* $H_n : \Sigma'_n \times [0, 1] \to \Sigma$ *is a homotopy relative to* $\partial \Sigma'_n$ *such that* $H_n(-, 0) = f|\Sigma'_n$ *and* $\mathrm{im}(H_n) \to \infty$. *Then* $\mathcal{H} : \Sigma' \times [0, 1] \to \Sigma$ *defined by*

$$\mathcal{H}(p, t) := \begin{cases} H_n(p, t) & \text{if } p \in \Sigma'_n \text{ and } t \in [0, 1], \\ f(p) & \text{if } p \in \Sigma' \setminus \left(\bigcup_{n \in \mathbb{N}} \Sigma'_n \right) \text{ and } t \in [0, 1], \end{cases}$$

is a proper map.

Proof Let $\mathcal{K}$ be a compact subset of Σ. By continuity of $\mathcal{H}$, $\mathcal{H}^{-1}(\mathcal{K})$ is closed in Σ'. Since $\mathrm{im}(H_n) \to \infty$, there exists $n_0 \in \mathbb{N}$ such that $\mathrm{im}(H_l) \cap \mathcal{K} = \varnothing$ for all integers $l \geq n_0 + 1$. Now, $f^{-1}(\mathcal{K})$ is compact as f is proper. Also, the domain of each H_n is compact. Hence, the closed subset $\mathcal{H}^{-1}(\mathcal{K})$ of Σ' is contained in the compact set $f^{-1}(\mathcal{K}) \cup \bigcup_{l=1}^{n_0} H_n^{-1}(\mathcal{K})$. □

To remove trivial components from the transverse preimage of an LFCS with infinitely many components, we need to impose some conditions on this LFCS. One such preferred LFCS is given in Theorem 3.1.5, but we will require other kinds of LFCS on the codomain, so here is the list of different preferred LFCS:

Definition 3.3.3 Let Σ be a noncompact surface such that $\Sigma \not\cong \mathbb{R}^2$. Suppose $\mathscr{A}$ is a given LFCS on Σ. We say $\mathscr{A}$ is a *preferred LFCS* on Σ if either

  (i) $\mathscr{A}$ is a finite collection of primitive circles on Σ, or

  (ii) $\mathscr{A}$ decomposes Σ into bordered subsurfaces, and a complementary component of this decomposition is homeomorphic to $S_{1,1}$, $S_{0,3}$, $S_{0,2}$, or $S_{0,1,1}$.

Remark 3.3.4 The only use of case (i) of Definition 3.3.3 is in Section 3.6, where we consider the process of removing unnecessary circles from the transverse preimage of the boundary of an essential pair of pants or an essential punctured disk. It is worth noting that by a finite LFCS, one can't decompose an infinite-type surface into finite-type bordered surfaces.

In the theorem below, we construct a proper homotopy which removes all trivial components keeping a neighborhood of each primitive component stationary from the transverse preimage of a preferred LFCS. Recall that a homotopy $H : X \times [0, 1] \to Y$ is said to be *stationary on a subset A of X* if $H(a, t) = H(a, 0)$ for all $(a, t) \in A \times [0, 1]$.

Theorem 3.3.5 *Let* $f : \Sigma' \to \Sigma$ *be a smooth proper map between two noncompact surfaces, where* $\Sigma' \not\cong \mathbb{R}^2 \not\cong \Sigma$, *and let* $\mathscr{A}$ *be a preferred LFCS on* Σ *such that* $f \pitchfork \mathscr{A}$. *Then we can properly homotope* f *to remove all trivial components of the LFCS* $f^{-1}(\mathscr{A})$ *such that each primitive component of* $f^{-1}(\mathscr{A})$ *has an open neighborhood on which this proper homotopy is stationary.*

Proof Since $\Sigma' \not\cong \mathbb{R}^2$ and $f^{-1}(\mathscr{A})$ is an LFCS (see Theorem 3.2.3), by Lemma 3.3.1 there don't exist infinitely many components $\mathcal{C}_1', \mathcal{C}_2', \ldots$ of $f^{-1}(\mathscr{A})$ bounding the disks $\mathcal{D}_1', \mathcal{D}_2', \ldots$, respectively, such that $\mathcal{C}_n'$ is contained in the interior of $\mathcal{D}_{n+1}'$ for each n. Thus, if $f^{-1}(\mathscr{A})$ has a trivial component, we can introduce the notion of an outermost disk bounded by a component of $f^{-1}(\mathscr{A})$ in the following way: A disk $\mathcal{D}' \subset \Sigma'$ bounded by a component of $f^{-1}(\mathscr{A})$ is called an *outermost disk* if, given another disk $\mathcal{D}'' \subset \Sigma$ bounded by a component of $f^{-1}(\mathscr{A})$, then either $\mathcal{D}'' \subseteq \mathcal{D}'$ or $\mathcal{D}' \cap \mathcal{D}'' = \varnothing$.

Let $\{\mathcal{D}_n'\}$ be the pairwise-disjoint collection (which may be an infinite collection) of all outermost disks. Assume $\mathcal{C}_n$ represents that component of $\mathscr{A}$ for which $f(\partial \mathcal{D}_n') \subseteq \mathcal{C}_n$. Note $\mathcal{C}_n$ may equal to $\mathcal{C}_m$ even if $m \neq n$.

For each integer n, we will construct a compact bordered subsurface $\mathcal{Z}_n$ with $f(\mathcal{D}_n') \subseteq \mathcal{Z}_n$ such that $\mathcal{Z}_n \to \infty$. Roughly, $\mathcal{Z}_n$ will be obtained from taking the union of all those complementary components of Σ (if a punctured disk appears, truncate it) which are hit by $f(\mathcal{D}_n')$.

Fix an integer n. Let $\mathcal{X}_{n,1}', \ldots, \mathcal{X}_{n,k_n}'$ be the all connected components of $\mathcal{D}_n' \setminus f^{-1}(\mathscr{A})$. By continuity of f, for each $\mathcal{X}_{n,l}'$, there exists a complementary component $\mathcal{Y}_{n,l}$ of Σ decomposed by $\mathscr{A}$ such that $f(\mathcal{X}_{n,l}') \subseteq \mathcal{Y}_{n,l}$ and $\partial \mathcal{X}_{n,l}' \subseteq f^{-1}(\partial \mathcal{Y}_{n,l})$; see Figure 4. For each l, define a compact bordered subsurface $\mathcal{Z}_{n,l}$ of Σ as follows: If $\mathcal{Y}_{n,l}$ is homeomorphic to $S_{1,1}$, $S_{0,3}$, or $S_{0,2}$; define $\mathcal{Z}_{n,l} := \mathcal{Y}_{n,l}$. On the other hand, if $\mathcal{Y}_{n,l}$ is homeomorphic to $S_{0,1,1}$, then let $\mathcal{Z}_{n,l}$ be an annulus in $\mathcal{Y}_{n,l}$ such that $\partial \mathcal{Z}_{n,l} \cap \partial \mathcal{Y}_{n,l} = \partial \mathcal{Y}_{n,l}$ and $f(\overline{\mathcal{X}_{n,l}'}) \subseteq \mathcal{Z}_{n,l}$. Finally, define $\mathcal{Z}_n := \mathcal{Z}_{n,1} \cup \cdots \cup \mathcal{Z}_{n,k_n}$.

Now, we show $\mathcal{Z}_n \to \infty$, so consider a compact subset $\mathcal{K}$ of Σ. Let $S_1, \ldots, S_m$ be a collection of complementary components of Σ decomposed by $\mathscr{A}$ such that $\mathcal{K} \subseteq \text{int}\left(\bigcup_{l=1}^m S_l\right)$. Define $S := \bigcup_{l=1}^m S_l$. Thus, for an integer n, $f(\mathcal{D}_n') \cap S \neq \varnothing$ if and only if $\mathcal{D}_n'$ contains at least one component of $\bigcup_{l=1}^m f^{-1}(\partial S_l)$. This is due to the construction of each $\mathcal{Z}_n$; see Figure 4. For each component $\mathcal{C}$ of $\mathscr{A}$, Theorem 3.2.3 says that $f^{-1}(\mathcal{C})$ has only finitely many components. So $\mathcal{D}_n'$ doesn't contain any component of $\bigcup_{l=1}^m f^{-1}(\partial S_l)$ for all sufficiently large n, ie $f(\mathcal{D}_n') \cap S = \varnothing$ for all sufficiently large n. Since $\mathcal{K} \subseteq \text{int}(S)$ and each $\mathcal{Z}_n$ is obtained from taking the union of all those complementary components of Σ (if a punctured disk appears, truncate it) which are hit by $f(\mathcal{D}_n')$, we can say that $\mathcal{Z}_n \cap \mathcal{K} = \varnothing$ for all sufficiently large n. Therefore $\mathcal{Z}_n \to \infty$ as $\mathcal{K}$ is an arbitrary compact subset of Σ.

For each n, adding a small external collar to one of the boundary components of $\mathcal{Z}_n$ (if needed), we can construct a compact bordered surface Σ_n with $\mathcal{C}_n \subseteq \text{int}(\Sigma_n)$, $f(\mathcal{D}_n') \subseteq \Sigma_n$ such that $\{\Sigma_n\}$ is a locally finite collection, ie $\Sigma_n \to \infty$; see Figure 4.

For each n, write $\mathcal{C}_n' := \partial \mathcal{D}_n'$. Thus $f(\mathcal{C}_n') \subseteq \mathcal{C}_n$. Since $\mathcal{C}_n \subseteq \text{int}(\Sigma_n)$, using Theorem A.2.1, choose a one-sided tubular neighborhood $\mathcal{C}_n \times [0, \varepsilon_n]$ of $\mathcal{C}_n$ in Σ with $\mathcal{C}_n \times 0 \equiv \mathcal{C}_n$ such that $f \pitchfork (\mathcal{C}_n \times t_n)$ for each $t_n \in [0, \varepsilon_n]$ and $\mathcal{C}_n \times [0, \varepsilon_n] \subseteq \Sigma_n$. Without loss of generality, we may further assume that $f(x') \in \mathcal{C}_n \times [0, \varepsilon_n]$ for each $x' \in \Sigma' \setminus \mathcal{D}_n'$ sufficiently near to $\mathcal{C}_n'$. Next, since $f^{-1}(\mathscr{A})$ is an LFCS, for each n, Theorem A.2.3 gives a one-sided compact tubular neighborhood $\mathcal{U}_n'$ of $\mathcal{C}_n'$ such that $\mathcal{U}_n' \cap \mathcal{D}_n' = \mathcal{C}_n' = \mathcal{U}_n' \cap f^{-1}(\mathscr{A})$, $f(\mathcal{U}_n') \subseteq \mathcal{C}_n \times [0, \varepsilon_n]$ for each n, and $\mathcal{U}_n' \cap \mathcal{U}_m' = \varnothing$ for $m \neq n$. Finally, Theorem A.2.5
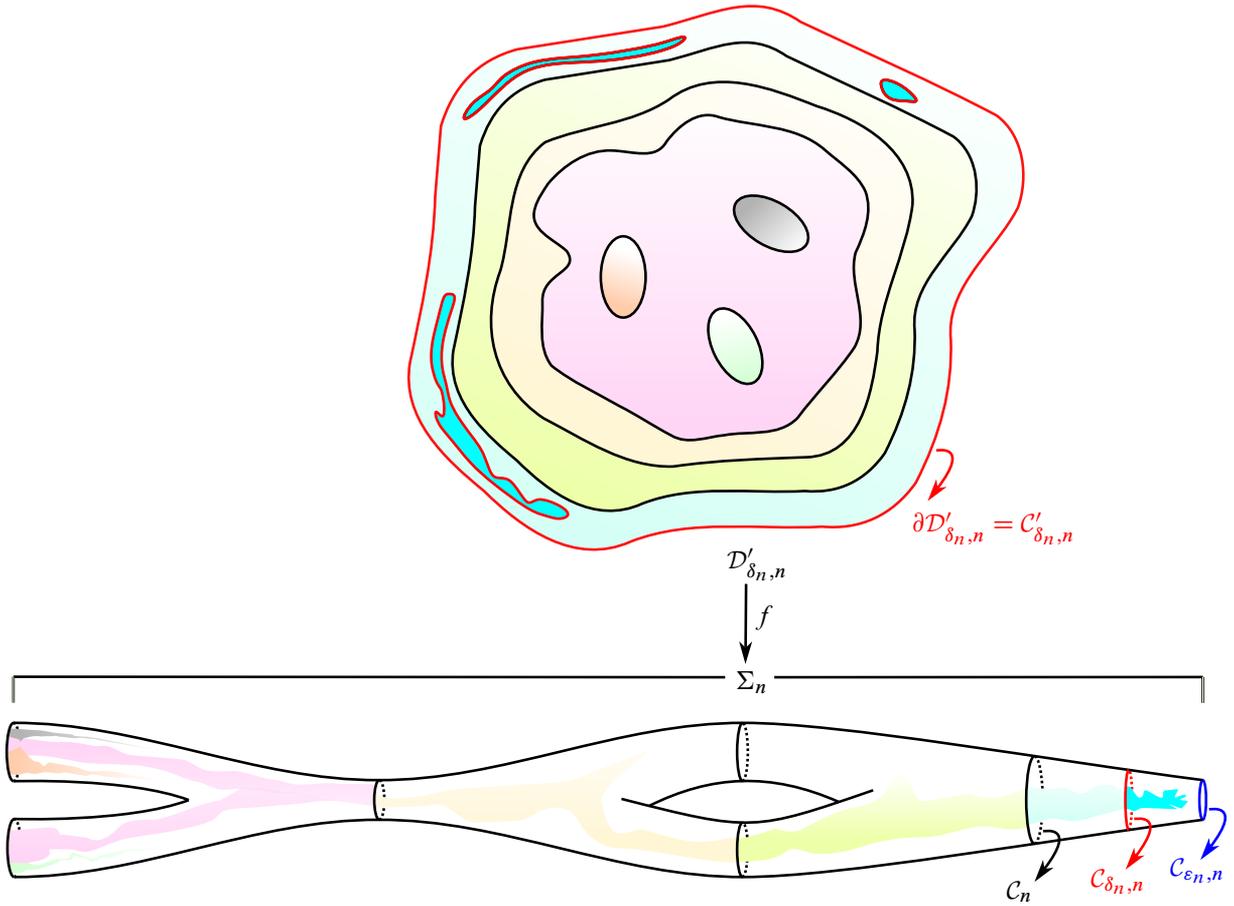
Figure 4: Each component of $\mathcal{D}'_n \setminus f^{-1}(\mathscr{A})$ maps into a component of $\Sigma \setminus \mathscr{A}$. This fact, together with Theorem A.2.5, provides Σ_n. A black circle denotes a component of either $\mathscr{A}$ or a component of $f^{-1}(\mathscr{A})$.

gives $\delta_n \in (0, \varepsilon_n)$ and a component $\mathcal{C}'_{\delta_n,n}$ of $f^{-1}(\mathcal{C}_{\delta_n,n})$ such that $\mathcal{C}'_{\delta_n,n}$ bounds a disk $\mathcal{D}'_{\delta_n,n}$ in Σ' with $(\mathcal{U}'_n \cup \mathcal{D}'_n) \supseteq \mathcal{D}'_{\delta_n,n} \supset \text{int}(\mathcal{D}'_{\delta_n,n}) \supset \mathcal{D}'_n$ (equivalently, $\mathcal{U}'_n$ contains the annulus cobounded by $\mathcal{C}'_{\delta_n,n}$ and $\mathcal{C}'_n$) and $f(\mathcal{D}'_{\delta_n,n} \setminus \text{int}(\mathcal{D}'_n)) \subseteq \mathcal{C}_n \times [0, \varepsilon_n]$. Thus $\mathcal{D}'_{\delta_n,n} \cap f^{-1}(\mathscr{A}) = \mathcal{D}'_n \cap f^{-1}(\mathscr{A})$, $f(\mathcal{D}'_{\delta_n,n}) \subseteq \Sigma_n$ for each n, and $\mathcal{D}'_{\delta_n,n} \cap \mathcal{D}'_{\delta_m,m} = \varnothing$ when $m \neq n$.

Since $\mathcal{C}_{\delta_n,n}$ cobounds an annulus with the primitive circle $\mathcal{C}_n$, the inclusion $\mathcal{C}_{\delta_n,n} \hookrightarrow \Sigma_n$ is π_1–injective (see Theorem 2.2.2). Also, Σ_n is homotopy equivalent to $\bigvee_{\text{finite}} \mathbb{S}^1$, which implies that the universal cover of Σ_n is contractible, and thus $\pi_2(\Sigma_n) = 0$. Therefore, exactness of

$$\cdots \to \pi_2(\Sigma_n) \to \pi_2(\Sigma_n, \mathcal{C}_{\delta_n,n}) \to \pi_1(\mathcal{C}_{\delta_n,n}) \to \pi_1(\Sigma_n) \to \cdots$$

gives $\pi_2(\Sigma_n, \mathcal{C}_{\delta_n,n}) = 0$, ie we have a homotopy $H_n \colon \mathcal{D}'_{\delta_n,n} \times [0, 1] \to \Sigma_n$ relative to $\mathcal{C}'_{\delta_n,n}$ from $f|(\mathcal{D}'_{\delta_n,n}, \mathcal{C}'_{\delta_n,n}) \to (\Sigma_n, \mathcal{C}_{\delta_n,n})$ to a map $\mathcal{D}'_{\delta_n,n} \to \mathcal{C}_{\delta_n,n}$ for each n; see [25, Lemma 4.6]. Now, to conclude, apply Lemma 3.3.2 on $\{H_n\}$. $\qquad\square$

Remark 3.3.6 In Theorem 3.3.5, the number of components of $\mathscr{A}$ can be infinite; thus the number of trivial components of $f^{-1}(\mathscr{A})$ can be infinite. That's why we have removed all trivial components of $f^{-1}(\mathscr{A})$ by a single proper homotopy upon considering all outermost disks simultaneously. This process is in contrast to the finite-type surface theory, where the number of decomposition circles is finite, and therefore all trivial circles in the collection of transverse preimages of all decomposition circles can be removed one by one, considering the notion of an innermost disk.

3.4 Homotope a degree 1 map between circles to a homeomorphism

Previously, we have removed all trivial components keeping a neighborhood of each primitive component stationary from the transverse preimage $f^{-1}(\mathscr{A})$ of a preferred LFCS $\mathscr{A}$. In this section, we properly homotope our pseudoproper homotopy equivalence $f : \Sigma' \to \Sigma$ to send each component $\mathcal{C}'$ of $f^{-1}(\mathscr{A})$ homeomorphically onto a component $\mathcal{C}$ of $\mathscr{A}$ so that the restriction of f to a small one-sided tubular neighborhood $\mathcal{C}' \times [1, 2]$ of $\mathcal{C}'$ (on either side of $\mathcal{C}'$) can be described by the following homeomorphism:

$$\mathcal{C}' \times [1, 2] \ni (z, t) \mapsto (f(z), t) \in \mathcal{C} \times [1, 2].$$

First, we fix some notation. Define $\partial_\varepsilon := \mathbb{S}^1 \times \varepsilon$ for $\varepsilon \in \mathbb{R}$ and $\boldsymbol{I} := [0, 1]$. Let $p : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{S}^1$ be the projection. The following lemma roughly says that a self-map of $\mathbb{S}^1 \times [0, 2]$ can be homotoped rel $\mathbb{S}^1 \times 0$ to map $\mathbb{S}^1 \times [1, 2]$ into itself by the product $\theta \times \mathrm{Id}_{[1,2]}$, where θ is a self-map of $\mathbb{S}^1$.

Lemma 3.4.1 *Let Φ be a self-map of $A := \mathbb{S}^1 \times [0, 2]$ such that $\Phi^{-1}(\partial_b) = \partial_b$ for each $b \in \{0, 2\}$. If we are given a map $\varphi_2 : \partial_2 \to \partial_2$ and a homotopy $h_{(2)} : \partial_2 \times \boldsymbol{I} \to \partial_2$ from $\Phi|_{\partial_2} \to \partial_2$ to φ_2, then Φ can be homotoped relative to ∂_0 to map $\mathbb{S}^1 \times [0, 1]$ into $\mathbb{S}^1 \times [0, 1]$ and to satisfy $\Phi(-, r) = (p \circ \varphi_2(-, 2), r)$ for each $r \in [1, 2]$.*

Remark 3.4.2 In Lemma 3.4.1, up to homotopy, φ_2 is either a constant map or a covering map.

Proof Homotope Φ relative to $\partial_0 \cup \partial_2$ so that $\Phi(\mathbb{S}^1 \times [0, 1]) \subseteq \mathbb{S}^1 \times [0, 1]$ and $\Phi(z, r) = (p \circ \Phi(z, 2), r)$ for all $(z, r) \in \mathbb{S}^1 \times [1, 2]$. For each $r \in [1, 2]$, $h_{(2)}$ provides a homotopy $h_{(r)} : \partial_r \times \boldsymbol{I} \to \partial_r$. Let $H : (\partial_0 \cup \partial_1) \times \boldsymbol{I} \to \partial_0 \cup \partial_1$ be the homotopy defined by $H|\partial_1 \times \boldsymbol{I} = h_{(1)}$ and $H(-, t)|\partial_0 = \Phi|\partial_0$ for any $t \in [0, 1]$. The homotopy extension theorem gives a homotopy $\widetilde{H} : \mathbb{S}^1 \times [0, 1] \times \boldsymbol{I} \to \mathbb{S}^1 \times [0, 1]$ such that $\widetilde{H}|(\partial_0 \cup \partial_1) \times \boldsymbol{I} = H$. Finally, paste $\widetilde{H}$ with the collection $h_{(r)}, 1 \le r \le 2$. $\square$

The following theorem is the simple modification (in the proper category) of the analogous theorem for closed surfaces:

Theorem 3.4.3 *Let $f : \Sigma' \to \Sigma$ be a smooth pseudoproper homotopy equivalence between two noncompact surfaces where $\Sigma' \not\cong \mathbb{R}^2 \not\cong \Sigma$, and let $\mathscr{A}$ be a preferred LFCS on Σ such that $f \pitchfork \mathscr{A}$. Then f can be properly homotoped to remove all trivial components of $f^{-1}(\mathscr{A})$ as well as to map each primitive component of $f^{-1}(\mathscr{A})$ homeomorphically onto a component of $\mathscr{A}$. Moreover, after this proper homotopy, near each component of $f^{-1}(\mathscr{A})$, the map f can be described as follows:*

Let $\mathcal{C}'_{\boldsymbol{p}}$ (resp. $\mathcal{C}$) be a component of $f^{-1}(\mathscr{A})$ (resp. $\mathscr{A}$) such that $f|\mathcal{C}'_{\boldsymbol{p}} \to \mathcal{C}$ is a homeomorphism. Then $\mathcal{C}'_{\boldsymbol{p}}$ (resp. $\mathcal{C}$) has two one-sided tubular neighborhoods $\mathcal{M}'$ and $\mathcal{N}'$ (resp. $\mathcal{M}$ and $\mathcal{N}$) with the specific identifications $(\mathcal{M}', \mathcal{C}'_{\boldsymbol{p}}) \cong (\mathcal{C}'_{\boldsymbol{p}} \times [1,2], \mathcal{C}'_{\boldsymbol{p}} \times 2) \cong (\mathcal{N}', \mathcal{C}'_{\boldsymbol{p}})$ (resp. $(\mathcal{M}, \mathcal{C}) \cong (\mathcal{C} \times [1,2], \mathcal{C} \times 2) \cong (\mathcal{N}, \mathcal{C})$) such that the following hold:

- $\mathcal{M}' \cup \mathcal{N}'$ is a (two-sided) tubular neighborhood of $\mathcal{C}'_{\boldsymbol{p}}$,
- $f|\mathcal{M}' \to \mathcal{M}$ and $f|\mathcal{N}' \to \mathcal{N}$ are homeomorphisms given by $\mathcal{C}'_{\boldsymbol{p}} \times [1,2] \ni (z,t) \mapsto (f(z),t) \in \mathcal{C} \times [1,2]$.

Remark 3.4.4 In Theorem 3.4.3, though $\mathcal{M}' \cup \mathcal{N}'$ is a (two-sided) tubular neighborhood of $\mathcal{C}'_{\boldsymbol{p}}$, both $\mathcal{M}$ and $\mathcal{N}$ may lie on the same side of $\mathcal{C}$, ie $\mathcal{M} \cup \mathcal{N}$ may not be a two-sided tubular neighborhood of $\mathcal{C}$.

Proof Let $\{\mathcal{C}'_{\boldsymbol{pn}}\}$ be the collection of all primitive components of $f^{-1}(\mathscr{A})$. Assume $\mathcal{C}_n$ represents that component of $\mathscr{A}$ for which $f(\mathcal{C}'_{\boldsymbol{pn}}) \subseteq \mathcal{C}_n$. Note $\mathcal{C}_n$ may equal to $\mathcal{C}_m$ even if $m \neq n$.

Claim 3.4.4.1 There are one-sided compact tubular neighborhoods $\mathcal{U}'_n, \mathcal{V}'_n (\subseteq \Sigma')$ of $\mathcal{C}'_{\boldsymbol{pn}}$, and there are one-sided compact tubular neighborhoods $\mathcal{U}_n, \mathcal{V}_n (\subseteq \Sigma)$ of $\mathcal{C}_n$ such that after defining $\mathcal{T}'_n := \mathcal{U}'_n \cup \mathcal{V}'_n$, the following hold:

(1) $\widetilde{\mathscr{A}} := \mathscr{A} \cup \{(\partial \mathcal{U}_n \cup \partial \mathcal{V}_n) \setminus \mathcal{C}_n\}_n$ is an LFCS and $f \pitchfork \widetilde{\mathscr{A}}$,

(2) $\partial \mathcal{U}'_n \setminus \mathcal{C}'_{\boldsymbol{pn}}$ (resp. $\partial \mathcal{V}'_n \setminus \mathcal{C}'_{\boldsymbol{pn}}$) is the only component of $f^{-1}(\partial \mathcal{U}_n \setminus \mathcal{C}_n) \cap \mathcal{U}'_n$ (resp. $f^{-1}(\partial \mathcal{V}_n \setminus \mathcal{C}_n) \cap \mathcal{V}'_n$) that cobounds an annulus with $\mathcal{C}'_{\boldsymbol{pn}}$ (see Figure 5),
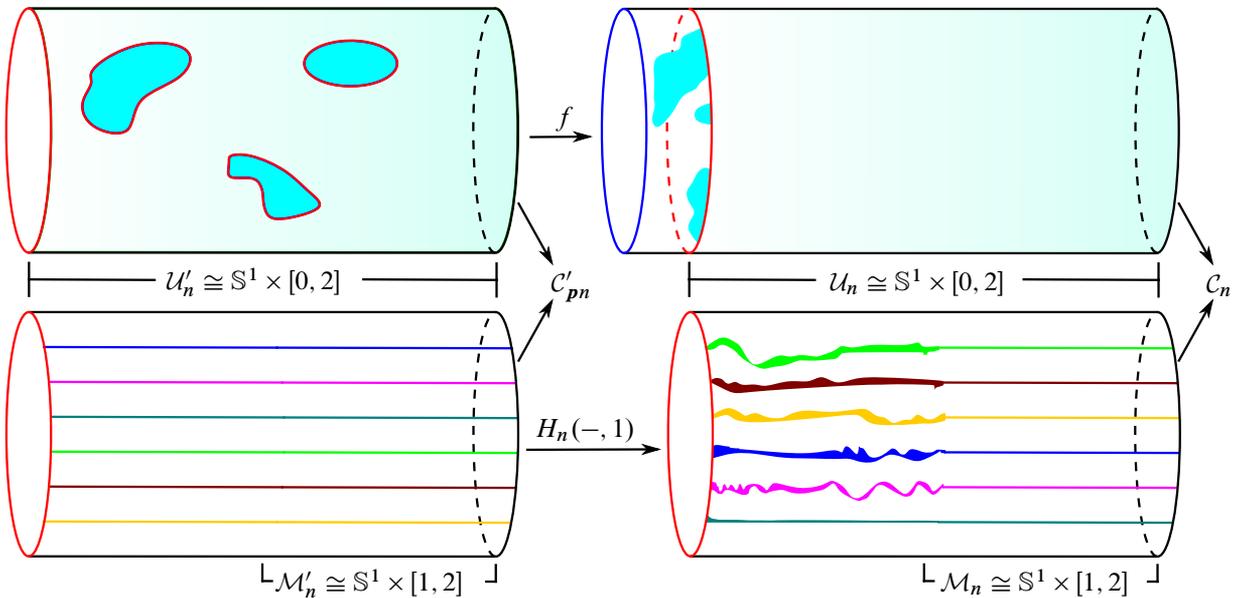

Figure 5: Top: description of $f|\mathcal{U}'_n \to \mathcal{U}_n$ using Theorem A.2.5. Bottom: after removing all trivial components of $f^{-1}(\partial \mathcal{U}_n \setminus \mathcal{C}_n)$ from $\mathcal{U}'_n$ and then applying Lemma 3.4.1 to $f|\mathcal{U}'_n \to \mathcal{U}_n$, we obtain $H_n(-,1)|\mathcal{U}'_n \to \mathcal{U}_n$.

(3) *each point of* $\text{int}(\mathcal{U}'_n)$ *(resp.* $\text{int}(\mathcal{V}'_n)$*) that is sufficiently near to* $\mathcal{C}'_{pn}$ *is mapped into* $\text{int}(\mathcal{U}_n)$ *(resp.* $\text{int}(\mathcal{V}_n)$*),*

(4) $\mathcal{T}'_n$ *is a two-sided tubular neighborhood of* $\mathcal{C}'_{pn}$ *with* $f^{-1}(\mathscr{A}) \cap \mathcal{T}'_n = \mathcal{C}'_{pn}$*, and*

(5) $\mathcal{T}'_n \cap \mathcal{T}'_m = \varnothing$ *if* $m \neq n$*, and* $(\mathcal{U}_n \cup \mathcal{V}_n) \to \infty$*.*

Proof For any positive integer n_0, Theorem 3.2.3 says that the set $\{m \in \mathbb{N} : \mathcal{C}_m = \mathcal{C}_{n_0}\}$ is finite. Also, $\mathscr{A}$ is locally finite. Thus $\{\mathcal{C}_n : n \in \mathbb{N}\}$ is locally finite. So, for each n, there exists a two-sided tubular neighborhood $\mathcal{C}_n \times [-\varepsilon_n, \varepsilon_n]$ of $\mathcal{C}_n$ with $\mathcal{C}_n \times 0 \equiv \mathcal{C}_n$ such that $\{\mathcal{C}_n \times [-\varepsilon_n, \varepsilon_n] : n \in \mathbb{N}\}$ is a locally finite collection. Further, for each $n \in \mathbb{N}$, we may assume that $f \pitchfork (\mathcal{C}_n \times t_n)$ whenever $t_n \in [-\varepsilon_n, \varepsilon_n]$ by Theorem A.2.1.

Since $f^{-1}(\mathscr{A})$ is a locally finite collection, for each n there are one-sided compact tubular neighborhoods $\mathcal{U}'_n$ and $\mathcal{V}'_n$ of $\mathcal{C}'_{pn}$ in Σ' such that after defining $\mathcal{T}'_n := \mathcal{U}'_n \cup \mathcal{V}'_n$, the following hold: $\mathcal{T}'_n$ is a two-sided tubular neighborhood of $\mathcal{C}'_{pn}$, $f^{-1}(\mathscr{A}) \cap \mathcal{T}'_n = \mathcal{C}'_{pn}$, and $\mathcal{T}'_n \cap \mathcal{T}'_m = \varnothing$ if $m \neq n$. Moreover, using Theorem A.2.3, $f(\mathcal{T}'_n) \subseteq \mathcal{C}_n \times [-\varepsilon_n, \varepsilon_n]$ can also be assumed for each n.

Next, by Theorem A.2.5, we may further assume $\partial \mathcal{U}'_n \setminus \mathcal{C}'_{pn}$ (resp. $\partial \mathcal{V}'_n \setminus \mathcal{C}'_{pn}$) is a component of $f^{-1}(\mathcal{C}_n \times x_n)$ (resp. $f^{-1}(\mathcal{C}_n \times y_n)$) for some $x_n, y_n \in (-\varepsilon_n, 0) \cup (0, \varepsilon_n)$ such that after defining $\mathcal{U}_n$ (resp. $\mathcal{V}_n$) as the annulus in $\mathcal{C}_n \times [-\varepsilon_n, \varepsilon_n]$ cobounded by $\mathcal{C}_n \times 0$ and $\mathcal{C}_n \times x_n$ (resp. $\mathcal{C}_n \times y_n$), Claim 3.4.4.1(2)–(3) hold. Finally $\mathcal{C}_n \times [-\varepsilon_n, \varepsilon_n] \to \infty$ implies $(\mathcal{U}_n \cup \mathcal{V}_n) \to \infty$. $\square$

Using Theorem 3.3.5, keeping stationary a neighborhood of each primitive component of $f^{-1}(\widetilde{\mathscr{A}})$, we can properly homotope f to remove all trivial components from $f^{-1}(\widetilde{\mathscr{A}})$. So, after this proper homotopy, Claim 3.4.4.1(2)–(3) imply that $f(\mathcal{U}'_n) \subseteq \mathcal{U}_n$, $f^{-1}(\partial \mathcal{U}_n) \cap \mathcal{U}'_n = \partial \mathcal{U}'_n$, $f(\mathcal{V}'_n) \subseteq \mathcal{V}_n$, and $f^{-1}(\partial \mathcal{V}_n) \cap \mathcal{V}'_n = \partial \mathcal{V}'_n$. Notice the abuse of notation: the initial and final maps of this proper homotopy both are denoted by f.

Now, let $h_n : \mathcal{C}'_{pn} \times [0, 1] \to \mathcal{C}_n$ be a homotopy from $f|\mathcal{C}'_{pn} \to \mathcal{C}_n$ such that $h_n(-, 1)$ is either a constant map or a covering map between two circles. Applying Lemma 3.4.1 on $f|\mathcal{U}'_n \to \mathcal{U}_n$ and $f|\mathcal{V}'_n \to \mathcal{V}_n$ separately upon considering h_n, a homotopy $H_n : \mathcal{T}'_n \times [0, 1] \to \mathcal{U}_n \cup \mathcal{V}_n$ relative to $\partial \mathcal{T}'_n$ exists such that $H_n(-, 0) = f|\mathcal{T}'_n$, $(H_n(-, 1))^{-1}(\mathcal{C}_n) = \mathcal{C}'_{pn}$, and $H_n(-, 1)|\mathcal{C}'_{pn} \to \mathcal{C}_n$ is the same as $h_n(-, 1)$; see Figure 5.

Next, Claim 3.4.4.1(5) tells us that we can apply Lemma 3.3.2 on $\{H_n\}$ to obtain a proper homotopy $\mathcal{H} : \Sigma' \times [0, 1] \to \Sigma$ starting from f. Next, being an isomorphism, $\pi_1(f) = \pi_1(\mathcal{H}(-, 1))$ preserves primitiveness, ie $h_n(-, 1) = \mathcal{H}(-, 1)|\mathcal{C}'_{pn} \to \mathcal{C}_n$ must be a homeomorphism. Thus $\mathcal{H}$ is our ultimate required homotopy.

Finally, we need to describe f near each component of $f^{-1}(\mathscr{A})$ after the proper homotopy $\mathcal{H}$. Abusing notation, the final map of $\mathcal{H}$ will be denoted by f. Since Lemma 3.4.1 is being used, we have $\mathcal{M}'_n \subseteq \mathcal{U}'_n$ and $\mathcal{M}_n \subseteq \mathcal{U}_n$ with the identifications $(\mathcal{M}'_n, \mathcal{C}'_{pn}) \cong (\mathcal{C}'_{pn} \times [1, 2], \mathcal{C}'_{pn} \times 2)$ and $(\mathcal{M}_n, \mathcal{C}_n) \cong (\mathcal{C}_n \times [1, 2], \mathcal{C}_n \times 2)$ such that after the proper homotopy $\mathcal{H} : \Sigma' \times [0, 1] \to \Sigma$, the map f sends $\mathcal{C}'_{pn} \times r$ onto $\mathcal{C}_n \times r$ using the homeomorphism $f|\mathcal{C}'_{pn} \to \mathcal{C}_n$ for all $r \in [1, 2]$; see Figure 5. The reasoning is similar for $f|\mathcal{V}'_n \to \mathcal{V}_n$. $\square$

The following proposition, *which we don't need to use anywhere*, tells what happens if we drop the phrase "homotopy equivalence" in the statement of Theorem 3.4.3. Its proof is almost the same.

Proposition 3.4.5 *Let $f : \Sigma' \to \Sigma$ be a smooth proper map between two noncompact surfaces, where $\Sigma' \not\cong \mathbb{R}^2 \not\cong \Sigma$, and let $\mathscr{A}$ be a preferred LFCS on Σ such that $f \pitchfork \mathscr{A}$. Then f can be properly homotoped to remove all trivial components of $f^{-1}(\mathscr{A})$ as well as to map each primitive component of $f^{-1}(\mathscr{A})$ into a component of $\mathscr{A}$ so that for any component $\mathcal{C}$ of $\mathscr{A}$ and any primitive component $\mathcal{C}'_p$ of $f^{-1}(\mathcal{C})$, after this proper homotopy, $f|\mathcal{C}'_p \to \mathcal{C}$ is either a constant map or a covering map.*

3.5 Annulus removal

In the previous two sections, after removing all trivial components from the transverse preimage of a decomposition circle, the remaining primitive circles have been mapped homeomorphically to that decomposition circle. This section aims to remove all these primitive circles except one from the inverse image of each decomposition circle using the following three steps: annulus bounding, then annulus compression, and finally annulus pushing.

We start with annulus bounding. Consider the collection of inverse images of all decomposition circles. The following lemma says that any two circles in this collection cobound an annulus in the domain if and only if their images are the same. In other words, in the domain, by pasting all small annuli, we get the outermost annulus corresponding to a decomposition circle.

Lemma 3.5.1 *Let $f : \Sigma' \to \Sigma$ be a homotopy equivalence between two noncompact surfaces, and let $\mathscr{A}'$ and $\mathscr{A}$ be two LFCS on Σ' and Σ, respectively, such that f maps each component of $\mathscr{A}'$ homeomorphically onto a component of $\mathscr{A}$. Suppose each component of $\mathscr{A}$ is primitive, and any two distinct components of $\mathscr{A}$ don't cobound an annulus in Σ. Let $\mathcal{C}'_0$ and $\mathcal{C}'_1$ be two distinct components of $\mathscr{A}'$. Then $\mathcal{C}'_0$ and $\mathcal{C}'_1$ cobound an annulus in Σ' if and only if $f(\mathcal{C}'_0) = f(\mathcal{C}'_1)$.*

Proof To prove the only if part, let $\Phi : \mathbb{S}^1 \times [0,1] \hookrightarrow \Sigma'$ be an embedding such that $\Phi(\mathbb{S}^1, k) = \mathcal{C}'_k$ for $k = 0, 1$. Note that f maps each component of $\mathscr{A}'$ homeomorphically onto a component of $\mathscr{A}$, and each component of $\mathscr{A}$ is a primitive circle on Σ. Thus the embeddings $f\Phi(-, 0), f\Phi(-, 1) : \mathbb{S}^1 \hookrightarrow \Sigma$ are freely homotopic, and hence $f\Phi(-, 0), f\Phi(-, 1) : \mathbb{S}^1 \hookrightarrow \Sigma$ represent the same nontrivial conjugacy class in $\pi_1(\Sigma, *)$. Since any two distinct components of $\mathscr{A}$ don't cobound an annulus in Σ, by Theorem 2.2.3, $f(\mathcal{C}'_0) = f(\mathcal{C}'_1)$.

To prove the if part, let $g : \Sigma \to \Sigma'$ be a homotopy inverse of f, and let $\mathcal{C}$ be the component of $\mathscr{A}$ defined by $\mathcal{C} := f(\mathcal{C}'_0) = f(\mathcal{C}'_1)$. Now, $f|\mathcal{C}'_k \to f(\mathcal{C}'_k)$ is a homeomorphism for $k = 0, 1$. Thus, for a homeomorphism $j : \mathbb{S}^1 \xrightarrow{\cong} \mathcal{C}$, there are homeomorphisms $\ell_0 : \mathbb{S}^1 \xrightarrow{\cong} \mathcal{C}'_0$ and $\ell_1 : \mathbb{S}^1 \xrightarrow{\cong} \mathcal{C}'_1$ such that $f\ell_0 = j = f\ell_1$. Since $\ell_0 \simeq gf\ell_0 = gj = gf\ell_1 \simeq \ell_1$, applying Theorem 2.2.3 to ℓ_0, ℓ_1, we are done. □

The following theorem, which will be used to compress each annulus bounded by two primitive circles of the domain, roughly says that most homotopies of a circle embedded in a surface are trivial:

Theorem 3.5.2 [35, Lemma 4.9.15] *Let S be a compact bordered surface other than the disk, and let Φ be a map from $A := \mathbb{S}^1 \times [0, 1]$ to S such that $\Phi(\text{int}(A)) \subseteq \text{int}(S)$ and there is a boundary component C of S for which $\Phi(-, 0), \Phi(-, 1) \colon \mathbb{S}^1 \xrightarrow{\cong} C$ are the same homeomorphisms. Then Φ can be homotoped relative to ∂A to map A onto C.*

The following theorem considers the last two steps — annulus compressing and annulus pushing. At first, by a proper homotopy, each outermost annulus will be mapped onto its decomposition circle; after that, by another proper homotopy, each outermost annulus will be pushed into a one-sided tubular neighborhood of one of its boundary components.

Theorem 3.5.3 *Let $f \colon \Sigma' \to \Sigma$ be a smooth pseudoproper homotopy equivalence between two noncompact surfaces, where $\Sigma' \not\cong \mathbb{R}^2 \not\cong \Sigma$, and let $\mathscr{A}$ be a preferred LFCS on Σ such that $f \pitchfork \mathscr{A}$. Suppose any two distinct components of $\mathscr{A}$ don't cobound an annulus in Σ. In that case, f can be properly homotoped to a proper map g such that for each component $\mathcal{C}$ of $\mathscr{A}$, either $g^{-1}(\mathcal{C})$ is empty or $g^{-1}(\mathcal{C})$ is a component of $f^{-1}(\mathscr{A})$ that is mapped homeomorphically onto $\mathcal{C}$ by g.*

Proof Using Theorem 3.4.3, we may assume each component of $f^{-1}(\mathscr{A})$ is primitive and also mapped homeomorphically onto a component of $\mathscr{A}$. So for simplicity, we may drop the subscript p to indicate a primitive component of $f^{-1}(\mathscr{A})$. Let $\{\mathcal{C}_n\}$ be the pairwise disjoint collection of all those components of $\mathscr{A}$ such that for each n, $f^{-1}(\mathcal{C}_n)$ has more than one component. By Lemma 3.5.1, for each n, an annulus $\mathcal{A}'_n$ (say the nth *outermost annulus*) exists with the following properties: $\partial \mathcal{A}'_n \subseteq f^{-1}(\mathcal{C}_n)$, and $\mathcal{A}'_n$ is not contained in the interior of an annulus bounded by any two components of $f^{-1}(\mathscr{A})$. Thus $\mathcal{A}'_n \cap f^{-1}(\mathscr{A}) = f^{-1}(\mathcal{C}_n)$ and $\mathcal{A}'_n \cap \mathcal{A}'_m = \varnothing$ for $m \neq n$. Now, using Theorem 2.2.3, find a parametrization $\tau_n \colon \mathbb{S}^1 \times [0, k_n] \xrightarrow{\cong} \mathcal{A}'_n$ for some integer $k_n \geq 1$ such that $\tau_n(\mathbb{S}^1 \times \{0, \ldots, k_n\}) = f^{-1}(\mathcal{C}_n)$ and $f \tau_n(-, l) \colon \mathbb{S}^1 \xrightarrow{\cong} \mathcal{C}_n$ represents the same homeomorphism of $\mathcal{C}_n$ for each $l = 0, \ldots, k_n$.

Claim 3.5.3.1 *The proper map $f \colon \Sigma' \to \Sigma$ can be properly homotoped relative to $\Sigma' \setminus \bigcup_n \text{int}(\mathcal{A}'_n)$ so that $f(\mathcal{A}'_n) = \mathcal{C}_n$ for each n.*

Proof For each integer n, we will construct a compact bordered subsurface Σ_n of Σ with $f(\mathcal{A}'_n) \subseteq \text{int}(\Sigma_n)$ such that $\Sigma_n \to \infty$. Roughly, Σ_n will be obtained from taking the union of all those complementary components of Σ (if a punctured disk appears, truncate it) which are hit by $f(\mathcal{A}'_n)$.

Using continuity of $f | \Sigma' \setminus f^{-1}(\mathscr{A}) \to \Sigma \setminus \mathscr{A}$, we can say that $f(\mathcal{A}'_n) \subseteq \mathcal{X}_n \cup \mathcal{Y}_n$, where $\mathcal{X}_n$ and $\mathcal{Y}_n$ are complementary components of Σ decomposed by $\mathscr{A}$ such that $\mathcal{C}_n \subseteq \partial \mathcal{X}_n \cap \partial \mathcal{Y}_n$.

(1) We define Σ_n as $\Sigma_n := \mathcal{X}_n \cup \mathcal{Y}_n$ if $\mathcal{X}_n \cong S_{0,3} \cong \mathcal{Y}_n$, $\mathcal{X}_n \cong S_{1,1}$ and $\mathcal{Y}_n \cong S_{0,3}$, or $\mathcal{Y}_n \cong S_{1,1}$ and $\mathcal{X}_n \cong S_{0,3}$; see Figure 6.

(2) If $\mathcal{X}_n \cong S_{0,1,1} \cong \mathcal{Y}_n$ (in this case Σ is homeomorphic to the punctured plane), then using compactness of $f(\mathcal{A}'_n)$, let Σ_n be an annulus in $\mathcal{X}_n \cup \mathcal{Y}_n$ such that $f(\mathcal{A}'_n) \subseteq \text{int}(\Sigma_n)$.
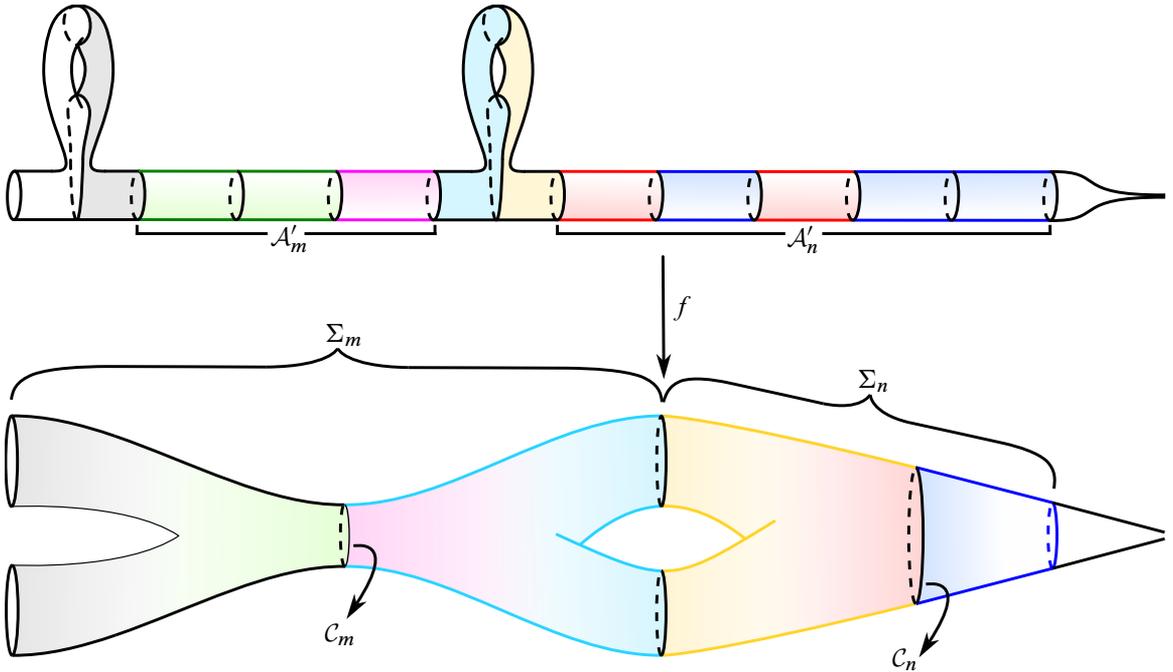
Figure 6: Illustration of parts (1) and (3) of the definition of Σ_n given in the proof of Claim 3.5.3.1. Only black circles denote a component of either $\mathscr{A}$ or a component of $f^{-1}(\mathscr{A})$.

(3) If $\mathcal{X}_n \cong S_{0,1,1}$ and $\mathcal{Y}_n$ is homeomorphic to either $S_{0,3}$ or $S_{1,1}$, then using compactness of $f(\mathcal{A}'_n)$, find an annulus $\mathcal{A}_n$ in $\mathcal{X}_n$ such that $f(\mathcal{A}'_n) \subseteq \mathrm{int}(\mathcal{A}_n \cup \mathcal{Y}_n)$. Define $\Sigma_n := \mathcal{A}_n \cup \mathcal{Y}_n$; see Figure 6.

(4) If $\mathcal{Y}_n \cong S_{0,1,1}$ and $\mathcal{X}_n$ is homeomorphic to either $S_{0,3}$ or $S_{1,1}$, define Σ_n similarly to (3).

Thus $f(\mathcal{A}'_n) \subseteq \mathrm{int}(\Sigma_n)$ for each n. Now we show $\Sigma_n \to \infty$, so consider a compact subset $\mathcal{K}$ of Σ. Let $S_1, \ldots, S_m$ be a collection of complementary components of Σ decomposed by $\mathscr{A}$ such that $\mathcal{K} \subseteq \mathrm{int}\big(\bigcup_{l=1}^m S_l\big)$. Define $S := \bigcup_{l=1}^m S_l$. Notice that for an integer n, $f(\mathcal{A}'_n) \cap S \neq \varnothing$ if and only if $\mathcal{C}_n$ is a component of $\bigcup_{l=1}^m \partial S_l$. This is due to the construction of each Σ_n; see Figure 6. Since $\mathcal{C}_n \to \infty$ and $\bigcup_{l=1}^m \partial S_l$ is compact, $f(\mathcal{A}'_n) \cap S = \varnothing$ for all sufficiently large n. Now, $\mathcal{K} \subseteq \mathrm{int}(S)$ and each Σ_n is obtained from taking the union of all those complementary components of Σ (if a punctured disk appears, truncate it) which are hit by $f(\mathcal{A}'_n)$. Thus $\Sigma_n \cap \mathcal{K} = \varnothing$ for all sufficiently large n. Therefore, $\Sigma_n \to \infty$, as $\mathcal{K}$ is an arbitrary compact subset of Σ.

Next, for each $l \in \{1, \ldots, k_n\}$, applying Theorem 3.5.2 to each $f\tau_n|\mathbb{S}^1 \times [l-1, l] \to \mathcal{Z}_n$ where $\mathcal{Z}_n$ can be either $\Sigma_n \cap \mathcal{X}_n$ or $\Sigma_n \cap \mathcal{Y}_n$, we have a homotopy $H_n : \mathcal{A}'_n \times [0,1] \to \Sigma_n$ relative to $\partial \mathcal{A}'_n$ such that $H_n(-, 0) = f|\mathcal{A}'_n$ and $H_n(\mathcal{A}'_n, 1) = \mathcal{C}_n$. Finally, apply Lemma 3.3.2 on $\{H_n\}$. $\qquad\square$

Consider Figure 7, where $\mathcal{M}'_{n\alpha}$ and $\mathcal{M}_n$ are provided by Theorem 3.4.3 so that after defining $\mathcal{A}'_{\varepsilon n}$ as $\mathcal{A}'_n \cup \mathcal{M}'_{n\alpha}$,

$$(\mathcal{A}'_{\varepsilon n}, \mathcal{M}'_{n\alpha}, \mathcal{A}'_n) \cong (\mathbb{S}^1 \times [1,3], \mathbb{S}^1 \times [1,2], \mathbb{S}^1 \times [2,3]) \quad \text{and} \quad (\mathcal{M}_n, \mathcal{C}_n) \cong (\mathbb{S}^1 \times [1,2], \mathbb{S}^1 \times 2),$$
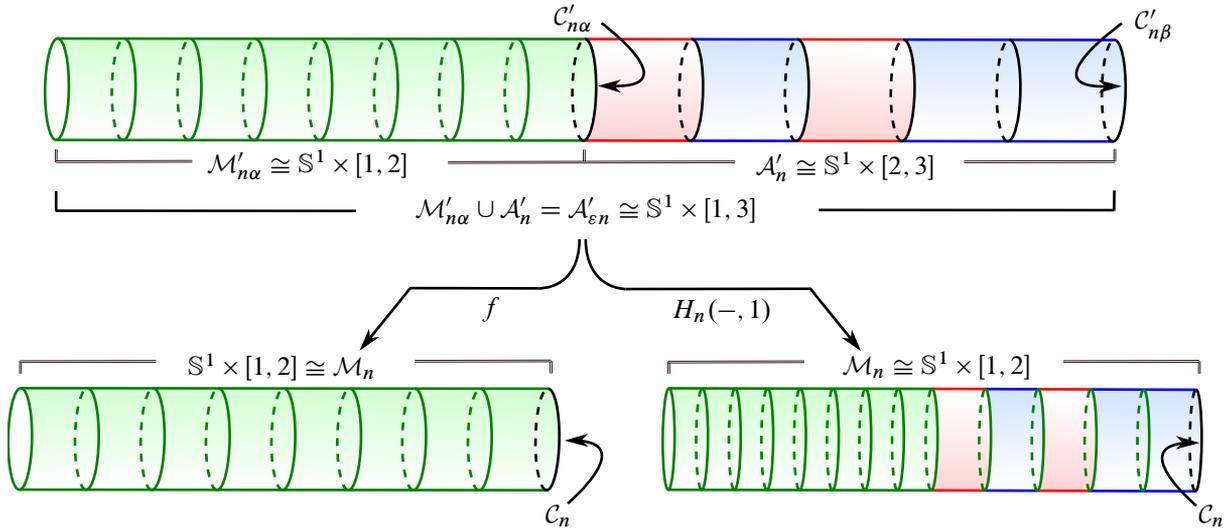
Figure 7: The description of $f|_{\mathcal{A}'_{\varepsilon n}} \to \mathcal{M}_n$ (resp. $H_n(-,1): \mathcal{A}'_{\varepsilon n} \to \mathcal{M}_n$) using Theorem 3.4.3 and Claim 3.5.3.1 (resp. Lemma 3.5.4). Only black circles denote a component of either $\mathscr{A}$ or a component of $f^{-1}(\mathscr{A})$.

resulting in the following description of f: if $\theta: \mathbb{S}^1 \to \mathbb{S}^1$ describes the homeomorphism $f|_{\mathcal{C}'_{n\alpha}} \to \mathcal{C}_n$ under the above identification, then $f(z,t) = (\theta(z), t)$ for $z \in \mathbb{S}^1 \times [1,2]$ and $f(z,t) \in \mathbb{S}^1 \times 2$ for $(z,t) \in \mathbb{S}^1 \times [2,3]$. Consider Claim 3.5.3.1 to see why $f(\mathbb{S}^1 \times [2,3]) = \mathbb{S}^1 \times 2$.

Now, use Lemma 3.5.4 to construct a homotopy $H_n: \mathcal{A}'_{\varepsilon n} \times [0,1] \to \mathcal{M}_n$ relative to $\partial \mathcal{A}'_{\varepsilon n}$ from $f|_{\mathcal{A}'_{\varepsilon n}} \to \mathcal{M}_n$ to the map $H_n(-,1)$ so that $(H_n(-,1))^{-1}(\mathcal{C}_n) = \mathcal{C}'_{n\beta}$ and $H_n(-,1)|_{\mathcal{C}'_{n\beta}} \to \mathcal{C}_n$ is a homeomorphism.

Notice that we are using the setup of the proof of Theorem 3.4.3. By Claim 3.4.4.1(4)–(5), $\mathcal{A}'_{\varepsilon n} \cap f^{-1}(\mathscr{A}) = f^{-1}(\mathcal{C}_n)$, $\mathcal{A}'_{\varepsilon n} \cap \mathcal{A}'_{\varepsilon m} = \varnothing$ if $m \neq n$, and $\mathcal{M}_n \to \infty$. Now consider Lemma 3.3.2 with $\{H_n\}$ to obtain the desired homotopy. $\qquad\square$

We prove the annulus-pushing lemma used in the proof of the previous theorem:

Lemma 3.5.4 *Any map $\varphi: \mathbb{S}^1 \times [1,3] \to \mathbb{S}^1 \times [1,2]$ which sends $\mathbb{S}^1 \times r$ into $\mathbb{S}^1 \times r$ for $1 \leq r \leq 2$ and sends $\mathbb{S}^1 \times r$ into $\mathbb{S}^1 \times 2$ for $2 \leq r \leq 3$ can be homotoped relative to $\mathbb{S}^1 \times \{1,3\}$ to satisfy $\varphi^{-1}(\mathbb{S}^1 \times 2) = \mathbb{S}^1 \times 3$.*

Proof Let $\varphi_1: \mathbb{S}^1 \times [1,3] \to \mathbb{S}^1$ and $\varphi_2: \mathbb{S}^1 \times [1,3] \to [1,2]$ be the components of φ. Consider a homeomorphism $\ell: [1,3] \to [1,2]$ with $\ell(1) = 1$ and $\ell(3) = 2$. Now $H: \mathbb{S}^1 \times [1,3] \times [0,1] \to \mathbb{S}^1 \times [1,2]$ defined by

$$H((z,s),t) := (\varphi_1(z,s), (1-t)\varphi_2(z,s) + t\ell(s)) \quad \text{for } (z,s) \in \mathbb{S}^1 \times [1,3] \text{ and } t \in [0,1]$$

is our required homotopy. $\qquad\square$

Remark 3.5.5 In Theorem 3.5.3, the number of components of $\mathscr{A}$ can be infinite; thus, the number of outermost annuli (one outermost annulus for each component of $\mathscr{A}$, if any) can be infinite. That's why we have removed all outermost annuli simultaneously by a single proper homotopy, not one by one. Also, to prove the topological rigidity of closed surfaces, one may ignore the annulus removal process considering induction on the genus; see [13, Theorem 3.1] or [35, Theorems 4.6.2 and 4.6.3]. But, since the genus of a noncompact surface can be infinite, we can't ignore the annulus removal process here.

3.6 Is pseudoproper homotopy equivalence a map of degree ±1?

Let $f : \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces, where surfaces are homeomorphic to neither the plane nor the punctured plane. Our aim in this section is to properly homotope f to obtain a closed disk $\mathcal{D} \subseteq \Sigma$ so that $f | f^{-1}(\mathcal{D}) \to \mathcal{D}$ becomes a homeomorphism, and thus show $\deg(f) = \pm 1$; see Theorem 2.6.1. Having got this and then using Lemma 3.6.4.1, it can be said that f is surjective, which further implies that after a proper homotopy for removing unnecessary components from the transverse preimage of a decomposition circle $\mathcal{C}$, a single circle will still be left that can be mapped onto $\mathcal{C}$ homeomorphically; see Theorem 3.6.4.4.

The argument for finding such a disk $\mathcal{D}$ is based on finding a finite-type bordered surface S in Σ such that for each component $\mathcal{C}$ of ∂S, we have $f^{-1}(\mathcal{C}) \neq \varnothing$, even after any proper homotopy of f. Once we get S, after a proper homotopy, we may assume that $f | f^{-1}(\partial S) \to \partial S$ is a homeomorphism; see Theorem 3.5.3. Now, since f is π_1–injective, by the topological rigidity of the pair of pants together with the proper rigidity of the punctured disk, after a proper homotopy, one can show that $f | f^{-1}(S) \to S$ is a homeomorphism. Therefore the required $\mathcal{D}$ can be any disk in $\operatorname{int}(S)$.

Now, to find such an S, we consider two cases: If Σ is either an infinite-type surface or any $S_{g,0,p}$ with high complexity (to us, high complexity always means $g + p \geq 4$ or $p \geq 6$), then using π_1–surjectivity of f, we can choose S as a pair of pants in Σ so that $\Sigma \setminus S$ has at least two components and every component of $\Sigma \setminus S$ has a nonabelian fundamental group. On the other hand, if Σ is a finite-type surface, then we choose a punctured disk in Σ as S so that the puncture of S is an end $e \in \operatorname{im}(\operatorname{Ends}(f)) \subseteq \operatorname{Ends}(\Sigma)$.

We can recall our earlier two examples to show that the plane and the punctured plane are the only surfaces for which our theory fails: consider the pseudoproper homotopy equivalences $\varphi : \mathbb{C} \ni z \mapsto z^2 \in \mathbb{C}$ and $\psi : \mathbb{S}^1 \times \mathbb{R} \ni (z, x) \mapsto (z, |x|) \in \mathbb{S}^1 \times \mathbb{R}$. The local homeomorphism φ is a map of $\deg = \pm 2$ by [16, Lemma 2.1b] (note that for any local homeomorphism $p : X \to Y$ between two manifolds, where Y is orientable, an orientation of Y can be pulled back to induce an orientation on X so that p becomes an orientation-preserving map). On the other hand, $\deg(\psi) = 0$ as ψ is not surjective; see Lemma 3.6.4.1.

3.6.1 Essential pairs of pants and the degree of a pseudoproper homotopy equivalence

Definition 3.6.1.1 A smoothly embedded pair of pants P in a surface Σ is said to be an *essential pair of pants* of Σ if $\Sigma \setminus P$ has at least two components and every component of $\Sigma \setminus P$ has a nonabelian fundamental group.
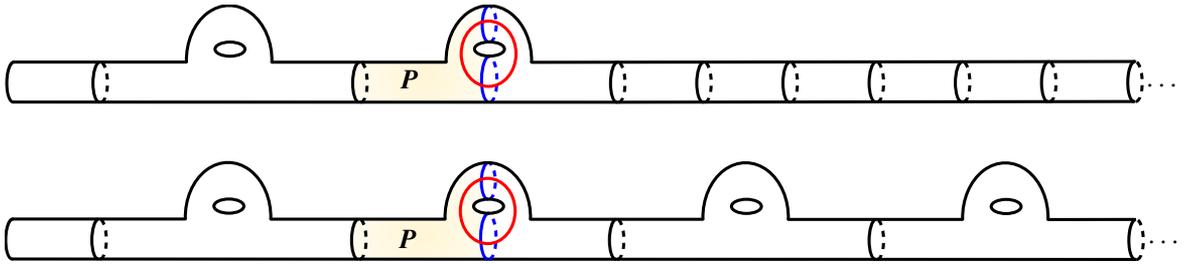
Figure 8: Finding an essential pair of pants P in each of $S_{2,0,1}$ and the Loch Ness monster by decomposing a two-holed torus into two copies of the pair of pants.

Finding an essential pair of pants in a noncompact surface will be divided into two cases: when the genus is at least two and when the space of ends has at least six elements.

Definition 3.6.1 Let P be a smoothly embedded copy of the pair of pants in a two-holed torus $\mathcal{S}$ ($\mathcal{S}$ is a copy of $S_{1,2}$). We say P *is obtained from decomposing* $\mathcal{S}$ *into two copies of the pair of pants* if there exists another smoothly embedded copy $\tilde{P}$ of the pair of pants in $\mathcal{S}$ such that $P \cup \tilde{P} = \mathcal{S}$ and $P \cap \tilde{P} = \partial P \cap \partial \tilde{P}$ is the union of two smoothly embedded disjoint circles in the interior of $\mathcal{S}$ (∂P shares exactly two of its components with $\partial \tilde{P}$).

The following lemma says that every noncompact surface with a genus of at least two has an essential pair of pants with some additional properties:

Lemma 3.6.1.2 *Let* Σ *be a noncompact surface of genus at least two. Then* Σ *has an essential pair of pants* P *with the following additional properties*: Σ *contains a smoothly embedded copy* $\mathcal{S}$ *of* $S_{1,2}$ *such that* $\Sigma \setminus \mathcal{S}$ *has precisely two components and each component of* $\Sigma \setminus \mathcal{S}$ *has a nonabelian fundamental group, and* P *is a smoothly embedded copy of the pair of pants in* $\mathcal{S}$ *obtained by decomposing* $\mathcal{S}$ *into two copies of the pair of pants.*

Proof Consider an inductive construction of Σ; see Theorem 2.5.1. Since $g(\Sigma) \geq 2$, at least two smoothly embedded copies of $S_{1,2}$ are used in this inductive construction. By Lemma 3.1.8, without loss of generality, we may assume that two smoothly embedded copies of $S_{1,2}$ are used successively just after the initial disk; see Figure 8. Among these two copies of $S_{1,2}$, breaking the last one (that copy of $S_{1,2}$ which we just used to construct K_3 from K_2) into two copies of the pair of pants, as illustrated in Figure 8, we get the required essential pair of pants. $\square$

Lemma 3.6.1.3 *Let* $f : \Sigma' \to \Sigma$ *be a* π_1*–surjective map between two noncompact surfaces, where* Σ *has genus at least two. Consider an essential pair of pants* P *in* Σ *given by Lemma 3.6.1.2. Then* $f^{-1}(\mathrm{int}\, P) \neq \varnothing$ *and* $f^{-1}(c) \neq \varnothing$ *for each component* c *of* ∂P.

Proof Let $\mathcal{S}$ be a smoothly embedded copy of $S_{1,2}$ in Σ such that P is obtained from decomposing $\mathcal{S}$ into two copies of the pair of pants. If possible, let $f^{-1}(\mathrm{int}\, P) \neq \varnothing$. By continuity of f, the image of f

is contained in precisely one of the two components of $\Sigma \setminus \text{int}(P)$. But each component of $\Sigma \setminus \text{int}(P)$ has a nonabelian fundamental group, ie $\pi_1(f) \colon \pi_1(\Sigma') \to \pi_1(\Sigma)$ is not surjective, a contradiction. Therefore $f^{-1}(\text{int } P)$ must be nonempty.

To prove the second part, let c_1, c_2, and c_3 denote all three components of P such that both $\Sigma \setminus c_1$ and $\Sigma \setminus (c_2 \cup c_2)$ are disconnected, but neither $\Sigma \setminus c_2$ nor $\Sigma \setminus c_3$ is disconnected. In Figure 8, c_2 and c_3 are blue circles, and c_1 is black. Notice that we have a smoothly embedded primitive circle $\mathcal{C} \subseteq \text{int}(\mathcal{S})$ (in Figure 8, each red circle denotes $\mathcal{C}$) so that for each $k = 2, 3$, $c_k \cap \mathcal{C}$ is a single point, where c_k intersects $\mathcal{C}$ transversally. Therefore, for each $k = 2, 3$, using the bigon criterion [17, Proposition 1.7], any loop belonging to class $[\mathcal{C}] \in \pi_1(\Sigma)$ must intersect c_k. That is, if $f^{-1}(c_2)$ or $f^{-1}(c_3)$ were empty, then $[\mathcal{C}]$ would not belong to the image of $\pi_1(f) \colon \pi_1(\Sigma') \to \pi_1(\Sigma)$. But f is π_1–surjective. Thus $f^{-1}(c_2) \neq \varnothing \neq f^{-1}(c_3)$. On the other hand, $\Sigma \setminus c_1$ has precisely two components, and each component of $\Sigma \setminus c_1$ has a nonabelian fundamental group, ie by continuity and π_1–surjectivity of f, we can say that $f^{-1}(c_1) \neq \varnothing$. $\square$

Now we consider the second case of finding an essential pair of pants in a noncompact surface, namely when the space of ends has at least six elements:

Lemma 3.6.1.4 *Let Σ be a noncompact surface with at least six ends. Then Σ has an essential pair of pants P such that $\Sigma \setminus P$ has precisely three components and each component of $\Sigma \setminus P$ has a nonabelian fundamental group.*

Proof Consider an inductive construction of Σ; see Theorem 2.5.1. Since $|\text{Ends}(\Sigma)| \geq 6$, at least five smoothly embedded copies of $S_{0,3}$ are used in this inductive construction. By Lemma 3.1.8, without loss of generality, we may assume that five smoothly embedded copies of $S_{0,3}$ are used successively just after the initial disk. Let P be the copy that shares all three boundary components with three other copies of this sequence of five copies of $S_{0,3}$; see Figure 9. Thus $\Sigma \setminus P$ has precisely three components, and each component of $\Sigma \setminus P$ has a nonabelian fundamental group.

In Figure 9, inductive constructions (up to a sufficient number of steps) of two surfaces have been given: the surface at the top contains a copy of $\{z \in \mathbb{C} : |z| \geq 1, z \notin \mathbb{N} \times 0\}$, and the bottom is the Cantor tree surface (the planar surface whose space of ends is homeomorphic to the Cantor set). In each surface, an essential pair of pants P is contained in the shaded compact bordered subsurface. $\square$

We can prove the following lemma by a similar argument as in the proof of Lemma 3.6.1.3:

Lemma 3.6.1.5 *Let $f \colon \Sigma' \to \Sigma$ be a π_1–surjective proper map between two noncompact surfaces, where Σ has at least six ends. Consider an essential pair of pants P in Σ given by Lemma 3.6.1.4. Then $f^{-1}(\text{int } P) \neq \varnothing$ and $f^{-1}(c) \neq \varnothing$ for each component c of ∂P.*

The following theorem completes the whole process of finding an essential pair of pants, which will be used to find the degree of a pseudoproper homotopy equivalence:
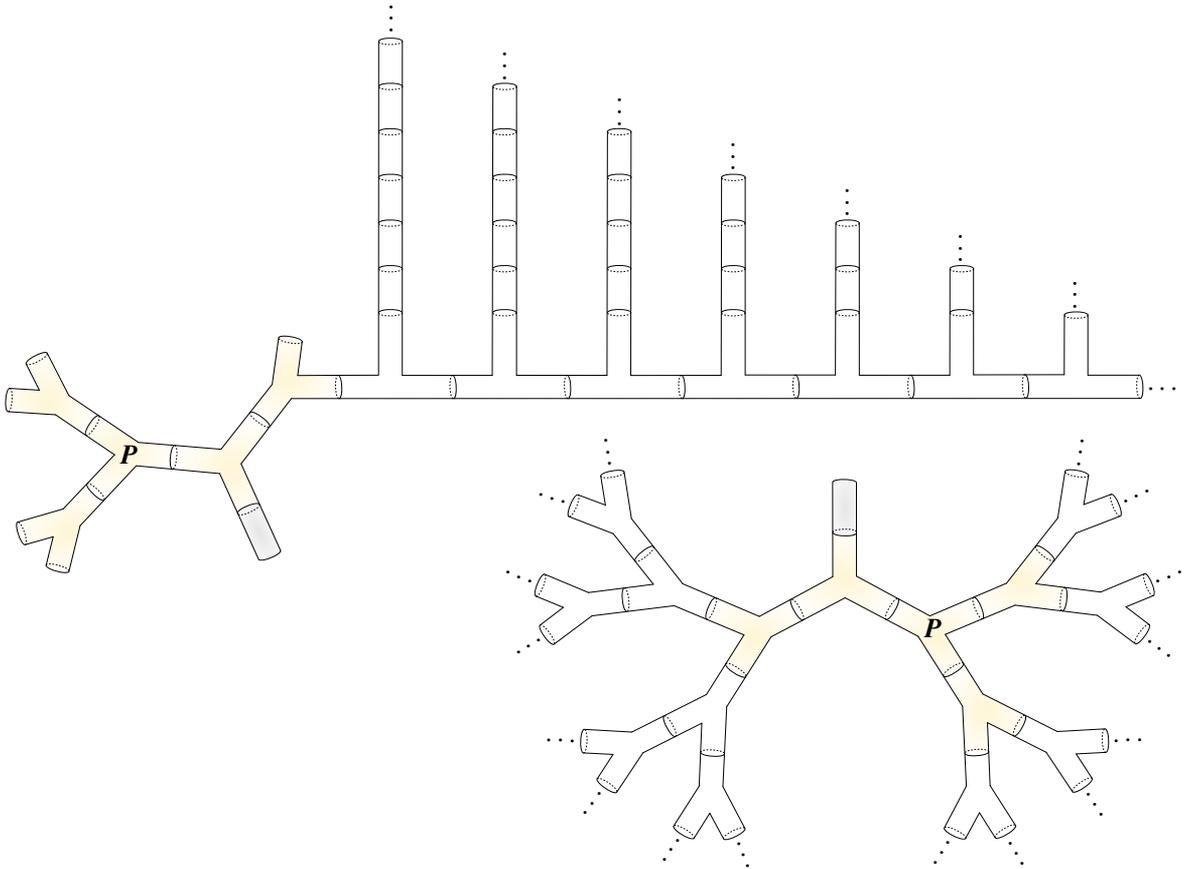
Figure 9: Finding an essential pair of pants P in a noncompact surface with at least six ends.

Theorem 3.6.1.6 *Let $f : \Sigma' \to \Sigma$ be a π_1–surjective proper map between two noncompact surfaces. Suppose Σ is either an infinite-type surface or a finite-type surface $S_{g,0,p}$ with high complexity (ie $g + p \geq 4$ or $p \geq 6$). Then Σ contains an essential pair of pants P such that $f^{-1}(\text{int } P) \neq \varnothing$ and $f^{-1}(c) \neq \varnothing$ for each component c of ∂P.*

Proof If an infinite-type surface has a finite genus, then it must have infinitely many ends; see Proposition 3.1.10. Thus using Lemmas 3.6.1.2, 3.6.1.3, 3.6.1.4, and 3.6.1.5, the proof is complete in all cases, except when Σ is homeomorphic to $S_{1,0,3}$, $S_{1,0,4}$, or $S_{1,0,5}$. We consider the case when $\Sigma \cong S_{1,0,3}$; the other cases are similar.

Define an inductive construction of $S_{1,0,3}$ by starting with a copy of $S_{0,1}$, consecutively adding two copies of $S_{0,3}$, then a copy of $S_{1,2}$, and finally three sequences of annuli to obtain three planar ends; see Figure 1. Therefore, in this inductive construction, K_4 is obtained from K_3, adding a copy $\mathcal{S}$ of $S_{1,2}$. Let P be a smoothly embedded copy of the pair of pants in $\mathcal{S}$ such that P is obtained from decomposing $\mathcal{S}$ into two copies of the pair of pants and $P \cap K_3 \neq \varnothing$. Now, an argument similar to that given in Lemma 3.6.1.3 completes the proof. □

At this stage, we need a couple of lemmas. The first one, Lemma 3.6.1.7, is well known; its proof has been given for reader convenience.

Lemma 3.6.1.7 *Let Σ be a surface, and let S be a smoothly embedded bordered subsurface of Σ. Then the inclusion-induced map $\pi_1(S) \to \pi_1(\Sigma)$ is injective if either of the following is satisfied:*

(1) *∂S is a separating primitive circle on Σ and S is one of the two sides of ∂S in Σ.*

(2) *S is compact and each component of ∂S is a primitive circle on Σ.*

Proof (1) Since $\pi_1(\Sigma) \cong \pi_1(S) *_{\pi_1(\partial S)} \pi_1(\Sigma \setminus \mathrm{int}\, S)$ by the Seifert–Van Kampen theorem and the inclusions $\partial S \hookrightarrow S$, $\Sigma \setminus \mathrm{int}(S)$ are π_1–injective, we are done.

(2) It is enough to construct a sequence $\Sigma = S_0 \supseteq S_1 \supseteq \cdots \supseteq S_n = S$ of subsurfaces of Σ, where n is the number of components of ∂S, such that for each $k = 1, \ldots, n$, the following hold:

(i) S_k is a bordered subsurface of S_{k-1} and the inclusion map $S_k \hookrightarrow S_{k-1}$ is π_1–injective.

(ii) $\partial S_k \setminus \partial S_{k-1}$ is either a component of ∂S_k or union of two components of ∂S_k. In either case, $\partial S_k \setminus \partial S_{k-1}$ shares only one component with ∂S.

We construct this sequence inductively as follows: To construct S_k from S_{k-1}, pick a component c of $\partial S \setminus \partial S_{k-1}$. If c separates S_{k-1}, define S_k as that side of c in S_{k-1} which contains S; then consider an argument similar to the proof of Lemma 3.6.1.7(1). If c doesn't separate S_{k-1}, pick a smoothly embedded annulus $A \subset \mathrm{int}(S_{k-1})$ such that $A \cap S = c$. Define $S_k := S_{k-1} \setminus \mathrm{int}(A)$. Now, S_{k-1} is obtained from S_k identifying c with $\partial A \setminus c$ by an orientation-reversing diffeomorphism $\varphi \colon c \to \partial A \setminus c$. By the HNN–Seifert–Van Kampen theorem, $\pi_1(S_{k-1}) \cong \pi_1(S_k) *_{\pi_1(\varphi)}$, where the map $\pi_1(S_k) \to \pi_1(S_{k-1})$ (which is inclusion induced) is injective due to Britton's lemma. $\square$

The following lemma roughly says that the degree of a map between two compact bordered surfaces can be determined from the degree of its restriction on the boundaries:

Lemma 3.6.1.8 *Let $\varphi \colon S_{g_1,b_1} \to S_{g_2,b_2}$ be a map between two compact bordered surfaces such that $\varphi|\partial S_{g_1,b_1} \hookrightarrow \partial S_{g_2,b_2}$ is an embedding. Then $\varphi(\partial S_{g_1,b_1}) = \partial S_{g_2,b_2}$ and $\deg(\varphi) = \pm 1$.*

Proof Notice that φ maps each component of $\partial S_{g_1,b_1}$ homeomorphically onto a component of $\partial S_{g_2,b_2}$, and any two distinct components of $\partial S_{g_1,b_1}$ have distinct φ–images. Now, naturality of homology long exact sequences of $(S_{g_1,b_1}, \partial S_{g_1,b_1})$ and $(S_{g_2,b_2}, \partial S_{g_2,b_2})$ give the following commutative diagram:

$$
\begin{array}{ccc}
H_2(S_{g_1,b_1}, \partial S_{g_1,b_1}) \cong \mathbb{Z} & \xrightarrow{\ 1 \mapsto \bigoplus^{b_1} 1\ } & \bigoplus^{b_1} \mathbb{Z} \cong H_1(\partial S_{g_1,b_1}) \\
{\scriptstyle \times \deg(\varphi)} \downarrow & & \downarrow {\scriptstyle \bigoplus^{b_1} 1 \mapsto \bigoplus^{b_1}(\pm 1) \oplus \bigoplus^{b_2 - b_1} 0} \\
H_2(S_{g_2,b_2}, \partial S_{g_2,b_2}) \cong \mathbb{Z} & \xrightarrow{\ 1 \mapsto \bigoplus^{b_2} 1\ } & \bigoplus^{b_2} \mathbb{Z} \cong H_1(\partial S_{g_2,b_2})
\end{array}
$$

The horizontal maps are the connecting homomorphisms for homology long exact sequences; for their description see [25, Exercise 31 of Section 3.3]. Commutativity of this diagram gives $b_2 = b_1$ (the integer $\deg(\varphi)$ can't be simultaneously 0 and ± 1), and thus $\deg(\varphi) = \pm 1$. $\square$

The proof of Theorem 3.6.1.9 can be found in [35, Theorem 4.6.2]. It also follows from the much more general result [13, Theorem 3.1]. Since compact bordered surfaces are aspherical, an application of the Whitehead theorem says that the assumption "$\varphi \colon S' \to S$ is a homotopy equivalence" in Theorem 3.6.1.9 is equivalent to the assumption "$\pi_1(\varphi)$ is an isomorphism".

Theorem 3.6.1.9 (rigidity of compact bordered surfaces) *Let $\varphi \colon S' \to S$ be a homotopy equivalence between two compact bordered surfaces such that $\varphi^{-1}(\partial S) = \partial S'$. If $\varphi|\partial S' \to \partial S$ is a homeomorphism, then φ is homotopic to a homeomorphism relative to $\partial S'$.*

The following lemma gives some sufficient conditions that ensure the preimage of a compact bordered subsurface under a proper map becomes a compact bordered subsurface of the same homeomorphism type. Its usage is twofold: firstly, in Theorem 3.6.1.11, to find the degree of a pseudoproper homotopy equivalence, and secondly in the proof of Theorem 1.

Lemma 3.6.1.10 *Let $f \colon \Sigma' \to \Sigma$ be a π_1–injective proper map between two noncompact oriented surfaces, and let S be a smoothly embedded compact bordered subsurface of Σ with $f^{-1}(\text{int } S) \neq \varnothing$. Suppose $f^{-1}(\partial S)$ is a pairwise-disjoint collection of smoothly embedded primitive circles on Σ' such that f sends $f^{-1}(\partial S)$ homeomorphically onto ∂S. Then $f^{-1}(S)$ is a copy of S in Σ' with $\partial f^{-1}(S) = f^{-1}(\partial S)$, and $\deg(f) = \pm 1$.*

Proof Since $f^{-1}(\text{int } S) \neq \varnothing$ and f is proper, the continuity of $f|\Sigma' \setminus f^{-1}(\partial S) \to \Sigma \setminus \partial S$ tells us that $\Sigma' \setminus f^{-1}(\partial S)$ is disconnected. Let $S' \subset \Sigma'$ be a bordered subsurface obtained as a complementary component of the decomposition of Σ' by $f^{-1}(\partial S)$ such that $f(S') \subseteq S$. That is, S' is the closure of one of the components of $\Sigma' \setminus f^{-1}(\partial S)$ and S' is contained in the compact set $f^{-1}(S)$. So S' is a compact bordered subsurface of Σ', and each component of $\partial S'$ is a component of $f^{-1}(\partial S)$. In the following few lines, we will show that each component of $f^{-1}(\partial S)$ is also a component of $\partial S'$. Since $f|f^{-1}(\partial S) \to \partial S$ is a homeomorphism, we can say that $f|\partial S' \hookrightarrow \partial S$ is an embedding. Now, by Lemma 3.6.1.8, $\partial S' = f^{-1}(\partial S)$ and $\deg(f|S' \to S) = \pm 1$. Next, by Theorem 2.6.3, $f|S' \to S$ is π_1–surjective. Since the inclusion $S' \hookrightarrow \Sigma'$ and f are π_1–injective, $f|S' \to S$ is also so; see Lemma 3.6.1.7(2). Thus $f|S' \to S$ is π_1–bijective, and so Theorem 3.6.1.9 tells that $S' \cong S$. Finally, if S'' is another bordered subsurface obtained as a complementary component of decomposition of Σ' by $f^{-1}(\partial S)$ with $f(S'') \subseteq S$, then similarly, $S'' \cong S$. Since $f|f^{-1}(\partial S) \to \partial S$ is a homeomorphism and Σ' is connected, $S'' = S'$ (otherwise Σ' would be the compact surface $S' \cup_{\partial S' = \partial S''} S''$). Therefore $f^{-1}(S) = S' \cong S$, and thus the proof of the first part is completed.

Now we will prove that $\deg(f) = \pm 1$. Since $\deg(f)$ remains invariant after any proper homotopy of f, we can properly homotope f as we want. So apply Theorem 3.6.1.9 to $f|S' \to S$. Thus $f \colon \Sigma' \to \Sigma$ can be properly homotoped relative to $\Sigma' \setminus \text{int}(S')$ to map $S' = f^{-1}(S)$ homeomorphically onto S. Then by Theorem 2.6.1, $\deg(f) = \pm 1$. $\qquad\square$

We are now ready to prove that a pseudoproper homotopy equivalence is a map of degree ± 1 if the codomain contains an essential pair of pants.

Theorem 3.6.1.11 *Let $f: \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces, where Σ is either an infinite-type surface or a finite-type noncompact surface $S_{g,0,p}$ with high complexity (to us, high complexity means $g + p \geq 4$ or $p \geq 6$). Then $\deg(f) = \pm 1$.*

Proof Since $\deg(f)$ remains invariant after any proper homotopy of f, we can properly homotope f as we want. Theorem 3.6.1.6 gives an essential pair of pants P in Σ such that $f^{-1}(\text{int}(P)) \neq \varnothing$ and $f^{-1}(c) \neq \varnothing$ for each component c of ∂P, even after any proper homotopy of f. Using Theorem 3.2.1 and then Theorem 3.5.3, after a proper homotopy, we may assume that $f^{-1}(\text{int } P) \neq \varnothing$ and $f^{-1}(\partial P)$ is a pairwise-disjoint collection of three smoothly embedded circles on Σ such that $f|f^{-1}(\partial P) \to \partial P$ is a homeomorphism.

Now, if possible, let c' be a component of $f^{-1}(\partial P)$ such that there is an embedding $i': \mathbb{D}^2 \hookrightarrow \Sigma'$ with $c' = i'(\mathbb{S}^1)$. Then the embedding $f \circ i'|\mathbb{S}^1 \hookrightarrow \Sigma$ is nullhomotopic and $c := f \circ i'(\mathbb{S}^1)$ is a component of ∂P. But P is an essential pair of pants in Σ, and so each component of ∂P is a primitive circle on Σ. Theorem 2.2.2 tells us we have reached a contradiction. Hence each component of $f^{-1}(\partial P)$ is a primitive circle on Σ'. Finally, applying Lemma 3.6.1.10, we complete the proof. $\square$

3.6.2 An essential punctured disk of a proper map and the degree of a pseudoproper homotopy equivalence

We first build up notation for Section 3.6.2. Let Σ be a noncompact surface. Since $\text{Ends}(\Sigma)$ is independent of the choice of efficient exhaustion of Σ by compacta, we will use Goldman's inductive construction to define $\text{Ends}(\Sigma)$; see Section 2.3. So consider an inductive construction of Σ. For each $i \geq 1$, define K_i to be the compact bordered subsurface of Σ after the ith step of the induction. Then $\{K_i\}_{i=1}^{\infty}$ is an efficient exhaustion of Σ by compacta. Also, notice that $\text{int}(K_1) \subseteq \text{int}(K_2) \subseteq \cdots$ is an increasing sequence of open subsets of Σ such that $\bigcup_{i=1}^{\infty} \text{int}(K_i) = \Sigma$, and thus every compact subset of Σ is contained in some $\text{int}(K_i)$.

Suppose Σ' is another noncompact surface and $f: \Sigma' \to \Sigma$ is a proper map. Let $(V_1, V_2, \ldots)$ be an end of Σ, ie V_i is a component of $\Sigma \setminus K_i$ and $V_1 \supseteq V_2 \supseteq \cdots$. With this setup, we are now ready to state a lemma that is more or less related to Proposition 2.3.1:

Theorem 3.6.2.1 *Assume that $f^{-1}(V_i) \neq \varnothing$ for each $i \geq 1$. Then for every proper map $g: \Sigma' \to \Sigma$ which is properly homotopic to f, we have $g^{-1}(V_i) \neq \varnothing$ for each $i \geq 1$.*

Proof Let $g: \Sigma' \to \Sigma$ be a proper map and let $\mathcal{H}: \Sigma' \times [0, 1] \to \Sigma$ be a proper homotopy from f to g. Notice that $V_i \to \infty$: if $\mathcal{X}$ is a compact subset of Σ, then $\mathcal{X} \subseteq \text{int}(K_{i_0})$ for some positive integer i_0, ie $\mathcal{X} \cap V_i = \varnothing$ for all $i \geq i_0$. Therefore $f^{-1}(V_i) \to \infty$: if $\mathcal{X}'$ is a compact subset of Σ', then $f(\mathcal{X}')$ is compact, so $f(\mathcal{X}') \cap V_i = \varnothing$ for all but finitely many i, ie $\mathcal{X}' \cap f^{-1}(V_i) = \varnothing$ for all but finitely many i.

Let n be any positive integer. Consider the compact subset $p(\mathcal{H}^{-1}(K_n))$ of Σ', where $p\colon \Sigma' \times [0,1] \to \Sigma'$ is the projection. Since $f^{-1}(V_i) \to \infty$, we have an integer $i_n > n$ such that $f^{-1}(V_{i_n}) \subseteq \Sigma' \setminus p(\mathcal{H}^{-1}(K_n))$. Now consider any $x_{i_n} \in f^{-1}(V_{i_n})$. Then $\mathcal{H}(x_{i_n} \times [0,1]) \subseteq \Sigma \setminus K_n$, that is, the connected set $\mathcal{H}(x_{i_n} \times [0,1])$ is contained in one of the components of $\Sigma \setminus K_n$. But $\mathcal{H}(x_{i_n}, 0) = f(x_{i_n}) \in V_{i_n} \subseteq V_n$, meaning $\mathcal{H}(x_{i_n} \times [0,1]) \subseteq V_n$. In particular, this means $g(x_{i_n}) = \mathcal{H}(x_{i_n}, 1) \in V_n$. Since n is an arbitrary positive integer, we are done. $\qquad\square$

Definition 3.6.2.2 Let $e = (V_1, V_2, \dots)$ be an end of Σ such that for some nonnegative integer i_e, $\overline{V}_i \cong S_{0,1,1}$ for all $i \geq i_e$ (e is an isolated planar end of Σ). If $f^{-1}(V_i) \neq \varnothing$ for all $i \geq 1$, then for each integer $i \geq i_e$, we say $\overline{V}_i$ is an *essential punctured disk* of f.

Theorem 3.6.2.1 says that the notion of an essential punctured disk is invariant under proper homotopy. In Theorem 3.6.2.6, we show that, after a proper homotopy, the preimage of the boundary of an essential punctured disk under a pseudoproper homotopy equivalence bounds a planar end of the domain. Before proving this, we need the following lemma, which gives some sufficient conditions so that the preimage of a punctured disk in the codomain under a proper map becomes a punctured disk in the domain.

Lemma 3.6.2.3 *Let $f\colon \Sigma' \to \Sigma$ be a π_1–injective proper map between two noncompact oriented surfaces, and let $\mathcal{C}$ be a smoothly embedded separating circle on Σ such that one of the two sides of $\mathcal{C}$ in Σ is a punctured disk $\mathcal{D}_*$. Also, let Σ' be homeomorphic to neither $\mathbb{S}^1 \times \mathbb{R}$ nor $\mathbb{R}^2$. If $f^{-1}(\mathcal{C})$ is a smoothly embedded primitive circle on Σ' such that $f|f^{-1}(\mathcal{C}) \to \mathcal{C}$ is a homeomorphism and $f^{-1}(\mathrm{int}\,\mathcal{D}_*) \neq \varnothing$, then $f^{-1}(\mathcal{D}_*)$ is a copy of the punctured disk in Σ' with $\partial f^{-1}(\mathcal{D}_*) = f^{-1}(\mathcal{C})$ and $\deg(f) = \pm 1$.*

Proof Notice that $\Sigma' \not\cong \mathbb{R}^2, \mathbb{S}^1 \times \mathbb{R}$, ie $\pi_1(\Sigma')$ is nonabelian by Theorem 3.1.9. Since $f^{-1}(\mathrm{int}\,\mathcal{D}_*) \neq \varnothing$ and $\pi_1(f)(\pi_1(\Sigma'))$ is nonabelian, by continuity of $f|\Sigma' \setminus f^{-1}(\mathcal{C}) \to \Sigma \setminus \mathcal{C}$ we can say that $\Sigma' \setminus f^{-1}(\mathcal{C})$ is disconnected. Let S' be a side of $f^{-1}(\mathcal{C})$ in Σ' for which $f(S') \subseteq \mathcal{D}_*$. Since f is π_1–injective, by Lemma 3.6.1.7(1), $f|S' \to \mathcal{D}_*$ is also so. Thus $\pi_1(S')$ is a subgroup of $\mathbb{Z}$. Now, $\mathrm{int}(S')$ is homotopy equivalent to S' and bounded by the primitive circle $f^{-1}(\mathcal{C})$ on Σ'; so, using Theorem 3.1.9, $S' \cong S_{0,1,1}$. Next, if S'' is another side of $f^{-1}(\mathcal{C})$ in Σ' for which $f(S'') \subseteq \mathcal{D}_*$, then similarly, $S'' \cong S_{0,1,1}$. Since $f|f^{-1}(\mathcal{C}) \to \mathcal{C}$ is a homeomorphism and Σ' is connected, $S'' = S'$; otherwise, Σ' would be $S' \cup_{f^{-1}(\mathcal{C})} S'' \cong \mathbb{S}^1 \times \mathbb{R}$. Therefore $f^{-1}(\mathcal{D}_*) = S' \cong \mathcal{D}_*$, and thus the proof of the first part is completed.

Since $\deg(f)$ remains invariant after any proper homotopy of f, we can properly homotope f as we want. So apply Theorem 3.6.2.4 to $f|S' \to \mathcal{D}_*$. Thus $f\colon \Sigma' \to \Sigma$ can be properly homotoped relative to $\Sigma' \setminus \mathrm{int}(S')$ to map $S' = f^{-1}(\mathcal{D}_*)$ homeomorphically onto $\mathcal{D}_*$. Now, by Theorem 2.6.1, $\deg(f) = \pm 1$. $\quad\square$

We prove a well-known theorem used in the previous lemma:

Theorem 3.6.2.4 (proper rigidity of the punctured disk) *Let $\boldsymbol{D}_*$ be a punctured disk and let $\varphi\colon \boldsymbol{D}_* \to \boldsymbol{D}_*$ be a proper map such that $\varphi^{-1}(\partial \boldsymbol{D}_*) = \partial \boldsymbol{D}_*$ and $\varphi|\partial \boldsymbol{D}_* \to \partial \boldsymbol{D}_*$ is a homeomorphism. Then φ is properly homotopic to a homeomorphism $\boldsymbol{D}_* \to \boldsymbol{D}_*$ relative to the boundary $\partial \boldsymbol{D}_*$.*

Proof Without loss of generality we may assume $\boldsymbol{D}_* = \{z \in \mathbb{C} : 0 < |z| \le 1\}$. Define $\mathcal{H} : \boldsymbol{D}_* \times [0,1] \to \boldsymbol{D}_*$ by

$$\mathcal{H}(z,t) := \begin{cases} (1-t)\varphi(z/(1-t)) & \text{if } 0 < |z| \le 1-t, \\ |z|\varphi(z/|z|) & \text{if } 1-t < |z| \le 1. \end{cases}$$

Notice that $\varphi \simeq \mathcal{H}(-,1)$ relative to $\partial \boldsymbol{D}_*$, and $\mathcal{H}(-,1) : \boldsymbol{D}_* \to \boldsymbol{D}_*$ is a homeomorphism.

Now we prove $\mathcal{H}$ is a proper map, so let $\{(z_n, t_n)\}$ be a sequence in $\boldsymbol{D}_* \times [0,1]$ with $z_n \to 0$. We need to show that $\mathcal{H}(z_n, t_n) \to 0$. Define $\mathscr{A} := \{n \in \mathbb{N} : 1 - t_n < |z_n|\}$ and $\mathscr{B} := \{n \in \mathbb{N} : |z_n| \le 1 - t_n\}$. Then $\mathbb{N} = \mathscr{A} \cup \mathscr{B}$. Therefore it is enough to show $\{\mathcal{H}(z_n, t_n) : n \in \mathscr{A}\} \to 0$ (resp. $\{\mathcal{H}(z_n, t_n) : n \in \mathscr{B}\} \to 0$) whenever $\mathscr{A}$ (resp. $\mathscr{B}$) is infinite.

If $\mathscr{A}$ is infinite, then $\{\mathcal{H}(z_n, t_n) : n \in \mathscr{A}\} \to 0$, since $|\mathcal{H}(z_n, t_n)| = |z_n| \cdot |\varphi(z_n/|z_n|)| \le |z_n|$ for all $n \in \mathscr{A}$.

Next, assume $\mathscr{B}$ is infinite. We will prove that $\{\mathcal{H}(z_n, t_n) : n \in \mathscr{B}\} \to 0$, so consider any $\varepsilon > 0$. We need to show $|\mathcal{H}(z_n, t_n)| < \varepsilon$ for all but finitely many $n \in \mathscr{B}$. Let $\mathscr{B}'_\varepsilon := \{n \in \mathscr{B} : 1 - t_n < \varepsilon\}$. Therefore $|\mathcal{H}(z_n, t_n)| = (1 - t_n)|\varphi(z_n/(1-t_n))| \le (1 - t_n) < \varepsilon$ for all $n \in \mathscr{B}'_\varepsilon$. Also, if $\mathscr{B} \setminus \mathscr{B}'_\varepsilon$ is infinite, then $\{z_n/(1-t_n) : n \in \mathscr{B} \setminus \mathscr{B}'_\varepsilon\} \to 0$, which implies $\{\varphi(z_n/(1-t_n)) : n \in \mathscr{B} \setminus \mathscr{B}'_\varepsilon\} \to 0$ (as φ is proper), and thus $|\mathcal{H}(z_n, t_n)| \le |\varphi(z_n/(1-t_n))| < \varepsilon$ for all but finitely many $n \in \mathscr{B} \setminus \mathscr{B}'_\varepsilon$. The previous two lines together imply that $|\mathcal{H}(z_n, t_n)| < \varepsilon$ for all but finitely many $n \in \mathscr{B}$. $\qquad\square$

Remark 3.6.2.5 Theorem 3.6.2.4 is obtained from a straightforward modification of the Alexander trick [17, Lemma 2.1].

Theorem 3.6.2.6 *Let $f : \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces. Suppose $\pi_1(\Sigma)$ is a finitely generated nonabelian group (equivalently $\Sigma \cong S_{g,0,p}$ for some $(g,p) \ne (0,1), (0,2)$). Then $\deg(f) = \pm 1$.*

Proof Since $\deg(f)$ remains invariant after any proper homotopy of f, we can properly homotope f as we want. The fact that Σ is a finite-type noncompact surface implies each end of Σ is an isolated planar end, that is, for every $e = (V_1, V_2, \dots) \in \text{Ends}(\Sigma)$, we have an integer i_e such that $\overline{V_i}$ is homeomorphic to the punctured disk for each $i \ge i_e$. Next, since f is proper there exists $\mathcal{E} = (\mathcal{W}_1, \mathcal{W}_2, \dots) \in \text{Ends}(\Sigma)$ such that $f^{-1}(\mathcal{W}_i) \ne \varnothing$ for each $i \ge 1$. Notice that $\overline{\mathcal{W}}_{i_\varepsilon}$ is an essential punctured disk and $\mathcal{C}_{i_\varepsilon} := \partial \overline{\mathcal{W}}_{i_\varepsilon}$ is a smoothly embedded separating circle on Σ. Also, $\mathcal{C}_{i_\varepsilon}$ is a primitive circle on Σ as $\mathcal{C}_{i_\varepsilon}$ bounds the punctured disk $\overline{\mathcal{W}}_{i_\varepsilon}$ on $\Sigma \not\cong \mathbb{R}^2$.

We aim to use Lemma 3.6.2.3, but some observations are needed before that. Let $g : \Sigma' \to \Sigma$ be a proper map such that g is properly homotopic to f (note that f is properly homotopic to itself, ie g can be f). If possible, assume $g^{-1}(\mathcal{C}_{i_\varepsilon}) = \varnothing$. Then continuity of g implies $g(\Sigma')$ is contained in one of the two components of $\Sigma \setminus \mathcal{C}_{i_\varepsilon}$. By Theorem 3.6.2.1, $g(\Sigma')$ must be contained in $\mathcal{W}_{i_\varepsilon}$. But then $\pi_1(f) = \pi_1(g)$ is nonsurjective as $\pi_1(\Sigma \setminus \mathcal{W}_{i_\varepsilon}) = \pi_1(\Sigma)$ is nonabelian. Therefore $g^{-1}(\mathcal{C}_{i_\varepsilon}) \ne \varnothing$. Also, by Theorem 3.6.2.1, $g^{-1}(\mathcal{W}_i) \ne \varnothing$ for each $i \ge 1$, and thus $g^{-1}(\mathcal{W}_{i_\varepsilon}) \ne \varnothing$.

We are ready to apply Lemma 3.6.2.3 after the observation given in the previous paragraph. At first, notice that Σ' is homeomorphic to neither the plane nor the punctured plane as $\pi_1(\Sigma') = \pi_1(\Sigma)$ is nonabelian. After a proper homotopy of f, we may assume that $f \pitchfork \mathcal{C}_{i_\varepsilon}$; see Theorem 3.2.1. By the previous paragraph, $f^{-1}(\mathcal{C}_{i_\varepsilon})$ is a pairwise-disjoint nonempty collection of finitely many smoothly embedded circles on Σ'. By Theorem 3.5.3 and the previous paragraph, after a proper homotopy of f, we may further assume that $\mathcal{C}'_{i_\varepsilon} := f^{-1}(\mathcal{C}_{i_\varepsilon})$ is a (single) smoothly embedded circle on Σ' and $f|\mathcal{C}'_{i_\varepsilon} \to \mathcal{C}_{i_\varepsilon}$ is a homeomorphism. The previous paragraph also tells that after all these proper homotopies, $f^{-1}(\mathcal{W}_{i_\varepsilon})$ remains nonempty.

We show that $\mathcal{C}'_{i_\varepsilon}$ is a primitive circle on Σ'. On the contrary, let there be an embedding $i' : \mathbb{D}^2 \hookrightarrow \Sigma'$ with $\mathcal{C}'_{i_\varepsilon} = i'(\mathbb{S}^1)$. Then the embedding $f \circ i'|\mathbb{S}^1 \hookrightarrow \Sigma$ is nullhomotopic and $\mathcal{C}_{i_\varepsilon} = f \circ i'(\mathbb{S}^1)$. But $\mathcal{C}_{i_\varepsilon}$ is a primitive circle on Σ. Now, Theorem 2.2.2 tells us we have reached a contradiction. Finally, applying Lemma 3.6.2.3, we can say that $\deg(f) = \pm 1$. $\qquad\square$

3.6.3 Most pseudoproper homotopy equivalences between noncompact surfaces are of degree ± 1

Theorem 3.6.3.1 *Let $f : \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces. If $\Sigma \ncong \mathbb{S}^1 \times \mathbb{R}, \mathbb{R}^2$ (equivalently $\Sigma' \ncong \mathbb{S}^1 \times \mathbb{R}, \mathbb{R}^2$), then $\deg(f) = \pm 1$.*

Proof Combining Theorems 3.6.1.11 and 3.6.2.6, we complete the proof. $\qquad\square$

The following proposition, *which we don't need to use anywhere*, says that if either of the integers 1 or -1 appears as the degree of a pseudoproper homotopy equivalence between two noncompact oriented surfaces, then the other also appears.

Proposition 3.6.3.2 *Let $f : \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces. Then there exists another pseudoproper homotopy equivalence $\bar{f} : \Sigma' \to \Sigma$ such that $\deg(\bar{f}) = -\deg(f)$.*

Proof Write Σ as the double of a bordered surface S; see Theorem 2.4.2. Define a homeomorphism $\varphi : \Sigma \to \Sigma$ by sending $[p, t] \in \Sigma$ to $[p, 1 - t] \in \Sigma$ for all $(p, t) \in S \times \{0, 1\}$. Then φ is an orientation-reversing homeomorphism. Therefore the degree of $\bar{f} := \varphi \circ f$ is $-\deg(f)$ as the degree is multiplicative; see Section 2.6. $\qquad\square$

3.6.4 An application of the nonvanishing degree of a pseudoproper homotopy equivalence Consider a nonsurjective map $\varphi : \mathcal{M} \to \mathcal{N}$ between two closed oriented connected n–manifolds. Then for any $p \in \mathcal{N} \setminus \mathrm{im}(\varphi)$, the map $H^n(\varphi)$ factors through the inclusion-induced zero map $H^n(\mathcal{N}) \cong \mathbb{Z} \to 0 \cong H^n(\mathcal{N} \setminus p)$ (recall that the top integral singular cohomology of any connected noncompact boundaryless manifold is zero), ie $\deg(\varphi) = 0$. The lemma below generalizes this phenomenon in the proper category:

Lemma 3.6.4.1 *Let $\Phi : M \to N$ be a proper map between two connected oriented boundaryless smooth k–dimensional manifolds. If $\deg(\Phi) \neq 0$, then Φ is surjective.*

Proof Being a proper map between two manifolds, Φ is a closed map; see [30]. Now, if possible, let Φ be nonsurjective. Therefore $N \setminus \Phi(M)$ is a nonempty open subset of N. Pick a point $y \in N \setminus \Phi(M)$. Since N is locally Euclidean, there is a smoothly embedded closed ball $B \subset N$ such that $B \subseteq N \setminus \Phi(M)$. Notice that $N \setminus \mathrm{int}(B)$ is a smoothly embedded codimension-zero submanifold of N with $\partial(N \setminus \mathrm{int}(B)) = \partial B$. By Poincaré duality (see [25, Exercise 35 of Section 3.3]), $H_c^k(N \setminus \mathrm{int}(B); \mathbb{Z}) \cong H_0(N \setminus \mathrm{int}(B), \partial B; \mathbb{Z})$. Also, $H_0(N \setminus \mathrm{int}(B), \partial B; \mathbb{Z}) = 0$ as N is path connected; see [25, Exercise 16(a) of Section 2.1]. Now, $\Phi \colon M \to N$ can be thought as the composition $M \xrightarrow{\Phi^\dagger} N \setminus \mathrm{int}(B) \xhookrightarrow{i} N$, where i is the inclusion map and $\Phi^\dagger(m) := \Phi(m)$ for all $m \in M$. Certainly, $\Phi^\dagger$ and i are both proper maps. Therefore $H_c^k(\Phi)$ is the composition

$$H_c^k(N; \mathbb{Z}) \xrightarrow{H_c^k(i)} H_c^k(N \setminus \mathrm{int}(B); \mathbb{Z}) = 0 \xrightarrow{H_c^k(\Phi^\dagger)} H_c^k(M; \mathbb{Z}),$$

ie $H_c^k(\Phi) = 0$, which contradicts $\deg(\Phi) \neq 0$. Thus Φ must be a surjective map. $\square$

The above lemma, together with Theorem 3.6.3.1, gives the following corollary:

Corollary 3.6.4.2 *A pseudoproper homotopy equivalence between two noncompact surfaces is a surjective map, provided the surfaces are homeomorphic to neither the plane nor the punctured plane.*

The following lemma tells that one way to achieve the surjectivity throughout a proper homotopy is to assume that the initial map of this proper homotopy is a map of nonzero degree. Note that any proper map $f \colon X \to Y$ is properly homotopic to itself due to the proper homotopy $X \times [0,1] \ni (x,t) \mapsto f(x) \in Y$.

Lemma 3.6.4.3 *Let $\Phi \colon M \to N$ be a proper map of nonzero degree between two connected oriented boundaryless smooth k–dimensional manifolds, and let $\Psi \colon M \to N$ be a proper map such that Ψ is properly homotopic to Φ. Then Ψ is a surjective map.*

Proof Since Ψ is properly homotopic to Φ, $\deg(\Psi) = \deg(\Phi) \neq 0$; see Section 2.6. Now, to conclude, consider Lemma 3.6.4.1. $\square$

Here is the main application of the nonvanishing degree of a pseudoproper homotopy equivalence:

Theorem 3.6.4.4 *Let $f \colon \Sigma' \to \Sigma$ be a smooth pseudoproper homotopy equivalence between two noncompact surfaces, where $\mathbb{S}^1 \times \mathbb{R} \ncong \Sigma \ncong \mathbb{R}^2$, and let $\mathscr{A}$ be a preferred LFCS on Σ such that $f \pitchfork \mathscr{A}$. Suppose any two distinct components of $\mathscr{A}$ don't cobound an annulus in Σ. In that case, f can be properly homotoped to a proper map g such that for each component $\mathcal{C}$ of $\mathscr{A}$, $g^{-1}(\mathcal{C})$ is a component of $f^{-1}(\mathscr{A})$ that is mapped homeomorphically onto $\mathcal{C}$ by g.*

Proof Theorem 3.5.3 gives a proper map $g \colon \Sigma' \to \Sigma$ such that g is properly homotopic to f, and for each component $\mathcal{C}$ of $\mathscr{A}$, if $g^{-1}(\mathcal{C}) \neq \varnothing$, then $g^{-1}(\mathcal{C})$ is a component of $f^{-1}(\mathscr{A})$ such that $g|g^{-1}(\mathcal{C}) \to \mathcal{C}$ is a homeomorphism. But $\deg(f) = \pm 1$ by Theorem 3.6.3.1. Thus the map g is surjective since it is properly homotopic to the nonzero degree map f; see Lemma 3.6.4.3. So, for each component $\mathcal{C}$ of $\mathscr{A}$, $g^{-1}(\mathcal{C})$ is a component of $f^{-1}(\mathscr{A})$ such that $g|g^{-1}(\mathcal{C}) \to \mathcal{C}$ is a homeomorphism. $\square$

Remark 3.6.4.5 For closed surfaces, the analog of Theorem 3.6.4.4 can be stated far before, in the "annulus removal" section, as every homotopy equivalence between two closed manifolds has a homotopy inverse, so is a map of degree ± 1, and hence is surjective. But, before Section 3.6, we didn't know the degree of a pseudoproper homotopy equivalence; even at this stage, we don't know whether a pseudoproper homotopy equivalence has a proper homotopy inverse or not.

4 Finishing the proofs of Theorems 1, 2 and 3

Proof of Theorem 1 Consider an LFCS $\mathscr{C}$ on Σ provided by Theorem 3.1.5. Using Theorem 3.2.1, assume f is smooth as well as $f \pitchfork \mathscr{C}$. Thus $f^{-1}(\mathscr{C})$ is a nonempty LFCS on Σ'; see Corollary 3.6.4.2 and Theorem 3.2.3. By Theorem 3.6.4.4, f can be properly homotoped to a proper map g such that for each component $\mathcal{C}$ of $\mathscr{C}$, $g^{-1}(\mathcal{C})$ is a component of $f^{-1}(\mathscr{C})$ that is mapped homeomorphically onto $\mathcal{C}$ by g. Thus $g^{-1}(\mathscr{C})$ decomposes Σ' into bordered subsurfaces and each component of $\Sigma \setminus \mathscr{C}$ has nonempty preimage; see Corollary 3.6.4.2. Let $S \subset \Sigma$ be a bordered subsurface obtained as a complementary component of the decomposition of Σ by $\mathscr{C}$. Now, $S \cong g^{-1}(S)$ by Lemmas 3.6.1.10 (see its proof also) and 3.6.2.3. Since g sends $\mathrm{int}(g^{-1}(S))$ onto $\mathrm{int}(S)$ and $\partial g^{-1}(S)$ homeomorphically onto ∂S, we can properly homotope $g|g^{-1}(S) \to S$ relative to $\partial g^{-1}(S)$ to a homeomorphism $g^{-1}(S) \to S$; see Theorems 3.6.1.9 and 3.6.2.4. Finally, vary S over different complementary components of Σ decomposed by $\mathscr{C}$ to collect these boundary-relative proper homotopies and then paste them to get a proper homotopy from g to a homeomorphism $\Sigma' \to \Sigma$. Since g is properly homotopic to f, we are done. □

The proof of Theorem 1 shows that we are using the nonzero degree assumption of the pseudoproper homotopy equivalence (which is given by Theorem 3.6.3.1) to ensure surjectivity after each proper homotopy. Thus, by a similar argument, we can prove Theorem 4.1.

Theorem 4.1 *Let $f \colon \Sigma' \to \Sigma$ be a pseudoproper homotopy equivalence between two noncompact oriented surfaces. Suppose Σ is not homeomorphic to $\mathbb{R}^2$ and $\deg(f) \neq 0$. Then Σ' is homeomorphic to Σ and f is properly homotopic to a homeomorphism.*

Theorem 4.2 *Let $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a proper map of degree ± 1. Then f is properly homotopic to a homeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$.*

Proof By Theorem 2.6.2, f can be properly homotoped to get smoothly embedded closed disks $D, D' \subseteq \mathbb{R}^2$ such that $D' = f^{-1}(D)$ and $f|D' \to D$ is a homeomorphism. Using the Jordan–Schönflies theorem, $f|\overline{\mathbb{R}^2 \setminus D'} \to \overline{\mathbb{R}^2 \setminus D}$ resembles a map between two punctured disks, on which we apply Theorem 3.6.2.4. □

Proof of Theorem 3 Since $\deg(f) = \pm 1$, by Theorem 2.6.3, $\pi_1(f)$ is surjective. Thus $\pi_1(f)$ is bijective. Both Σ' and Σ are homotopy equivalent to $\bigvee_{\mathscr{I}} \mathbb{S}^1$ for some index set $\mathscr{I}$ with $|\mathscr{I}| \leq \aleph_0$, ie $\pi_k(\Sigma') = 0 = \pi_k(\Sigma)$ for all $k \geq 2$. So, by the Whitehead theorem, f is a homotopy equivalence (note that each surface

has a CW–complex structure due to its C^∞–smooth structure). Now, a simply connected noncompact surface is homeomorphic to $\mathbb{R}^2$; see Theorem 3.1.9. So, combining Theorems 4.1 and 4.2, we are done. □

Proof of Theorem 2   A proper homotopy equivalence is a π_1–injective map of degree ± 1. Now apply Theorem 3. □

The following proposition is an application of Theorem 1:

Proposition 4.3   *Let Σ be a noncompact surface such that $\mathbb{S}^1 \times \mathbb{R} \not\cong \Sigma \not\cong \mathbb{R}^2$. Suppose $f, g : \Sigma \to \Sigma$ are two pseudoproper homotopy equivalences. If f is homotopic to g, then f is properly homotopic to g.*

Proof   By applying Theorem 1 up to proper homotopy, we may assume both f and g are homeomorphisms without loss of generality. Since $\mathbb{S}^1 \times \mathbb{R} \not\cong \Sigma \not\cong \mathbb{R}^2$ and $f^{-1}g$ is homotopic to Id_Σ, by [15, Theorem 6.4], there exists a level-preserving homeomorphism $\mathcal{H} : \Sigma \times [0, 1] \to \Sigma \times [0, 1]$ which agrees with $f^{-1}g$ on $\Sigma \times 0$ and with Id_Σ on $\Sigma \times 1$. The fact that the projection $\Sigma \times [0, 1] \to \Sigma$ is proper implies $f^{-1}g$ is properly homotopic to Id_Σ, so we are done. □

Appendix

A.1   Approximation and transversality in the proper category

Throughout Section A.1, M and N will denote two smooth boundaryless manifolds, possibly noncompact. Let $F : N \to M$ be a smooth map, and let X be a smoothly embedded boundaryless submanifold of M. We say F *is transverse to* X, and write $F \pitchfork X$, if for every $p \in F^{-1}(X)$ we have $T_{F(p)}X + dF_p(T_pN) = T_{F(p)}M$. If F is transverse to X and $F(N) \cap X \neq \varnothing$, then $F^{-1}(X)$ is a smoothly embedded boundaryless submanifold of N such that $\dim(N) - \dim(F^{-1}(X)) = \dim(M) - \dim(X)$; see [28, Theorem 6.30(a)].

The Whitney approximation theorem [28, Theorem 6.26] says that any continuous map $N \to M$ is homotopic to a smooth map. The transversality homotopy theorem [28, Theorem 6.36] says that for any smooth map $F : N \to M$ and for any smoothly embedded boundaryless submanifold X of M, the smooth map F can be homotoped to another smooth map $\widetilde{F} : N \to M$ such that $\widetilde{F} \pitchfork X$. We modify these two theorems in the proper category. Our interest is in the properness of homotopies; the extra stuff not related to properness is in [28, Theorems 6.26 and 6.36].

Theorem A.1.1   (proper Whitney approximation theorem)   *Let $f : N \to M$ be a continuous proper map. Then f is properly homotopic to a smooth proper map.*

Theorem A.1.2   (proper transversality homotopy theorem)   *Let $f : N \to M$ be a smooth proper map, and let X be a smoothly embedded boundaryless submanifold of M. Then f is properly homotopic to a smooth proper map $g : N \to M$ which is transverse to X.*

We start by summarizing key facts in and around the tubular neighborhood theorem. Let $M \hookrightarrow \mathbb{R}^l$ be a smooth proper embedding; see [28, Theorems 6.15]. For each $x \in M$, define the *normal space* $\mathcal{N}_x M$ to M at x as $\mathcal{N}_x M := \{v \in \mathbb{R}^l : v \perp T_x M\}$. Then $\mathcal{N}M := \{(x, v) \in \mathbb{R}^l \times \mathbb{R}^l : x \in M, v \perp T_x M\}$ is a smoothly embedded l–dimensional submanifold of $\mathbb{R}^l \times \mathbb{R}^l$ and $\pi \colon \mathcal{N}M \ni (x, v) \mapsto x \in M$ is vector bundle of rank $l - \dim(M)$, called the *normal bundle* of M in $\mathbb{R}^l$; see [28, Corollary 10.36].

Consider the smooth map $E \colon \mathcal{N}M \ni (x, v) \mapsto x + v \in \mathbb{R}^l$. One can show that $dE_{(x,0)}$ is bijective for each $x \in M$. Thus, for each $x \in M$, we have $\delta > 0$ such that E maps

$$V_\delta(x) := \{(x', v') \in \mathcal{N}M : |x - x'| < \delta, |v'| < \delta\}$$

diffeomorphically onto an open neighborhood of x in $\mathbb{R}^l$. Now, the map $\rho \colon M \to (0, 1]$ defined by

$$\rho(x) := \sup\{\delta \le 1 : E \text{ maps } V_\delta(x) \text{ diffeomorphically onto an open neighborhood of } x \text{ in } \mathbb{R}^l\}$$

is continuous. Further, $V := \{(x, v) \in \mathcal{N}M : |v| < \frac{1}{2}\rho(x)\}$ is an open subset of NM and E maps V diffeomorphically onto an open subset U of $\mathbb{R}^l$ with $M \subseteq U$, ie U is a tubular neighborhood of M in $\mathbb{R}^l$; see [28, Theorem 6.24]. Note that the map $r \colon U \to M$ defined by $r := \pi \circ (E|V \to U)^{-1}$ is a retraction and submersion; see [28, Proposition 6.25]. Denote $\{y \in \mathbb{R}^l : |y - x| < \varepsilon\}$ by $B_\varepsilon(x)$. By an argument similar to showing the continuity of ρ, one can prove that $\delta \colon M \to (0, 1]$, defined by $\delta(x) := \sup\{\varepsilon \le 1 : B_\varepsilon(x) \subseteq U\}$ for any $x \in M$, is also continuous.

With this setup, we are now ready to state a crucial lemma, which in particular says that if two points are at most unit distance apart, then the distance between their images under the tubular neighborhood retraction can be at most 2.

Lemma A.1.3 *Let $\varepsilon > 0$. If $y, y' \in U$ with $|y - y'| < \varepsilon$, then $|r(y) - r(y')| \le \varepsilon + 1$.*

Proof Notice $|r(y) - r(y')| - |y - y'| \le |y - r(y)| + |y' - r(y')| \le \frac{1}{2}\rho \circ r(y) + \frac{1}{2}\rho \circ r(y')$. $\qquad\square$

Consider another smooth proper embedding $N \hookrightarrow \mathbb{R}^k$ for the proof of the following three facts. The following lemma says that a homotopy lying in a λ–neighborhood (where λ is a fixed positive number) of a proper map is a proper homotopy:

Lemma A.1.4 *Let $h \colon N \to M$ be a continuous proper map, and let $\mathcal{H} \colon N \times [0, 1] \to M$ be a homotopy. If there exists a constant λ such that $|\mathcal{H}(p, t) - h(p)| \le \lambda$ for each $(p, t) \in N \times [0, 1]$, then $\mathcal{H}$ is proper.*

Proof Note that the embeddings $M \hookrightarrow \mathbb{R}^l$ and $N \hookrightarrow \mathbb{R}^k$ are closed maps as they are proper maps; see [30]. Consider the induced metric d_M on M inherited from $\mathbb{R}^l$, ie $d_M(m, m') = |m - m'|$ for all $m, m' \in M$. Also, we have the induced metric $d_{N \times [0,1]}$ on $N \times [0, 1]$ inherited from $\mathbb{R}^k \times [0, 1]$, ie $d_{N \times [0,1]}((n, t), (n', t')) = |n - n'| + |t - t'|$ for all $(n, t), (n', t') \in N \times [0, 1]$. Thus a subset of $N \times [0, 1]$ (resp. M) is compact if and only if it is closed and bounded in $N \times [0, 1]$ (resp. M).

Let C be a compact subset of M. Continuity of $\mathcal{H}$ implies $\mathcal{H}^{-1}(C)$ is closed in $N \times [0, 1]$. Also, if there were an unbounded sequence $\{(n_i, t_i)\} \subseteq \mathcal{H}^{-1}(C)$, then $\{n_i\}$, and hence $\{h(n_i)\}$, would be unbounded (as h is proper); thus the unbounded set $\{h(n_i)\}$ would be inside the λ–neighborhood of the bounded set C, a contradiction. Therefore $\mathcal{H}^{-1}(C)$ is closed and bounded in $N \times [0, 1]$, and hence $\mathcal{H}^{-1}(C)$ is compact. Since C is an arbitrary compact subset of M, we are done. $\qquad \square$

Now we are ready to prove the analogs of the Whitney approximation theorem and transversality homotopy theorem in the proper category:

Proof of Theorem A.1.1 The Whitney approximation theorem gives a smooth function $\tilde{f} \colon N \to \mathbb{R}^l$ such that $|\tilde{f}(y) - f(y)| < \delta(f(y))$ for each $y \in N$; see [28, Theorem 6.21]. Now define $\mathcal{H} \colon N \times [0, 1] \to M$ as $\mathcal{H}(p, t) := r((1-t)f(p) + t\tilde{f}(p))$ for all $(p, t) \in N \times [0, 1]$. If $(p, t) \in N \times [0, 1]$, then

$$|(1-t)f(p) + t\tilde{f}(p) - f(p)| \leq t|\tilde{f}(p) - f(p)| \leq 1.$$

Therefore, for all $(p, t) \in N \times [0, 1]$, we have that $|\mathcal{H}(p, t) - f(p)| = |\mathcal{H}(p, t) - r \circ f(p)| \leq 2$ by Lemma A.1.3. Now Lemma A.1.4 tells us that $\mathcal{H}$ is proper. Therefore $\mathcal{H}(-, 1) = r \circ \tilde{f}$ is a smooth proper map that is properly homotopic to f (recall that r is a smooth retraction). $\qquad \square$

Proof of Theorem A.1.2 The Whitney approximation theorem gives a smooth function $e \colon N \to (0, \infty)$ with $0 < e < \delta \circ f$; see [28, Corollary 6.22]. Let $\mathbb{B}^l := \{s \in \mathbb{R}^l : |s| < 1\}$. Define $F \colon N \times \mathbb{B}^l \to M$ as $F(p, s) := r(f(p) + e(p)s)$ for any $(p, s) \in N \times \mathbb{B}^l$. If $p \in N$, the restriction of F to $\{p\} \times \mathbb{B}^l$ is the composition of the local diffeomorphism $s \mapsto f(p) + e(p)s$ with the smooth submersion r, so F is a smooth submersion and hence transverse to X.

By the parametric transversality theorem [28, Theorem 6.35], $F(-, s_0)$ is transverse to X for some $s_0 \in \mathbb{B}^l$. Now define $\mathcal{H} \colon N \times [0, 1] \to M$ as $\mathcal{H}(p, t) := r(f(p) + te(p)s_0)$ for all $(p, t) \in N \times [0, 1]$. If $(p, t) \in N \times [0, 1]$, then

$$|(f(p) + te(p)s_0) - f(p)| \leq te(p)|s_0| < \delta(f(p)) \leq 1.$$

Therefore $|\mathcal{H}(p, t) - f(p)| = |\mathcal{H}(p, t) - r \circ f(p)| \leq 2$ for all $(p, t) \in N \times [0, 1]$ by Lemma A.1.3. Lemma A.1.4 tells that $\mathcal{H}$ is proper. Define $g := \mathcal{H}(-, 1)$, ie $g = r(f(-) + e(-)s_0) = F(-, s_0)$ is properly homotopic to f (recall that r is a smooth retraction) as well as transverse to X. $\qquad \square$

A.2 Transversality of a proper map between two surfaces with respect to a circle

Here is some notation that will be used throughout Section A.2. Let $f \colon \Sigma' \to \Sigma$ be a smooth *proper* map between two surfaces, and let $\mathcal{C}$ be a smoothly embedded circle on Σ such that f is transverse to $\mathcal{C}$. Also, let $\varphi \colon \mathcal{C} \times [-1, 1] \hookrightarrow \Sigma$ be a smooth embedding with $\varphi(\mathcal{C}, 0) = \mathcal{C}$, that is, $\mathrm{im}(\varphi)$ is a *two-sided (trivial) tubular neighborhood* of $\mathcal{C}$. We call each of $\varphi(\mathcal{C} \times [-1, 0])$ and $\varphi(\mathcal{C} \times [0, 1])$ a *one-sided tubular neighborhood of $\mathcal{C}$* (in short, *a side of $\mathcal{C}$*). By scaling, we may replace $[-1, 0]$ and $[0, 1]$ with other closed intervals.

The following theorem says that f is transverse to all circles which are parallel to and sufficiently near $\mathcal{C}$.

Theorem A.2.1 *There exists $\varepsilon_0 \in (0, 1)$ such that f is transverse to $\mathcal{C}_\varepsilon := \varphi(\mathcal{C}, \varepsilon)$ for each $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$. Thus for any $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$, $f^{-1}(\mathcal{C}_\varepsilon)$ is either empty or a pairwise-disjoint collection of finitely many smoothly embedded circles on Σ'.*

First we need a lemma:

Lemma A.2.2 *Let $g: \mathbb{R}^2 \to \mathbb{R}^2$ be a smooth map and $x_n \to x$ in $\mathbb{R}^2$ with $r_n := |g(x_n)| \to 1$. Write $S_r := \{z \in \mathbb{R}^2 : |z| = r\}$ and assume $\operatorname{im}(dg_{x_n}) = T_{g(x_n)}(S_{r_n})$ for all n. If $dg_x \neq 0$, then $\operatorname{im}(dg_x) = T_{g(x)}(S_1)$.*

Proof The derivative map $dg: \mathbb{R}^2 \to L(\mathbb{R}^2, \mathbb{R}^2)$ is continuous so $dg_{x_n} \to dg_x$, and this convergence can be thought as convergence of (2×2)–matrices. In particular, if $\boldsymbol{i}, \boldsymbol{j} \in \mathbb{R}^2$ are two perpendicular unit vectors, then $dg_{x_n}(\boldsymbol{i}) \to dg_x(\boldsymbol{i})$ and $dg_{x_n}(\boldsymbol{j}) \to dg_x(\boldsymbol{j})$.

Recall that the tangent space at any point of a circle is the vector space of all points perpendicular to this point. So $\langle dg_{x_n}(\boldsymbol{i}), g(x_n) \rangle = 0 = \langle dg_{x_n}(\boldsymbol{j}), g(x_n) \rangle$ by hypothesis, and now $\langle dg_x(\boldsymbol{i}), g(x) \rangle = 0 = \langle dg_x(\boldsymbol{j}), g(x) \rangle$ by the convergence of the inner product. Hence $\operatorname{im}(dg_x) \subseteq T_{g(x)}(S_1)$. Since $dg_x \neq 0$ and $\dim T_{g(x)}(S_1) = 1$, we are done. $\qquad\square$

Proof of Theorem A.2.1 Suppose not. So, a sequence $\varepsilon_n \to 0$ and points $x_n \in f^{-1}(\mathcal{C}_{\varepsilon_n})$ exist such that $\operatorname{im}(df_{x_n}) + T_{f(x_n)}\mathcal{C}_{\varepsilon_n} \subsetneq T_{f(x_n)}\Sigma$ for all n. Hence $\operatorname{im}(df_{x_n}) \subseteq T_{f(x_n)}\mathcal{C}_{\varepsilon_n}$ as $T_{f(x_n)}\mathcal{C}_{\varepsilon_n} \oplus \mathcal{N}_{f(x_n)}\mathcal{C}_{\varepsilon_n} = T_{f(x_n)}\Sigma$ for all n. Now, $\{x_n\}$ is contained in the compact set $f^{-1}(\operatorname{im}(\varphi))$ (recall that f is a proper map), ie passing to subsequence, if needed, assume $x_n \to x \in f^{-1}(\mathcal{C})$.

The continuity of the derivative map says $df_{x_n} \to df_x$. After discarding the first few terms, we may assume $df_{x_n} \neq 0$ for all n (otherwise $df_x = 0$, which would mean $T_{f(x)}\mathcal{C} + \operatorname{im}(df_x) = T_{f(x)}\mathcal{C}$ wouldn't be equal to $T_{f(x)}\Sigma$ and so f wouldn't be transverse to $\mathcal{C}$). So $\operatorname{im}(df_{x_n}) = T_{f(x_n)}(\mathcal{C}_{\varepsilon_n})$ for all n (a nonzero vector subspace of a 1–dimensional vector space is equal to the whole space).

Now, restricting f to a coordinate ball containing x and then postcomposing with φ^{-1}, we can consider Lemma A.2.2, which gives $\operatorname{im}(df_x) = T_{f(x)}(\mathcal{C})$, a contradiction to the assumption $f \pitchfork \mathcal{C}$. $\qquad\square$

The previous theorem guarantees transversality near $\mathcal{C}$. In the rest of Section A.2, we aim to prove that every small one-sided tubular neighborhood of a component of $f^{-1}(\mathcal{C})$ maps into a small one-sided tubular neighborhood of $\mathcal{C}$.

We first fix some notation. Let $\mathcal{C}'$ be a component of $f^{-1}(\mathcal{C})$. Also, consider an $\varepsilon_0 \in (0, 1)$ such that $f \pitchfork \mathcal{C}_\varepsilon$ for every $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$; see Theorem A.2.1.

Theorem A.2.3 *Let $\varepsilon \in (0, \varepsilon_0]$, and let $\mathcal{T}'$ be a two-sided compact tubular neighborhood of $\mathcal{C}'$ in Σ'. Then there exist two one-sided compact tubular neighborhoods $\mathcal{U}'_l$ and $\mathcal{U}'_r$ of $\mathcal{C}'$ in Σ' such that $\mathcal{U}'_l \cup \mathcal{U}'_r$ is a two-sided tubular neighborhood of $\mathcal{C}'$ with $\mathcal{U}'_l \cup \mathcal{U}'_r \subseteq \mathcal{T}'$, and for each $s \in \{l, r\}$ the following hold: $f^{-1}(\mathcal{C}) \cap \mathcal{U}'_s = \mathcal{C}'$, and either $f(\mathcal{U}'_s) \subseteq \varphi(\mathcal{C} \times [0, \varepsilon])$ or $f(\mathcal{U}'_s) \subseteq \varphi(\mathcal{C} \times [-\varepsilon, 0])$.*

Proof By Theorem A.2.1, $f^{-1}(\mathcal{C}_{-\varepsilon}) \cup f^{-1}(\mathcal{C}) \cup f^{-1}(\mathcal{C}_\varepsilon)$ is a pairwise-disjoint collection of finitely many smoothly embedded circles on Σ'. Now, consider two one-sided compact tubular neighborhoods $\mathcal{U}'_l$ and $\mathcal{U}'_r$ of $\mathcal{C}'$ in Σ' such that $\mathcal{U}'_l \cup \mathcal{U}'_r$ is a two-sided tubular neighborhood of $\mathcal{C}'$ with $\mathcal{U}'_l \cup \mathcal{U}'_r \subseteq \mathcal{T}'$, and for each $s \in \{l, r\}$ the following hold: $f^{-1}(\mathcal{C}) \cap \mathcal{U}'_s = \mathcal{C}'$, and $\mathcal{U}'_s \cap f^{-1}(\mathcal{C}_\varepsilon) = \varnothing = \mathcal{U}'_s \cap f^{-1}(\mathcal{C}_{-\varepsilon})$.

Now fix $s \in \{l, r\}$. Since $\mathcal{U}'_s \setminus \mathcal{C}'$ is connected and f is continuous, $f(\mathcal{U}'_s \setminus \mathcal{C}')$ is contained in one of the components of $\Sigma \setminus (\mathcal{C}_{-\varepsilon} \cup \mathcal{C} \cup \mathcal{C}_\varepsilon)$. But $f(\mathcal{C}') \subseteq \mathcal{C}$ implies either $f(\mathcal{U}'_s) \subseteq \varphi(\mathcal{C} \times [0, \varepsilon])$ or $f(\mathcal{U}'_s) \subseteq \varphi(\mathcal{C} \times [-\varepsilon, 0])$. □

Remark A.2.4 In Theorem A.2.3, it is possible that $f(\mathcal{U}'_l \cup \mathcal{U}'_r)$ is contained in either $\varphi(\mathcal{C} \times [0, \varepsilon])$ or $\varphi(\mathcal{C} \times [-\varepsilon, 0])$, ie f may map both sides of $\mathcal{C}'$ into one of the two sides of $\mathcal{C}$.

Consider the one-sided compact tubular neighborhoods $\mathcal{U}'_l$ and $\mathcal{U}'_r$ of $\mathcal{C}'$ in Σ' given by Theorem A.2.3. Notice that for some $s \in \{l, r\}$, it is possible that $f((\partial \mathcal{U}'_s) \setminus \mathcal{C}') \not\subseteq \varphi(\mathcal{C} \times t)$ for any $t \in [-\varepsilon, \varepsilon]$. A remedy for this is given in the following theorem:

Theorem A.2.5 *Let $\varepsilon \in (0, \varepsilon_0]$, and let $\mathcal{U}'$ be a one-sided compact tubular neighborhood of $\mathcal{C}'$ such that $f^{-1}(\mathcal{C}) \cap \mathcal{U}' = \mathcal{C}'$ and $f(\mathcal{U}') \subseteq \varphi(\mathcal{C} \times [0, \varepsilon])$. Then there is a $\delta \in (0, \varepsilon)$ and a component $\mathcal{C}'_\delta$ of $f^{-1}(\mathcal{C}_\delta)$ such that the following hold*:

(1) *$\mathcal{C}'_\delta$ together with $\mathcal{C}'$ cobound an annulus $\mathcal{A}' \subseteq \mathcal{U}'$ such that any other component of $f^{-1}(\mathcal{C}_\delta)$ in $\mathrm{int}(\mathcal{A}')$, if any, bounds a disk inside $\mathcal{A}'$.*

(2) *The map f sends $\mathcal{A}'$ into $\varphi(\mathcal{C} \times [0, \varepsilon])$. Also, after removing the interiors of all disks bounded by components of $f^{-1}(\mathcal{C}_\delta)$ from $\mathcal{A}'$, we can send it to $\varphi(\mathcal{C} \times [0, \delta])$ by f.*

Proof of Theorem A.2.5(1) Choose a $\delta \in (0, \varepsilon)$ such that $\varphi(\mathcal{C} \times [0, \delta]) \cap f((\partial \mathcal{U}') \setminus \mathcal{C}') = \varnothing$. Note that such a δ exists; otherwise, using the compactness of $(\partial \mathcal{U}') \setminus \mathcal{C}'$, we would have a sequence $\{x'_n\} \subseteq (\partial \mathcal{U}') \setminus \mathcal{C}'$ converging to some $x' \in (\partial \mathcal{U}') \setminus \mathcal{C}'$ such that $f(x'_n) \in \varphi(\mathcal{C} \times [0, 1/n])$, ie $f(x')$ would belong to $\mathcal{C}$, a contradiction to the assumption $f^{-1}(\mathcal{C}) \cap \mathcal{U}' = \mathcal{C}'$. Define an open set $\mathcal{W}'$ by

$$\mathcal{W}' := \mathrm{int}(\mathcal{U}') \cap f^{-1}(\varphi(\mathcal{C} \times (0, \delta))).$$

Notice that no sequence in $\mathcal{W}'$ converges to some point of $(\partial \mathcal{U}') \setminus \mathcal{C}'$. Otherwise, if we assume $\mathcal{W}'_n \ni w'_n \to x' \in (\partial \mathcal{U}') \setminus \mathcal{C}'$, then $\varphi(\mathcal{C} \times (0, \delta)) \ni f(w'_n) \to f(x')$. Since $\varphi(\mathcal{C} \times [0, \delta])$ is a closed set containing the sequence $\{f(w'_n)\}$, we can say that $f(x') \in f((\partial \mathcal{U}') \setminus \mathcal{C}') \cap \varphi(\mathcal{C} \times [0, \delta])$, which is impossible by our choice of δ.

So $\overline{\mathcal{W}}' \subseteq \mathcal{U}'$ (as $\mathcal{U}'$ is compact) but $((\partial \mathcal{U}') \setminus \mathcal{C}') \cap \overline{\mathcal{W}}' = \varnothing$. In particular, $\partial \mathcal{W}' \subseteq \mathcal{U}'$ but $((\partial \mathcal{U}') \setminus \mathcal{C}') \cap \partial \mathcal{W}' = \varnothing$.

Claim A.2.5.1 *We have $\partial \mathcal{W}' \subseteq \mathcal{C}' \cup f^{-1}(\mathcal{C}_\delta)$. Thus $\partial \mathcal{W}'$ is contained in a finite union of pairwise-disjoint circles.*

Proof of Claim A.2.5.1 Let $y' \in \partial \mathcal{W}'$ and consider a sequence $\{y'_n\} \subseteq \mathcal{W}'$ converging to y'. Then $\varphi(\mathcal{C} \times (0, \delta)) \ni f(y'_n) \to f(y') \in \varphi(\mathcal{C} \times [0, \delta])$. If $f(y') \in \varphi(\mathcal{C} \times \{0, \delta\}) = \mathcal{C} \cup \mathcal{C}_\delta$, then we are done since $f^{-1}(\mathcal{C}) \cap \mathcal{U}' = \mathcal{C}'$. On the other hand, if $f(y') \in \varphi(\mathcal{C} \times (0, \delta))$, then the definition of $\mathcal{W}'$ and $\mathcal{W}' \cap \partial \mathcal{W}' = \varnothing$ (as $\mathcal{W}'$ is open) together imply $y' \in \mathcal{U}' \setminus \operatorname{int}(\mathcal{U}') = \partial \mathcal{U}'$, ie $y' \in \mathcal{C}'$ as $((\partial \mathcal{U}') \setminus \mathcal{C}') \cap \partial \mathcal{W}' = \varnothing$. Since $y' \in \partial \mathcal{W}'$ is arbitrary, we are done. $\square$

The definition of $\mathcal{W}'$ tells us that each point of $\operatorname{int}(\mathcal{U}')$ that is sufficiently near to $\mathcal{C}'$ must belong to $\mathcal{W}'$. Now, using Claim A.2.5.1, we can say that there is at least one component of $f^{-1}(\mathcal{C}_\delta)$ which cobounds an annulus with $\mathcal{C}'$ inside $\mathcal{U}'$. Of all the $\mathcal{C}'$–parallel components of $f^{-1}(\mathcal{C}_\delta)$, we consider the closest to $\mathcal{C}'$ as $\mathcal{C}'_\delta$. $\square$

Proof of Theorem A.2.5(2) Certainly $f(\mathcal{A}') \subseteq f(\mathcal{U}') \subseteq \varphi(\mathcal{C} \times [0, \varepsilon])$. The rest follows, once we observe that, removing the interiors of all disks bounded by components of $f^{-1}(\mathcal{C}_\delta)$ from $\mathcal{A}'$, $\mathcal{A}'$ remains connected, so by continuity of $f|\Sigma' \setminus f^{-1}(\mathcal{C} \cup \mathcal{C}_\delta) \to \Sigma \setminus (\mathcal{C} \cup \mathcal{C}_\delta)$ it maps into $\varphi(\mathcal{C} \times (0, \delta))$. $\square$

Acknowledgments

References

[1] **Y Algom-Kfir**, **M Bestvina**, *Groups of proper homotopy equivalences of graphs and Nielsen realization* (2021) arXiv 2109.06908 To appear in Contemp. Math., Amer. Math. Soc.

[2] **V Álvarez**, **J M Rodríguez**, *Structure theorems for Riemann and topological surfaces*, J. Lond. Math. Soc. 69 (2004) 153–168 MR Zbl

[3] **R Ayala**, **E Dominguez**, **A Márquez**, **A Quintero**, *Proper homotopy classification of graphs*, Bull. Lond. Math. Soc. 22 (1990) 417–421 MR Zbl

[4] **M Barros**, *Classificação das superfícies não compactas cujo bordo é reunião de curvas de Jordan*, Anais da Faculdade de Ciências 57, Univ. Porto (1974) 33–41 Zbl

[5] **M Brittenham**, π_1*–injective*, *proper maps of open surfaces*, unpublished (1989)

[6] **M Brittenham**, *Essential laminations and deformations of homotopy equivalences: from essential pullback to homeomorphism*, Topology Appl. 60 (1994) 249–265 MR Zbl

[7] **E M Brown**, *Proper homotopy theory in simplicial complexes*, from "Topology Conference" (R F Dickman, Jr, P Fletcher, editors), Lecture Notes in Math. 375, Springer (1974) 41–46 MR Zbl

[8] **E M Brown**, **R Messer**, *The classification of two-dimensional manifolds*, Trans. Amer. Math. Soc. 255 (1979) 377–402 MR Zbl

[9] **E M Brown**, **T W Tucker**, *On proper homotopy theory for noncompact 3–manifolds*, Trans. Amer. Math. Soc. 188 (1974) 105–126 MR Zbl

[10] **S Chang**, **S Weinberger**, *Modular symbols and the topological nonrigidity of arithmetic manifolds*, Comm. Pure Appl. Math. 68 (2015) 2022–2051 MR Zbl

[11] **S Chang**, **S Weinberger**, *A course on surgery theory*, Ann. of Math. Stud. 211, Princeton Univ. Press (2021) MR Zbl

[12] **M Dehn**, *Papers on group theory and topology*, Springer (1987) MR Zbl

[13] **A L Edmonds**, *Deformation of maps to branched coverings in dimension two*, Ann. of Math. 110 (1979) 113–125 MR Zbl

[14] **C H Edwards, Jr**, *Open 3–manifolds which are simply connected at infinity*, Proc. Amer. Math. Soc. 14 (1963) 391–395 MR Zbl

[15] **D B A Epstein**, *Curves on 2–manifolds and isotopies*, Acta Math. 115 (1966) 83–107 MR Zbl

[16] **D B A Epstein**, *The degree of a map*, Proc. Lond. Math. Soc. 16 (1966) 369–383 MR Zbl

[17] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[18] **F T Farrell**, **L E Jones**, *Topological rigidity for compact non-positively curved manifolds*, from "Differential geometry: riemannian geometry" (R Greene, S-T Yau, editors), Proc. Sympos. Pure Math. 54.3, Amer. Math. Soc., Providence, RI (1993) 229–274 MR Zbl

[19] **F T Farrell**, **L R Taylor**, **J B Wagoner**, *The Whitehead theorem in the proper category*, Compos. Math. 27 (1973) 1–23 MR Zbl

[20] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR Zbl

[21] **D Gabai**, **G R Meyerhoff**, **N Thurston**, *Homotopy hyperbolic 3–manifolds are hyperbolic*, Ann. of Math. 157 (2003) 335–431 MR Zbl

[22] **M E Goldman**, *Open surfaces and an algebraic study of ends*, PhD thesis, Yale University (1967) MR Available at https://www.proquest.com/docview/288355577

[23] **M E Goldman**, *An algebraic classification of noncompact 2–manifolds*, Trans. Amer. Math. Soc. 156 (1971) 241–258 MR Zbl

[24] **C R Guilbault**, *Ends, shapes, and boundaries in manifold topology and geometric group theory*, from "Topology and geometric group theory" (M W Davis, J Fowler, J-F Lafont, I J Leary, editors), Springer Proc. Math. Stat. 184, Springer (2016) 45–125 MR Zbl

[25] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[26] **B Hughes**, **A Ranicki**, *Ends of complexes*, Cambridge Tracts in Math. 123, Cambridge Univ. Press (1996) MR Zbl

[27] **B Kerékjártó**, *Vorlesungen über Topologie, I: Flächentopologie*, Grundl. Math. Wissen. 8, Springer (1923) JFM

[28] **J M Lee**, *Introduction to smooth manifolds*, 2nd edition, Graduate Texts in Math. 218, Springer (2013) MR Zbl

[29] **D R McMillan, Jr**, *Some contractible open 3–manifolds*, Trans. Amer. Math. Soc. 102 (1962) 373–382 MR Zbl

[30] **R S Palais**, *When proper maps are closed*, Proc. Amer. Math. Soc. 24 (1970) 835–836 MR Zbl

[31] **T Porter**, *Proper homotopy theory*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 127–167 MR Zbl

[32] **M Reichbach**, *The power of topological types of some classes of 0–dimensional sets*, Proc. Amer. Math. Soc. 13 (1962) 17–23 MR Zbl

[33] **I Richards**, *On the classification of noncompact surfaces*, Trans. Amer. Math. Soc. 106 (1963) 259–269 MR Zbl

[34] **J Rosenberg**, *C*–algebras, positive scalar curvature, and the Novikov conjecture, III*, Topology 25 (1986) 319–336 MR Zbl

[35] **P Scott**, *An introduction to 3–manifolds*, course notes, University of Maryland (1974) Available at https://www.ams.org/open-math-notes/omn-view-listing?listingId=110690

[36] **L C Siebenmann**, *On detecting Euclidean space homotopically among topological manifolds*, Invent. Math. 6 (1968) 245–261 MR Zbl

[37] **E H Spanier**, *Algebraic topology*, McGraw-Hill, New York (1966) MR Zbl

[38] **R J Tondra**, *Homeotopy groups of surfaces whose boundary is the union of 1–spheres*, Fund. Math. 105 (1980) 79–85 MR Zbl

[39] **F Waldhausen**, *On irreducible 3–manifolds which are sufficiently large*, Ann. of Math. 87 (1968) 56–88 MR Zbl

Department of Mathematics, Indian Institute of Science
Bangalore, India

sumantadas@iisc.ac.in

Combinatorial proof of Maslov index formula in Heegaard Floer theory

ROMAN KRUTOWSKI

We prove Lipshitz's Maslov index formula in Heegaard Floer homology via the combinatorics of Heegaard diagrams.

1 Introduction

Peter Ozsváth and Zoltán Szabó [2004b; 2004c] introduced Heegaard Floer homology, a collection of invariants for closed oriented 3–manifolds. Since then, Heegaard Floer homology has emerged as an extremely powerful invariant, producing strong results in low-dimensional topology (see [Ghiggini 2008; Ni 2009a; 2009b; Ozsváth and Szabó 2004a; 2004b; 2004c; 2006]). All of the numerous versions of Heegaard Floer homology involve counting the number of points in the unparametrized moduli space of J–holomorphic disks of a certain Maslov index joining some intersection points of two half-dimensional tori in a certain symmetric product of a surface.

The most notable advantage of Heegaard Floer homology when compared to other types of Floer homology is its combinatorial nature which allows computation of these invariants in various cases. One of the ingredients is the combinatorial formula for the Maslov index $\mu(\varphi)$ of a Whitney disk φ, which is the Fredholm index of some differential operator, or alternatively, a homology class of a certain loop in the Lagrangian Grassmannian. Jacob Rasmussen [2003] gave a formula that relates the intersection number of the disk with the fat diagonal in the symmetric product, with its Maslov index. Later on, Robert Lipshitz [2006], in a paper devoted to the cylindrical reformulation of the whole theory, gave a purely combinatorial formula for the Maslov index of these disks.

The proof of this formula in [Lipshitz 2006] is based on an elegant geometric approach. In this paper, we provide a combinatorial proof of this formula which is inspired by the proof of the index formula for Maslov n–gons due to Sucharit Sarkar [2011].

Let $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ be a Heegaard diagram. Any Whitney disk $\varphi \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$ connecting $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{Sym}^g(\Sigma)$ has a *shadow* $D(\varphi)$ (see Definition 2.1), a certain 2–chain in Σ with boundary satisfying $\partial(\partial D(\varphi) \cap \boldsymbol{\alpha}) = \boldsymbol{y} - \boldsymbol{x}$ and $\partial(\partial D(\varphi) \cap \boldsymbol{\beta}) = \boldsymbol{x} - \boldsymbol{y}$. We denote the set of such 2–chains by $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ and call them *domains*. We denote by $\mathcal{D}$ the set of all domains in all Heegaard diagrams (see Section 2 for details).

In this paper, we assume that $\boldsymbol{\alpha}$ curves intersect $\boldsymbol{\beta}$ curves perpendicularly with respect to some metric on Σ. Among the domains, there are two special types that serve us as building blocks. *Bigons* and *rectangles* are 2–sided and 4–sided domains, respectively, which are homeomorphic to disks with all angles equal to $90°$ and which do not contain any $\boldsymbol{x}$– and $\boldsymbol{y}$–coordinates in the interior.

Given two domains $D \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ and $D' \in \mathcal{D}(\boldsymbol{y}, \boldsymbol{z})$ the composition of these domains $D * D'$ is a domain $D + D' \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{z})$.

In this paper any map $\bar{\mu} \colon \mathcal{D} \to \mathbb{Z}$ will be called an *index*. An index $\bar{\mu}$ is said to be *additive* if for any Heegaard diagram and any two of its domains $D \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ and $D' \in \mathcal{D}(\boldsymbol{y}, \boldsymbol{z})$,

$$\bar{\mu}(D * D') = \bar{\mu}(D) + \bar{\mu}(D').$$

In Section 2 we introduce two types of transformations of any Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ which assign to each of its domain D a new domain D' in the new Heegaard diagram. These transformations are called *finger moves* and *empty stabilizations*. An index $\bar{\mu}$ is said to be *stable* if for any such transformation $\bar{\mu}(D) = \bar{\mu}(D')$.

Define the *combinatorial index* $\widetilde{\mu}$ of a domain $D \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ via the formula due to Lipshitz [2006],

$$(1) \qquad\qquad \widetilde{\mu}(D) := \widetilde{\mu}_{\boldsymbol{x}}(D) + \widetilde{\mu}_{\boldsymbol{y}}(D) + e(D),$$

where $\widetilde{\mu}_{\boldsymbol{x}}(D)$ is a point measure of D at $\boldsymbol{x}$ and $e(D)$ is the Euler measure of D (see Section 2.1).

We are now ready to formulate our main results.

Theorem 1.1 *There exists a unique index $\bar{\mu} \colon \mathcal{D} \to \mathbb{Z}$ satisfying the following axioms:*

(1) *$\bar{\mu}$ is additive;*

(2) *$\bar{\mu}$ is stable;*

(3) *$\bar{\mu}(B) = 1$ for any bigon $B \in \mathcal{D}$;*

(4) *$\bar{\mu}(R) = 1$ for any rectangle $R \in \mathcal{D}$.*

Moreover, this index agrees with the combinatorial index $\widetilde{\mu}$.

Theorem 1.2 *Let $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ be a Heegaard diagram and φ be a Whitney disk connecting two generators of the corresponding Heegaard Floer chain complex. Then*

$$\mu(\varphi) = \widetilde{\mu}(D(\varphi)),$$

where $\mu(\varphi)$ is the Maslov index of φ.

2 Notations and preliminary results

2.1 Heegaard Floer theory preliminaries

To start, recall the definition of a Heegaard diagram. Let Σ be an oriented closed Riemannian surface (ie there is a fixed Riemannian metric) and $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_g\}$ and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_g\}$ be two sets of nonintersecting *oriented*[1] simple closed curves such that both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ generate half-dimensional subspaces of $H_1(\Sigma)$ (in particular this means that g is not smaller than the genus $g(\Sigma)$ of the surface). The last part is equivalent to assuming that the complement of $\boldsymbol{\alpha}$ consists of $g - g(\Sigma) + 1$ components (the same holds for $\boldsymbol{\beta}$). We also assume that the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves intersect perpendicularly. We call by *regions* closures of connected components of $\Sigma \setminus (\boldsymbol{\alpha} \cup \boldsymbol{\beta})$. The collection $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is usually called an *unpointed Heegaard diagram*, but in the text we refer to it simply as a *Heegaard diagram*. We need not assume these diagrams to be pointed, because it is not relevant for the Maslov index calculations.

The Heegaard Floer homology chain complex of a diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is generated by g–tuples of points of the form $\boldsymbol{x} = \{x_1, \ldots, x_g\}$ where $x_i \in \alpha_i \cap \beta_{\sigma(i)}$ and $\sigma \in S_g$ is arbitrary. We may regard $\boldsymbol{x}$ as a point in $\mathrm{Sym}^g(\Sigma)$ which belongs to $T_{\boldsymbol{\alpha}} \cap T_{\boldsymbol{\beta}}$ where $T_{\boldsymbol{\alpha}} = \alpha_1 \times \alpha_2 \times \cdots \times \alpha_g$ and $T_{\boldsymbol{\beta}} = \beta_1 \times \beta_2 \times \cdots \times \beta_g$.

Let us consider the unit disk D^2 in $\mathbb{C}$ with the usual orientation. Let $s_1 \subset \partial D^2$ be the portion of the oriented boundary that connects i to $-i$ and let $s_2 \subset \partial D^2$ be the remaining portion of the boundary connecting $-i$ to i. The differentials and chain maps are signed counts of J–holomorphic maps

$$u \colon D^2 \setminus \{i, -i\} \to \mathrm{Sym}^g(\Sigma)$$

representing Whitney disks in $\pi_2(\boldsymbol{x}, \boldsymbol{y})$. Here a Whitney disk is a homotopy type of maps from D^2 to $\mathrm{Sym}^g(\Sigma)$ such that s_1 is mapped to $T_{\boldsymbol{\alpha}}$, s_2 is mapped to $T_{\boldsymbol{\beta}}$, and i and $-i$ are sent to $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

Let D be a 2–chain obtained as a sum of regions of a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ with integer coefficients, $D = \sum_R a(R) R$ where $a(R) \in \mathbb{Z}$. Such a 2–chain D is called a *domain* if there exist two generators $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\partial(\partial D|_{\boldsymbol{\alpha}}) := \partial(\partial D \cap \boldsymbol{\alpha}) = \boldsymbol{y} - \boldsymbol{x}$ (here we regard $\boldsymbol{x}$ and $\boldsymbol{y}$ as 0–chains in Σ) and $\partial(\partial D|_{\boldsymbol{\beta}}) = \boldsymbol{x} - \boldsymbol{y}$. We say that D *connects* $\boldsymbol{x}$ and $\boldsymbol{y}$ and denote the set of these domains by $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$. We also denote by $\mathcal{D}$ the set of all pairs of $((\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}), D)$, where the first element in the pair is some Heegaard diagram and the second is some domain in this Heegaard diagram.

Definition 2.1 Given a Whitney disk $\varphi \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$ we associate a domain $D(\varphi) \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ called the *shadow* of φ as follows: to a region R we assign the number $n_R(\varphi)$ which is equal to the intersection number $\varphi \cdot Z_r$, where $Z_r = \{r\} \times \mathrm{Sym}^{g-1}(\Sigma)$ for some r in the interior of R (note that $\varphi \cdot Z_r$ does not depend on a choice of r); then set

$$D(\varphi) = \sum_R n_R(\varphi) R \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y}).$$

---

[1]This is nonstandard but useful to assume for the entirety of the paper.
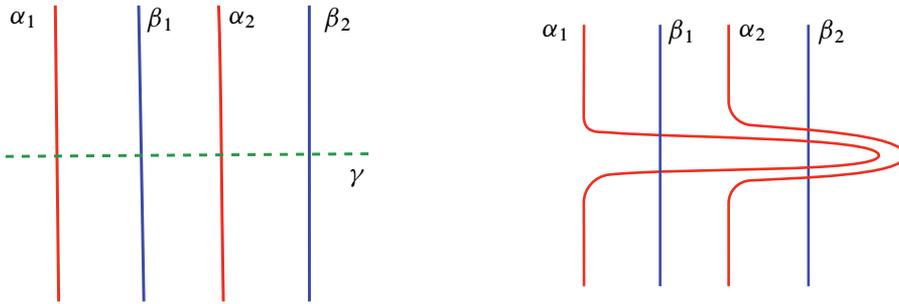
Figure 1: Finger move.

Let us recall the definition of the Maslov index of a Whitney disk $\varphi \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are some generators of the chain complex associated with a given Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$. The generators $\boldsymbol{x}$ and $\boldsymbol{y}$ are regarded as points belonging to $T_{\boldsymbol{\alpha}} \cap T_{\boldsymbol{\beta}} \subset \mathrm{Sym}^g(\Sigma)$. Let Gr_g be a space of totally real g–dimensional subspaces of $\mathbb{C}^g$. Since φ is regarded as a map from $D^2 \setminus \{i, -i\} = [0, 1] \times \mathbb{R}$ to $\mathrm{Sym}^g(\Sigma)$ we take the pullback of $T \, \mathrm{Sym}^g(\Sigma)$ to D^2 which is a trivial bundle $\mathbb{C}^g$. One may then identify pullbacks of $T_{\boldsymbol{x}}(T_{\boldsymbol{\alpha}})$ and $T_{\boldsymbol{x}}(T_{\boldsymbol{\beta}})$ with two points $V_{\boldsymbol{x}, \boldsymbol{\alpha}}, V_{\boldsymbol{x}, \boldsymbol{\beta}} \in \mathrm{Gr}_g$. We pick a "short" path $\gamma_{\boldsymbol{x}}$ between $V_{\boldsymbol{x}, \boldsymbol{\alpha}}$ and $V_{\boldsymbol{x}, \boldsymbol{\beta}}$ in Gr_g and also construct in the same fashion a path $\gamma_{\boldsymbol{y}}$ for $\boldsymbol{y}$ (by "short" we mean a path in $\mathscr{P}^-(\mathrm{Gr}_g)$ in Seidel's terminology [2008, Section 11]). The map from $0 \times \mathbb{R}$ to $T_{\boldsymbol{\alpha}}$ assigns a path γ_0 in Gr_g by considering pullback of $T(T_{\boldsymbol{\alpha}}) \subset T \, \mathrm{Sym}^g(\Sigma)$ applying the trivialization of $\varphi^*(T \, \mathrm{Sym}^g(\Sigma))$ once again. Analogously, we get a path γ_1. Finally, the Maslov index $\mu(\varphi)$ is equal to a composition of all these four paths

$$\mu(\varphi) = [\gamma_{\boldsymbol{x}} \circ \gamma_0 \circ \gamma_{\boldsymbol{y}} \circ \gamma_1] \in H_1(\mathrm{Gr}_g) = \mathbb{Z}.$$

We also introduce the *combinatorial index* of any domain $D \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$. For $x_i \in \boldsymbol{x}$ define $\widetilde{\mu}_{x_i}(D)$ as the average of the coefficients of D in the 4 regions to which x_i belongs. Then the *point measure* of D at $\boldsymbol{x}$ is $\widetilde{\mu}_{\boldsymbol{x}}(D) = \sum_{i=1}^g \widetilde{\mu}_{x_i}(D)$. For a 2–chain D the Euler measure D is $\frac{1}{2\pi}$ of the integral of the curvature of the metric of Σ over D. It is equal to the 2–cochain that assigns $\frac{1}{2}(2 - n)$ to a $2n$–gon region. For a domain $D \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ we assign its combinatorial index by

$$\widetilde{\mu}(D) = \widetilde{\mu}_{\boldsymbol{x}}(D) + \widetilde{\mu}_{\boldsymbol{y}}(D) + e(D).$$

2.2 Transformations of Heegaard diagrams

We introduce two types of transformations on a given Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ that are used in this paper.

Definition 2.2 Let γ be an oriented path in Σ which is transverse to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and is disjoint from $\boldsymbol{\alpha} \cap \boldsymbol{\beta}$; see Figure 1. Also, assume the endpoints p_0 and p_1 of γ do not belong to $\boldsymbol{\alpha} \cup \boldsymbol{\beta}$. Let U be a small neighborhood of γ. Let ψ be an isotopy of Σ supported on U which moves $\boldsymbol{\alpha}$ curves in the direction of γ as given in Figure 1. The *finger move* on the $\boldsymbol{\alpha}$ curves along the curve γ is a restriction $\psi|_{\alpha}$; this does not move the $\boldsymbol{\beta}$ curves. Analogously we define the *finger move on the $\boldsymbol{\beta}$ curves*.
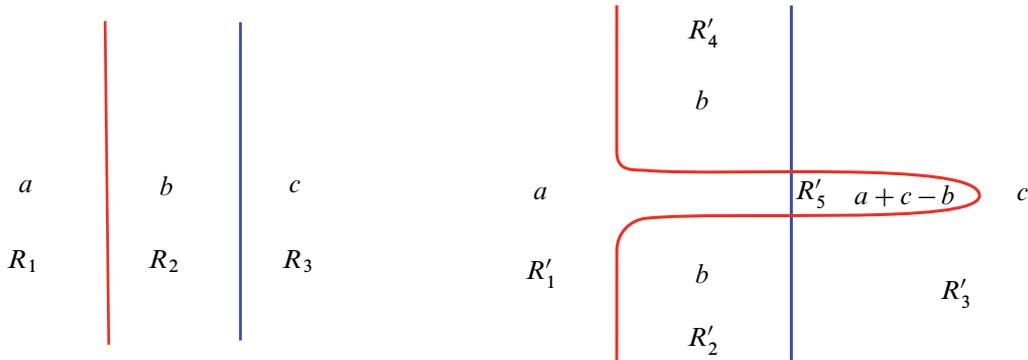
Figure 2: Image of a domain.

A transition from a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ to the diagram $(\Sigma, \psi(\boldsymbol{\alpha}), \boldsymbol{\beta})$ will be called by *performing a finger move on $\boldsymbol{\alpha}$ along γ*. Finger moves were introduced in [Sarkar and Wang 2010].

Definition 2.3 Given a domain D, its *image $D' = \psi(D)$* under a finger move ψ is defined as follows. First decompose a finger move ψ into a sequence $\psi = \psi_1 \circ \cdots \circ \psi_m$ of finger moves where under any ψ_i only two new points of intersection between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ appear (see Figure 2). Let R_1, R_2 and R_3 be regions at which $\psi_1 \circ \cdots \circ \psi_{i-1}(D)$ has coefficients a, b and c, respectively, and let the finger move ψ_i be as shown in the picture. Then for new regions R'_1, R'_2, R'_3, R'_4 and R'_5 shown in Figure 2 the coefficients are a, b, c, b and $a + c - b$ as shown. Repeating this procedure gives us $\psi(D)$.

Definition 2.4 Let D be a domain of a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$. We call an *empty stabilization* of this Heegaard diagram with respect to D a stabilization which is obtained by taking the connected sum with the standard genus 1 Heegaard diagram for S^3 where the attaching disk belongs to a region of $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and the coefficient of D is equal to 0. The new Heegaard diagram $(\Sigma', \boldsymbol{\alpha}', \boldsymbol{\beta}')$ is also called an *empty stabilization* of $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$. The image of D under an empty stabilization is D itself in the new diagram.

If we are given a Whitney disk $\varphi \in \pi_2(x, y)$ we may define its *image φ'* under a finger move on $\boldsymbol{\alpha}$ by isotoping φ in accordance with the induced isotopy of $T_{\boldsymbol{\alpha}}$ inside $\mathrm{Sym}^g(\Sigma)$. Analogously, one defines the image of φ under a finger move on $\boldsymbol{\beta}$.

We say that the stabilization is empty with respect to a Whitney disk φ if it is empty with respect to $D(\varphi)$. For the empty stabilization with respect to φ the image of φ is just $\varphi \times z$ where z is the intersection point of added α_{g+1} and β_{g+1}.

We will extensively use the invariance of the Maslov index and of $\widetilde{\mu}$ under these two types of transformations of a Heegaard diagram which is shown in the lemma below.

Lemma 2.5 *Let $D \in \mathcal{D}(x, y)$ be a domain in a Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and let $\varphi \in \pi_2(x, y)$ be a Whitney disk. Let D' be the image of D under a finger move or an empty stabilization with respect*

to D. Let φ' be the image of φ under a finger move or an empty stabilization with respect to φ. Then $\mu(\varphi) = \mu(\varphi')$ and $\widetilde{\mu}(D) = \widetilde{\mu}(D')$.

Proof First, consider the case when this transformation is an empty stabilization with respect to D (or φ). Let α_{g+1} and β_{g+1} be two new curves of $(\Sigma, \alpha', \beta')$ and z be their point of intersection. We have $x' = x \cup \{z\}$, $y' = y \cup \{z\}$ and $D' = D \in \mathscr{D}(x', y')$ (or $\varphi' \in \pi_2(x', y')$). First, $\widetilde{\mu}(D) = \widetilde{\mu}(D')$, because $\widetilde{\mu}_z(D') = 0$. As for the Maslov index, notice that paths γ_0 and γ_1 in Gr_g rise to paths γ_0' and γ_1' in Gr_{g+1} obtained by taking direct sum with one-dimensional real subspaces of $\mathbb{R}^{2(g+1)} = \mathbb{C}^{g+1}$ corresponding to $T_z\alpha_{g+1}$ and $T_z\beta_{g+1}$ respectively. Hence $\mu(\varphi) = \mu(\varphi')$.

If the transformation is a finger move then the Maslov indices are equal because finger move keeps α and β unchanged near x and y. Hence, inspecting the definition of Maslov indices we see that only the paths in Gr_g coming from the boundary ∂D^2 differ, and they are isotopic to the initial ones, so the class in $H_1(\mathrm{Gr}_g)$ does not change. Finally, $\widetilde{\mu}(D) = \widetilde{\mu}(D')$ is immediate from the formula (see also [Sarkar 2011, Theorem 3.4]). $\square$

2.3 Additivity of indices

Our strategy for proving Lipshitz's formula is based on decomposing any domain into the composition of trivial pieces: bigons and rectangles. To apply this decomposition we need the additivity of the Maslov index μ and of the combinatorial index $\widetilde{\mu}$ under the composition of domains. It is generally known that the Maslov index μ is additive. It was shown in more generality in [Sarkar 2011, Theorem 3.3] that $\widetilde{\mu}$ is additive and here we repeat the proof in our terms.

Definition 2.6 Let γ_1 and γ_2 be oriented 1–chains in Σ supported on $\alpha \cup \beta$ and intersecting transversely. Denote by $\gamma_1 \cdot \gamma_2$ the intersection number of these 1–chains, which is equal to the signed count of intersection points where the sign is defined by comparing the orientation of Σ and the orientation coming from γ_1 and γ_2. A contribution to the intersection number at an endpoint is given by a fraction $\pm\frac{1}{2}$ or $\pm\frac{1}{4}$ as in [Sarkar 2011, Section 2].

Lemma 2.7 Let $D \in \mathscr{D}(x, y)$ and D' be any other 2–chain in the Heegaard diagram (Σ, α, β). Then $\widetilde{\mu}_y(D') - \widetilde{\mu}_x(D') = \partial D|_\alpha \cdot \partial D'|_\beta = \partial D'|_\alpha \cdot \partial D|_\beta$.

Proof Given orientations on α at each intersection point p between α and β we give numbers to regions from I to IV. Namely, quadrants I and II are in the upper half of β (the half in the positive direction of α) and quadrant I is the one for which the orientation induced from Σ is opposite to the orientation of α at p. Quadrants III and IV are defined following II in the counterclockwise order on Σ.

Let us pick a point p in the first quadrant near the point $x_i \in x$ lying on some α_k and travel parallel to $\partial D|_{\alpha_k}$ until we reach a first quadrant near $y_j \in y$ where $\partial(\partial D|_{\alpha_k}) = y_j - x_i$. Then let us track the coefficient of D' at each point along this path. Each time we change the region passing through β this

coefficient changes by the intersection number of this small portion of our path with $\partial D'|_{\beta}$. Hence, the difference between the coefficients is equal to $\partial D|_{\alpha_k} \cdot \partial D'|_{\beta}$.

Now, by taking such paths for all points in x and all 4 quadrants for each of these points and then averaging we get $\widetilde{\mu}_y(D') - \widetilde{\mu}_x(D') = \partial D|_{\alpha} \cdot \partial D'|_{\beta}$. $\qquad \square$

Corollary 2.8 *Let $D \in \mathscr{D}(x, y)$ and $D' \in \mathscr{D}(y, z)$. Then $\widetilde{\mu}(D * D') = \widetilde{\mu}(D) + \widetilde{\mu}(D')$.*

Proof First, we easily see that

$$\widetilde{\mu}(D * D') - \widetilde{\mu}(D) - \widetilde{\mu}(D') = \widetilde{\mu}_x(D') + \widetilde{\mu}_z(D) - \widetilde{\mu}_y(D') - \widetilde{\mu}_y(D).$$

Second, applying Lemma 2.7 we get $\widetilde{\mu}_y(D') - \widetilde{\mu}_x(D') = \partial D|_{\alpha} \cdot \partial D'|_{\beta}$ and $\widetilde{\mu}_z(D) - \widetilde{\mu}_y(D) = \partial D|_{\alpha} \cdot \partial D'|_{\beta}$.

Hence,

$$\widetilde{\mu}(D * D') - \widetilde{\mu}(D) - \widetilde{\mu}(D') = 0. \qquad \square$$

3 Main theorem

Theorem 3.1 *For a given domain D in a Heegaard diagram (Σ, α, β) there is a sequence of finger moves and empty stabilizations such that in the new Heegaard diagram the image of D can be represented as a composition of bigons, rectangles and their negatives.*

Before proving this theorem we show how Theorems 1.1 and 1.2 follow from Theorem 3.1.

Proof of Theorem 1.1 Let $D \in \mathscr{D}(x, y)$. Apply Theorem 3.1 to obtain the image D' of D and the decomposition $D' = D_1 * \cdots * D_k$ of D' in the new Heegaard diagram. Here each D_i is either a bigon, a rectangle, or the negative of a bigon or a rectangle. Since $\overline{\mu}$ is additive we may compute $\overline{\mu}(D') = \overline{\mu}(D_1) + \cdots + \overline{\mu}(D_k)$. For each D_i the value $\overline{\mu}(D_i)$ can be easily inferred from the 4 axioms.

The fact that $\overline{\mu}$ coincides with $\widetilde{\mu}$ then follows since $\widetilde{\mu}(R) = \widetilde{\mu}(B) = 1$ by direct computation, and $\widetilde{\mu}$ is additive by Lemma 2.7 and stable by Lemma 2.5. $\qquad \square$

Proof of Theorem 1.2 We are given a Whitney disk $\varphi \in \pi_2(x, y)$ and we may assume that $g(\Sigma) > 1$ by applying an empty stabilization if necessary. By Theorem 3.1 there is a transformation of the given Heegaard diagram such that there is a decomposition $D(\varphi') = D_1 * \cdots * D_k$, where φ' is an image of φ under this transformation; equivalently $0 = D(\varphi') * (-D_1) * \cdots * (-D_k)$. Since each D_i is either a bigon or a rectangle (possibly negative) there exists a corresponding Whitney disk φ_i such that $D(\varphi_i) = -D_i$. Then $\varphi' * \varphi_1 * \cdots * \varphi_k \in \pi_2(x, x)$ and, moreover, $D(\varphi' * \varphi_1 * \cdots * \varphi_k) = 0$. From [Ozsváth and Szabó 2004b, Proposition 2.15] it now follows that $\varphi' * \varphi_1 * \cdots * \varphi_k = 0 \in \pi_2(x, x)$, so $\mu(\varphi' * \varphi_1 * \cdots * \varphi_k) = 0$. Hence, $\mu(\varphi') = -(\mu(\varphi_1) + \cdots + \mu(\varphi_k))$.
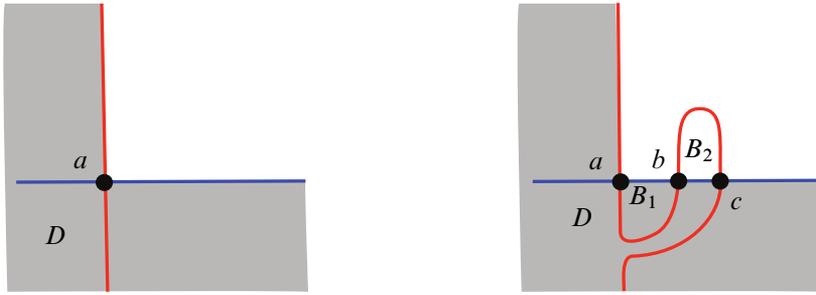
Figure 3: 270° angle.

It now suffices to show that $\mu(B) = \tilde{\mu}(B) = 1$ and $\mu(R) = \tilde{\mu}(R) = 1$ for any bigon B and any rectangle R. For a bigon $\mu(B) = 1$ since it agrees with the dimension of the (regular) moduli space of biholomorphisms from B to the strip $[0, 1] \times \mathbb{R}$. The fact that $\mu(R) = 1$ is shown in [Ozsváth and Szabó 2004b, Section 8.4].   □

Now we prove the main theorem.

Proof of Theorem 3.1   In the proof we abuse notation and denote the image of a domain D under any empty stabilization or a finger move by same letter D.

We break our proof into several steps starting with a general domain in Step 1 and simplifying it gradually by transformations of Heegaard diagrams. Before we start dealing with general domains we need additional preparation which we perform at Step 0.

Step 0: quadrilaterals   Before considering the most general domains we need to take care of another type of a "building block". Namely, let D be an embedded domain with nonnegative coefficients which is homeomorphic to a disk and has a boundary consisting of four sides. We would call such D a *quadrilateral*. If all angles of D are 90° then it is already a rectangle. Otherwise, an angle at one of its vertices a is 270°. Consider a finger move shown in Figure 3 which creates two bigons $B_1 \in \pi_2(a, b)$ and $B_2 \in \pi_2(b, c)$. Then $D * (-B_1)$ is a quadrilateral with one more 90° angle than D. Proceeding in the same fashion we decompose D into a rectangle and several bigons with possibly some x– or y–coordinates in the interiors. We refer the reader to Step 5 for the treatment of these inner points.

Henceforth, in the following steps it is enough to decompose any domain into quadrilaterals and bigons with arbitrary angles.

Step 1: making the boundary embedded   First, we show that after several finger moves we may compose D with some bigons (or their negatives) and obtain a domain D' with embedded boundary.

Assume without loss of generality that $\partial(\partial D|_{\alpha_1}) = y_1 - x_1$ with respect to the chosen orientation on α_1. Here x_1 lies on β_1 and y_1 lies on β_2 (where β_2 may be equal to β_1). We may assume that the positively oriented embedded arc from x_1 to y_1 in α_1 is covered $k + 1$ times by ∂D and the positively oriented embedded arc from y_1 to x_1 is covered k times. For now assume $k > 0$.
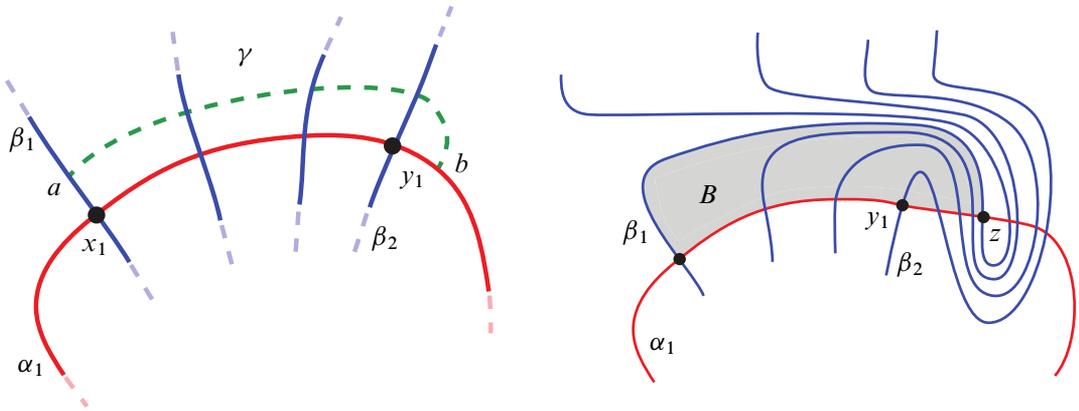
Figure 4: Making the boundary embedded.

Since the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves are oriented, for any region we may distinguish whether it lies to the left of a given $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ curve. Let a be a point very close to x_1 on β_1 to the left of α_1 (see Figure 4). Also let b be a point on α_1 near y_1 not belonging to the positive arc from x_1 to y_1. Let γ be (a slight extension of) a curve starting at a and parallel to α_1 until it hits β_2 and then we connect it with b. We make a finger move on $\boldsymbol{\beta}$ curves along the curve γ creating new points of intersection of $\boldsymbol{\beta}$ curves intersecting γ (including β_1 and β_2) with α_1. Let z be the new intersection point of β_1 and α_1 which is the closest to x_1 on β_1.

As a result, we created a bigon B connecting x_1 to z. Then

$$D' = (-B) * D \in \mathcal{D}(\{z, x_2, \ldots, x_g\}, \{y_1, y_2, \ldots, y_g\}).$$

Note that in $\partial D'$ the arc from z to y_1 is covered k times and the arc from y_1 to z is covered $k - 1$ times. Additionally, we notice that $\partial D'|_{\boldsymbol{\beta}}$ only changed by replacing x_1 with z which may be assumed to be very close to x_1 on β_1 and hence the number of times $\partial D'_{\boldsymbol{\beta}}$ covers z doesn't differ from that of x_1 for the former D. Hence performing isotopies as above only changes the geometry of the α_1–portion of ∂D.

In the case $k < 0$ we would draw γ parallel to the negative embedded arc from x_1 to y_1 and proceed analogously. Repeating this procedure we first make sure that $\partial D \cap \boldsymbol{\alpha}$ is embedded and then we repeat it for $\boldsymbol{\beta}$ and end up with a domain having an embedded boundary.

Step 2: making the boundary connected Assume the boundary of a domain D consists of more than one closed curve as shown in Figure 5. Suppose $\partial D \cap (\alpha_1 \cap \beta_1) = y_1$, $\partial D \cap (\beta_1 \cap \alpha_2) = x_2$ and $\partial(\partial D|_{\alpha_2}) = y_2 - x_2$. Let us pick a point on α_1 close to y_1 and let γ be an arc connecting it to a point on the restriction of some other component of ∂D to $\boldsymbol{\beta}$ such that $\text{int}(\gamma)$ does not intersect ∂D. We denote by β_1' the curve to which the endpoint of γ belongs and $\partial(\partial D|_{\beta_1'}) = x_2' - y_1'$ while $\partial(\partial D|_{\alpha_1'}) = y_1' - x_1'$. We also can make sure that γ intersects β_1' at a point near y_1'.

Let us decompose $\gamma = \gamma_1 \gamma_2 \gamma_3$ where γ_1 and γ_3 are small portions near the ends and γ_2 is the remaining portion in the middle. We draw a curve γ_1' connecting a point near y_1 on β_1 with some point near the end of γ_1 such that γ_1' does not intersect γ_1. Starting at the end of γ_1', we draw a curve γ_2' parallel to γ_2. Then
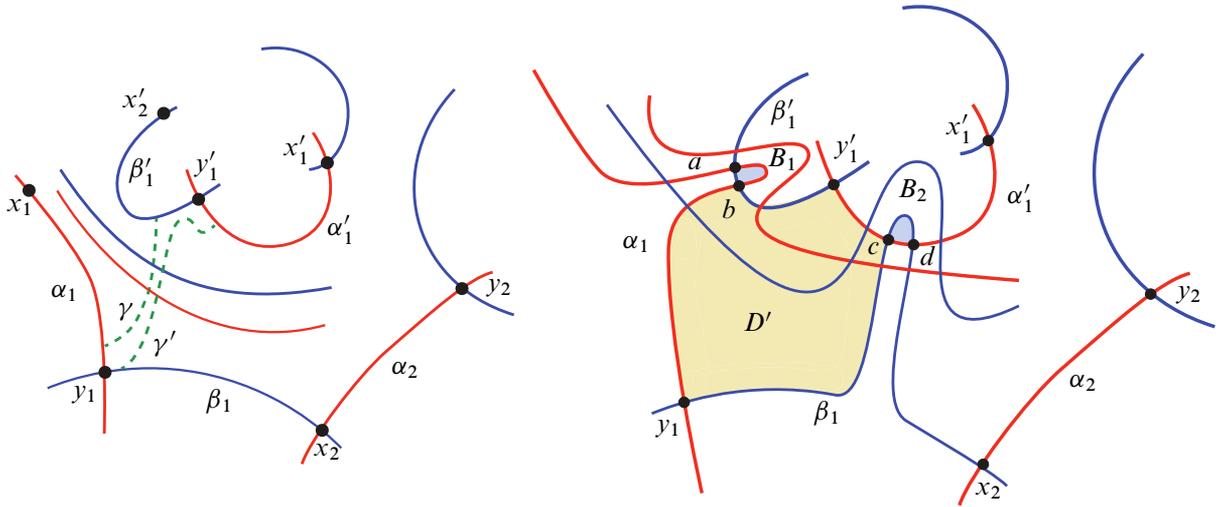
Figure 5: Making the boundary connected.

we connect the end of γ_2' by a curve γ_3' parallel to β_1' with a point on α_1' near y_1'. We set $\gamma' = \gamma_1'\gamma_2'\gamma_3'$. In short, γ' is a parallel copy of γ on the complement of neighborhoods of points y_1 and y_1'.

Now we make a finger move on $\boldsymbol{\alpha}$ curves along γ and we also make a finger move on $\boldsymbol{\beta}$ curves along γ'. Denote the new points of intersection between α_1 and β_1' by a and b, and analogously denote the points of intersection between β_1 and α_1' by c and d.

Denote the new quadrilateral by $D' \in \mathscr{D}(\{b, c\}, \{y_1, y_1'\})$ and the two new bigons by $B_1 \in \mathscr{D}(\{a\}, \{b\})$ and $B_2 \in \mathscr{D}(\{c\}, \{d\})$.

Then consider
$$D * (-D') * B_1 * B_2 \in \mathscr{D}(\boldsymbol{x}, \{a, y_2', \ldots, d, y_2, \ldots\})$$

whose boundary is embedded and contains one fewer component than D. From here we may proceed inductively on the number of boundary components.

Step 3: reducing to a quadrangle or a bigon boundary Let D be a domain. Applying Steps 1 and 2, we assume that ∂D is embedded and connected.

Let the boundary of a domain D be $2n$–sided, ie of the form $x_1 y_1 x_2 y_2 x_3 \cdots y_n$ with $n > 2$ where $x_i y_i$ is an arc on $\boldsymbol{\alpha}$ and $y_i x_{i+1}$ is an arc on $\boldsymbol{\beta}$. Assume that the first 4 sides lie on $\alpha_1, \beta_1, \alpha_2$ and β_2 (see Figure 6). Let us pick a point y_1' on $x_1 y_1$ near y_1 and a point y_2' on β_2 near y_2. Let γ be an arc connecting y_1' and y_2' such that $\mathrm{int}(\gamma) \cap \partial D = \varnothing$ and the arcs γ, $y_1 y_1'$, $y_1 x_2$, $x_2 y_2$ and $y_2 y_2'$ bound an embedded disk. We make a finger move on $\boldsymbol{\alpha}$ curves along γ and denote by a and b two new points of intersection between α_1 and β_2. Notice that we created two new domains: a bigon $B \in \mathscr{D}(\{a\}, \{b\})$ and a quadrilateral $D' \in \mathscr{D}(\{x_2, b\}, \{y_1, y_2\})$.
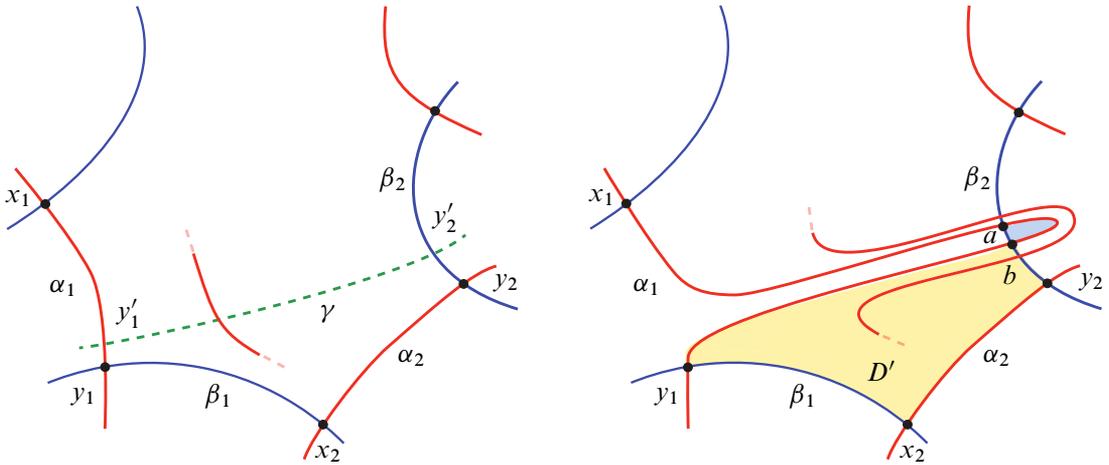
Figure 6: Decreasing the number of sides.

We may replace D with a domain

$$D * (-D') * B \in \pi_2(\{x_1, x_2, \ldots, x_n\}, \{a, x_2, y_3, \ldots, y_n\}),$$

which has an embedded and connected boundary with $2n - 2$ sides.

Step 4a: 4–sided boundary Given two domains which have the same 4–sided curve as the boundary, their difference is an element in $H_2(\Sigma)$ which is generated by $[\Sigma]$. We will prove in Claim 3.2 that Σ can be decomposed into bigons and rectangles after some finger moves and empty stabilizations. Therefore we may add $k[\Sigma]$ to a given domain D to ensure that we get a domain represented by an embedded 2–chain. This domain, which we also call D, is isotopic to a disk with m handles.

We will now show how to decompose this handlebody D into bigons and quadrilaterals.

We depict handles as pairs of circles with opposite orientations and one of the handles is drawn in Figure 7. We now draw two "palm trees" as in the picture by making two finger moves on $\boldsymbol{\alpha}$ curves and we assume that the other $m - 1$ handles are in the white region "between" two palm trees. Let the 8 new points of intersection between the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves be labeled as in Figure 7.

We now have 4 new bigons B_{11}, B_{12}, B_{21} and B_{22}, and 2 new quadrilaterals R_x and R_y. Then the domain $D * (-R_x) * (-R_y) * B_{11} * B_{12} * B_{21} * B_{22} \in \mathcal{D}(\{z_1', z_2'\}, \{w_1', w_2'\})$ has 4–sided boundary and $m - 1$ handles and from here proceed inductively on the number of handles.

Step 4b: 2–sided boundary Here we reduce this case to the 4–sided boundary case treated in Step 4a.

As in the 4–sided boundary case, we may assume D to be an embedded 2–chain represented by a handlebody as shown in Figure 8. By the method of Step 0 we may assume that both angles are 90°.

We may assume that $g \geqslant 2$ by applying an empty stabilization with respect to D if necessary. Let $D \in \mathcal{D}(\{x, z_1, \ldots, z_{g-1}\}, \{y, z_1, \ldots, z_{g-1}\})$. We may suppose $z_1 \in \alpha_2 \cap \beta_2$ lies outside of D by applying
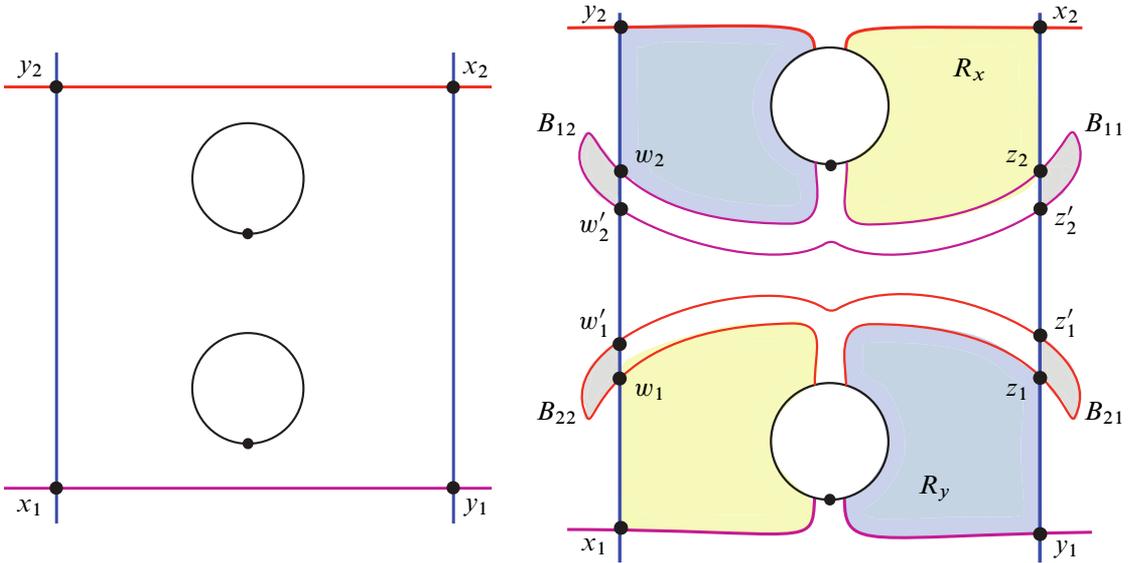
Figure 7: 4–sided boundary.

an empty stabilization with respect to a given D. Let γ be an arc connecting the point z_1 with some point a on the boundary of D such that γ intersects D once at the point a. We may assume a to be somewhere on β_1. We also alter the path γ slightly so that it starts at some point on α_2 and β_2 is to its left near this endpoint.

Now draw γ' almost parallel to γ so that one of its endpoints is on β_2 near z_1. When γ' reaches a neighborhood of a we extend it parallel to $-\beta_2$ until it hits α_1 at some point outside of D near y.

We make a finger move on $\boldsymbol{\alpha}$ along γ and a finger move on $\boldsymbol{\beta}$ along γ' creating points $b \in \alpha_2 \cap \beta_1$ and $c \in \beta_2 \cap \alpha_1$. There is now a new quadrilateral $R \in \mathcal{D}(\{y, z_1, z_2, \ldots, z_{g-1}\}, \{b, c, \ldots, z_{g-1}\})$.

Therefore, we may replace D with $D * R \in \mathcal{D}(\{x, z_1, z_2, \ldots, z_{g-1}\}, \{b, c, \ldots, z_{g-1}\})$ reducing to Step 4a since it has connected and embedded 4–sided boundary.
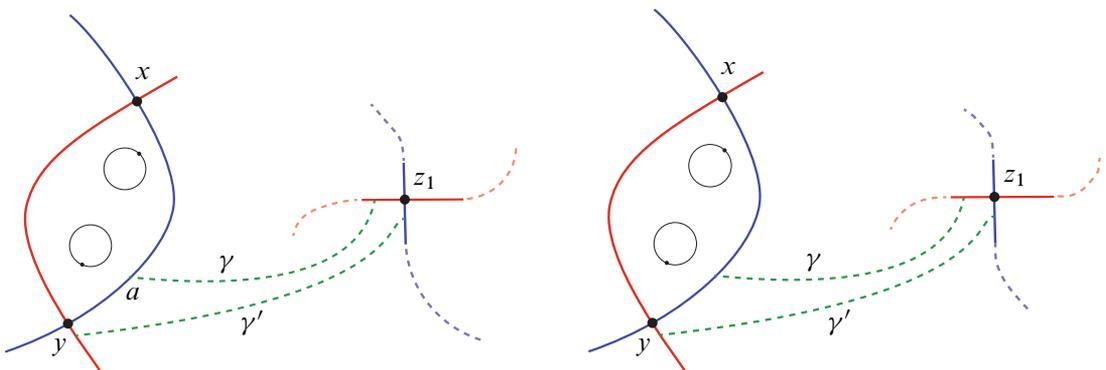
Figure 8: 2–sided boundary.
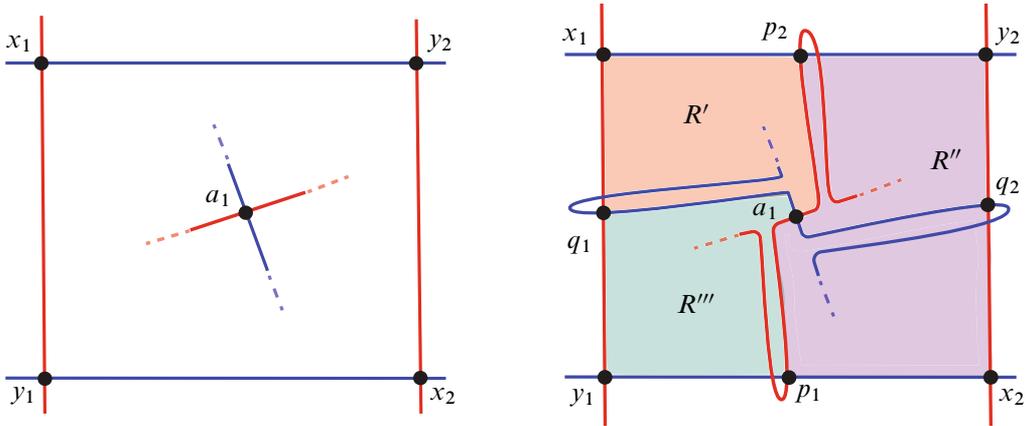
Figure 9: Inner points.

Step 5: inner points By now we have represented the initial D as a composition of domains that geometrically look like bigons and rectangles, but some of the rectangles may contain some points from generators inside of them. Namely, let $R \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x} = \{x_1, x_2, a_1, \ldots, a_k, b_{k+1}, \ldots, b_{g-2}\}$ and $\boldsymbol{y} = \{y_1, y_2, a_1, \ldots, a_k, b_{k+1}, \ldots, b_{g-2}\}$ where the a_i lie inside of R and the b_j lie outside of R.

Let $a_1 \in \alpha_3 \cap \beta_3$. We make finger moves on α_3 and β_3, creating 2 points of intersection between α_3 and β_1, β_2, and 2 points of intersection between β_3 and α_1, α_2 (actually, we created twice as many points, but we stress our attention on these 4). We call these points p_1, p_2, q_1 and q_2, respectively. This is illustrated in Figure 9.

Then we may see that $R = R' * R'' * R'''$ where

$$R' \in \mathcal{D}(\boldsymbol{x}, \{q_1, x_2, p_2, a_2, \ldots, a_k, b_{k+1}, \ldots, b_g\}),$$
$$R'' \in \mathcal{D}(\{q_1, x_2, p_2, a_2, \ldots, a_k, b_{k+1}, \ldots, b_g\}, \{q_1, y_2, p_1, \ldots, a_k, b_{k+1}, \ldots, b_g\}),$$
$$R''' \in \mathcal{D}(\{q_1, x_2, p_1, a_2, \ldots, a_k, b_{k+1}, \ldots, b_g\}, \boldsymbol{y}).$$

Each of these 3 rectangles has fewer points from generators inside than R and we may proceed by induction on the number of points.

This completes the proof of Theorem 3.1, assuming Claim 3.2 below. □

Claim 3.2 *The surface $\Sigma \in \mathcal{D}(\boldsymbol{x}, \boldsymbol{x})$ for $\boldsymbol{x} = \{x_1, \ldots, x_g\}$ is a domain that can be decomposed into bigons and rectangles after applying a sequence of finger moves.*

Proof Let us assume $g > 1$. We may apply a diffeomorphism φ to the Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ such that the image of $\boldsymbol{x}$ is a collection $\boldsymbol{x}'$ such that all x_i' are located in some small disk region and x_1' and x_2' are on the boundary of the convex hull of $\boldsymbol{x}'$. Let us denote by $(\Sigma, \boldsymbol{\alpha}', \boldsymbol{\beta}')$ the image of the initial diagram under φ, ie $\boldsymbol{\alpha}' = \varphi(\boldsymbol{\alpha})$ and $\boldsymbol{\beta}' = \varphi(\boldsymbol{\beta})$. We may assume that $x_1' \in \alpha_1' \cap \beta_2'$ and $x_2' \in \alpha_2' \cap \beta_1'$. Then we make finger moves on α_1', α_2' and on β_1', β_2' as in Figure 10, creating a rectangle R with boundary
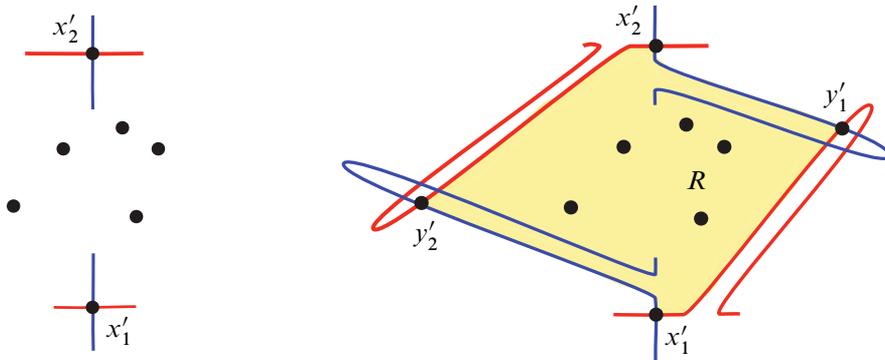
Figure 10: Claim 3.2.

$x_1' y_1' x_2' y_2'$ such that all other points of $\boldsymbol{x}'$ are inside of R. Then $\Sigma * (-R)$ connects $\boldsymbol{x}'$ to $\{y_1', y_2', \ldots, x_g'\}$ and it is a region with quadrangle boundary and g handles. Now we may proceed as in Step 4a of the proof of the theorem.

Let $g = 1$. We can make a finger move on α_1 which creates a bigon B connecting given point x to some point x'. Then $(-B) * \Sigma \in \mathcal{D}(\{x'\}, \{x\})$. Now we make a stabilization inside B and can proceed as in the proof of Step 4b. $\qquad\square$

References

[Ghiggini 2008]   **P Ghiggini**, *Knot Floer homology detects genus-one fibred knots*, Amer. J. Math. 130 (2008) 1151–1169 MR Zbl

[Lipshitz 2006]   **R Lipshitz**, *A cylindrical reformulation of Heegaard Floer homology*, Geom. Topol. 10 (2006) 955–1096 MR Zbl Correction in 18 (2014) 17–30

[Ni 2009a]   **Y Ni**, *Heegaard Floer homology and fibred 3–manifolds*, Amer. J. Math. 131 (2009) 1047–1063 MR Zbl

[Ni 2009b]   **Y Ni**, *Link Floer homology detects the Thurston norm*, Geom. Topol. 13 (2009) 2991–3019 MR Zbl

[Ozsváth and Szabó 2004a]   **P Ozsváth**, **Z Szabó**, *Holomorphic disks and genus bounds*, Geom. Topol. 8 (2004) 311–334 MR Zbl

[Ozsváth and Szabó 2004b]   **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR Zbl

[Ozsváth and Szabó 2004c]   **P Ozsváth**, **Z Szabó**, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. 159 (2004) 1159–1245 MR Zbl

[Ozsváth and Szabó 2006]   **P Ozsváth**, **Z Szabó**, *Holomorphic triangles and invariants for smooth four-manifolds*, Adv. Math. 202 (2006) 326–400 MR Zbl

[Rasmussen 2003]   **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) arXiv math/0306378

[Sarkar 2011]   **S Sarkar**, *Maslov index formulas for Whitney n–gons*, J. Symplectic Geom. 9 (2011) 251–270 MR Zbl

[Sarkar and Wang 2010] **S Sarkar**, **J Wang**, *An algorithm for computing some Heegaard Floer homologies*, Ann. of Math. 171 (2010) 1213–1236 MR Zbl

[Seidel 2008] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

Department of Mathematics, University of California, Los Angeles
Los Angeles, CA, United States

romankrut@ucla.edu

The $H\underline{\mathbb{F}}_2$–homology of C_2–equivariant Eilenberg–MacLane spaces

SARAH PETERSEN

We extend Ravenel–Wilson Hopf ring techniques to C_2–equivariant homotopy theory. Our main application and motivation is a computation of the $RO(C_2)$–graded homology of C_2–equivariant Eilenberg–MacLane spaces. The result we obtain for C_2–equivariant Eilenberg–MacLane spaces associated to the constant Mackey functor $\underline{\mathbb{F}}_2$ gives a C_2–equivariant analogue of the classical computation due to Serre. We also investigate a twisted bar spectral sequence computing the homology of these equivariant Eilenberg–MacLane spaces and suggest the existence of another twisted bar spectral sequence with E^2–page given in terms of a twisted Tor functor.

1 Introduction

Computations of invariants in equivariant homotopy theory have powerful applications contributing to solutions of outstanding classification problems in geometry, topology, and algebra. A primary example is Hill, Hopkins, and Ravenel's solution [Hill et al. 2016] to the Kervaire invariant one problem, which used computations in equivariant homotopy theory to answer the question of when a framed $(4k+2)$–dimensional manifold can be surgically converted into a sphere. Despite the success of numerous applications, many equivariant computations remain difficult to access due to their rich structure. This is especially true for (unstable) equivariant spaces, for which many computations have not yet been completed, despite their analogous nonequivariant results being well known.

This paper extends Ravenel–Wilson Hopf ring techniques [Ravenel and Wilson 1977; 1980; Wilson 1982] to C_2–equivariant homotopy theory. Our main application and motivation is a computation of the $RO(C_2)$–graded homology of C_2–equivariant Eilenberg–Mac Lane spaces. The result, stated over the course of Theorems 5.6, 6.6, and 6.7, is a C_2–equivariant analogue of the classical cohomology computation completed by Serre [1953].

Nonequivariantly, Serre applied the Borel theorem (see, for instance [Mosher and Tangora 1968, page 88, Theorem 1]) to the path space fibration

$$K(\mathbb{F}_p, n) \simeq \Omega K(\mathbb{F}_p, n+1) \to P(K(\mathbb{F}_p, n+1)) \to K(\mathbb{F}_p, n+1),$$

to calculate the cohomology of $K(\mathbb{F}_p, n+1)$ given $H^* K(\mathbb{F}_p, n)$. In C_2–equivariant homotopy theory, the constant Mackey functor $\underline{\mathbb{F}}_2$ is the analogue of the group $\mathbb{F}_2$ and the Eilenberg–Mac Lane spaces $K_V = K(\underline{\mathbb{F}}_2, V)$ are graded on the real representations V of the group C_2 rather than on the integers. Since the group C_2 has two irreducible real representations, the trivial representation and the sign representation σ, the analogous equivariant computation would require computing the cohomology of $K_{V+\sigma}$ from $H^\star K_V$ in addition to $H^\star K_{V+1}$ from $H^\star K_V$. This would necessitate having a so called signed or twisted version of the Borel theorem. However, no such theorem is known to exist, making it difficult to study the cohomology of the spaces $K_{V+\sigma}$ with these techniques. We call $K_{V+\sigma}$ a signed delooping of K_V since the space of signed loops $\Omega^\sigma K_{V+\sigma} \simeq K_V$.

While direct extension of Serre's original argument does not allow for the computation of the cohomology of signed deloopings, it has been successfully applied to study trivial representation deloopings of K_σ, whose cohomology is known [Hu and Kriz 2001]. This approach is described in Ugur Yigit's thesis [2019], where it is noted that the $RO(C_2)$–graded cohomology of all C_2–equivariant Eilenberg–Mac Lane spaces $K_{\sigma+*}$ can be computed using this method. Throughout, we use $*$ to denote integer grading and reserve $\star$ to denote grading by finite-dimensional real representations.

A major reason to study Ravenel–Wilson Hopf ring techniques in C_2–equivariant homotopy theory is that they provide a way to study σ–deloopings. These techniques, which investigate multiplicative structures coming from H–space maps on spaces having a graded multiplication, lend additional structure that can be exploited to complete computations.

An important tool in classical applications of Ravenel–Wilson Hopf ring techniques is the bar construction B. This construction plays a significant role in computation because B is a trivial representation delooping functor with $BK_V \simeq K_{V+1}$. In the C_2–equivariant world, there is a twisted bar construction B^σ, which is a sign representation delooping functor with $B^\sigma K_V \simeq K_{V+\sigma}$ [Liu 2020]. We use these two constructions to explicitly model multiplicative structures on the spaces K_V at the point set level (Theorem 5.4), directly extending work by Ravenel and Wilson [1980]. We also describe our approach to using this structure to investigate signed and trivial representation deloopings in Section 5.

Whereas Ravenel and Wilson use a collapsing integer-graded bar spectral sequence to compute by induction on n the homology of classical nonequivariant Eilenberg–Mac Lane spaces [Wilson 1982], we

deduce many of our equivariant computations from nonequivariant ones using a computational method introduced by Behrens and Wilson [2018, Lemma 2.8]. Starting with the $RO(C_2)$–graded homology of K_σ, we use the graded multiplication on the spaces K_V coming from the genuine equivariant ring structure on $H\mathbb{F}_2$, to produce elements of the $RO(C_2)$–graded homology of $K_{*\sigma}$. We then use the point set level understanding of multiplicative structures on the spaces $K_{*\sigma}$ developed in Theorem 5.4 to verify that these elements in fact form a free basis for the homology.

Once we have computed $H_\star K_{*\sigma}$ (Theorem 5.6), we use Hopf ring structures in $RO(C_2)$–graded bar spectral sequences to compute $H_\star K_{i\sigma+j}$ (Theorem 6.6) by induction on j. In the case where $i = 1$, that is for the spaces $K_{\sigma+*}$, we name all homology generators in terms of the Hopf ring structure (Theorem 6.7). The task of naming homology generators for the spaces K_V, where $\sigma + 1 \subset V$, increases in complexity as the number of sign representations increases. We illustrative this phenomenon in Section 6.

Knowing the $RO(C_2)$–graded homology of the C_2–equivariant Eilenberg–MacLane spaces K_V, we turn to investigating the $RO(C_2)$–graded twisted bar spectral sequence. Much like the classical integer graded bar spectral sequence, the $RO(C_2)$–graded twisted bar spectral sequence arises from a filtered complex. However, computations with this twisted spectral sequence are more complicated than in the classical case. For example, in contrast to the classical case where the integer-graded bar spectral sequence computing the nonequivariant mod p homology of the classical Eilenberg–MacLane spaces $K_* = K(\mathbb{F}_p, *)$ collapses on the E^2–page [Wilson 1982], we find there are arbitrarily long equivariant degree shifting differentials, similar to those observed in Kronholm's study [2010] of the cellular spectral sequence, in the $RO(C_2)$–graded twisted bar spectral sequences computing the homology of the signed representation spaces $K_{n\sigma}$, where $n \geq 2$.

While the $RO(C_2)$–graded twisted bar spectral sequence is quite complicated in general, the differentials and extensions appear to arise in an extremely structured way, governed by a norm structure. We use our knowledge of $H_\star K_{*\sigma}$ and the E^∞–page to deduce information about the $RO(C_2)$–graded twisted bar spectral sequences computing the homology of $K_{*\sigma}$. This allows us to write down conjectures concerning many of the differentials in Section 6. Our equivariant computations show that, unlike in the nonequivariant integer graded situation, the $RO(C_2)$–graded twisted bar spectral sequences computing $H_\star K_{n\sigma}$, where $n \geq 2$, have a rich structure quite distinct from the collapsing bar spectral sequence in the classical nonequivariant case [Wilson 1982]. Differences between integer graded and $RO(C_2)$–graded bar and twisted bar spectral sequences are discussed in Section 6.

In parallel with calculating the homology of a space, the corresponding computational tools are worth investigating in a purely algebraic setting. This study of the homological algebra involved produces tools which can also be applied in settings outside of topology. One example of this are Tor functors, the derived functors of the tensor product of modules over a ring. Besides playing a central role within algebraic topology theorems such as the Künneth theorem and coefficient theorem, Tor functors can also be used to calculate the homology of groups, Lie algebras, and associative algebras. Within the context of the classical

Ravenel–Wilson Hopf ring method, the identification of the E^2–page of the bar spectral sequence with Tor allows for the computations $\mathrm{Tor}^{E[x]}(\mathbb{F}_p, \mathbb{F}_p) \simeq \Gamma[sx]$ and $\mathrm{Tor}^{T[x]}(\mathbb{F}_p, \mathbb{F}_p) \simeq E[sx] \otimes \Gamma[\phi x]$, where sx is the suspension of x, ϕx is the transpotent, and $T[x]$ is the truncated polynomial ring $\mathbb{F}_p[x]/(x^p)$, to be used inductively in the calculations of the mod p homology of Eilenberg–Mac Lane spaces [Wilson 1982] and the Morava K–theory of Eilenberg–Mac Lane spaces [Ravenel and Wilson 1980].

In the C_2–equivariant setting, the $RO(C_2)$–graded homology of each signed delooping, $K_{V+\sigma}$, of an equivariant Eilenberg–Mac Lane space, K_V, also independently arises as the result of a C_2–equivariant twisted Tor computation. Thus under favorable circumstances, we believe it should be possible to formulate a twisted bar spectral sequence with E^2–page a twisted Tor functor arising as a derived functor of the twisted product of $H\mathbb{F}_2$–modules and use this to compute the E^2–page. However, we have not yet constructed such a spectral sequence.

Additionally, twisted Tor calculations are not yet well understood, with a complete lack of known examples. Theorems 5.6, 6.6, and 6.7 provide a countably infinite number of initial examples, which in turn lend insight on how such calculations might proceed in general. We discuss how the homology $H_\star K_{V+\sigma}$ arises as a result of twisted Tor and give evidence for $\mathrm{Tor}_{\mathrm{tw}}^{E[x]}(H_\star, H_\star) \simeq E[\sigma x] \otimes \Gamma[\mathcal{N}_e^{C_2}(x)]$, where σx is the signed suspension of x and $\mathcal{N}_e^{C_2}$ is the norm, under favorable circumstances in Section 7.

1.1 Statement of theorems

We state our main results. Recall that $H\mathbb{F}_2$ has distinguished elements $a \in H\mathbb{F}_{2\{-\sigma\}}$ and $u \in H\mathbb{F}_{2\{1-\sigma\}}$.

To describe our answer for $H_\star K_{*\sigma}$, we need notation for $H_\star K_\sigma$. Let

$$e_\sigma \in H_\sigma K_\sigma, \quad \bar{\alpha}_i \in H_{\rho i} K_\sigma \quad (i \geq 0).$$

Then the homology, $H_\star K_\sigma$, is exterior on generators

$$e_\sigma, \quad \bar{\alpha}_{(i)} = \bar{\alpha}_{2^i} \quad (i \geq 0)$$

with coproduct

$$\psi(e_\sigma) = 1 \otimes e_\sigma + e_\sigma \otimes 1 + a(e_\sigma \otimes e_\sigma),$$

$$\psi(\bar{\alpha}_n) = \sum_{i=0}^{n} \bar{\alpha}_{n-i} \otimes \bar{\alpha}_i + \sum_{i=0}^{n-1} u(e_\sigma \bar{\alpha}_{n-1-i} \otimes e_\sigma \bar{\alpha}_i).$$

For finite sequences

$$J = (j_\sigma, j_0, j_1, \ldots), \quad j_k \geq 0,$$

define

$$(e_\sigma \bar{\alpha})^J = e_\sigma^{\circ j_\sigma} \circ \bar{\alpha}_{(0)}^{\circ j_0} \circ \bar{\alpha}_{(1)}^{\circ j_1} \circ \cdots$$

where the ∘–product comes from the pairing $\circ \colon K_V \wedge K_W \to K_{V+W}$.

Then

$$H_\star K_{*\sigma} \cong \otimes_J E[(e_\sigma \bar{\alpha})^J]$$

as an algebra, where the tensor product is over all J and the coproduct follows by Hopf ring properties from the $\bar{\alpha}$'s.

Interestingly, this answer mirrors the classical nonequivariant answer at the prime 2 [Ravenel and Wilson 1980].

From there, we use the $RO(C_2)$–graded bar spectral sequence to compute $H_\star K_{i\sigma+j}$ be induction on j, starting with $H_\star K_{i\sigma}$. We show:

The $RO(C_2)$–graded homology of K_V, where $\sigma + 1 \subset V$, is exterior on generators given by the cycles on the E^2–page of the $RO(C_2)$–graded spectral sequence computing $H_\star BK_{V-1}$.

For the spaces $K_{\sigma+*}$, we name all homology generators in terms of the Hopf ring structure. To describe these rings, we need notation for $H_\star K_1$, $H_\star K_2$, and $H_\star K_\rho$. Let

$$e_1 \in H_1 K_1, \quad \alpha_i \in H_{2i} K_1, \quad \beta_i \in H_{2i}\mathbb{C}P^\infty, \quad i \geq 0.$$

This gives generators

$$e_1, \quad \alpha_{(i)} = \alpha_{p^i}, \quad \beta_{(i)} = \beta_{p^i}$$

of $H_\star K_1$ and $H_\star K_2$ with coproducts

$$\psi(\alpha_n) = \sum_{i=0}^{n} \alpha_{n-i} \otimes \alpha_i, \quad \psi(\beta_n) = \sum_{i=0}^{n} \beta_{n-i} \otimes \beta_i.$$

Also let

$$\bar{\beta}_i \in H_{\rho i} K(\mathbb{Z}, \rho) \quad (i \geq 0).$$

This gives additional generators,

$$\bar{\beta}_{(i)} = \bar{\beta}_{2^i} \quad (i \geq 0)$$

of $H_\star K_\rho$ with coproduct

$$\psi(\bar{\beta}_n) = \sum_{i=0}^{n} \bar{\beta}_{n-i} \otimes \bar{\beta}_i.$$

Then for finite sequences

$$
\begin{aligned}
I &= (i_1, i_2, \ldots, i_k), & & 0 \leq i_1 < i_2 < \cdots, \\
W &= (w_1, w_2, \ldots, w_q), & & 0 \leq w_1 < w_2 < \cdots, \\
J &= (j_{-1}, j_0, j_1, \ldots, j_\ell), & & \text{where } j_{-1} \in \{0, 1\} \text{ and all other } j_n \geq 0, \\
Y &= (y_{-1}, y_0, y_1, \ldots, y_r), & & \text{where } y_{-1} \in \{0, 1\} \text{ and all other } y_n \geq 0,
\end{aligned}
$$

define

$$(e_1\alpha\beta)^{I,J} = e_1^{\circ j_{-1}} \circ \alpha_{(i_1)} \circ \alpha_{(i_2)} \circ \cdots \circ \alpha_{(i_k)} \circ \beta_{(0)}^{\circ j_0} \circ \beta_{(1)}^{\circ j_1} \circ \cdots \circ \beta_{(\ell)}^{\circ j_\ell},$$

$$(e_1\alpha\beta)^{W,Y} = e_1^{\circ y_{-1}} \circ \alpha_{(w_1)} \circ \alpha_{(w_2)} \circ \cdots \circ \alpha_{(w_q)} \circ \beta_{(0)}^{\circ y_0} \circ \beta_{(1)}^{\circ y_1} \circ \cdots \circ \beta_{(r)}^{\circ y_r},$$

$$|I| = k, \quad |W| = q \quad \|J\| = \Sigma j_n, \quad \|Y\| = \Sigma y_n.$$

Theorem 6.7 *We have*

$$H_\star K_{\sigma+i} \cong E[(e_1\alpha\beta)^{I,J} \circ \bar{\alpha}_{(m)}, (e_1\alpha\beta)^{W,Y} \circ \bar{\beta}_{(t)}]$$

where $m > i_k$ and $m \geq \ell$, $t > w_q$ and $t \geq y_r$, $|I| + 2\|J\| = i$ and $|W| + 2\|Y\| = i - 1$, and the coproduct follows by Hopf ring properties from the $\alpha_{(i)}$'s, $\beta_{(i)}$'s, $\bar{\alpha}_{(i)}$'s, and $\bar{\beta}_{(i)}$'s.

We observe that this equivariant answer mirrors the classical nonequivariant answer for odd primes [Ravenel and Wilson 1980]. For the reader's convenience, we explicitly write some low-dimensional instances of the theorem. In particular,

$$H_\star K_\rho \cong E[e_1 \circ \bar{\alpha}_{(i)}, \alpha_{(i_1)} \circ \bar{\alpha}_{(i_2)}, \bar{\beta}_{(i)}]$$

and

$$H_\star K_{\sigma+2} \cong E[e_1 \circ \alpha_{(i_1)} \circ \bar{\alpha}_{(i_2)}, \alpha_{(i_1)} \circ \alpha_{(i_2)} \circ \bar{\alpha}_{(i_3)}, e_1 \circ \bar{\beta}_{(i_1)}, \beta_{(j_1)} \circ \bar{\alpha}_{(j_2)}, \alpha_{(i_1)} \circ \bar{\beta}_{(i_2)}]$$

where $i_1 < i_2$, $j_1 \leq j_2$; and the coproduct follows by Hopf ring properties from the $\alpha_{(i)}$'s, $\beta_{(i)}$'s, $\bar{\alpha}_{(i)}$'s, and $\bar{\beta}_{(i)}$'s.

Having computed the homology of the C_2–equivariant Eilenberg–Mac Lane spaces K_V, we turn to using the results to investigate the twisted bar spectral sequence arising from the twisted bar construction. Unlike the nonequivariant bar spectral sequence, the twisted bar spectral sequence E^2 page lacks an explicit homological description. This makes computations difficult in general. However, for the spaces $B^\sigma \underline{\mathbb{F}}_2 \simeq K_\sigma \simeq \mathbb{R}P_{\mathrm{tw}}^\infty$, $B^\sigma S^1 \simeq K(\mathbb{Z}, \rho) \simeq \mathbb{C}P_{\mathrm{tw}}^\infty$, and $B^\sigma S^\sigma \simeq K(\mathbb{Z}, 2\sigma)$, there is a gap in the spectral sequence forcing all differentials d^r for $r > 1$ to be zero. Further for these spaces, if there were a nonzero d^1 differential, we would end up killing a known generator of the underlying nonequivariant integer graded homology and arrive at a contradiction. Thus we can calculate the additive $RO(C_2)$–graded homology of these spaces completely. The multiplicative structure can also be deduced from the twisted bar spectral sequence.

Example 6.10 *We have*

$$H_\star \mathbb{R}P_{\mathrm{tw}}^\infty = E[e_\sigma, \bar{\alpha}_{(0)}, \bar{\alpha}_{(1)}, \ldots] = E[e_\sigma] \otimes \Gamma[\bar{\alpha}_{(0)}], \quad |e_\sigma| = \sigma, \ |\bar{\alpha}_{(i)}| = \rho 2^i,$$
$$H_\star \mathbb{C}P_{\mathrm{tw}}^\infty = E[\bar{\beta}_{(0)}, \bar{\beta}_{(1)}, \ldots] = \Gamma[e_\rho] \quad \text{where } |\bar{\beta}_{(i)}| = \rho 2^i.$$

Theorem 6.11 *We have*

$$H_\star K(\mathbb{Z}, 2\sigma) = E[e_{2\sigma}] \otimes \Gamma[\bar{x}_{(0)}] \quad \text{where } |e_{2\sigma}| = 2\sigma, \ |\bar{x}_{(0)}| = 2\rho.$$

Remark 1.1 The spaces $B^\sigma \underline{\mathbb{F}}_2 \simeq K_\sigma \simeq \mathbb{R}P_{\mathrm{tw}}^\infty$ and $B^\sigma S^1 \simeq K(\mathbb{Z}, \rho) \simeq \mathbb{C}P_{\mathrm{tw}}^\infty$ have well-known models arising as colimits of C_2–equivariant Grassmanian manifolds. In particular, if $\mathbb{R}^{i+j\sigma}$ is the real C_2–representation composed of a direct sum of i copies of the trivial representation and j copies of the sign representation, and the complex C_2–representation $\mathbb{C}^{i+j\sigma}$ is defined similarly, then $\mathbb{R}P_{\mathrm{tw}}^\infty$ is the colimit of the natural cellular inclusions

$$\cdots \hookrightarrow \mathbb{P}(\mathbb{R}^{1+\sigma}) \hookrightarrow \mathbb{P}(\mathbb{R}^{2+\sigma}) \hookrightarrow \mathbb{P}(\mathbb{R}^{2+2\sigma}) \hookrightarrow \cdots$$

and $\mathbb{C}P_{\mathrm{tw}}^\infty$ is the colimit of the natural cellular inclusions

$$\cdots \hookrightarrow \mathbb{P}(\mathbb{C}^{1+\sigma}) \hookrightarrow \mathbb{P}(\mathbb{C}^{2+\sigma}) \hookrightarrow \mathbb{P}(\mathbb{C}^{2+2\sigma}) \hookrightarrow \cdots.$$

In contrast, the space $B^\sigma S^\sigma \simeq K(\mathbb{Z}, 2\sigma)$ remains more mysterious. The author does not know of any models for this space besides applying the twisted bar construction to S^σ.

In forthcoming work, we will use the homology of $H_\star K_V$ to deduce differentials in the twisted bar spectral sequence. The beginning stages of this work are described in Section 6.

1.2 Paper structure

This paper has two primary aims: extending Ravenel–Wilson Hopf ring techniques [Ravenel and Wilson 1977; 1980; Wilson 1982] to C_2–equivariant homotopy theory, and computing the $RO(C_2)$–graded homology of C_2–equivariant Eilenberg–Mac Lane spaces associated to the constant Mackey functor $\underline{\mathbb{F}}_2$. These topics are investigated in several sections.

The first section consists of an introduction providing context for the main results, a description of the paper structure, and a list of notational conventions.

The second section recalls classical Ravenel–Wilson Hopf ring methods.

The third section recollects material from equivariant homotopy theory necessary for understanding our proof and computations.

The fourth section details the bar and twisted bar constructions, which are trivial and sign representation delooping functors respectively.

The fifth section applies the preliminaries of the previous sections to study multiplicative structures on C_2–equivariant Eilenberg–Mac Lane spaces. This section contains some primary extensions of Ravenel–Wilson Hopf ring methods to C_2–equivariant homotopy theory (Theorem 5.4). It also contains our calculation of the $RO(C_2)$–graded homology of many C_2–equivariant Eilenberg–Mac Lane spaces K_V associated to the constant Mackey functor $\mathbb{F}_2$ (Theorems 5.6, 6.6, and 6.7).

The sixth section details a number of computations and observations regarding the $RO(C_2)$–graded bar and twisted bar spectral sequences. The examples we provide should be a useful stepping stone towards further computations.

The seventh section describes a few questions of immediate interest given the results of this paper.

1.3 Notational conventions

- The asterisk $*$ denotes integer grading.
- The star $\star$ denotes representation grading.

- By the classical or nonequivariant Eilenberg–Mac Lane space K_n, we mean the classical nonequivariant Eilenberg–Mac Lane space $K_n = K(\mathbb{F}_p, n)$, where p is prime.
- C_2 is the cyclic group of order two with $C_2 = \langle \gamma \rangle$.
- σ denotes the one-dimensional sign representation of C_2.
- ρ is the regular representation of C_2.
- S^V is the one-point compactification of a finite-dimensional real representation V where the point at infinity is given a trivial group action and taken as the base point.
- $\Sigma^V(-) = S^V \wedge -$.
- $\Omega^V(-)$ is the space of continuous based maps $\mathrm{Map}_*(S^V, -)$ where the group action is given by conjugation.
- $\mathscr{S}$ is the category of spectra.
- $\mathscr{S}^G$ is the category of G–spectra indexed on a complete universe.

Acknowledgements

2 Classical Ravenel–Wilson Hopf ring methods

Classically, one place Hopf rings arise in homotopy theory is in the study of Ω–spectra. Consider an Ω–spectrum

$$G = \{G_k\}$$

and a multiplicative homology theory $E_*(-)$ with a Künneth isomorphism for the spaces G_k. The Ω–spectrum G represents a generalized cohomology theory with

$$G^* X \simeq [X, G_*].$$

Since $G^k X$ is an abelian group, G_k must be a homotopy commutative H–space (in fact G_k is an infinite loop space). This H–space structure

$$*: G_k \times G_k \to G_k$$

gives rise to a product in homology

$$*: E_* G_k \otimes E_* G_k \cong E_*(G_k \times G_k) \to E_* G_k$$

and the Künneth isomorphism implies the homology is in fact a Hopf algebra.

If G is a ring spectrum, then $G^* X$ is a graded ring and the graded abelian group object G_* becomes a graded ring object in the homotopy category. The multiplication

$$G^k X \times G^n X \to G^{k+n} X$$

has a corresponding multiplication in G_*,

$$\circ \colon G_k \times G_n \to G_{k+n},$$

and applying $E_*(-)$ we have

$$\circ \colon E_* G_k \otimes_{E_*} E_* G_n \to E_* G_{k+n}$$

turning $E_* G$ into a graded ring object in the category of coalgebras.

As a ring, $E_* G$ has a distributive law,

(2.1) $$x \circ (y * z) = \sum \pm (x' \circ y) * (x'' \circ z) \quad \text{where } \psi(x) = \sum x' \otimes x'',$$

coming from the distributive law in $G^* X$.

Ravenel and Wilson pursued the idea that these two products could be used to construct many elements in homology from just a few. They successfully applied this approach to compute the Hopf ring for complex cobordism [Ravenel and Wilson 1977], the Morava K–theory of nonequivariant Eilenberg–Mac Lane spaces [Ravenel and Wilson 1980], and the mod p homology of classical Eilenberg–Mac Lane spaces [Wilson 1982].

In the case of classical Eilenberg–Mac Lane spaces, the Eilenberg–Mac Lane spectrum

$$H\mathbb{F}_p = \{K(\mathbb{F}_p, n)\} = \{K_n\}$$

is a ring spectrum with $\Omega K_{n+1} \simeq K_n$. Further, $H_*(-) := H_*(-; \mathbb{F}_p)$, ordinary homology with mod p coefficients, has a Künneth isomorphism and thus the homology $H_* K_*$ has the structure of a Hopf ring.

A key computational insight of Ravenel and Wilson was that the bar spectral sequence

$$E^2_{*,*} \simeq \operatorname{Tor}^{E_* G_k}_{*,*}(E_*, E_*) \Rightarrow E_* G_{k+1}$$

is in fact a spectral sequence of Hopf algebras. The additional structure of the $\circ$ multiplication in the bar spectral sequence meant that they could inductively deduce the homology of Eilenberg–Mac Lane spaces using standard homological algebra. Starting with elements in $H_* K_1$ and $H_* \mathbb{C}P^\infty$ and identifying circle products in the bar spectral sequence, Ravenel and Wilson computed the Hopf ring associated to the mod p Eilenberg–Mac Lane spectrum [Wilson 1982].

To describe their answer, let

$$e_1 \in H_1 K_1, \quad \alpha_i \in H_{2i} K_1, \quad \beta_i \in H_{2i} \mathbb{C}P^\infty, \quad i \geq 0.$$

The generators are

$$e_1, \quad \alpha_{(i)} = \alpha_{p^i}, \quad \beta_{(i)} = \beta_{p^i}$$

with coproduct

$$\psi(\alpha_n) = \sum_{i=0}^{n} \alpha_{n-i} \otimes \alpha_i, \quad \psi(\beta_n) = \sum_{i=0}^{n} \beta_{n-i} \otimes \beta_i.$$

For finite sequences,

$$I = (i_1, i_2, \ldots), \quad 0 \le i_1 < i_2 < \cdots,$$
$$J = (j_0, j_1, \ldots), \quad j_k \ge 0,$$

define

$$\alpha_I = \alpha_{(i_1)} \circ \alpha_{(i_2)} \circ \cdots, \quad \beta^J = \beta_{(0)}^{\circ j_0} \circ \beta_{(1)}^{\circ j_1} \circ \cdots,$$

and let $T(x)$ denote the truncated polynomial algebra $\mathbb{F}_p[x]/(x^p)$.

Theorem A (Ravenel and Wilson [Wilson 1982]) *We have*

$$H_* K_* \simeq \otimes_{I,J} E(e_1 \circ \alpha_I \circ \beta^J) \otimes_{I,J} T(\alpha_I \circ \beta^J)$$

as an algebra where the tensor product is over all I and J and the coproduct follows by Hopf ring properties from the α's and β's.

When the prime $p = 2$, there are additional relations $e_1 \circ e_1 = \beta_{(0)}$ and $\alpha_{(i-1)} \circ \alpha_{(i-1)} = \beta_{(i)}$. In this case, the theorem can be stated using only circle products of generators of $\mathbb{R}P^\infty$.

For finite sequences

$$I = (i_{(-1)}, i_0, i_1, i_2, \ldots), \quad i_k \ge 0,$$

define

$$(e_1 \alpha)^I = e_1^{\circ i_{(-1)}} \circ \alpha_{(0)}^{\circ i_0} \circ \alpha_{(1)}^{\circ i_1} \circ \cdots.$$

Theorem B (Ravenel and Wilson [Wilson 1982]) *Then*

$$H_* K_n \cong \otimes_I E[(e_1 \alpha)^I],$$

where $\sum i_k = n$, and considering all spaces at once,

$$H_* K_* \simeq \otimes_I E[(e_1 \alpha)^I]$$

as an algebra where the tensor product is over all I and the coproduct follows by Hopf ring properties from the α's.

Ravenel and Wilson also show that homology suspending $\beta_{(i)}$ to define

$$\xi_i \in H_{2(p^i-1)} H,$$

and $\alpha_{(i)}$ to define

$$\tau_i \in H_{2p^i-1} H.$$

Theorem A then implies that stably,

$$H_* H \simeq E[\tau_0, \tau_1, \ldots] \otimes P[\xi_1, \xi_2, \ldots].$$

3 Equivariant preliminaries

We set notation and recall equivariant foundations. Throughout, the group $G = C_2$.

Given an orthogonal real G–representation V, S^V denotes the representation sphere given by the one-point compactification of V. For a p–dimensional real C_2–representation V, we write

$$V \cong \mathbb{R}^{(p-q,0)} \oplus \mathbb{R}^{(q,q)}$$

where $\mathbb{R}^{(1,0)}$ is the trivial 1–dimensional real representation of C_2 and $\mathbb{R}^{(1,1)}$ is the sign representation. We allow p and q to be integers, so V may be a virtual representation. The integer p is called the topological dimension while q is the weight or twisted dimension of $V \cong \mathbb{R}^{(p,q)}$.

The Vth graded component of the ordinary $RO(C_2)$–graded Bredon equivariant homology of a C_2–space X with coefficients in the constant Mackey functor $\underline{\mathbb{F}}_2$ is denoted $H_V^{C_2}(X; \underline{\mathbb{F}}_2) = H_{p,q}(X; \underline{\mathbb{F}}_2)$. To consider all representations at once we write $H_\star(X)$, and when working nonequivariantly $H_*(X^e)$ denotes the singular homology of the underlying topological space with $\mathbb{F}_2$ coefficients.

It is often convenient to plot the bigraded homology in the plane. Our plots have topological dimension p on the horizontal axis and weight q on the vertical axis.

The homology of a point with coefficients in the constant Mackey functor $\underline{\mathbb{F}}_2$, is the bigraded ring

$$H_\star(\text{pt}, \underline{\mathbb{F}}_2) = \mathbb{F}_2[a, u] \oplus \frac{\mathbb{F}_2[a, u]}{(a^\infty, u^\infty)}\{\theta\}$$

where $|a| = -\sigma$, $|u| = 1 - \sigma$, and $|\theta| = 2\sigma - 2$. A bigraded plot of $H_\star(\text{pt}, \underline{\mathbb{F}}_2)$ appears in Figure 1. The image on the left is more detailed with each lattice point within the two cones representing a copy of $\mathbb{F}_2$. The image on the right is a more succinct representation and appears in figures illustrating our spectral sequence computations.

The genuine equivariant Eilenberg–Mac Lane spectrum representing $H_\star(-)$ is $H\underline{\mathbb{F}}_2$, the Eilenberg–Mac Lane spectrum for the C_2 constant Mackey functor $\underline{\mathbb{F}}_2$. It has underlying nonequivariant spectrum $H\mathbb{F}_2$. We denote the spaces of $H\underline{\mathbb{F}}_2$ by

$$H\underline{\mathbb{F}}_2 = \{K(\underline{\mathbb{F}}_2, V)\}_{V \cong k\sigma+l} = \{K_V\}_{V \cong k\sigma+l}.$$
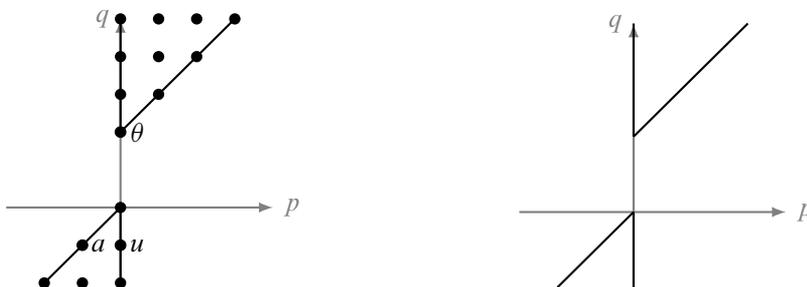
Figure 1: $H_\star(\text{pt}, \underline{\mathbb{F}}_2)$ with axis gradings determined by $V \simeq \mathbb{R}^{p-q} \oplus \mathbb{R}^{q\sigma}$.

Analogously to the nonequivariant case, $H\mathbb{F}_2$ is characterized up to C_2–equivariant homotopy by $H^V(X; \mathbb{F}_2) = [X, K_V]$ naturally for all C_2–spaces X.

We recall a computational lemma due to Behrens and Wilson [2018], which allows us to check whether a set of elements in the $RO(C_2)$–homology in fact forms a free basis for $H_\star(X)$, greatly simplifying our computations. To state this lemma, we first define two homomorphisms, Φ^e and Φ^{C_2}. Let Ca be the cofiber of the Euler class $a \in \pi^{C_2}_{-\sigma} S$ given geometrically by the inclusion

$$S^0 \hookrightarrow S^\sigma.$$

Applying $\pi^{C_2}_V$ to the map

$$H \wedge X \to H \wedge X \wedge Ca,$$

we get a homomorphism

$$\Phi^e \colon H_V(X) \to H_{|V|}(X^e).$$

Taking geometric fixed points of a map

$$S^V \to H \wedge X$$

gives a map

$$S^{V^{C_2}} \to H^{\Phi C_2} \wedge X^{\Phi C_2}.$$

Using the equivalence $H^\Phi_* X \simeq H_*(X^{\Phi C_2})[a^{-1}u]$ coming from $H^{\Phi C_2} \simeq \bigvee_{i \geq 0} \Sigma^i H\mathbb{F}_2$ and passing to the quotient by the ideal generated by $a^{-1}u$ gives the homomorphism

$$\Phi^{C_2} \colon H_V(X) \to H_{|V^{C_2}|}(X^{\Phi C_2}).$$

Lemma 3.1 [Behrens and Wilson 2018] *Suppose $X \in Sp^{C_2}$ and $\{b_i\}$ is a set of elements of $H_\star(X)$ such that*

(1) $\{\Phi^e(b_i)\}$ *is a basis of $H_*(X^e)$ and*

(2) $\{\Phi^{C_2}(b_i)\}$ *is a basis of $H_*(X^{\Phi C_2})$.*

Then $H_\star(X)$ is free over $H_\star$ and $\{b_i\}$ is a basis.

We use the following notation for $H^\star K_\sigma$.

Theorem 3.2 [Hu and Kriz 2001] $H^\star(\mathbb{R}P^\infty_{\mathrm{tw}}) = H^\star(\mathrm{pt})[\alpha, \beta]/(\alpha^2 = a\alpha + u\beta)$ *where* $|\alpha| = \sigma$, $|\beta| = \rho$, $|a| = \sigma$, *and* $|u| = \sigma - 1$.

Since this cohomology is free, the homology $H_\star K_\sigma$ immediately follows. In our notation we have elements

$$e_\sigma \in H_\sigma K_\sigma, \quad \bar{\alpha}_i \in H_{\rho i} K_\sigma \quad (i \geq 0).$$

The generators are

$$e_\sigma, \quad \bar{\alpha}_{(i)} = \bar{\alpha}_{2i} \quad (i \geq 0)$$

with coproduct

$$\psi(e_\sigma) = 1 \otimes e_\sigma + e_\sigma \otimes 1 + a(e_\sigma \otimes e_\sigma),$$

$$\psi(\bar{\alpha}_n) = \sum_{i=0}^{n} \bar{\alpha}_{n-i} \otimes \bar{\alpha}_i + \sum_{i=0}^{n-1} u(e_\sigma \bar{\alpha}_{n-1-i} \otimes e_\sigma \bar{\alpha}_i),$$

and ring structure $H_\star K_\sigma \simeq E[e_\sigma, \bar{\alpha}_{(i)}]$ which can be deduced from the twisted bar spectral sequence computing $H_\star B^\sigma \mathbb{F}_2 \cong H_\star \mathbb{R}P^\infty_{\mathrm{tw}}$.

We also require notation for $H_\star K(\mathbb{Z}, \rho)$. This can be deduced by applying the $RO(C_2)$–graded bar spectral sequence to S^σ. Let

$$\bar{\beta}_i \in H_{\rho i} K(\mathbb{Z}, \rho) \quad (i \geq 0).$$

The generators are

$$\bar{\beta}_{(i)} = \bar{\beta}_{2^i} \quad (i \geq 0)$$

with coproduct

$$\psi(\bar{\beta}_n) = \sum_{i=0}^{n} \bar{\beta}_{n-i} \otimes \bar{\beta}_i$$

and ring structure

$$H_\star K(\mathbb{Z}, \rho) \simeq E[\beta_{(i)}].$$

3.1 The fixed point spaces of C_2–equivariant Eilenberg–Mac Lane spaces

It is useful to understand the C_2 fixed points of the C_2–equivariant Eilenberg–Mac Lane spaces K_V in applications of the Behrens–Wilson computational lemma. We state a relevant proposition due to Caruso.

Proposition 3.3 [Caruso 1999] *Let $G = C_p$ and V be an n–dimensional fixed point free virtual representation of G with $n > 0$ and m an integer. Then*

$$K(\underline{\mathbb{F}}_p, m + V)^{C_p} \simeq K(\mathbb{F}_p, m) \times \cdots \times K(\mathbb{F}_p, m + n).$$

3.2 Notation for the underlying nonequivariant homology of $K_V^{C_2}$

To use the Behrens–Wilson lemma, we also need to understand the homology of the fixed point spaces. Applying Theorem B to the nonequivariant homology of $(K_{n\sigma})^{C_2}$ gives

$$H_*(K_{n\sigma}^{C_2}) \simeq E[e_0, a_{(i_1)}, a_{(i_1)} \circ a_{(i_2)}, \ldots, a_{(i_1)} \circ \cdots \circ a_{(i_n)}]$$

where $0 \leq i_1 \leq i_2 \leq \cdots \leq i_n$, $|e_0| = 0$, and $|a_{(i)}| = 2^i$.

4 Bar and twisted bar constructions

A first task in implementing the Ravenel–Wilson Hopf ring approach is to generalize the bar spectral sequence to the C_2–equivariant case. In the classical story, the bar spectral sequence is used to inductively compute the homology of $K_n \simeq BK_{n-1}$ from $H_* K_{n-1}$. In the C_2–equivariant setting, our spaces K_V are bigraded on the trivial and sign representations of C_2. Due to this new grading, we should now

additionally compute the homology of $K_{V+\sigma}$ inductively from $H_\star K_V$. In order to do so, we need a good model of σ–delooping. We begin by reviewing the classical bar construction which is a trivial representation delooping functor.

Construction 4.1 (Classical bar construction) *For a topological monoid A, the pointed space BA is defined as a quotient*

$$BA = \coprod_n \Delta^n \times A^{\times n} / \sim$$

where the relation $\sim$ is generated by

(1) $(t_1, \ldots t_n, a_1, \ldots, a_n) \sim (t_1, \ldots, \hat{t}_i, \ldots, t_n, a_1, \ldots, \hat{a}_i, (a_i a_{i+1}), \ldots, a_n)$ *if $t_i = t_{i+1}$ or $a_i = *$;*

(2) *for $i = n$, delete the last coordinate if $t_n = 1$ or $a_n = *$; for $i = 0$, delete the first coordinate if $t_0 = -1$ or $a_0 = *$; and Δ^n denotes the topological simplex*

$$\Delta^n = \{(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n \mid -1 \leq t_1 \leq \cdots \leq t_n \leq 1\}.$$

Remark 4.2 We use the slightly nonstandard topological n–simplex

$$\Delta^n = \{(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n \mid -1 \leq t_1 \leq \cdots \leq t_n \leq 1\}$$

so that when we introduce a C_2 action, the simplex rotates around the origin. This makes writing down a model for the H–space structure on the C_2–equivariant Eilenberg–Mac Lane spaces K_V more straightforward.

Given a commutative monoid A, we observe that BA is also a commutative monoid via the pairing

$$*: BX \times BX \to BX$$

defined by

$$(t_1, \ldots, t_n, x_0, \ldots, x_n) *_\sigma (t_{n+1}, \ldots, t_{n+m}, x_{n+1}, \ldots, x_{n+m}) = (t_{\tau(1)}, \ldots, t_{\tau(n+k)}, x_{\tau(1)}, \ldots, x_{\tau(n+m)}),$$

where τ is any element of the symmetric group on $n + k$ letters such that $t_{\tau(i)} \leq t_{\tau(i+1)}$. This pairing was first described by Milgram [1967].

Definition 4.3 [Liu 2020] A C_2–space A is a twisted monoid if it is a topological monoid in the nonequivariant sense with the product satisfying $\gamma(xy) = \gamma(y)\gamma(x)$ where $C_2 \simeq \langle \gamma \rangle$.

Construction 4.4 [Liu 2020] *For any twisted monoid A, construct $B_*^\sigma A$ in the same way as the nonequivariant bar construction, that is such that $B_n^\sigma A = \Delta^n \times A^n$. However, define a C_2–action on A^n by*

$$\gamma(a_1, a_2, \ldots, a_n) = (\gamma a_n, \gamma a_{n-1}, \ldots, \gamma a_1).$$

Then the C_2–actions commute with the face and degeneracy maps as $\gamma \circ s_i = s_{n-i} \circ \gamma$ and $\gamma \circ d_i = d_{n-i} \circ \gamma$. Further, define the C_2–action on each

$$\Delta^n = \{(t_1, t_2, \ldots, t_n) \in \mathbb{R}^{t+1} \mid -1 \leq t_1 \leq \cdots \leq t_n \leq 1\}.$$

by $\gamma(t_1, t_2, \ldots, t_n) = (-t_n, -t_{n-1}, \ldots, -t_1)$). *Then define $B^\sigma A$ to be the geometric realization*

$$\coprod \Delta^n \times A^n / \sim .$$

Example 4.5 The space $B^\sigma K_0 \simeq \mathbb{R}P^\infty_{\text{tw}}$ is the space of lines through a direct sum of an infinite number of copies of the C_2–regular representation ρ.

We can inductively define an H–space pairing on $B^l B^{k\sigma} \mathbb{F}_2$, similar to the one given by Milgram in the nonequivariant case. Define a mapping

$$*_\sigma \colon B^\sigma X \times B^\sigma X \to B^\sigma X$$

by

$$(t_0, \ldots, t_n, x_0, \ldots, x_n) *_\sigma (t_{n+1}, \ldots, t_{n+m}, x_{n+1}, \ldots, x_{n+m}) = (t_{\tau(1)}, \ldots, t_{\tau(n+k)}, x_{\tau(1)}, \ldots, x_{\tau(n+m)}),$$

where τ is any element of the symmetric group on $n + k$ letters such that $t_{\tau(i)} \le t_{\tau(i+1)}$. Then $*_\sigma$ is well defined, continuous, and C_2–equivariant. Going forward, we suppress the σ notation in $*_\sigma$, using only $*$ to denote the H–space pairing. The relevant C_2–action is deduced from context.

Definition 4.6 A G–space X is said to be G–*connected* if X^H is connected for each subgroup H of G.

Proposition 4.7 [Liu 2020] *For any commutative monoid A in the category of based C_2–spaces, the V–degree bar construction $B^V A$ is defined by applying the ordinary bar construction l times and the twisted bar construction m times for $V = l + m\sigma$. There exists a natural map $A \to \Omega^V B^V A$. When A is C_2–connected, this map is a C_2–equivalence.*

5 Multiplicative structures on C_2–equivariant Eilenberg–Mac Lane spaces

We describe multiplicative structures on C_2–equivariant Eilenberg–Mac Lane spaces, extending Ravenel and Wilson's description of similar structures on classical nonequivariant Eilenberg–Mac Lane spaces. We use our understanding of these structures to compute the $RO(C_2)$–graded homology of many C_2–equivariant Eilenberg–Mac Lane spaces K_V associated to the constant Mackey functor $\mathbb{F}_2$. In particular, we compute the $RO(C_2)$–graded homology of all C_2–equivariant Eilenberg–Mac Lane spaces $K_{*\sigma}$ and $K_{\sigma+*}$.

5.1 Multiplicative structures on K_V

The $RO(C_2)$–graded cup product is induced by a map

(5.1) $$\circ = \circ_{V,W} \colon K_V \wedge K_W \to K_{V+W}.$$

We will construct $\circ_{V,W}$ explicitly within the framework of trivial and σ–representation delooping given by B and B^σ. We will also discuss how $\circ_{V,W}$ descends to a product on the fixed points.

Given a real C_2 representation $V \cong l + k\sigma$, the Eilenberg–Mac Lane space K_V is a V–fold delooping of $\mathbb{F}_2$ and therefore can be constructed iteratively by taking $B^l B^{k\sigma} \mathbb{F}_2$ where l and k are nonnegative integers. The following construction extends exposition by Ravenel and Wilson [1980] in their computation of the Morava K–theory of Eilenberg–Mac Lane spaces.

We construct the map (5.1) inductively on V. Assuming $\circ_{V,W}$ has been defined, we define $\circ_{V+1,W}$ and $\circ_{V+\sigma,W}$ by replacing K_{V+1}, K_{V+W+1} and $K_{V+\sigma}, K_{V+W+\sigma}$ with their bar and twisted bar constructions respectively. In both cases this is denoted as follows; there is a notationally suppressed C_2–action each case:

$$\textbf{(5.2)} \qquad \left\{ \coprod_n \Delta^n \times K_V^n / \sim \right\} \wedge K_W \to \left\{ \coprod_n \Delta^n \times K_{V+W} / \sim \right\}.$$

Let $t \in \Delta^n$, $x = (x_0, \ldots, x_n) \in K_V$, and $y \in K_W$. The image of $x_i \wedge y \in K_V \wedge K_W$ under the map (5.1) is denoted $x_i \circ y$. We use the notation $x \circ y$ to mean $(x_0 \circ y, \ldots, x_n \circ y)$. Define (5.2) by

$$\textbf{(5.3)} \qquad \{(t, x)\} \circ y = \{(t, x \circ y)\}.$$

Theorem 5.4 *The above construction is well defined and gives the cup product pairings*

$$\circ \colon K_{V+1} \wedge K_W \to K_{V+W+1}, \quad \circ \colon K_{V+\sigma} \wedge K_W \to K_{V+W+\sigma}.$$

Lemma 5.5 *The map $\circ \colon K_0 \times K_V \to K_V$ is given by $(q) \circ x = x^{*q}$ where $q \in \mathbb{F}_2$.*

Proof This map multiplies $\pi_V^{C_2} K_V \simeq \mathbb{F}_2$ by q which is what $\circ$ should do restricted to $(q) \times K_V \simeq K_V$. $\square$

Proof of Theorem 5.4 We must show the map (5.2) defined by (5.3) is well defined and in fact gives the cup product pairings $\circ \colon K_{V+1} \wedge K_W \to K_{V+W+1}$ and $\circ \colon K_{V+\sigma} \wedge K_W \to K_{V+W+\sigma}$. Our proof is a direct extension of the nonequivariant argument of Ravenel and Wilson [1980]. We prove our result by induction on i in the σ direction noting that the result also holds and is similar in the trivial representation direction (that is we assume the statement holds for V, and show it for $V + \sigma$). Assume we have proved Theorem 5.4 for $K_V \wedge K_W \to K_{V+W}$ with Lemma 5.5 beginning the induction. We need our construction to satisfy

$$(z_1 * z_2) \circ y = (z_1 \circ y) * (z_2 \circ y).$$

For $i = 0$, $z_i = q_i \in \mathbb{F}_2 = K_0$. So,

$$(q_1 * q_2) \circ y = (q_1 + q_2) \circ y = y^{q_1 + q_2} = y^{*q_1} * y^{*q_2} = (q_1 \circ y) * (q_2 \circ y).$$

For $i > 0$,

$$\begin{aligned}
[z_1 * z_2] \circ y &= [(t, x) * (t_{n+1}, \ldots, t_{n+k}; x_{n+1}, \ldots, x_{n+k})] \circ y \\
&= (t_{\tau(1)}, \ldots, t_{\tau(n+k)}; x_{\tau(1)}, \ldots, x_{\tau(n+k)}) \circ y \\
&= (t_{\tau(1)}, \ldots, t_{\tau(n+k)}; x_{\tau(1)} \circ y, \ldots, x_{\tau(n+k)} \circ y) \\
&= (t; x \circ y)) * (t_{n+1}, \ldots, t_{n+k}; x_{n+1} \circ y, \ldots, x_{n+k} \circ y) \\
&= (z_1 \circ y) * (z_2 \circ y),
\end{aligned}$$

where the second line is due to the definition of $*$, the third is due to the induction hypothesis and (5.3), and the fourth is due to the definition of $*$. □

We must show (5.3) gives well-defined maps $K_{V+1} \wedge K_W \to K_{V+W+1}$ and $K_{V+\sigma} \wedge K_W \to K_{V+W+\sigma}$. The relations in the (twisted) bar construction make this the case. We show the main case, leaving the others to the reader. Assume $0 \le q < n$ with $t_q = t_{q+1}$ or $x_q = *$. Then

$$(t, x) \circ y = (t, x \circ y)$$
$$\sim (t_1, \ldots, \hat{t}_q, \ldots, t_n; x_1 \circ y, \ldots, (x_q \circ y) * (x_{q+1} \circ y), \ldots, x_n \circ y)$$
$$= (t_1, \ldots, \hat{t}_q, \ldots, t_n; x_1 \circ y, \ldots, (x_q * x_{q+1}) \circ y, \ldots, x_n \circ y)$$
$$= (t_1, \ldots, \hat{t}_q, \ldots, t_n; x_1, \ldots, x_q * x_{q+1}, \ldots, x_n) \circ y,$$

which is the necessary relation. That this map factors through the smash product is straightforward to verify using induction.

The remaining task is to show that this is the cup product pairing map. This follows by induction from the observation that $\circ$ commutes with (signed) suspension on the first factor since $B_1 K_V \simeq S^1 \wedge K_V$ and $B_1^\sigma K_V \simeq S^\sigma \wedge K_V$, and following diagrams commute:

$$
\begin{array}{ccc}
S^1 \wedge K_V \wedge K_W & \longrightarrow & S^1 \wedge K_{V+W} \\
\downarrow & & \downarrow \\
K_{V+1} \wedge K_W & \longrightarrow & K_{V+W+1}
\end{array}
\qquad
\begin{array}{ccc}
S^\sigma \wedge K_V \wedge K_W & \longrightarrow & S^\sigma \wedge K_{V+W} \\
\downarrow & & \downarrow \\
K_{V+\sigma} \wedge K_W & \longrightarrow & K_{V+W+\sigma}
\end{array}
$$

5.2 Multiplicative structures on $K_V^{C_2}$

We turn to understanding the $\circ$–product on the fixed points of the spaces K_V. Notice $(B^\sigma A)^{C_2}$ consists of points of the form

$$(t_1, \ldots, t_n, 0, -t_n, \ldots, -t_1, a_1, \ldots, a_n, a, \gamma(a_n), \ldots, \gamma(a_1)) \in (B^\sigma A)^{[2n+1]}$$

where $a \in A^{C_2}$ since for

$$(t_1, \ldots, t_m, -t_m, \ldots, -t_1, a_1, \ldots, a_m, \gamma(a_m), \ldots, \gamma(a_1)) \in (B^\sigma A)^{[2n]},$$

there is a degeneracy map inducing an equivalence to

$$(t_1, \ldots, t_n, 0, -t_n, \ldots, -t_1, a_1, \ldots, a_n, *, \gamma(a_n), \ldots, \gamma(a_1)) \in (B^\sigma A)^{[2n+1]}.$$

Taking the fixed points in the construction of map (5.2) we recover the classical nonequivariant $\circ$ product on the fixed point spaces.

5.3 Circle product generators for $H_\star K_{n\sigma}$

Recall that $H\mathbb{F}_2$ has generators $a \in H\mathbb{F}_{2\{-\sigma\}}$ and $u \in H\mathbb{F}_{2\{1-\sigma\}}$. To describe our answer, we recall our notation for $H_\star K_\sigma$. Let

$$e_\sigma \in H_\sigma K_\sigma, \quad \bar{\alpha}_i \in H_{\rho i} K_\sigma \quad (i \ge 0).$$

The homology, $H_\star K_\sigma$, is exterior on generators

$$e_\sigma, \quad \bar{\alpha}_{(i)} = \bar{\alpha}_{2^i} \quad (i \geq 0),$$

with coproduct

$$\psi(e_\sigma) = 1 \otimes e_\sigma + e_\sigma \otimes 1 + a(e_\sigma \otimes e_\sigma),$$

$$\psi(\bar{\alpha}_n) = \sum_{i=0}^{n} \bar{\alpha}_{n-i} \otimes \bar{\alpha}_i + \sum_{i=0}^{n-1} u(e_\sigma \bar{\alpha}_{n-1-i} \otimes e_\sigma \bar{\alpha}_i).$$

For finite sequences

$$J = (j_\sigma, j_0, j_1, \ldots), \quad j_k \geq 0,$$

define

$$(e_\sigma \bar{\alpha})^J = e_\sigma^{\circ j_\sigma} \circ \bar{\alpha}_{(0)}^{\circ j_0} \circ \bar{\alpha}_{(1)}^{\circ j_1} \circ \cdots,$$

where the $\circ$ product comes from the pairing $\circ \colon K_V \wedge K_W \to K_{V+W}$.

Theorem 5.6 *Then*

$$H_\star K_{\ast\sigma} \cong \otimes_J E[(e_\sigma \bar{\alpha})^J]$$

as an algebra, where the tensor product is over all J and the coproduct follows by Hopf ring properties from the $\bar{\alpha}$'s.

Proof For finite sequences

$$J = (j_\sigma, j_0, j_1, \ldots), \quad j_k \geq 0,$$

define $\|J\| = \sum j_k$ (including the σ subscript) and

$$(e_\sigma \bar{\alpha})^J = e_\sigma^{\circ j_\sigma} \circ \bar{\alpha}_{(0)}^{\circ j_0} \circ \bar{\alpha}_{(1)}^{\circ j_1} \circ \cdots.$$

Consider elements $(e_\sigma \bar{\alpha})^J$ with $\|J\| = n$ in the homology of $B^\sigma K_{(n-1)\sigma}$.

To show these elements in fact form a free basis for the homology, we show that they satisfy the conditions of the Behrens–Wilson computational lemma. The map to the underlying homology, $H_\star K_{n\sigma} \to H_\ast K_n$, the underlying homology of $H_\star K_{n\sigma}$, is given by

$$(e_\sigma \bar{\alpha})^J \mapsto (e_1 \alpha)^J.$$

The map on fixed points $H_\star K_{n\sigma} \to H_\ast K_{n\sigma}^{C_2}$ is given by

$$(e_\sigma \bar{\alpha})^J \mapsto e_0^{\circ j_\sigma} \circ a_{(0)}^{\circ j_0} \circ a_{(1)}^{\circ j_1} \circ \cdots.$$

Thus these elements from a free basis for $H_\star K_{n\sigma}$.

We deduce the multiplicative ring structure using a Hopf ring argument due to Ravenel and Wilson [Wilson 1982]. Each $(e_\sigma \bar{\alpha})^J$ can be written as $e_\sigma^{\circ j_\sigma} \circ \bar{\alpha}_{(0)}^{\circ j_0} \circ \bar{\alpha}_{(1)}^{\circ j_1} \circ \cdots \circ \bar{\alpha}_{(n)}^{\circ j_n}$ where n is some nonnegative integer or $n = \sigma$. By the distributive law (2.1),

$$(e_\sigma \bar{\alpha})^J \ast (e_\sigma \bar{\alpha})^J = e_\sigma^{\circ j_\sigma} \circ \bar{\alpha}_{(0)}^{\circ j_0} \circ \bar{\alpha}_{(1)}^{\circ j_1} \circ \cdots \circ (\bar{\alpha}_{(n)} \ast \bar{\alpha}_{(n)}) = 0.$$

The coproduct is induced by the map $K_\sigma \times \cdots \times K_\sigma \to K_{n\sigma}$ which is a map of coalgebras on $H_\star$. □

Remark 5.7 Note that $e_0^{\circ k} = e_0$ for $k > 0$ by Lemma 5.5.

6 Bar and twisted bar spectral sequence computations

The first half of this section focuses on the $RO(C_2)$–graded bar spectral sequence. We describe the d_1–differentials, the Tor term coinciding with the E^2–page, and Hopf ring structure present in the spectral sequences computing $H_\star K_V$ when $\sigma + 1 \subset V$.

In the second half of this section, we study the analogous twisted spectral sequence giving evidence of arbitrarily long equivariant degree shifting differentials appearing computations of the $RO(C_2)$–graded homology of the spaces $K_{*\sigma}$. We describe how these differentials appear to arise in a structured way involving the norm.

6.1 The $RO(C_2)$–graded bar spectral sequence

The $RO(C_2)$–graded bar spectral sequences arises via a filtered complex in the same way as the ordinary integer graded version. The bar construction B on a topological monoid A, is filtered by

$$B^{[t]}A \simeq \coprod_{t \geq n \geq 0} \Delta^n \times A^n / \sim \; \subset BA$$

with associated graded pieces

$$(B^{[t]}A / B^{[t-1]})A \simeq S^t \wedge A^{\wedge t}.$$

Applying $H_\star(-)$ to these filtered spaces gives the $RO(C_2)$–graded bar spectral sequence with E^1–page

$$E_{t,\star}^1 = H_\star(S^t) \otimes H_\star(A)^{\otimes t},$$

computing $H_\star(BA)$. This $RO(C_2)$–graded bar spectral sequence has

$$E_{*,\star}^2 \simeq \mathrm{Tor}_{*,\star}^{H_\star K_V}(H\mathbb{F}_{2\star}, H\mathbb{F}_{2\star}) \Rightarrow H_\star BK_V \cong H_\star K_{V+1}$$

and behaves similarly to the integer graded version in many examples. In particular, the spectral sequences computing the $RO(C_2)$–graded homology of $BS^1 \simeq \mathbb{C}P^\infty$, $BS^\sigma \simeq \mathbb{C}P_{\mathrm{tw}}^\infty$, and $BK_0 \simeq \mathbb{R}P^\infty$ (Example 6.1) collapse for degree reasons.

Example 6.1 We have

$$
\begin{aligned}
H_\star \mathbb{C}P^\infty &= E[\beta_{(0)}, \beta_{(1)}, \ldots] = \Gamma[e_2] &&\text{where } |\beta_{(i)}| = 2^{i+1}, \\
H_\star \mathbb{C}P_{\mathrm{tw}}^\infty &= E[\bar\beta_{(0)}, \bar\beta_{(1)}, \ldots] = \Gamma[e_\rho] &&\text{where } |\bar\beta_{(i)}| = \rho 2^i, \\
H_\star \mathbb{R}P^\infty &= E[e_1, \alpha_{(0)}, \alpha_{(1)}, \ldots] &&\text{where } |e_1| = 1 \text{ and } |\alpha_{(i)}| = 2^{i+1}.
\end{aligned}
$$

Remark 6.2 The relations $e_1 \circ e_1 = e_2 = \beta_1 = \beta_{(0)}$ and $e_1 \circ e_\sigma = e_\rho = \bar\beta_1 = \bar\beta_{(0)}$ in $RO(C_2)$–graded homology are analogous to the classical relation $e_1 \circ e_1 = \beta_1 = \beta_{(0)}$ in nonequivariant integer graded homology (see [Wilson 1982, Proof of 8.5]).
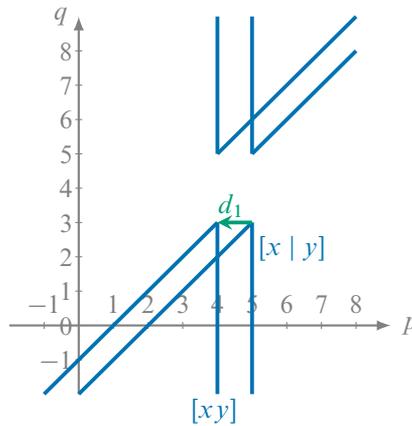
Figure 2: Example: a d_1–differential in the $RO(C_2)$–graded bar spectral sequence.

6.2   The $RO(C_2)$–graded bar spectral sequence: d_1–differentials

The classical bar construction does not introduce any group action; hence the d_1–differentials in the $RO(C_2)$–graded bar spectral sequence behave in almost the same as those in the underlying integer-graded spectral sequence. The difference is that the cycles supporting d_1–differentials in the $RO(C_2)$–graded spectral sequence are representation degree shifted copies of the $RO(C_2)$–graded homology of the point and their targets are the same. This is in contrast with the integer-graded case where the differentials are maps of nongraded rings. For example, all d_1 differentials in the $RO(C_2)$–graded case look and behave like those shown in Figure 2, where the bigraded homology is plotted and the filtration degree is suppressed. We follow this convention for all remaining figures.

In greater specificity, Figure 3 shows a d_1 differential in the $RO(C_2)$–graded bar spectral sequence

$$E^2_{*,\star} \simeq \mathrm{Tor}^{H_\star K_\sigma}_{*,\star}(H\underline{\mathbb{F}}_{2\star}, H\underline{\mathbb{F}}_{2\star}) \Rightarrow H_\star BK_\sigma \cong H_\star K_{\sigma+1}$$
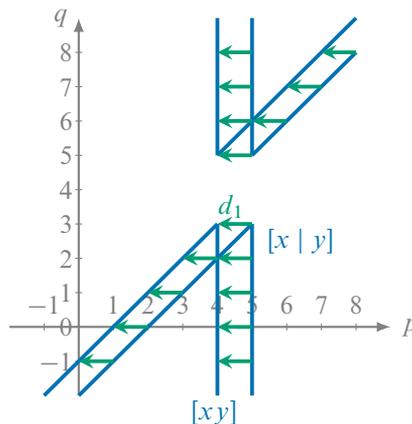
Figure 3: A more detailed picture of a d_1–differential in the $RO(C_2)$–graded bar spectral sequence.

computing the $RO(C_2)$–graded homology of K_ρ. In the figure, $x := e_\sigma$ with $|x| = \sigma$ and $y := \bar{\alpha}_{(0)}$ with $|y| = \rho$. The two double cones shown are supported by the bar representatives $[x\,y]$ and $[x \mid y]$. The d_1–differential maps from the unit of the infinite-dimensional graded ring $H\mathbb{F}_{2\,\star}$ supported by $[x \mid y]$ onto the unit of the $RO(C_2)$–graded homology of a point supported by the bar representative $[x\,y]$. Figure 3 depicts that this map of units in fact induces a map of graded rings surjecting onto the copy of the $RO(C_2)$–graded homology of a point supported by $[x\,y]$.

6.3 Hopf ring structure in the $RO(C_2)$–graded bar spectral sequence and $H_\star K_V$, where $\sigma + 1 \subset V$

In Theorem 5.6, we computed $H_\star K_{n\sigma}$, showing that it is free over $H_\star$. To compute $H_\star K_V$ for real representations $V \cong i + j\sigma$, we consider $\circ$–product structure in the $RO(C_2)$–graded bar spectral sequence

$$E^2_{*,*} \simeq \mathrm{Tor}^{H_\star K_V}_{*,*}(H_\star, H_\star) \Rightarrow H_\star K_{V+1},$$

and observe that theorems of Thomason and Wilson extend directly from the nonequivariant integer graded setting to the C_2–equivariant $RO(C_2)$–graded setting. In Theorem 6.4, we need an additional flatness hypothesis to account for $H_\star(X; \mathbb{F}_2)$ not necessarily being flat, unlike $H_*(X; \mathbb{F}_2)$.

Theorem 6.3 [Thomason and Wilson 1980] *The $\circ$ product factors as*

$$
\begin{array}{ccc}
B_t K_V \times K_W & \longrightarrow & B_t K_{V+W} \\
\cap & & \cap \\
\circ\colon BK_V \times K_W & \longrightarrow & BK_{V+W}
\end{array}
$$

and the map

$$
\begin{array}{ccc}
(B_t K_V / B_{t-1} K_V) \times K_W & \longrightarrow & (B_t K_{V+W} / B_{t-1} K_{V+W}) \\
\wr\vert & & \wr\vert \\
S^t \wedge K_V^{\wedge t} \times K_W & \longrightarrow & S^t \wedge K_{V+W}^{\wedge t}
\end{array}
$$

is described inductively as $(k_1, \ldots, k_t) \circ k = (k_1 \circ k, \ldots, k_t \circ k)$.

Theorem 6.4 [Thomason and Wilson 1980] *Let* $E^r_{*,\star}(E_\star K_V) \Rightarrow E_\star K_{V+1}$ *be the bar spectral sequence and suppose* E^r *is* $H_\star$–*flat for all* $i \leq r$. *Compatible with*

$$\circ\colon E_\star K_{V+1} \otimes_{H_\star} E_\star K_W \to E_\star K_{V+W+1},$$

there is a pairing

(6.5) $$E^r_{t,\star}(E_\star K_V) \otimes_{H_\star} E_\star K_W \to E^r_{t,\star}(E_\star K_{V+W})$$

where $d^r(x) \circ y = d^r(x \circ y)$. *When* $r = 1$ *this pairing is given by*

$$(k_1 | \cdots | k_t) \circ k = \sum \pm (k_1 \circ k' | k_2 \circ k'' | \cdots | k_s \circ k^{(t)})$$

where $k \to \sum k' \otimes k'' \otimes \cdots \otimes k^{(t)}$ *is the iterated reduced coproduct.*

Theorem 6.6 *The $RO(C_2)$–graded homology of K_V, where $\sigma + 1 \subset V$, is exterior on generators given by the cycles on the E^2–page of the $RO(C_2)$–graded bar spectral sequence.*

Proof Let $E^r_{*,\star}(E_\star K_V) \Rightarrow E_\star K_{V+1}$ be the bar spectral sequence and $\Delta \colon K_V \to K_V \times K_V$ be the diagonal map. If E^r is $H_\star$–flat for all $i \leq r$, then there is a natural transformation

$$\mu \colon E^r(X) \otimes E^r(Y) \to E^r(X \times Y)$$

and the coalgebra structure on E^r is given by $\mu^{-1} \Delta_*$.

Suppose $E^r_{*,\star}$ where $r > 2$ is the first page after the $E^2_{*,\star}$–page with a nonzero differential. Then $E^r_{*,\star} = E^2_{*,\star} \cong \mathrm{Tor}^{H_\star K_V}_{*,*}(H_\star, H_\star)$ which is a coalgebra, so μ is an isomorphism and the differentials d_r satisfy the Leibniz and co-Leibniz rules.

Consider the shortest nonzero differential d_r in lowest topological degree. If such a differential exists, it must map from an algebra indecomposable to a coalgebra primitive. To see this, we recount a classical Hopf ring argument, which also appears in [Ravenel and Wilson 1980] and [Angeltveit and Rognes 2005]. Suppose $d_r(xy) \neq 0$ and xy is in lowest topological degree. Then

$$d_r(xy) = d^r(x)y \pm xd_r(y)$$

so $d_r(x)$ or $d_r(y)$ are nonzero, contradicting that xy is in lowest topological degree. Dually, if $d_r(z)$ is not a coalgebra primitive, then

$$\psi(z) = z|1 + 1|z + \Sigma z'_i|z''_i$$

and the co-Leibniz formula

$$\psi \circ d_r = (d_r|1 \pm 1|d_r)\psi$$

implies $d_r(z'_i)$ or $d_r(z''_i)$ is nonzero, contradicting that z is in lowest topological degree.

There are no coalgebra primitives on $E^2_{*,\star} = E^r_{*,\star}$ due to the coproduct structure on $H_\star K_\sigma$. Thus there are no nontrivial differentials and the spectral sequence collapses.

Let x be a cycle on $E^2_{*,\star}$. To show there are no extension problems, we only need to show

$$x * x = 0.$$

The multiplication by 2 map $2 \colon K_V \to K_V$, which factors as the composition

$$K_V \xrightarrow{\Delta} K_V \times K_V \xrightarrow{*} K_V,$$

is homotopically trivial so

$$0 = 2_\star \colon H_\star K_V \to H_\star K_V.$$

Consider the coproduct structure on $H_\star K_{*\sigma}$ and $E^2_{*,\star}$. There is a cycle y on $E^2_{*,\star}$, with the symmetric term of the coproduct $\psi(y)$ equal to $x \otimes x$. This means there is y such that $2_\star y = x * x$, so $x * x = 0$ as desired. □

6.4 Circle product names for the generators of $H_\star K_{\sigma+i}$

We give names to the generators of $H_\star K_{\sigma+i}$ and indicate how the bookkeeping becomes increasingly complicated as the number of sign representations in V where $1 + \sigma \subset V$ increases (Example 6.8).

To write these answers, we recall our notation for $H_\star K_\rho$. Let

$$\bar{\beta}_i \in H_{\rho i} K(\mathbb{Z}, \rho) \quad (i \geq 0).$$

This gives additional generators,

$$\bar{\beta}_{(i)} = \bar{\beta}_{2i} \quad (i \geq 0),$$

of $H_\star K_\rho$ with coproduct

$$\psi(\bar{\beta}_n) = \sum_{i=0}^{n} \bar{\beta}_{n-i} \otimes \bar{\beta}_i.$$

Then for finite sequences

$$
\begin{aligned}
I &= (i_1, i_2, \ldots, i_k), & 0 &\leq i_1 < i_2 < \cdots, \\
W &= (w_1, w_2, \ldots, w_q), & 0 &\leq w_1 < w_2 < \cdots, \\
J &= (j_{-1}, j_0, j_1, \ldots, j_\ell), & \text{where } &j_{-1} \in \{0, 1\} \text{ and all other } j_n \geq 0, \\
Y &= (y_{-1}, y_0, y_1, \ldots, y_r), & \text{where } &y_{-1} \in \{0, 1\} \text{ and all other } y_n \geq 0,
\end{aligned}
$$

define

$$
\begin{aligned}
(e_1\alpha\beta)^{I,J} &= e_1^{\circ j_{-1}} \circ \alpha_{(i_1)} \circ \alpha_{(i_2)} \circ \cdots \circ \alpha_{(i_k)} \circ \beta_{(0)}^{\circ j_0} \circ \beta_{(1)}^{\circ j_1} \circ \cdots \circ \beta_{(\ell)}^{\circ j_\ell}, \\
(e_1\alpha\beta)^{W,Y} &= e_1^{\circ y_{-1}} \circ \alpha_{(w_1)} \circ \alpha_{(w_2)} \circ \cdots \circ \alpha_{(w_q)} \circ \beta_{(0)}^{\circ y_0} \circ \beta_{(1)}^{\circ y_1} \circ \cdots \circ \beta_{(r)}^{\circ y_r}, \\
&\qquad |I| = k, \quad |W| = q, \quad \|J\| = \Sigma j_n, \quad \|Y\| = \Sigma y_n.
\end{aligned}
$$

Then:

Theorem 6.7 *We have*

$$H_\star K_{\sigma+i} \cong E[(e_1\alpha\beta)^{I,J} \circ \bar{\alpha}_{(m)}, (e_1\alpha\beta)^{W,Y} \circ \bar{\beta}_{(t)}]$$

where $m > i_k$ and $m \geq l$, $t > w_q$ and $t \geq y_r$, $|I| + 2\|J\| = i$ and $|W| + 2\|Y\| = i - 1$, and the coproduct follows by Hopf ring properties from the $\alpha_{(i)}$'s, $\beta_{(i)}$'s, $\bar{\alpha}_{(i)}$'s and $\bar{\beta}_{(i)}$'s.

We offer a proof distinct from that of Theorem 6.6.

Proof Apply the Behrens–Wilson lemma to the generators $(e_1\alpha\beta)^{I,J} \circ \bar{\alpha}_{(m)}$ and $(e_1\alpha\beta)^{W,Y} \circ \bar{\beta}_{(t)}$ defined in the theorem. The map to the underlying homology is clear as the generators have no a–torsion. On fixed points,

$$(e_1\alpha\beta)^{I,J} \circ \bar{\alpha}_{(m)} \mapsto (e_1\alpha\beta)^{I,J} \circ a_{(m)}, \quad (e_1\alpha\beta)^{W,Y} \circ \bar{\beta}_{(t)} \mapsto (e_1\alpha\beta)^{W,Y} \circ a_{(t)},$$

giving a basis for $K_{\sigma+i}^{C_2} \simeq K_{i+1} \times K_i$ where the $a_{(i)}$ are notation for the underlying nonequivariant homology of K_σ (see Section 3.2). The multiplicative and comultiplicative structures are deduced similarly to Theorem 5.6. $\qquad \square$

Example 6.8 Consider the $RO(C_2)$–graded bar spectral sequence

$$E^2_{*,\star} \simeq \mathrm{Tor}^{H_\star K_{2\sigma}}_{*,\star}(H\mathbb{F}_{2\star}, H\mathbb{F}_{2\star}) \Rightarrow H_\star BK_{2\sigma} \cong H_\star K_{2\sigma+1}.$$

The indecomposable cycles on the E^2–page are

$$[e_\sigma \circ e_\sigma], \quad [e_\sigma \circ \bar{\alpha}_{(i)}], \quad [\bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}],$$
$$\phi^{(k)}(e_\sigma \circ e_\sigma), \quad \phi^{(k)}(e_\sigma \circ \bar{\alpha}_{(i)}), \quad \phi^{(k)}(\bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}),$$

where $\phi^{(k)}(x)$ is notation for the bar representative

$$\underbrace{[x|\cdots|x]}_{2^i \text{ copies}}.$$

Since trivial representation suspension is $\circ$ multiplication with e_1, we can identify $[e_\sigma \circ e_\sigma]$ with $e_1 \circ e_\sigma \circ e_\sigma$, $[e_\sigma \circ \bar{\alpha}_{(i)}]$ with $e_1 \circ e_\sigma \circ \bar{\alpha}_{(i)}$, and $[\bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}]$ with $e_1 \circ \bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}$. Using the bar spectral sequence pairing (6.5) compatible with

$$\circ: H_\star(BK_\sigma) \otimes_{H_\star} H_\star(K_\sigma) \to H_\star(BK_{2\sigma}),$$

we can identify $\phi^{(k)}(e_\sigma \circ \bar{\alpha}_{(i)})$ with $\bar{\beta}_{(j_1+1)} \circ \bar{\alpha}_{(j_2+1)}$, and using the bar spectral sequence pairing (6.5) compatible with

$$\circ: H_\star(BK_0) \otimes_{H_\star} H_\star(K_{2\sigma}) \to H_\star(BK_{2\sigma}),$$

we can identify $\phi^{(k)}(\bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)})$ with $\alpha_{(i_1)} \circ \bar{\alpha}_{(i_2)} \circ \bar{\alpha}_{(j_3)}$. By Theorem 6.6, the $\phi^{(k)}(e_\sigma \circ e_\sigma)$ are also permanent cycles. However, degree reasons make it impossible to identify them in terms of circle products (there are too many sign representations) and thus we have a new family of generators which are not circle products of elements in K_1, K_σ, K_ρ, or $K_{2\sigma}$.

Corollary 6.9 *We have*

$$H_\star K_{2\sigma+1} \cong E[e_1 \circ e_\sigma \circ e_\sigma, e_1 \circ e_\sigma \circ \bar{\alpha}_{(i)}, e_1 \circ \bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}, \bar{\beta}_{(j_1+1)} \circ \bar{\alpha}_{(j_2+1)}, \alpha_{(i_1)} \circ \bar{\alpha}_{(i_2)} \circ \bar{\alpha}_{(j_3)}, \phi^{(k)}(e_\sigma \circ e_\sigma)]$$

as an algebra, where the coproduct follows by Hopf ring properties from the $\alpha_{(i)}$'s, $\beta_{(i)}$'s, $\bar{\alpha}_{(i)}$'s, $\bar{\beta}_{(i)}$'s, and coproduct structure on $\mathrm{Tor}^{HK_{2\sigma}}_{*,\star}(H\mathbb{F}_{2\star}, H\mathbb{F}_{2\star})$.

As the number of sign representations in V where $1 + \sigma \subset V$ increases, the number of additional generators grows, making bookkeeping and identifying homology generators in terms of the bar spectral sequence pairing (6.5) an increasingly complicated task.

6.5 The $RO(C_2)$–graded twisted bar spectral sequence

We now turn to the twisted analogue of the $RO(C_2)$–graded bar spectral sequence. Similar to the classical case, the twisted bar construction $B^\sigma A$ is filtered by

$$(B^\sigma A)^{[t]} \simeq \coprod_{t \geq n \geq 0} \Delta^n \times A^n / \sim \; \subset B^\sigma A$$

with associated graded pieces

$$(B^\sigma A)^{[t]}/(B^\sigma A)^{[t-1]} \simeq S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor} \wedge A^{\wedge t},$$

where the C_2–action on A^t is given by $\gamma(a_1 \wedge \cdots \wedge a_n) = (\gamma a_n \wedge \cdots \wedge \gamma a_1)$. Applying $H_\star(-)$ to these filtered spaces gives the twisted bar spectral sequence

$$E^1_{t,\star} = \tilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor} \wedge A^t) \Rightarrow H_\star B^\sigma A,$$

with differentials

$$d_r \colon E^r_{t,\star} \to E^r_{t-r,\star-1},$$

computing $H_\star(B^\sigma A)$.

In general, this spectral sequence lacks an explicit E^2–page and can be difficult to compute. We give some readily computable examples which collapse on the E^1–page and then turn to analyzing the structure of the twisted bar spectral sequence in examples computing the $RO(C_2)$–graded homology of C_2–equivariant Eilenberg–MacLane spaces.

Example 6.10 The $RO(C_2)$–graded twisted bar spectral sequences computing the homology of

$$B^\sigma \mathbb{F}_2 \simeq K(\underline{\mathbb{F}}_2, \sigma) \simeq \mathbb{R}P^\infty_{\mathrm{tw}}, \quad B^\sigma S^1 \simeq K(\underline{\mathbb{Z}}, \rho) \simeq \mathbb{C}P^\infty_{\mathrm{tw}}$$

collapse on the E^1–page. As rings,

$$H_\star \mathbb{R}P^\infty_{\mathrm{tw}} = E[e_\sigma, \bar{\alpha}_{(0)}, \bar{\alpha}_{(1)}, \ldots] = E[e_\sigma] \otimes \Gamma[\bar{\alpha}_{(0)}], \quad |e_\sigma| = \sigma, |\bar{\alpha}_{(i)}| = \rho 2^i,$$

$$H_\star \mathbb{C}P^\infty_{\mathrm{tw}} = E[\bar{\beta}_{(0)}, \bar{\beta}_{(1)}, \ldots] = \Gamma[e_\rho] \quad \text{where } |\bar{\beta}_{(i)}| = \rho 2^i.$$

We write the proof for $H_\star \mathbb{R}P^\infty_{\mathrm{tw}}$ as the computation for $H_\star \mathbb{C}P^\infty_{\mathrm{tw}}$ is similar.

Proof We first prove the additive statement that $H_\star \mathbb{R}P^\infty_{\mathrm{tw}}$ is a free $H_\star$–module with a single generator in each degree $\lceil \frac{n}{2} \rceil \sigma + \lfloor \frac{n}{2} \rfloor$. We then show $H_\star \mathbb{R}P^\infty$ has ring structure $E[e_\sigma, \bar{\alpha}_{(0)}, \bar{\alpha}_{(1)}, \ldots] = E[e_\sigma] \otimes \Gamma[\bar{\alpha}_{(0)}]$ where $|e_\sigma| = \sigma$ and $|\bar{\alpha}_{(i)}| = 2^i \rho$. We start with the twisted bar spectral sequence

$$E^1_{t,\star} = \tilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor} \wedge \mathbb{F}_2^t) \Rightarrow H_\star B^\sigma \mathbb{F}_2.$$

Specifically,

$$
\begin{aligned}
E^1_{t,\star} &\cong \tilde{H}_\star((B^\sigma_t \mathbb{F}_2/B^\sigma_{t-1})\mathbb{F}_2) \\
&\cong \tilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor} \wedge \mathbb{F}_2^{\wedge t}) && \text{(by definition)} \\
&\cong \tilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor}) \otimes \tilde{H}_\star(\mathcal{N}_e^{C_2}(\mathbb{F}_2^{\wedge \lfloor t/2 \rfloor} \wedge \mathbb{F}_2^\epsilon)) && \text{(freeness \& properties of } \mathcal{N}_e^{C_2}) \\
&\cong \tilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor}) \otimes \tilde{H}_\star(\mathbb{F}_2)^{\wedge t} && \text{(homology of norm of underlying free space)}
\end{aligned}
$$

where in the last step, since the homology of $\mathbb{F}_2$ splits as the homology of induced representation spheres, the homology of the norm is the norm of the homology of the underlying space [Hill 2022].

Because the filtration degree t corresponds to the topological degree p and differentials d^r shift topological degree down by one, there are no nonzero d^r for $r > 1$. There can be no nonzero d^1 because if there
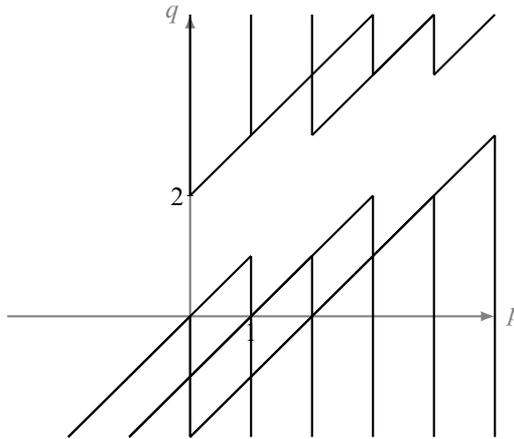
Figure 4: The E^1–page of the twisted bar spectral sequence computing $H_\star K_\sigma$.

were, on passing to the nonequivariant homology of the underlying space, $H_* \mathbb{R}P^\infty$, we would be killing a known generator which is a contradiction. Hence the homology is free with a single generator in each degree $\lceil \frac{n}{2} \rceil \sigma + \lfloor \frac{n}{2} \rfloor$. This E^1–page is depicted in Figure 4.

We deduce the multiplicative structure. There is no element in degree 2σ so e_σ must be exterior. The remaining exterior structure can also be deduced without appealing to Hopf rings. The multiplication by 2 map $2 \colon K_\sigma \to K_\sigma$, which factors as the composition

$$K_\sigma \xrightarrow{\Delta} K_\sigma \times K_\sigma \xrightarrow{*} K_\sigma$$

is homotopically trivial, so

$$0 = 2_\star \colon H_\star K_\sigma \to H_\star K_\sigma.$$

Since $2_\star(\bar{\alpha}_{(i+1)}) = \bar{\alpha}_{(i)} * \bar{\alpha}_{(i)}$, this proves the exterior multiplication. $\square$

Theorem 6.11 *The $RO(C_2)$–graded twisted bar spectral sequence computing the homology of*

$$B^\sigma S^\sigma \simeq K(\mathbb{Z}, 2\sigma)$$

collapses on the E^1–page. As a ring,

$$H_\star K(\mathbb{Z}, 2\sigma) = E[e_{2\sigma}] \otimes \Gamma[\bar{x}_{(0)}] \quad \text{where } |e_{2\sigma}| = 2\sigma, \ |\bar{x}_{(0)}| = 2\rho.$$

The proof of Theorem 6.11 is analogous to the computation of $H_\star \mathbb{R}P_{\mathrm{tw}}^\infty$ given in Example 6.10.

6.6 Higher differentials in the $RO(C_2)$–graded twisted bar spectral sequence

In this section, we use our understanding of $H_\star K_{*\sigma}$ to analyze the structure of the twisted bar spectral sequence and find evidence of arbitrarily long equivariant degree shifting differentials.

Consider the $RO(C_2)$–graded twisted bar spectral sequence

$$E^1_{t,\star} = \widetilde{H}_\star(S^{\lceil t/2 \rceil \sigma + \lfloor t/2 \rfloor} \wedge K_\sigma^{\wedge t}) \Rightarrow H_\star B^\sigma K_\sigma$$
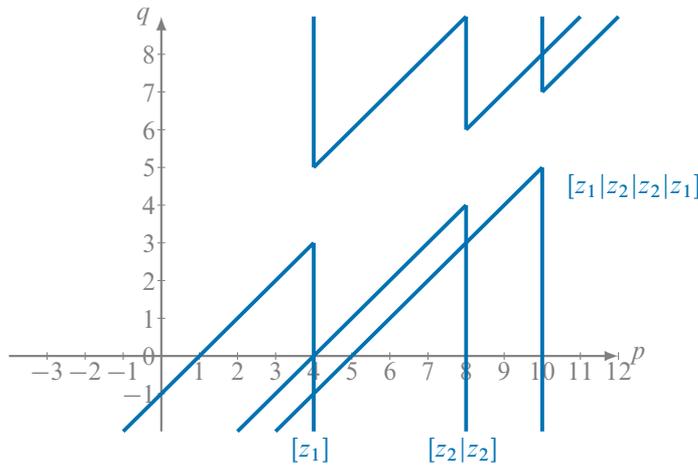
Figure 5: Twisted bar representatives fixed under the C_2–action support full double cones.

computing $H_\star K_{2\sigma}$. There are two basic building blocks in this spectral sequence. Twisted bar representatives $[z_1 \mid \cdots \mid z_n]$, where $z_i \in H_\star K_\sigma$, that are fixed under the C_2–action of the twisted bar construction and those that possess nontrivial C_2–action. The twisted bar representatives which are fixed support a full double cone, that is an $RO(C_2)$–graded representation degree shifted copy of the homology of the point. An example where $|z_1| = \sigma$ and $|z_2| = \rho$ is shown in Figure 5. Let γ denote the generator of C_2. The remaining twisted bar representatives come in pairs $[z_1 \mid \cdots \mid z_n]$ and $\gamma \cdot [z_1 \mid \cdots \mid z_n]$. Each pair gives a copy of C_{2+} and we choose a single twisted bar representative to represent each copy. In the twisted bar spectral sequence, the representatives $[z_1 \mid \cdots \mid z_n]$ with nontrivial C_2–action support shifted degree copies of $H_\star C_{2+}$ as depicted in Figure 6.
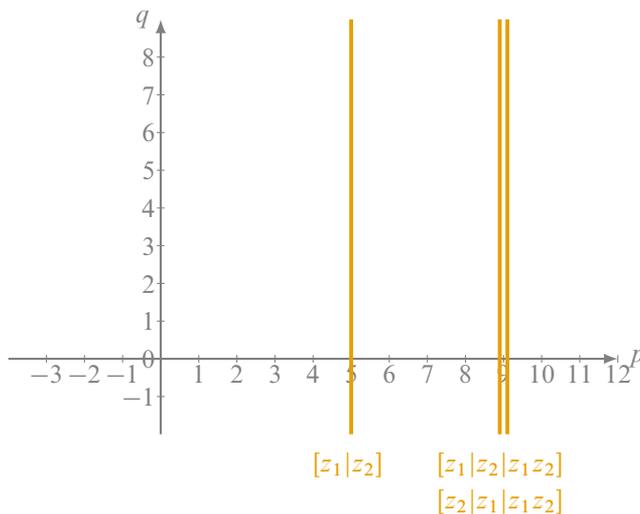

Figure 6: Twisted bar representatives with nontrivial C_2 action support copies of $H_\star C_{2+}$.
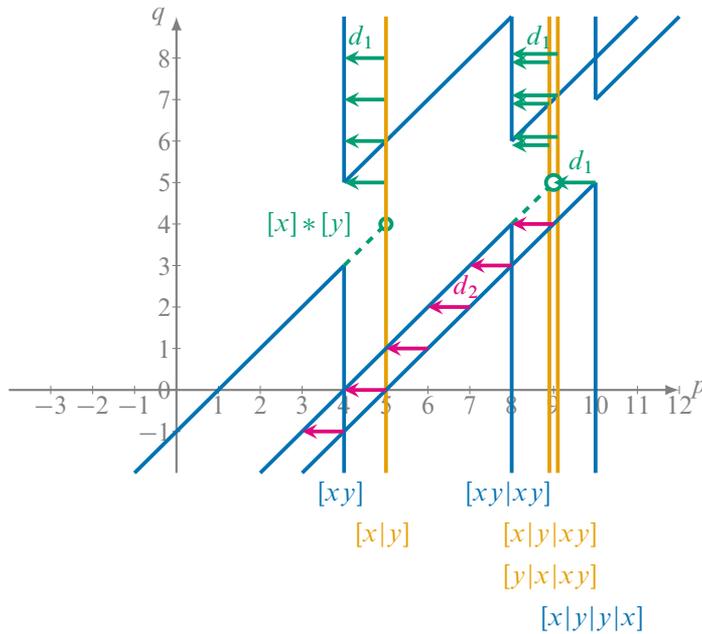
Figure 7: Differentials in the twisted bar spectral sequence computing $H_\star K_{2\sigma}$.

A portion of the twisted bar spectral sequence computing $H_\star K_{2\sigma}$ appears in Figure 7, where x represents e_σ and y represents $\bar{\alpha}_{(0)}$. To compute the d_1–differential in this spectral sequence, consider the cofiber sequence

$$S^0 \xrightarrow{a} S^\sigma \to Ca \simeq \Sigma C_{2+}.$$

This induces a long exact sequence in homology involving

$$H_\star S^0 \xrightarrow{\cdot a} H_\star S^\sigma \to H_\star(C_{2+}),$$

as shown in Figure 8. The map

$$H_\star(C_{2+}) \to H_\star(S^{\sigma-1})$$

is the map depicted in Figure 9.
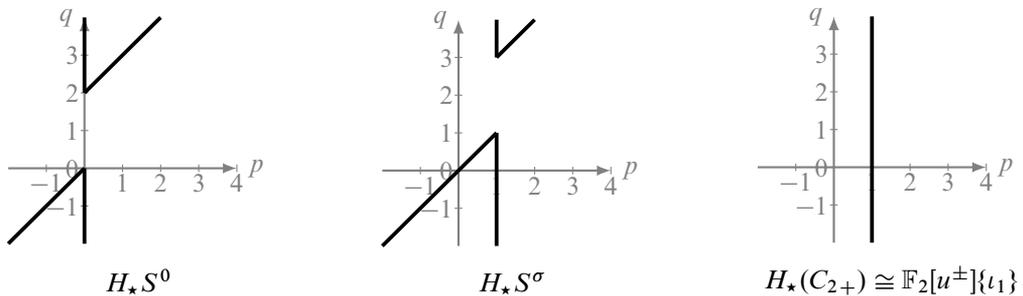
Figure 8: Computing a d_1–differential in the twisted bar spectral sequence.
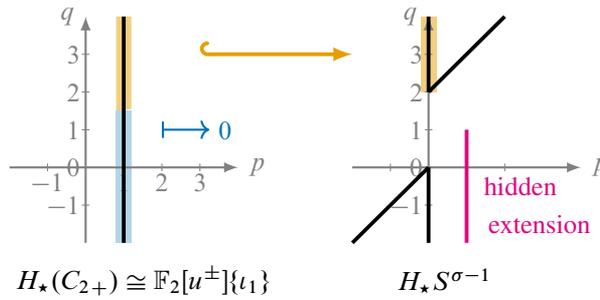
Figure 9: $RO(C_2)$–graded twisted bar spectral sequence d_1–differential with hidden extension.

We have shown that the d_1–differentials marked in green in Figure 7 both exist and have the behavior of the map in Figure 8. We also know from Theorem 5.6 that

$$H_\star K_{2\sigma} \cong E[e_{2\sigma}, \bar{\alpha}_{(j_1)} \circ \bar{\alpha}_{(j_2)}]$$

where $j_1 \leq j_2$.

Since the $RO(C_2)$–graded homology of $K_{2\sigma}$ is free over the $RO(C_2)$–graded homology of a point, all copies of $H_\star C_{2+}$ appearing on the E^1–page must either be killed off or used in shifting the representation degree of the $RO(C_2)$–graded homology of a point, similar to the equivariant degree shifting differential d_1 and hidden extension of Figure 9.

We also know the underlying integer-graded homology of K_2, and have both the forgetful map

$$H_\star K_{2\sigma} \to H_2 K_2$$

and the fixed point map

$$H_\star K_{2\sigma} \to H_*(K_{2\sigma})^{C_2} \cong H_*(K_2 \times K_1 \times K_0).$$

Given that $H_\star K_{2\sigma}$ is free and in the underlying nonequivariant case $[xy \mid xy]$ is killed by a d_1 differential (all generators of $H_\star K_{2\sigma}$ have nontrivial underlying homology), the entire double cone supported by the twisted bar representative $[xy \mid xy]$ must be hit by a differential.

There is a d_1–differential and hidden extension shifting the double cone supported by $[xy \mid xy]$ up by representation degree σ so that by the E^2–page the double cone is in fact in representation degree $\rho(|x| + |y|) + \rho + \sigma = \rho(\rho + \sigma) + \rho + \sigma = 4\rho + \sigma$. We hypothesize there is a d_2–differential induced by a d_1–differential supported by $[x \mid y \mid y \mid x]$. We notice that $[x \mid y \mid y \mid x]$ is a norm of $[xy \mid xy]$. We expect such norms play an important role in governing the structure of all the higher nontrivial differentials.

As one goes farther along in the spectral sequence, considering cycles supported by twisted bar representatives such as $[xy \mid xy \mid xy]$ and $[xyz \mid xyz]$, which must all be killed off in order to recover the correct underlying homology, we see that arbitrarily long equivariant degree shifting differentials are required in order to arrive at the answer given by Theorem 5.6. We conjecture all such cycles are killed by differentials induced by a norm structure on the twisted bar spectral sequence.

7 Related questions

We describe a few questions of immediate interest given the results of this paper.

7.1 Twisted Tor and the $RO(C_2)$–graded twisted bar spectral sequence

In the C_2–equivariant setting, the $RO(C_2)$–graded homology of each signed delooping, $K_{V+\sigma}$, of an equivariant Eilenberg–Mac Lane space, K_V, also independently arises as the result of a C_2–equivariant twisted Tor computation. This can be seen by taking the model of σ–delooping defined in [Hill 2022]. In this model, A is an E_σ–algebra and

$$B^\sigma(A) = B(A, \mathrm{Map}(C_2, A), \mathrm{Map}(C_2, *)),$$

where the action of $\mathrm{Map}(C_2, A)$ on A is via the E_σ–structure [Hill 2022, Definition 5.10]. In the case that A has R–free homology, Hill [2022, Theorem 5.11] constructs yet another twisted bar spectral sequence with E^2–page

$$E_2^{s,\star} = \mathrm{Tor}_{-s}^{N_e^{C_2}(i_e^* R_*(i_e^* A))}(R_\star(\mathrm{Map}(C_2, X)), R_\star(A)) \Rightarrow R_{\star-s}(B^\sigma(A)).$$

Computations with this spectral sequence are complicated and the literature lacks substantial examples. However, it does have a twisted Tor functor as its E^2–page and thus it would be interesting to compare with our computations.

One notable feature of the nonequivariant computation of $H_* K(\mathbb{F}_p, *)$ is that the integer graded bar spectral sequences collapse on the E^2–page [Wilson 1982]. In contrast, we saw that the $RO(C_2)$–graded twisted bar spectral sequences computing $H_\star K_{*\sigma}$ have arbitrarily long differentials in Section 6.6. Thus under favorable circumstances, we hope to formulate a twisted bar spectral sequence with E^2–page a twisted Tor functor arising as a derived functor of the twisted product of $H\mathbb{F}_2$–modules, which collapses in the relevant cases of $H_\star K_{*\sigma}$.

Given our computation of $H_\star K_{*\sigma}$, such a twisted Tor over an exterior algebra should have the property that

$$\mathrm{Tor}_{\mathrm{tw}}^{E[x]} \cong E[\sigma x] \otimes \Gamma[\mathcal{N}_e^{C_2} x].$$

7.2 Global Hopf rings

In their work computing the integer graded homology of classical nonequivariant Eilenberg–Mac Lane spaces, Ravenel and Wilson obtain a global statement. Specifically:

Theorem C (Ravenel and Wilson [Wilson 1982]) $H_* K_*$ *is the free Hopf ring on* $H_* K_0 = H_*[\mathbb{F}_p]$, $H_* K_1$, *and* $H_* \mathbb{C} P^\infty \subset H_* K_2$ *subject to the relation* $e_1 \circ e_1 = \beta_1$.

It is natural to ask if a similar statement be obtained in the C_2–equivariant case, and in that case, what specifically, is the global structure of the Hopf rings that do arise. One may also ask how the Hopf rings

here relate to Hill and Hopkins' work [2018] extending Ravenel and Wilson's construction of a universal Hopf ring over MU^* to C_2–equivariant homotopy theory.

7.3 Stabilizing to the C_2–dual Steenrod algebra

Besides understanding a global version of the unstable story, it also remains to fully understand how the unstable answer for $H_\star K_V$ stabilizes to give the C_2–equivariant dual Steenrod algebra,

$$\mathcal{A}^{C_2}_\star = H\mathbb{F}_2[\tau_0, \tau_1, \ldots, \xi_1, \xi_2, \ldots]/(\tau_i^2 = (u + a\tau_0)\xi_{i+1} + a\tau_{i+1}).$$

By Hu and Kriz's construction [2001] of the C_2–equivariant dual Steenrod algebra, we should homology suspend $\bar{\beta}_{(i)}$ to define

$$\xi_i \in H_{(2^i-1)\rho}H$$

and $\bar{\alpha}_{(i)}$ to define

$$\tau_i \in H_{2^i\rho-\sigma}H.$$

However, it is not at all clear what an arbitrary element in $H_\star K_V$ should stabilize to in the C_2–equivariant dual Steenrod algebra. Additionally, there is the interesting problem of understanding how the stable relation $\tau_i^2 = (u + a\tau_0)\xi_{i+1} + a\tau_{i+1}$ arises unstably. We look forward to studying these questions in forthcoming work.

References

[Angeltveit and Rognes 2005]　**V Angeltveit**, **J Rognes**, *Hopf algebra structure on topological Hochschild homology*, Algebr. Geom. Topol. 5 (2005) 1223–1290　MR Zbl

[Behrens and Wilson 2018]　**M Behrens**, **D Wilson**, *A C_2–equivariant analog of Mahowald's Thom spectrum theorem*, Proc. Amer. Math. Soc. 146 (2018) 5003–5012　MR Zbl

[Caruso 1999]　**J L Caruso**, *Operations in equivariant $\mathbb{Z}/p$–cohomology*, Math. Proc. Cambridge Philos. Soc. 126 (1999) 521–541　MR Zbl

[Hill 2022]　**M A Hill**, *Freeness and equivariant stable homotopy*, J. Topol. 15 (2022) 359–397　MR Zbl

[Hill and Hopkins 2018]　**M A Hill**, **M J Hopkins**, *Real Wilson spaces, I*, preprint (2018)　arXiv 1806.11033

[Hill et al. 2016]　**M A Hill**, **M J Hopkins**, **D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262　MR Zbl

[Hu and Kriz 2001]　**P Hu**, **I Kriz**, *Real-oriented homotopy theory and an analogue of the Adams–Novikov spectral sequence*, Topology 40 (2001) 317–399　MR Zbl

[Kronholm 2010]　**W C Kronholm**, *A freeness theorem for $RO(\mathbb{Z}/2)$–graded cohomology*, Topology Appl. 157 (2010) 902–915　MR Zbl

[Liu 2020]　**Y Liu**, *Twisted bar construction*, preprint (2020)　arXiv 2003.06856

[Milgram 1967]　**R J Milgram**, *The bar construction and abelian H–spaces*, Illinois J. Math. 11 (1967) 242–250　MR Zbl

[Mosher and Tangora 1968] **R E Mosher**, **M C Tangora**, *Cohomology operations and applications in homotopy theory*, Harper & Row, New York (1968) MR Zbl

[Ravenel and Wilson 1977] **D C Ravenel**, **W S Wilson**, *The Hopf ring for complex cobordism*, J. Pure Appl. Algebra 9 (1977) 241–280 MR Zbl

[Ravenel and Wilson 1980] **D C Ravenel**, **W S Wilson**, *The Morava K–theories of Eilenberg–MacLane spaces and the Conner–Floyd conjecture*, Amer. J. Math. 102 (1980) 691–748 MR Zbl

[Serre 1953] **J-P Serre**, *Cohomologie modulo 2 des complexes d'Eilenberg–Mac Lane*, Comment. Math. Helv. 27 (1953) 198–232 MR Zbl

[Thomason and Wilson 1980] **R W Thomason**, **W S Wilson**, *Hopf rings in the bar spectral sequence*, Q. J. Math. 31 (1980) 507–511 MR Zbl

[Wilson 1982] **W S Wilson**, *Brown–Peterson homology*: *an introduction and sampler*, CBMS Reg. Conf. Ser. Math. 48, Amer. Math. Sci., Providence, RI (1982) MR Zbl

[Yigit 2019] **U Yigit**, *The C_2–equivariant unstable homotopy theory*, PhD thesis, University of Rochester (2019) Available at https://www.proquest.com/docview/2305942280

Department of Mathematics, University of Colorado Boulder
Boulder, CO, United States

sarahllpetersen@gmail.com

Simple balanced three-manifolds, Heegaard Floer homology and the Andrews–Curtis conjecture

Neda Bagherifard

Eaman Eftekhary

The first author introduced a notion of equivalence on a family of 3–manifolds with boundary, called (simple) balanced 3–manifolds in an earlier paper and discussed the analogy between the Andrews–Curtis equivalence for group presentations and the aforementioned notion of equivalence. Motivated by the Andrews–Curtis conjecture, we use tools from Heegaard Floer theory to prove that there are simple balanced 3–manifolds which are not in the trivial equivalence class (ie the equivalence class of $S^2 \times [-1, 1]$).

57K18, 57R58

1 Introduction

Suppose that $R = \{b_1, \ldots, b_m\}$ is a finite subset of the free group $F(X)$ generated by the finite set $X = \{a_1, \ldots, a_n\}$. We denote by $(X \mid R)$ the quotient G of $F(X)$ by the normal subgroup generated by R. The pair (X, R) is then called a *presentation* of G with *generators* X and *relators* R, which is balanced if $|X| = |R|$. An *extended Andrews–Curtis transformation* (EAC transformation for short) on (X, R) is defined as one of the following transformations, or its inverse, which of course results in another presentation of G [Wright 1975] (see also [Hog-Angeloni and Metzler 1993]):

(1) **Composition** Replace $b \in R$ with bb' for some $b' \neq b$ in R.

(2) **Inversion** Replace $b \in R$ with b^{-1}.

(3) **Cancellation** Replace $b = b'aa^{-1}b'' \in R$ with $b'b''$, where $a \in X$ or $a^{-1} \in X$.

(4) **Stabilization** Add a new element a to both X and R.

(5) **Replacement** Replace $a'a$ or $a'a^{-1}$ for a' in all the relators for some $a \neq a'$ in X.

Stable Andrews–Curtis transformations (or SAC transformations) consist of the first 4 transformations and their inverses. The presentations $P' = (X', R')$ and $P = (X, R)$ are called *EAC equivalent* (resp. *SAC equivalent*) if P' is obtained from P by a finite sequence of EAC transformations (resp. SAC transformations). For the trivial group, the SAC equivalence class of a presentation is the same as its EAC equivalence class [Wright 1975]. The stable Andrews–Curtis conjecture (or SAC conjecture) states that every balanced presentation of the trivial group is SAC equivalent to the trivial presentation, ie (a, a) (see [Andrews and Curtis 1965]). Most experts expect that the SAC conjecture is not true and there are potential counterexamples [Brown 1984; Burns and Macedońska 1993; Miller and Schupp 1999; Myasnikov et al. 2002]. One of the simplest potential counterexamples for the SAC conjecture is given by $P_0 = (X_0, R_0)$, where

(1) $$X_0 = \{x, y\} \quad \text{and} \quad R_0 = \{r = x^{-1}y^2xy^{-3}, s = y^{-1}x^2yx^{-3}\}$$

(see [Myasnikov et al. 2002]). The group presentation P_0 is considered in this paper in correspondence with a notion of equivalence for balanced 3–manifolds, as explained below.

A compact oriented 3–manifold N with boundary is called *balanced* if each component of N has two boundary components of the same genus. Let $\partial^{\pm}N$ denote boundary components of N where the orientation of ∂^+N (resp. ∂^-N) matches with (resp. is the opposite of) the orientation inherited as the boundary of N. Let $\iota^{\pm} : \partial^{\pm}N \to N$ denote the inclusion maps and $H^{\pm}$ denote the normalizer of $\iota^{\pm}_*(\pi_1(\partial^{\pm}N))$ in $\pi_1(N)$. A balanced 3–manifold is called *simple* if for each connected component N of it as above, both quotient groups $\pi_1(N)/H^{\pm}$ are trivial. Associated with each Heegaard diagram $\mathcal{H} = (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ of N, there are two balanced presentations $P_{\alpha}(\mathcal{H})$ and $P_{\beta}(\mathcal{H})$ for the latter quotient groups where for $P_{\alpha}(\mathcal{H})$ (resp. $P_{\beta}(\mathcal{H})$) the generators are in correspondence with the $\boldsymbol{\alpha}$ (resp. $\boldsymbol{\beta}$) and the relators are in correspondence with the $\boldsymbol{\beta}$ (resp. $\boldsymbol{\alpha}$) (see [Bagherifard 2021]). Let $\mathfrak{p}_{\alpha}(N)$ and $\mathfrak{p}_{\beta}(N)$ denote the EAC equivalence classes of the presentations $P_{\alpha}(\mathcal{H})$ and $P_{\beta}(\mathcal{H})$, respectively. Note that these EAC equivalence classes are independent of the choice of the Heegaard diagram $\mathcal{H}$ for N. Similarly, we may define $\mathfrak{p}_{\alpha}(N)$ and $\mathfrak{p}_{\beta}(N)$ for a balanced 3–manifold N which is not connected. If N is a simple balanced 3–manifold, $\mathfrak{p}_{\alpha}(N)$ and $\mathfrak{p}_{\beta}(N)$ are both EAC equivalence classes of presentations for the trivial group.

A notion of equivalence in the family of balanced 3–manifolds was introduced in [Bagherifard 2021]. We say that a balanced 3–manifold N simplifies to another balanced 3–manifold N' if there is an embedded cylinder $C \sim S^1 \times [-1, 1]$ in N, with $\partial^{\pm}C \sim S^1 \times \{\pm 1\} \subset \partial^{\pm}N$, such that N' is obtained by cutting N along C and gluing two copies of $D^2 \times [-1, 1]$ to the resulting boundary cylinders in $N \setminus C$. We then write $N \xrightarrow{C} N'$. We say that a balanced 3–manifold N admits a *simplifier* if there is a sequence of simplifications

$$N = N_n \xrightarrow{C_n} N_{n-1} \xrightarrow{C_{n-1}} \cdots \xrightarrow{C_2} N_1 \xrightarrow{C_1} N_0$$

such that N_0 is a disjoint union of copies of $S^2 \times [-1, 1]$. The inverse of a simplification is called an *antisimplification*. Two balanced 3–manifolds are called *equivalent* if they may be changed to one another by a finite sequence of simplifications, antisimplifications and homeomorphisms. The equivalence of the balanced 3–manifolds N and N' implies that $\mathfrak{p}_\alpha(N) = \mathfrak{p}_\alpha(N')$ and $\mathfrak{p}_\beta(N) = \mathfrak{p}_\beta(N')$. Therefore, a pair of well-defined EAC equivalence classes (of group presentations) are assigned to each equivalence class of balanced 3–manifolds and in this sense, the equivalence notion between balanced 3–manifolds is *weaker* than the EAC equivalence for group presentations. In the family of simple balanced 3–manifolds, both EAC equivalence classes are presentations of the trivial group. Motivated by the SAC conjecture, it is thus natural to ask if there is a simple balanced 3–manifold N which is not equivalent to the trivial simple balanced 3–manifold $S^2 \times [-1, 1]$. In this paper, we combine the main result of [Bagherifard 2021] with tools from Heegaard Floer theory (see [Ozsváth and Szabó 2004c]) to prove the following theorem.

Theorem 1.1 *There is a simple balanced 3–manifold N with*

$$\mathfrak{p}_\alpha(N) = \mathfrak{p}_\beta(N) = P_0 = [(X_0, R_0)],$$

where P_0 is given in (1), *which is not equivalent to $S^2 \times I$.*

As mentioned above, besides Heegaard Floer theory, the main tool used in proving Theorem 1.1 is a fundamental result about the equivalence class of the simple balanced 3–manifold $S^2 \times [-1, 1]$, which is proved in [Bagherifard 2021] and may be stated as follows.

Theorem 1.2 [Bagherifard 2021, Theorem 1.6] *Every balanced 3–manifold N which is equivalent to $S^2 \times I$ admits a simplifier.*

The group presentation P_0 of (1) is realized by the Heegaard diagram

$$\overline{\mathcal{H}} = (\overline{\Sigma}, \bar{\boldsymbol{\alpha}} = \{\alpha_1, \alpha_2\}, \bar{\boldsymbol{\beta}} = \{\beta_1, \beta_2\}),$$

illustrated in Figure 1. In fact, the Heegaard diagram $\overline{\mathcal{H}}$ determines a simple balanced 3–manifold N with $\mathfrak{p}_\alpha(N) = \mathfrak{p}_\beta(N) = [P_0]$. If N is equivalent to $S^2 \times I$, Theorem 1.2 implies that N admits a simplifier. We have $\partial N = \partial^+ N \amalg -\partial^- N$ where $\partial^\pm N$ are surfaces of genus 1. If N admits a simplifier, there is a nontrivial cylinder C in N such that $\partial^\pm C$ in $\partial^\pm N$ are essential curves. Let $f : \partial^+ N \to \partial^- N$ be the homeomorphism from $\partial^+ N$ to $\partial^- N$ which makes the following diagram commutative:

$$
\begin{array}{ccc}
H_1(\partial^+ N, \mathbb{Z}) & \xrightarrow{\quad f_* \quad} & H_1(\partial^- N, \mathbb{Z}) \\
& \searrow{\scriptstyle \iota_*^+} \quad \swarrow{\scriptstyle \iota_*^-} & \\
& H_1(N, \mathbb{Z}) &
\end{array}
$$

This criteria determines f up to isotopy. Since $\partial^+ C$ is homologous to $\partial^- C$, we may further assume that f maps $\partial^+ C$ to $\partial^- C$. Let N_f denote the closed 3–manifold obtained from N by identifying $\partial^+ N$ with
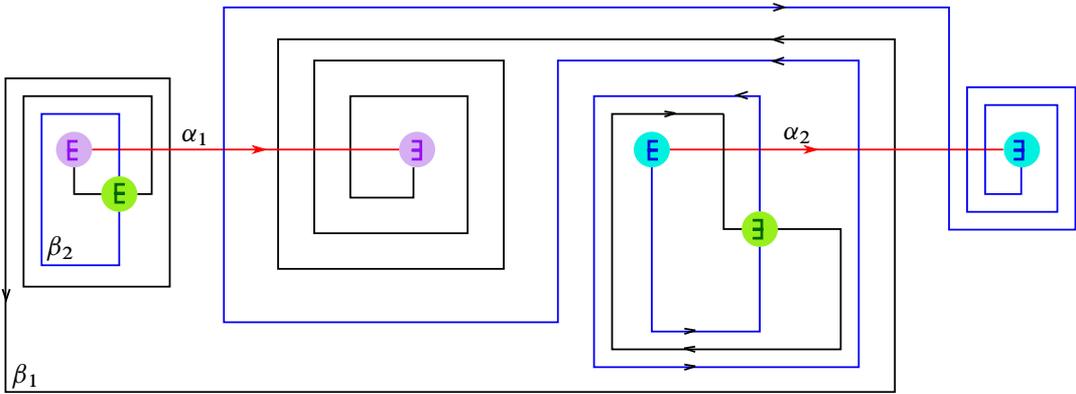
Figure 1: The Heegaard surface is a surface of genus three which is obtained by identifying the boundaries of disks with the same color. The curves are oriented in a way that the balanced presentation associated with this Heegaard diagram is P_0.

$\partial^- N$ using f. Let $\overline{C}$ denote the torus in M which is obtained from C by identifying $\partial^+ C$ with $\partial^- C$. Thus $\overline{C}$ and $\partial^+ N \sim_f \partial^- N$ represent linearly independent homology classes in $H_2(N_f, \mathbb{Z}) = \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ with zero Thurston seminorm. Recall that the Thurston seminorm of a closed 3–manifold M is defined on $H_2(M, \mathbb{Z})$ by

$$\Theta \colon H_2(M, \mathbb{Z}) \to \mathbb{Z}^{\geq 0}, \quad \Theta(\xi) := \min\{\chi_+(\Sigma) \mid \Sigma \hookrightarrow M \text{ and } [\Sigma] = \xi\},$$

where the minimum is taken over all compact oriented surfaces $\Sigma = \amalg_i \Sigma_i$ embedded in M and representing the homology class ξ, while $\chi_+(\Sigma)$ is defined by $\sum_{g(\Sigma_i)>0}(2g(\Sigma_i) - 2)$ (see [Thurston 1986]). Heegaard Floer homology groups with twisted coefficients detect the Thurston seminorm. More precisely, for a closed 3–manifold M, let $\widehat{HF}(M)$ denote the Heegaard Floer homology group of M with twisted coefficients, which is a $\mathbb{Z}/2\mathbb{Z}$–graded $\mathbb{Z}_2[H^1(M, \mathbb{Z})]$–module defined in [Ozsváth and Szabó 2004c]. There is a decomposition of this group by Spin^c structures,

$$\widehat{HF}(M) = \bigoplus_{\mathfrak{s} \in \mathrm{Spin}^c(M)} \widehat{HF}(M, \mathfrak{s}).$$

Theorem 1.3 [Ozsváth and Szabó 2004a, Theorem 1.1] *For a closed 3–manifold M and $\xi \in H_2(M, \mathbb{Z})$,*

$$\Theta(\xi) = \max_{\{\mathfrak{s} \in \mathrm{Spin}^c(M) \mid \widehat{HF}(M, \mathfrak{s}) \neq 0\}} |\langle c_1(\mathfrak{s}), \xi \rangle|.$$

Let us consider the case where $M = N_f$ is given as above. Extend $[\partial^+ N]$ to a basis for $H_2(N_f, \mathbb{Z}) \cong \mathbb{Z}^3$ and consider a corresponding identification of Spin$^c(M)$ with $\mathbb{Z}^3$ (by evaluation of the first Chern class of the Spin^c structures over the generators of the homology group $H_2(M, \mathbb{Z}) \cong \mathbb{Z}^3$). In order to prove Theorem 1.1, we show that there are two linearly independent Spin^c structures $\mathfrak{s}_1$ and $\mathfrak{s}_2$, with the property that

$$\langle c_1(\mathfrak{s}_i), [\partial^+ N] \rangle = 0 \quad \text{and} \quad \widehat{HF}(M, \mathfrak{s}_i) \neq 0 \quad \text{for } i = 1, 2.$$

Since $\Theta([\bar{C}]) = 0$, we thus have $[\bar{C}] = \lambda[\partial^+ N]$, for some integer λ, which contradicts our assumption. This shows that N does not have a simplifier and is thus not equivalent to $S^2 \times I$.

A remark about the above argument may be appropriate here. Let us assume that N_1 and N_2 are balanced 3–manifolds and that N_1 simplifies to N_2, ie $N_1 \xrightarrow{C} N_2$. For $k = 1, 2$, let M_k denote the closed 3–manifold obtained by taking two copies N_k^1 and N_k^2 of N_k, identifying $\partial^+ N_k^1$ with $\partial^+ N_k^2$ and identifying $\partial^- N_k^1$ with $\partial^- N_k^2$. Then the cylinder C gives the torus $T \subset M_1$, while M_2 is obtained by cutting M_1 along T and gluing two solid tori to the resulting boundary components. Theorem 1.3 is then helpful in detecting T. Nevertheless, the equivalence of N_1 and N_2 is yet not well translated to Heegaard Floer theory, eg to a practical correspondence between $\widehat{HF}(M_1)$ and $\widehat{HF}(M_2)$. If T is 2–sided, the problem is studied in [Eftekhary 2015; 2018], and a relatively powerful machinery is developed in [Hanselman et al. 2024]. For nonseparating T, it is interesting to develop such a correspondence.

About the proof In Section 2, we construct a Heegaard diagram $\mathcal{H}$ for the closed manifold M from $\overline{\mathcal{H}}$, following the approach of [Lekili 2013]. The number of generators for the Heegaard diagram $\mathcal{H}$ is 7936, and it is thus not feasible to find the Spin^c structures $\mathfrak{s}_1$ and $\mathfrak{s}_2$ and compute the groups $\widehat{HF}(M, \mathfrak{s}_i)$ without computational assistance (from computers). We prove a simple lemma from linear algebra in Section 3, in the spirit of the general discussion in [Eftekhary 2015, Section 2]. The lemma is used, in combination with a computer program, to obtain a shortlist of potential Spin^c structures $\mathfrak{s}$ with $\widehat{HF}(M, \mathfrak{s}) \neq 0$ (although obtaining the shortlist is not an official part of our argument). Among the potential candidates, two specific Spin^c structures $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are considered in Sections 5 and 6. The chain complexes associated with these Spin^c structures are 8–dimensional and 72–dimensional, respectively. The homology groups of the chain complexes $\widehat{CF}(M, \mathfrak{s}_i)$ (for $i = 1, 2$) are studied using the lemma proved in Section 3, a series of computer assisted computations and explicit computations of the contribution of moduli spaces associated with certain classes of Whitney disks. Since the Heegaard diagram is not nice (in the sense of [Sarkar and Wang 2010]), such explicit computations are necessary and appear in Section 4.

2 A Heegaard diagram for the mapping torus

In this section, we obtain a Heegaard diagram for $M = N_f$, using the construction of [Lekili 2013]. Let us assume that the diagram $\overline{\mathcal{H}}$ is obtained from a Morse function $h: N \to [-1, 1]$. Then h gives a circle-valued Morse function $\bar{h}: M \to S^1$ with two critical points x_1 and x_2 of index 1 and two critical points y_1 and y_2 of index 2, such that

$$N_{x_1}^u(h) \cap \Sigma = \alpha_1, \quad N_{x_2}^u(h) \cap \Sigma = \alpha_2, \quad N_{y_1}^s(h) \cap \Sigma = \beta_1, \quad N_{y_2}^s(h) \cap \Sigma = \beta_2,$$

$$\bar{h}^{-1}(1) = \partial^+ N \sim_f \partial^- N = \Sigma_{\min}, \qquad \bar{h}^{-1}(-1) = \bar{\Sigma} = \Sigma_{\max}.$$
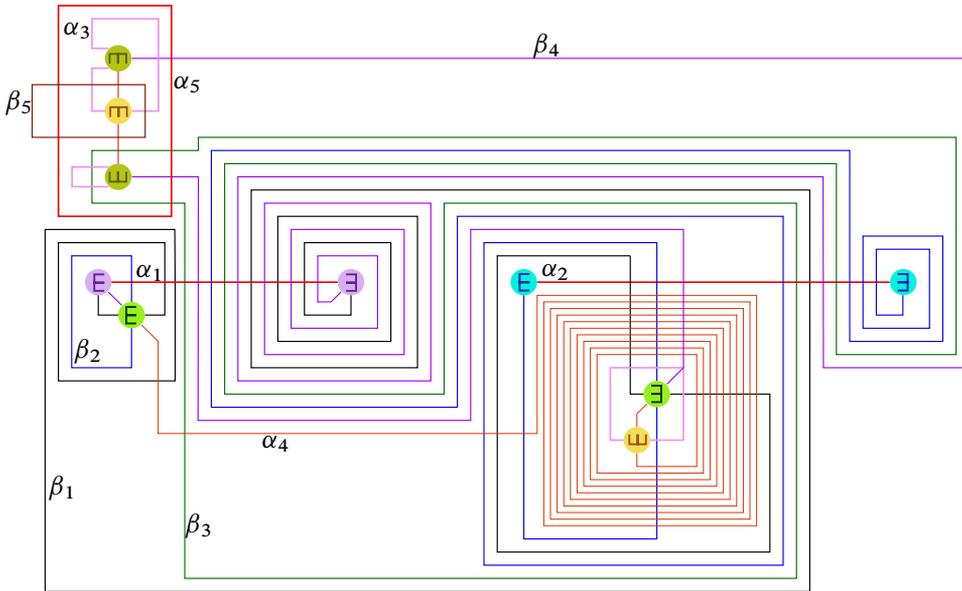
Figure 2: A weakly admissible Heegaard diagram of N_f with 21392 generators.

Here N_x^s and N_x^u denote the stable and unstable manifold of $x \in N$ with respect to the flow of a gradient-like vector field for h. Following [Lekili 2013], let p_1 and p_2 be disjoint points in $\bar{\Sigma} \setminus \boldsymbol{\alpha} \cup \boldsymbol{\beta}$ and γ_1 and γ_2 denote two gradient flow lines disjoint from $N_{x_i}^u$ and $N_{y_i}^s$ such that

$$\gamma_i \cap \bar{\Sigma} = \{p_i\} \quad \text{and} \quad \gamma_i \cap \partial^+ N = \{\bar{p}_i\} \quad \text{for } i = 1, 2.$$

Furthermore, γ_1 (resp. γ_2) is mapped onto the northern (resp. southern) semicircle of S^1. Let $N(\gamma_i)$, for $i = 1, 2$, denote the normal neighborhood of γ_i that intersects $\bar{\Sigma}$ and $\partial^+ N$ in the small disks D_{p_i} and $D_{\bar{p}_i}$, respectively. By removing $D_{p_i}^\circ$ and $D_{\bar{p}_i}^\circ$ and gluing ∂D_{p_i} to $\partial D_{\bar{p}_i}$ along $\partial N(\gamma_i)$ we obtain the Heegaard surface Σ. Let $\alpha_5 = \partial D_{\bar{p}_1}$ and $\beta_5 = \partial D_{\bar{p}_2}$. Let α_3' and α_4' (resp. β_3' and β_4') be disjoint arcs in $\partial^+ N$ such that $\partial \alpha_3'$ and $\partial \alpha_4'$ are disjoint points on β_5 and $\partial \beta_3'$ and $\partial \beta_4'$ are disjoint points on α_5, while $|\alpha_3' \cap \beta_4'| = |\alpha_4' \cap \beta_3'| = 1$ and $|\alpha_3' \cap \beta_3'| = |\alpha_4' \cap \beta_4'| = 0$. Flowing the arcs β_3' and β_4' through the gradient flow of $\bar{h}$ above the northern semicircle, we obtain disjoint arcs β_3'' and β_4'' in $\Sigma \setminus \partial^+ N$ which are disjoint from β_1 and β_2. Similarly, flowing the arcs α_3' and α_4', we obtain α_3'' and α_4'' which are disjoint from α_1 and α_2. This determines the sets of α and β curves,

$$\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_3 = \alpha_3' \cup \alpha_3'', \alpha_4 = \alpha_4' \cup \alpha_4'', \alpha_5\}, \quad \boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3 = \beta_3' \cup \beta_3'', \beta_4 = \beta_4' \cup \beta_4'', \beta_5\}.$$

Having fixed a marked point z, finger-move isotopies may be used to make $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ weakly admissible. If we apply the procedure to the Heegaard diagram of Figure 1, we arrive at the admissible Heegaard diagram illustrated in Figure 2 with 21392 generators. Handle-slides of α_4 over α_3 (10 times) and isotopies on α_3 give an alternative (more suitable) weakly admissible Heegaard diagram $\mathcal{H}$ with 7936 generators, as illustrated in Figure 3. We use $x_{i,j,k}$ to label the intersection point of α_i and β_j which is labeled k in the diagram of Figure 3.
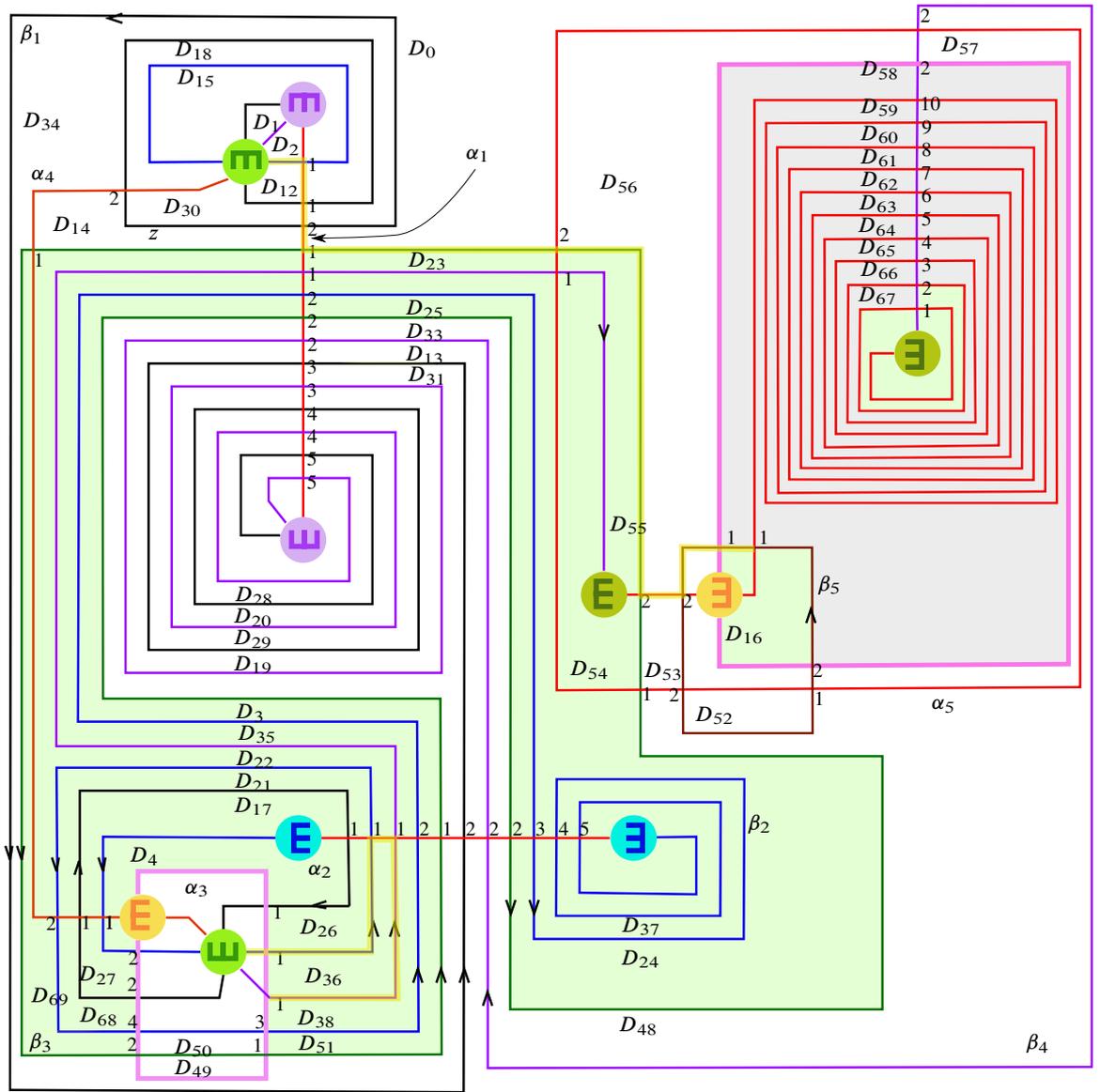
Figure 3: A weakly admissible Heegaard diagram for M with 7936 generators. The connected components of $\Sigma \setminus \boldsymbol{\alpha} \cup \boldsymbol{\beta}$ are labeled D_i, for $i = 0, \ldots, 67$. The periodic domains are generated by $P_1 = -D_{52} + \sum_{i=53}^{57} D_i + \sum_{i \in I_1} D_i$, $P_2 = -D_{49} + \sum_{i \in I_2} D_i + \sum_{i \in I_1} D_i$ and a third periodic domain P_3, where D_i is colored gray for $i \in I_1$ and green for $i \in I_2$. We have $\partial_b(P_1) = -\beta_5$ and $\partial_b(P_2) = \beta_3$. The periodic domain P_3 may be chosen so that $\partial_b(P_3) = \beta_4 + 2\beta_3 - 3\beta_1 + 2\beta_2$.

The set of periodic domains for $\mathcal{H}$ is generated by three domains P_1, P_2 and P_3. The first two generators are shown in Figure 3. The periodic domains P_1 and P_2 are of the form

$$P_1 = -D_{52} + \sum_{i=53}^{57} D_i + \sum_{i \in I_1} D_i, \quad P_2 = -D_{49} + \sum_{i \in I_2} D_i + \sum_{i \in I_1} D_i,$$

where the domains D_i with i in I_1 and I_2 are colored gray and green in Figure 3, respectively. If $\partial_b P$ denote the β–boundary of a periodic domain P, we then have $\partial_b(P_1) = -\beta_5$ and $\partial_b(P_2) = \beta_3$. We may choose the third generator P_3 of the space of periodic domains so that $\partial_b(P_3) = \beta_4 + 2\beta_3 - 3\beta_1 + 2\beta_2$. Let $H(P_i) \in H_2(M, \mathbb{Z})$ denote the homology classes associated with the periodic domains P_i for $i = 1, 2, 3$, which form a basis for $H_2(M, \mathbb{Z})$ (see [Ozsváth and Szabó 2004c, Proposition 2.15]). Correspondingly, we obtain a bijection

$$c \colon \mathrm{Spin}^c(M) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}, \quad c(\mathfrak{s}) := \tfrac{1}{2}\big(\langle c_1(\mathfrak{s}), H(P_1)\rangle, \langle c_1(\mathfrak{s}), H(P_2)\rangle, \langle c_1(\mathfrak{s}), H(P_3)\rangle\big),$$

which gives an identification of $\mathrm{Spin}^c(M)$ with $\mathbb{Z}^3$. To compute $\mathfrak{s}_z \colon \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}} \to \mathrm{Spin}^c(M) = \mathbb{Z}^3$ under this identification, define $\mathfrak{s}_z^r \colon \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}} \to \mathbb{Z}^3$ by setting $\mathfrak{s}_z^r(x_0) = (0, 0, 0)$ for

$$x_0 = (x_i^0)_{i=1}^5 = (x_{1,1,1}, x_{2,2,1}, x_{3,4,1}, x_{4,5,1}, x_{5,3,2}).$$

Let $(y_i^0)_{i=1}^5$ denote a permutation of $(x_i^0)_{i=1}^5$ such that $y_i^0 \in \beta_i$. Fix a connected path γ_0 on $\boldsymbol{\alpha} \cup \boldsymbol{\beta}$ in the diagram such that for each $\alpha \in \boldsymbol{\alpha}$ and $\beta \in \boldsymbol{\beta}$, $\gamma_0 \cap \alpha$ and $\gamma_0 \cap \beta$ are connected and $x_i^0 \in \gamma_0$ for $1 \leq i \leq 5$ (the yellow path in Figure 3 satisfies these properties). Fix

$$x = (x_i)_{i=1}^5 = (y_i)_{i=1}^5 \in \mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}} \quad \text{with } x_i \in \alpha_i \text{ and } y_i \in \beta_i,$$

ie $(y_i)_i$ is just a permutation of $(x_i)_i$. Let $\epsilon(x_0, x)$ denote the closed 1–cycle in Σ obtained by connecting y_i to y_i^0 through β_i, connecting y_i^0 to x_i^0 through γ_0, and connecting x_i^0 to x_i through α_i for $i = 1, \dots, 5$. Note that for $j = 1, 2, 3$, the evaluation $\langle PD[\epsilon(x_0, x)], H(P_j)\rangle$ is the algebraic intersection number of $\epsilon(x_0, x)$ with $\partial_b P_j$. Therefore, if we set

$$\mathfrak{s}_z^r(x) := \big(-\langle \epsilon(x_0, x), \beta_5\rangle, \langle \epsilon(x_0, x), \beta_3\rangle, \langle \epsilon(x_0, x), \beta_4 + 2\beta_3 - 3\beta_1 + 2\beta_2\rangle\big),$$

there is a fixed triple $(a, b, c) = (0, -1, -4) \in \mathbb{Z}^3$ such that $\mathfrak{s}_z = \mathfrak{s}_z^r(x) + (a, b, c)$. In the definition of $\mathfrak{s}_z^r$, note that the intersection numbers take place over the Heegaard surface. The map $\mathfrak{s}_z^r$ is used instead of $\mathfrak{s}_z$ for the purposes of this paper.

3 Simplifying computations using algorithmic calculations

All our computations are performed with coefficients in $\mathbb{Z}_2[H^1(M, \mathbb{Z})]$. In the discussions of this section, we have the diagram $\mathscr{H} = (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$ from Figure 3 in mind. Nevertheless, the strategy works for many of the chain complexes associated with sutured manifold diagrams in the sense of [Juhász 2006], or even [Alishahi and Eftekhary 2015]. Since there is a large number of generators associated with $\mathscr{H}$, we break the computation of $\widehat{HF}(M)$ into a computer-assisted part and a human part using the following observation.

Let $z_2 \subset z_1$ denote two sets of marked points containing z. Most of the time, we take $z_2 = \{z\}$. If z_1 is sufficiently large that it contains a marked point in each one of the periodic domains, we may choose

a decomposition $\widehat{\underline{CF}}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}_1) = A \oplus B \oplus H$ such that the differentials $d_{\boldsymbol{z}_1}$ and $d_{\boldsymbol{z}_2} = d_{\boldsymbol{z}_1} + d'$ are determined by the matrices

$$
(2) \qquad d_{\boldsymbol{z}_1} = \begin{pmatrix} 0 & 0 & 0 \\ I & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad d' = \begin{pmatrix} f & h & m \\ k & g & n \\ p & q & l \end{pmatrix}.
$$

Lemma 3.1 *Suppose that, with the above notation in place, $I + k$ is invertible. Then*

$$
H_*(\widehat{\underline{CF}}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}_2)) = H_*(H, l + p(I+k)^{-1}n).
$$

Proof The proof follows from two base changes. The first base change is given by

$$
\begin{pmatrix} I+k & (I+k)f(I+k)^{-1} & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} f & h & m \\ I+k & g & n \\ p & q & l \end{pmatrix} \begin{pmatrix} (I+k)^{-1} & f(I+k)^{-1} & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}
$$
$$
= \begin{pmatrix} 0 & * & m' \\ I & * & n \\ p(I+k)^{-1} & * & l \end{pmatrix} = \begin{pmatrix} 0 & m'A & m' \\ I & nA & n \\ A & lA & l \end{pmatrix},
$$

where $A = p(I+k)^{-1}$ and the last equality follows from $d_{\boldsymbol{z}_2}^2 = 0$. The second base change is

$$
\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & A & I \end{pmatrix} \begin{pmatrix} 0 & m'A & m' \\ I & nA & n \\ A & lA & l \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & A & I \end{pmatrix} = \begin{pmatrix} 0 & 0 & m' \\ I & 0 & n \\ 0 & 0 & l + An \end{pmatrix}. \qquad \square
$$

In applications of Lemma 3.1, we choose an area assignment $\mathcal{A}$ for the regions in $\Sigma \setminus \boldsymbol{\alpha} \cup \boldsymbol{\beta}$ such that $\mathcal{A}(P_i) = 0$ for $i = 1, 2, 3$. Moreover, $\boldsymbol{z}_1$, $\boldsymbol{z}_2$ and $\mathcal{A}$ are chosen so that the regions not touched by $\boldsymbol{z}_1$ have very small areas and the regions containing marked points from $\boldsymbol{z}_1 \setminus \boldsymbol{z}_2$ have very large areas. Under these assumptions, in each Spin^c class $\mathfrak{s}$, $\mathcal{A}$ descends to an energy filtration on $\widehat{\underline{CF}}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}_2, \mathfrak{s})$ (see [Ozsváth and Szabó 2004c]). We may further assume that

$$
A = \langle a_1, \dots, a_r \rangle, \quad B = \langle b_1, \dots, b_r \rangle, \quad \text{with } \mathcal{A}(a_1) < \mathcal{A}(a_2) < \dots < \mathcal{A}(a_r),
$$

while the differential $d_{\boldsymbol{z}_1}$ is given by sending a_i to b_i. With respect to the energy filtration, $\mathcal{A}(a_i) - \mathcal{A}(b_i)$ is then a small positive number and k is a lower triangular matrix with zeros on the diagonal. Therefore, $I + k$ is an invertible matrix with $(I + k)^{-1} = \sum_{i=0}^{\infty} k^i$. This allows us to use Lemma 3.1. Of course, the use of Lemma 3.1 is not restricted to the aforementioned situation.

In our search for the Spin^c classes $\mathfrak{s}$ with the property that $\widehat{HF}(M, \mathfrak{s}) \neq 0$, we may first restrict our attention to the Spin^c classes which satisfy $\langle c_1(\mathfrak{s}), H(P_1) \rangle = 0$, since P_1 corresponds to $\partial^+ N$ and is represented by an embedded surface of genus 1. We may then enlarge the set $\boldsymbol{z}_2 = \{z\}$ of punctures in the Heegaard diagram to a bigger set $\boldsymbol{z}_1$, so that $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}_1)$ is nice, while the criteria discussed in the previous two paragraphs is satisfied. If the group

$$
\widehat{\underline{HF}}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}_1, \mathfrak{s})
$$
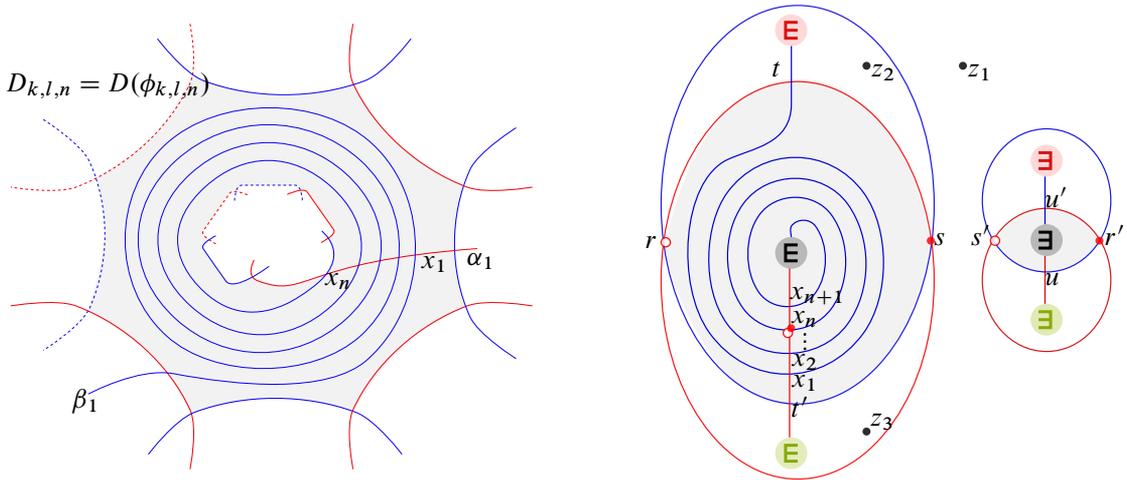
Figure 4: Part of a Heegaard diagram which illustrates the domain associated with the disk $\phi_{k,1,n}$ is illustrated (left). The red curves are the α curves and the blue curves are the β curves. A Heegaard diagram of genus 3 containing the domain $D_{1,1,n}$ is illustrated on the right. The domain of the Whitney disk $\psi \in \pi_2(R_n, U_n)$ is shaded.

is trivial, it follows that $\widehat{HF}(M, \mathfrak{s})$ is also trivial. Trying different sets z_1 of marked points allows us to exclude many of the Spin^c classes $\mathfrak{s}$ from the Thurston polytope, consisting of Spin^c structures $\mathfrak{t}$ with $\widehat{HF}(M, \mathfrak{t}) \neq 0$. Among the remaining Spin^c classes, we combine Lemma 3.1, computer assisted computations and the study of certain classes of Whitney disks (from next section) to show that $\widehat{HF}(M, \mathfrak{s}_i) \neq 0$ for $i = 1, 2$, where $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are the classes of generators x_1 and x_2 of $\widehat{CF}(\mathcal{H})$ with

$$\mathfrak{s}_z^r(x_1) = (0, 1, 7) \quad \text{and} \quad \mathfrak{s}_z^r(x_2) = (0, -1, -8),$$

respectively. As we will see in Sections 5 and 6, there are 8 generators x_1 and 72 generators x_2 of the above type.

4 Nonpolygonal disks with holomorphic representatives

In this section, we study the moduli spaces associated with three classes of Whitney disks with nonpolygonal domains, which will be encountered in Section 6. First, let $D_{k,l,n} = D(\phi_{k,l,n})$ denote the genus zero domain of a Whitney disk $\phi_{k,l,n}$, with two boundary components having $2k$–edges and $2l$–edges, respectively. The edges on each boundary component consist of alternating arcs from distinct α and β curves. For such a disk to have Maslov index 1, it is necessary that all the $2(k + l)$ angles on the boundary are acute angles, except for precisely one of them. We further assume that the obtuse angle is on the boundary component with $2l$ edges, where α_1 and β_1 meet at x_n and enter the interior of $D_{k,l,n}$, and intersect each other at $x_{n-1}, \ldots, x_1$ in $D_{k,l,n}^\circ$. There is some extra freedom in choosing the domain $D_{k,l,n}$ (up to isotopy of the curves) which corresponds to the edges where α_1 and β_1 exit $D_{k,l,n}$ and are dropped from the notation (see Figure 4, left).

Lemma 4.1 Let $\phi_{k,l,n}$ be a disk with a domain as described above. Then $\#\widehat{M}(\phi_{k,l,n}) = 1$.

Proof First, consider the case $k = l = 1$. Consider the triply punctured Heegaard diagram

$$\mathscr{H}_1 = \mathscr{H}_1^n = (\Sigma_1, \boldsymbol{\alpha}_1 = \{\alpha_1, \alpha_2, \alpha_3\}, \boldsymbol{\beta}_1 = \{\beta_1, \beta_2, \beta_3\}, z_1, z_2, z_3)$$

illustrated in Figure 4, right. Here Σ_1 is a surface of genus three and the sutured manifold determined by $\mathscr{H}_1$ is the same as the sutured manifold determined by a Heegaard diagram

$$(T = S^1 \times S^1, \alpha, \beta = \{b\} \times S^1, z_1, z_2, z_3),$$

where α is homotopically trivial and cuts β twice, one of the punctures is located in one of the two bigons in $T - \alpha - \beta$, and two of the punctures are located in the cylindrical component of $T - \alpha - \beta$. Therefore, the Heegaard Floer group associated with $\mathscr{H}_1$ is trivial. With the notation of Figure 4, the generators in $\mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$ are

$$R_i = (x_i, r, r'), \quad S_i = (x_i, r, s'), \quad T_i = (x_i, s, r'), \quad U_i = (x_i, s, s') \quad i = 1, \ldots, n+1,$$
$$V = (t, t', r'), \quad W = (t, t', s'), \quad X = (u, r, u'), \quad Y = (u, s, u').$$

Most Whitney disks with positive domain and index 1 which contribute to the differential are of the form $\phi_{1,1,k}$ for some $k = 1, \ldots, n$. In fact, there are Whitney disks

$$\psi_k^1 \in \pi_2(U_k, S_{k+1}), \quad \psi_k^2 \in \pi_2(T_k, R_{k+1}), \quad \psi_{n+1-k}^3 \in \pi_2(R_{k+1}, S_k), \quad \psi_{n+1-k}^4 \in \pi_2(T_{k+1}, U_k)$$

for $k = 1, \ldots, n$, where each ψ_k^i is of type $\phi_{1,1,k}$. Other than these classes, there are also disks

$$\psi_0^1 \in \pi_2(V, R_1), \quad \psi_0^2 \in \pi_2(W, S_1), \quad \psi_0^3 \in \pi_2(X, S_{n+1}), \quad \psi_0^4 \in \pi_2(Y, U_{n+1}),$$

with Maslov index one, and the domain of every one of them is a rectangle. Therefore, $\#\widehat{M}(\psi_0^i) = 1$ for $i = 1, \ldots, 4$. Moreover, there are disks $\phi \in \pi_2(T_1, W)$ and $\phi' \in \pi_2(T_{n+1}, X)$ with domains of type $\phi_{1,2,n}$. If we set

$$m = \#\widehat{M}(\phi), \quad m' = \#\widehat{M}(\phi'), \quad m_k^i = \#\widehat{M}(\psi_k^i) \quad \text{for } i = 1, \ldots, 4 \text{ and } k = 0, \ldots, n,$$

it follows that $m_0^i = m_1^i = 1$ (see [Ozsváth and Szabó 2004b, Lemma 3.4; Sarkar and Wang 2010, Theorem 3.4]) and that the differential of the chain complex is given by

$$
\begin{array}{ccccccccc}
T_k \xrightarrow{m_k^2} R_{k+1} & & T_1 \xrightarrow{1} R_2 & & T_{n+1} \xrightarrow{m'} X & & Y & & V \\
\Big\downarrow{m_{n+2-k}^4} \quad \Big\downarrow{m_{n+1-k}^3} & & m\Big\downarrow \quad \Big\downarrow{m_n^3} & & 1\Big\downarrow \quad \Big\downarrow 1 & & 1\Big\downarrow & & 1\Big\downarrow \\
U_{k-1} \xrightarrow{m_{k-1}^1} S_k & & W \xrightarrow{1} S_1 & & U_n \xrightarrow{m_n^1} S_{n+1} & & U_{n+1} & & R_1
\end{array}
$$

for $k = 2, \ldots, n$. Therefore, we conclude that $m = m_n^3$, $m' = m_n^1$ and

(3) $$m_k^2 \cdot m_{n+1-k}^3 = m_{n+2-k}^4 \cdot m_{k-1}^1 \quad \text{for } k = 2, \ldots, n.$$
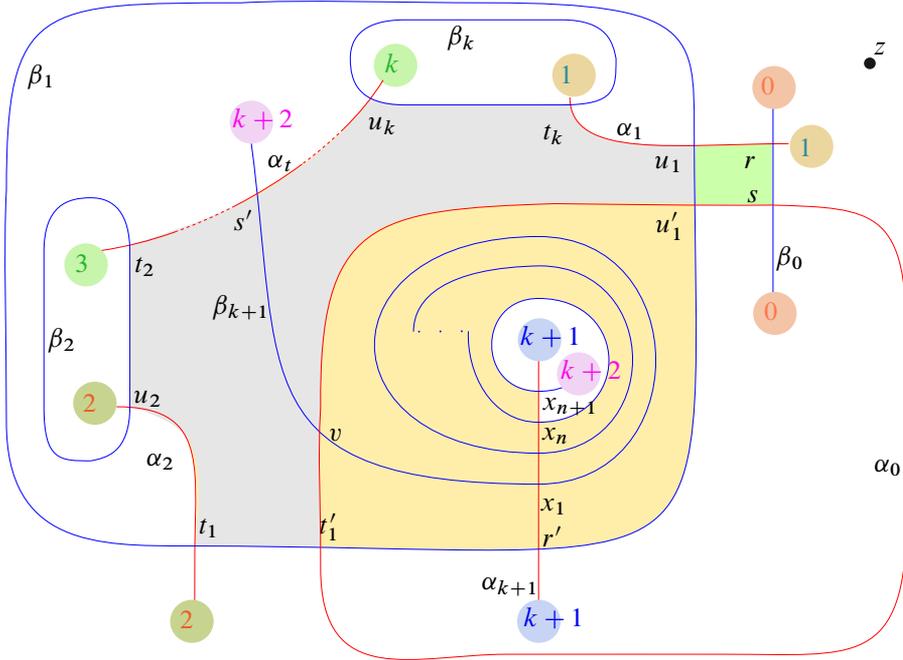
Figure 5: A Heegaard surface of genus $k + 3$.The colored region denotes the domain associated with ψ, which degenerates in two ways as $\phi_{1,1,n} * \phi$ and $\phi' * \phi_{k,1,n}$, where $D(\phi')$ is the region colored green and $D(\phi_{1,1,n})$ is the union of regions colored yellow.

For $n = 2$ and $k = 2$, (3) implies $m_2^2 = m_2^4$. Moreover, since the homology is trivial, $m_2^2 = m_2^4 = 1$. This proves the claim for $\phi_{1,1,2}$. Having established the proof for $\phi_{1,1,j}$ with $j = 1, \ldots, n-1$ (where $n > 2$), equation (3) for $k = 2$ implies that $m_n^4 = 1$, proving the claim for $\phi_{1,1,n}$.

Next, we consider the case $l = 1$ while k is arbitrary. Let

$$\mathcal{H}_2 = (\Sigma_2, \boldsymbol{\alpha}_2 = \{\alpha_0, \alpha_1, \ldots, \alpha_{k+1}\}, \boldsymbol{\beta}_2 = \{\beta_0, \beta_1, \ldots, \beta_{k+1}\}, z)$$

be the Heegaard diagram shown in Figure 5. Here Σ_2 is a surface of genus $k + 3$. With the notation of Figure 5 in place and refreshing the notation set for the case $k = l = 1$, the generators in $\mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$ are

$$
\begin{aligned}
W &= (v, r, u_2, \ldots, u_k, r'), & V &= (s, t_k, u_2, \ldots, u_{t-1}, s', t_t, \ldots, t_{k-1}, r'), \\
R_i &= (u_1', r, u_2, \ldots, u_k, x_i), & S_i &= (t_1', r, u_2, \ldots, u_k, x_i), \\
T_i &= (s, t_k, t_1, \ldots, t_{k-1}, x_i), & U_i &= (s, u_1, u_2, \ldots, u_k, x_i),
\end{aligned}
$$

where i belongs to $\{1, \ldots, n+1\}$. Consider the Whitney disks

$$\phi_{1,1,n} \in \pi_2(R_n, S_{n+1}), \quad \phi \in \pi_2(S_{n+1}, T_{n+1}), \quad \phi' \in \pi_2(R_n, U_n), \quad \phi_{k,1,n} \in \pi_2(U_n, T_{n+1})$$

such that $D(\phi')$ is the green domain, $D(\phi_{1,1,n})$ is the union of yellow domains, $D(\phi)$ is the union of gray and green domains and $D(\phi_{k,1,n})$ is the union of gray and yellow domains. Then

$$\psi = \phi_{1,1,n} * \phi = \phi' * \phi_{k,1,n}$$
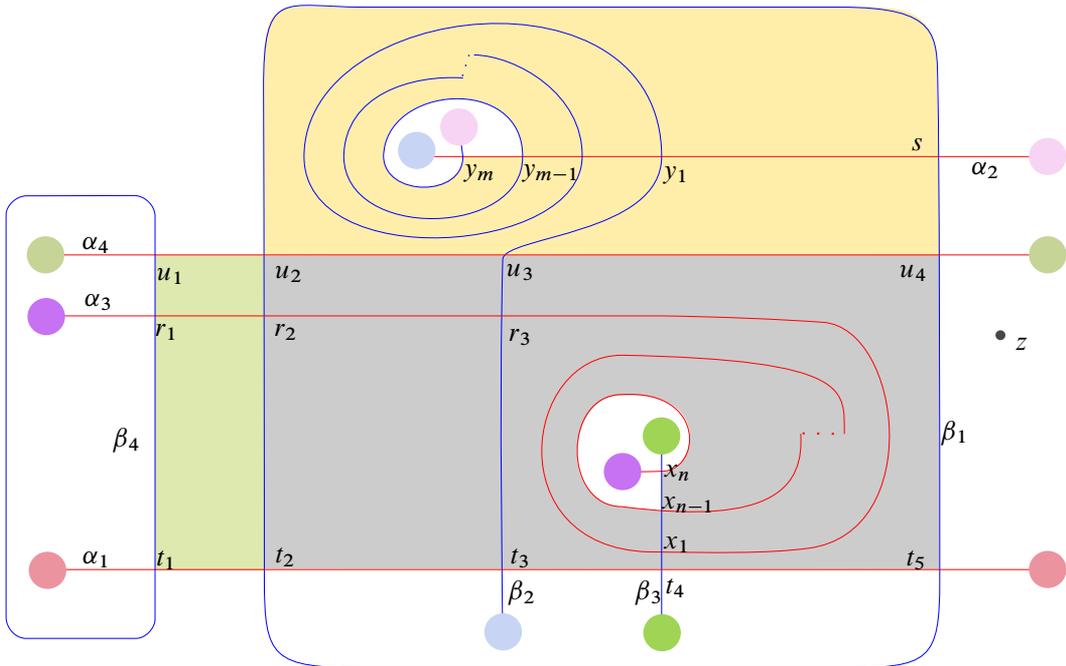
Figure 6: A Heegaard surface of genus 6. The colored regions denote the domain associated with a Whitney disk ψ of index 2 which degenerates in two ways.

has index 2, while these are the only degenerations of ψ as a juxtaposition of two positive Whitney disks of Maslov index 1. This implies

$$\#\widehat{M}(\phi_{k,1,n}) = \#\widehat{M}(\phi_{k,1,n}) \cdot \#\widehat{M}(\phi') = \#\widehat{M}(\phi_{1,1,n}) \cdot \#\widehat{M}(\phi) = \#\widehat{M}(\phi_{1,1,n}) = 1,$$

completing the proof for the case where $l = 1$, while k and n are arbitrary. Similarly, the argument above may be used to conclude $\#\widehat{M}(\phi_{k,l,n}) = 1$ for arbitrary values of k, l and n. $\qquad\square$

Let $D(\phi_{n,m})$ denote the genus zero domain of a Whitney disk $\phi_{n,m}$ which has three boundary components, each consisting of 2 edges on α_i and β_i for $i = 1, 2, 3$. Let α_3 have n intersection points $\{x_1, \ldots, x_n\}$ with β_3 and α_2 have m intersection points $\{y_1, \ldots, y_m\}$ with β_2 in $D(\phi_{n,m})$. The union of the yellow regions and the gray regions in Figure 6 illustrates the domain of such a disk. We assume that all the corners of the boundary edges in $D(\phi_{n,m})$ are acute except for two, where α_2 intersects β_2 in an obtuse angle in y_{m-1} and α_3 intersects β_3 in an obtuse angle in x_{n-1} (see Figure 6).

Lemma 4.2 *If the domain of $\phi_{n,m}$ is as described above, then $\#\widehat{M}(\phi_{n,m}) = 1$.*

Proof Consider the Heegaard diagram

$$\mathcal{H}_3 = (\Sigma, \boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}, \boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3, \beta_4\}, z)$$

which is illustrated in Figure 6. Here Σ is a surface of genus six which is obtained by attaching six one-handles such that each one connects the boundary circles of disks with the same color. There are
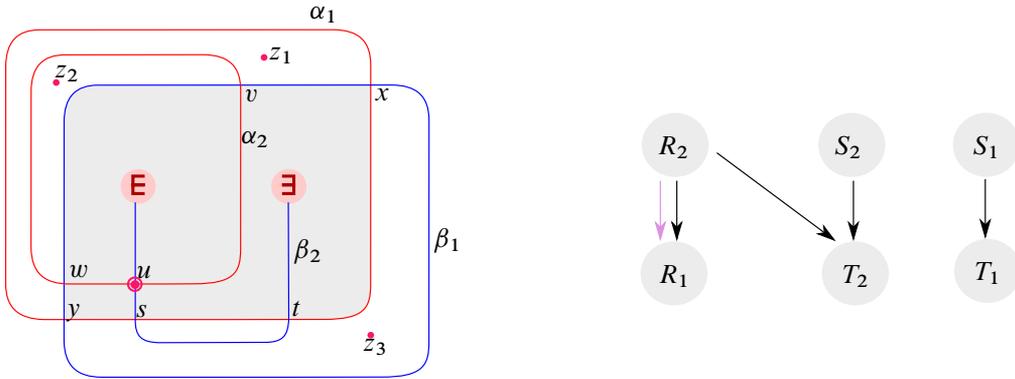
Figure 7: Left: a Heegaard diagram with three marked points and a Heegaard surface of genus one. Right: the differential associated with this diagram.

$4nm + 3m + 2n + 2$ intersection points in $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$. With the notation of Figure 6 in place, these intersection points are

$$P = (t_4, u_1, r_3, s), \qquad Q = (t_4, u_3, r_1, s), \qquad R_i = (t_1, u_3, x_i, s), \qquad S_i = (t_3, u_1, x_i, s),$$

$$T_{i,j} = (t_2, u_1, x_i, y_j), \quad U_{i,j} = (t_5, u_1, x_i, y_j), \quad V_{i,j} = (t_1, u_2, x_i, y_j), \quad W_{i,j} = (t_1, u_4, x_i, y_j),$$

$$X_j = (t_4, u_2, r_1, y_j), \qquad Y_j = (t_4, u_4, r_1, y_j), \qquad Z_j = (t_4, u_1, r_2, y_j),$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Consider the Whitney disks of index 1

$$\phi_{1,1,m-1} \in \pi_2(V_{n-1,m-1}, W_{n-1,m}), \qquad \phi_{2,1,n-1} \in \pi_2(W_{n-1,m}, U_{n,m}),$$

$$\phi \in \pi_2(V_{n-1,m-1}, T_{n-1,m-1}), \qquad \phi_{n,m} \in \pi_2(T_{n-1,m-1}, U_{n,m}).$$

Here, the domains $D(\phi_{1,1,m-1})$ is the union of the regions colored yellow, $D(\phi)$ is the union of the regions colored green, $D(\phi_{2,1,n-1})$ is the union of the regions colored gray and green, and $D(\phi_{n,m})$ is the union of the regions colored yellow and gray in Figure 6. Then

$$\psi = \phi_{1,1,m-1} * \phi_{2,1,n-1} = \phi * \phi_{n,m},$$

determined by $D(\psi)$ which is the union of all colored regions, is a Whitney disk of Maslov index 2 in $\pi_2(V_{n-1,m-1}, U_{n,m})$. The disk ψ degenerates as the juxtaposition of two disks of Maslov index 1 only in the above two ways. Therefore, we conclude that

$$\#\widehat{M}(\phi_{n,m}) = \#\widehat{M}(\phi_{n,m}) \cdot \#\widehat{M}(\phi) = \#\widehat{M}(\phi_{1,1,m-1}) \cdot \#\widehat{M}(\phi_{2,1,n-1}) = 1.$$

The last equality, which follows from Lemma 4.1, completes the proof of the lemma.   □

For $\phi \in \pi_2(x, y)$, let $D(\phi)$ be a surface of genus one with one boundary component consisting of 2 edges that contains a unique intersection point u in the interior which belongs to both x and y (see Figure 7). The gray domains on the left illustrate $D(\phi)$.

Lemma 4.3 *Let ϕ be a disk with a domain as described above. Then $\#\widehat{M}(\phi) = 1$.*

Proof Consider the triply punctured Heegaard diagram

$$\mathcal{H}_4 = (\Sigma, \boldsymbol{\alpha} = \{\alpha_1, \alpha_2\}, \boldsymbol{\beta} = \{\beta_1, \beta_2\}, \boldsymbol{z} = \{z_1, z_2, z_3\})$$

of genus 1 which is illustrated in Figure 7, left. With the notation of Figure 7 in place, there are 6 intersection points in $\mathbb{T}_{\boldsymbol{\alpha}} \cap \mathbb{T}_{\boldsymbol{\beta}}$, which may be listed as

$$R_1 = (x, u), \quad R_2 = (y, u), \quad S_1 = (t, v), \quad S_2 = (t, w), \quad T_1 = (s, v), \quad T_2 = (s, w).$$

The differential is shown in Figure 7, right. A black arrow which connects a generator X to a generator Y denotes that there is a disk from X to Y with a unique holomorphic representative. The arrow in purple denotes the disk with the domain $D(\phi)$. By doing an isotopy which removes the two intersection points of β_2 with α_1 and then doing a destabilization which removes α_2 and β_2, we obtain the standard genus zero Heegaard diagram for the closed three manifold $S^1 \times S^2$. Therefore the Heegaard Floer homology group associated with $\mathcal{H}$ is $\mathbb{Z}_2^2$. This proves that $\#\hat{M}(\phi) = 1$. $\qquad\square$

5 The first nontrivial Heegaard–Floer group

Let us assume that $\mathfrak{s}_1$ corresponds to the triple $(0, 1, 7)$. The chain complex $\widehat{CF}(M, \mathfrak{s}_1)$ is then generated by the following 8 generators:

$$\begin{aligned}
&\text{①} = \{x_{1,1,2}, x_{2,2,5}, x_{3,3,1}, x_{4,5,2}, x_{5,4,2}\}, \quad &&\text{②} = \{x_{1,1,2}, x_{2,2,5}, x_{3,3,2}, x_{4,5,2}, x_{5,4,2}\}, \\
&\text{③} = \{x_{1,1,2}, x_{2,4,2}, x_{3,2,1}, x_{4,5,2}, x_{5,3,2}\}, \quad &&\text{④} = \{x_{1,1,3}, x_{2,4,2}, x_{3,2,2}, x_{4,5,2}, x_{5,3,2}\}, \\
&\text{⑤} = \{x_{1,2,1}, x_{2,1,2}, x_{3,5,1}, x_{4,3,1}, x_{5,4,2}\}, \quad &&\text{⑥} = \{x_{1,2,1}, x_{2,4,2}, x_{3,5,1}, x_{4,1,2}, x_{5,3,2}\}, \\
&\text{⑦} = \{x_{1,3,1}, x_{2,1,2}, x_{3,2,1}, x_{4,5,2}, x_{5,4,2}\}, \quad &&\text{⑧} = \{x_{1,4,2}, x_{2,1,2}, x_{3,2,2}, x_{4,5,2}, x_{5,3,2}\}.
\end{aligned}$$

Let z_1 consist of marked points in all domains except for D_{12}, D_{13}, D_{30} and D_{49}. The differentials for the Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1)$ along with the domains of the connecting disks are shown in Figure 8, left. In this figure, a black arrow from a generator x to a generator y indicates that there is a disk from x to y with a unique holomorphic representative. In fact, the domains associated with all the
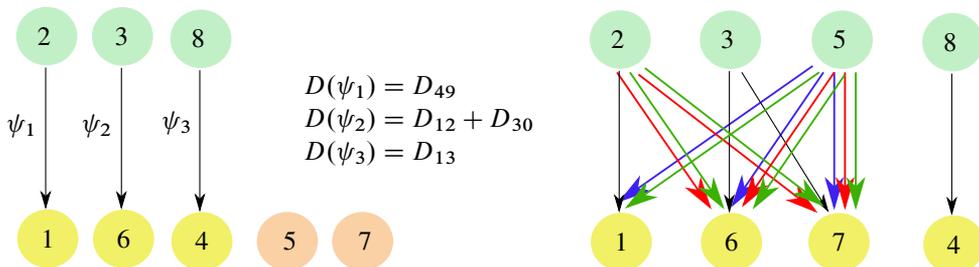

Figure 8: Left: the differential for $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1)$. Right: the differential for $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$; the contributions from the disks ϕ_i for $i = 1, \ldots, 5$, the disks $\phi_i - P_1$ for $i = 1, \ldots, 4$ and the disks $\phi_i + P_1$ for $i = 1, 2, 5$ are denoted with green, red and blue arrows, respectively.

disks are polygons. The group $H_*(\widehat{CF}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1), \mathfrak{s}_1)$ is thus isomorphic to $(\mathbb{Z}_2[H^1(M; \mathbb{Z})])^2$ and is generated by $C = \{\circled{5}, \circled{7}\}$. To compute $\widehat{HF}(M, \mathfrak{s}_1)$, we need to determine the matrices l, n, p and k in Lemma 3.1. Define the disks $\phi \in \pi_2(\circled{3}, \circled{7})$ and

$$\phi_1 \in \pi_2(\circled{5}, \circled{7}), \quad \phi_2 \in \pi_2(\circled{5}, \circled{6}), \quad \phi_3 \in \pi_2(\circled{2}, \circled{7}), \quad \phi_4 \in \pi_2(\circled{2}, \circled{6}), \quad \phi_5 \in \pi_2(\circled{5}, \circled{1})$$

of Maslov index 1 by specifying their domains. If $I = \{4, 16, 18, 26, 34, 36, 38, 51, 52\}$ and D is the formal sum $\sum_{i \in I} D_i$, we have

$$\begin{aligned} D(\phi) &= D_0, & D(\phi_1) &= \Sigma - \{D_{12} + D_{14} + D_{30}\}, \\ D(\phi_2) &= \Sigma - \{D_0 + D_{14}\}, & D(\phi_3) &= \Sigma - (D + D_{12} + D_{14} + D_{30}), \\ D(\phi_4) &= \Sigma - (D + D_0 + D_{14}), & D(\phi_5) &= D + D_{49}. \end{aligned}$$

Then all the disks of index 1 and positive domain between the generators of this complex are the disks ψ_i for $i = 1, 2, 3$, the disk ϕ, the disks ϕ_i for $i = 1, \ldots, 5$, the disks $\phi_i - P_1$ for $i = 1, \ldots, 4$ and the disks $\phi_i + P_1$ for $i = 1, 2, 5$; see Figure 8, right. Let

$$\begin{aligned} b_i &= \#\widehat{M}(\phi_i) \quad \text{for } i = 1, \ldots, 5, & c_i &= \#\widehat{M}(\phi_i - P_1) \quad \text{for } i = 1, \ldots, 4, \\ d_i &= \#\widehat{M}(\phi_i + P_1) \quad \text{for } i = 1, 2, & c_5 &= \#\widehat{M}(\phi_5 + P_1). \end{aligned}$$

Setting $K = b_4 + c_4 e^{-P_1}$, $N_1 = b_5 + c_5 e^{P_1}$, $N_2 = b_2 + c_2 e^{-P_1} + d_2 e^{P_1}$, $P = b_3 + c_3 e^{-P_1}$ and $L = b_1 + c_1 e^{-P_1} + d_1 e^{P_1}$, it then follows that

$$k = \begin{pmatrix} 0 & 0 & 0 \\ K & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad n = \begin{pmatrix} N_1 & 0 \\ N_2 & 0 \\ 0 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 0 & 0 & 0 \\ P & 1 & 0 \end{pmatrix}, \quad l = \begin{pmatrix} 0 & 0 \\ L & 0 \end{pmatrix}$$

$$\implies \quad l + p(I + k)^{-1}n = \begin{pmatrix} 0 & 0 \\ L + N_1(P + K) + N_2 & 0 \end{pmatrix}.$$

Note that in this matrix,

$$\begin{aligned} (4) \quad \star &= L + N_1(P + K) + N_2 \\ &= (b_1 + b_2 + b_5 b_3 + b_5 b_4 + c_5 c_3 + c_5 c_4) + (c_1 + c_2 + b_5 c_3 + b_5 c_4)e^{-P_1} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad + (d_1 + d_2 + c_5 b_3 + c_5 b_4)e^{P_1}. \end{aligned}$$

The computation of $\widehat{HF}(M, \mathfrak{s}_1)$ is thus reduced to a computation of $\star$. Consider the disks

$$\begin{aligned} \lambda_1 &\in \pi_2(\circled{7}, \circled{5}), & \lambda_2 &\in \pi_2(\circled{6}, \circled{5}), & \lambda_3 &\in \pi_2(\circled{7}, \circled{2}), \\ \lambda_4 &\in \pi_2(\circled{6}, \circled{2}), & \lambda_5 &\in \pi_2(\circled{1}, \circled{5}), & \lambda_6 &\in \pi_2(\circled{1}, \circled{2}), \end{aligned}$$

which correspond to the domains

$$D(\lambda_i) = \Sigma - D(\phi_i) \quad \text{for } i = 1, \ldots, 5, \qquad D(\lambda_6) = \Sigma - D(\psi_1).$$

The domains of λ_1 and λ_2 are polygons. Then all the positive disks λ of index 1 in $\pi_2(\circled{1}, x)$, $\pi_2(x, \circled{5})$ and $\pi_2(x, \circled{2})$, with x a generator of $\widehat{CF}(M, \mathfrak{s}_1)$ and with $n_z(\lambda) > 0$, are $\lambda = \lambda_i$ for $i = 1, \ldots, 6$,

$\lambda = \lambda_i + P_1$ for $i = 3, 4, 6$, and $\lambda = \lambda_i - P_1$ for $i = 5, 6$. Let

$$b_i' = \#\widehat{M}(\lambda_i) \quad \text{for } i = 3, \dots, 6, \qquad c_i' = \#\widehat{M}(\lambda_i + P_1) \quad \text{for } i = 3, 4,$$
$$c_i' = \#\widehat{M}(\lambda_i - P_1) \quad \text{for } i = 5, 6, \qquad d_6' = \#\widehat{M}(\lambda_6 + P_1).$$

Consider the Whitney disk classes of index 2,

$$\eta_1, \eta_1', \eta_1'' \in \pi_2(\text{⑤}, \text{⑤}), \quad \eta_2, \eta_2', \eta_2'' \in \pi_2(\text{①}, \text{①}), \quad \eta_3, \eta_3', \eta_3'' \in \pi_2(\text{②}, \text{②}),$$

which correspond to the periodic domains

$$D(\eta_i) = \Sigma, \quad D(\eta_i') = \Sigma - P_1, \quad D(\eta_i'') = \Sigma + P_1 \quad \text{for } i = 1, 2, 3.$$

The possible degenerations of η_1, η_1' and η_1'' to positive disks of Maslov index 1 are

$$\eta_1 = \phi_j * \lambda_j = (\phi_5 + P_1) * (\lambda_5 - P_1) \quad \text{for } j = 1, 2, 5,$$
$$\eta_1' = (\phi_i - P_1) * \lambda_i = \phi_5 * (\lambda_5 - P_1) \quad \text{for } i = 1, 2,$$
$$\eta_1'' = (\phi_i + P_1) * \lambda_i = (\phi_5 + P_1) * \lambda_5 \quad \text{for } i = 1, 2.$$

Therefore,

$$(5) \qquad b_1 + b_2 + b_5'b_5 + c_5'c_5 = 0, \quad c_1 + c_2 + c_5'b_5 = 0, \quad d_1 + d_2 + b_5'c_5 = 0.$$

The possible degenerations of η_2, η_2' and η_2'' into positive disks of Maslov index 1 are

$$\eta_2 = \lambda_6 * \psi_1 = \lambda_5 * \phi_5 = (\lambda_5 - P_1) * (\phi_5 + P_1),$$
$$\eta_2' = (\lambda_6 - P_1) * \psi_1 = (\lambda_5 - P_1) * \phi_5,$$
$$\eta_2'' = (\lambda_6 + P_1) * \psi_1 = \lambda_5 * (\phi_5 + P_1).$$

Therefore,

$$(6) \qquad b_6' + b_5'b_5 + c_5'c_5 = 0, \quad c_6' + c_5'b_5 = 0, \quad d_6' + b_5'c_5 = 0.$$

Similarly, the possible degeneration of η_3, η_3' and η_3'' into positive disks of Maslov index 1 are

$$\eta_3 = \phi_i * \lambda_i = (\phi_i - P_1) * (\lambda_i + P_1) = \psi_1 * \lambda_6,$$
$$\eta_3' = (\phi_i - P_1) * \lambda_i = \psi_1 * (\lambda_6 - P_1),$$
$$\eta_3'' = \phi_i * (\lambda_i + P_1) = \psi_1 * (\lambda_6 + P_1)$$

for $i = 3, 4$. Therefore,

$$(7) \qquad b_3'b_3 + b_4'b_4 + c_3'c_3 + c_4'c_4 + b_6' = 0, \quad b_3'c_3 + b_4'c_4 + c_6' = 0, \quad c_3'b_3 + c_4'b_4 + d_6' = 0.$$

Let z_1' contain a marked point in all the regions of $\Sigma - \boldsymbol{\alpha} - \boldsymbol{\beta}$ except for those appearing in $D(\lambda_3)$, $D(\lambda_4)$, $D(\phi_5)$ and D_{13}, and let ∂_1 denote the corresponding differential. Note that P_1, P_2 and $P_3 - \Sigma$ may still be considered as a basis for the space of periodic domains. Therefore, the diagram remains admissible for this choice of marked points. Then

$$(8) \qquad \partial_1^2 \text{③} = (b_3' + b_4') \text{②} \quad \text{and} \quad \partial_1^2 \text{⑦} = (b_5 + b_3') \text{①} \quad \Longrightarrow \quad b_3' = b_4' = b_5.$$

Similarly, let z_2' contain a marked point in all the regions of $\Sigma - \boldsymbol{\alpha} - \boldsymbol{\beta}$ except for those appearing in $D(\lambda_3 + P_1)$, $D(\lambda_4 + P_1)$, $D(\phi_5 + P_1)$ and D_{13}, and let ∂_2 denote the corresponding differential. Then

$$\text{(9)} \qquad \partial_2^2 \text{③} = (c_3' + c_4')\text{②} \quad \text{and} \quad \partial_2^2 \text{⑦} = (c_5 + c_3')\text{①} \quad \Longrightarrow \quad c_3' = c_4' = c_5.$$

If follows from (4)–(9) that the matrix $l + p(I + k)^{-1}n = 0$. Thus $\widehat{HF}(M, \mathfrak{s}_1) \neq 0$.

6 The second nontrivial Heegaard–Floer group

Let $\mathfrak{s}_2$ be the Spin^c class which corresponds to $(0, -1, -8)$. The chain complex $\widehat{CF}(M, \mathfrak{s}_2)$ is bigger, in comparison with $\widehat{CF}(M, \mathfrak{s}_1)$, and is generated by the following 72 generators:

$$\text{①} = \{x_{1,1,2}, x_{2,2,4}, x_{3,4,1}, x_{4,5,1}, x_{5,3,2}\}, \qquad \text{②} = \{x_{1,1,3}, x_{2,2,5}, x_{3,4,1}, x_{4,5,1}, x_{5,3,2}\},$$

$$\text{③} = \{x_{1,1,2}, x_{2,2,5}, x_{3,5,1}, x_{4,3,1}, x_{5,4,1}\}, \qquad \text{④} = \{x_{1,1,2}, x_{2,3,2}, x_{3,2,2}, x_{4,5,1}, x_{5,4,1}\},$$

$$\text{⑤} = \{x_{1,1,2}, x_{2,3,2}, x_{3,5,2}, x_{4,2,1}, x_{5,4,1}\}, \qquad \text{⑥} = \{x_{1,1,2}, x_{2,4,1}, x_{3,2,2}, x_{4,5,1}, x_{5,3,1}\},$$

$$\text{⑦} = \{x_{1,1,2}, x_{2,4,1}, x_{3,5,2}, x_{4,2,1}, x_{5,3,1}\}, \qquad \text{⑧} = \{x_{1,1,3}, x_{2,4,1}, x_{3,5,1}, x_{4,2,1}, x_{5,3,2}\},$$

$$\text{⑨} = \{x_{1,2,2}, x_{2,3,2}, x_{3,1,1}, x_{4,5,1}, x_{5,4,2}\}, \qquad \text{⑩} = \{x_{1,2,2}, x_{2,4,2}, x_{3,1,1}, x_{4,5,1}, x_{5,3,1}\},$$

$$\text{⑪} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,1}, x_{4,4,7}, x_{5,5,1}\}, \qquad \text{⑫} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,1}, x_{4,4,7}, x_{5,5,2}\},$$

$$\text{⑬} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,2}, x_{4,4,5}, x_{5,5,1}\}, \qquad \text{⑭} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,2}, x_{4,4,5}, x_{5,5,2}\},$$

$$\text{⑮} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,3}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{⑯} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,3}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{⑰} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,4}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{⑱} = \{x_{1,3,1}, x_{2,1,1}, x_{3,2,4}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{⑲} = \{x_{1,3,2}, x_{2,1,1}, x_{3,2,1}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{⑳} = \{x_{1,3,2}, x_{2,1,1}, x_{3,2,1}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{㉑} = \{x_{1,3,2}, x_{2,1,1}, x_{3,2,2}, x_{4,4,8}, x_{5,5,1}\}, \qquad \text{㉒} = \{x_{1,3,2}, x_{2,1,1}, x_{3,2,2}, x_{4,4,8}, x_{5,5,2}\},$$

$$\text{㉓} = \{x_{1,3,1}, x_{2,2,3}, x_{3,1,2}, x_{4,5,1}, x_{5,4,2}\}, \qquad \text{㉔} = \{x_{1,3,2}, x_{2,2,3}, x_{3,1,1}, x_{4,5,1}, x_{5,4,2}\},$$

$$\text{㉕} = \{x_{1,3,1}, x_{2,2,1}, x_{3,1,1}, x_{4,4,7}, x_{5,5,1}\}, \qquad \text{㉖} = \{x_{1,3,1}, x_{2,2,1}, x_{3,1,1}, x_{4,4,7}, x_{5,5,2}\},$$

$$\text{㉗} = \{x_{1,3,1}, x_{2,2,1}, x_{3,1,2}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{㉘} = \{x_{1,3,1}, x_{2,2,1}, x_{3,1,2}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{㉙} = \{x_{1,3,1}, x_{2,2,2}, x_{3,1,1}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{㉚} = \{x_{1,3,1}, x_{2,2,2}, x_{3,1,1}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{㉛} = \{x_{1,3,1}, x_{2,2,3}, x_{3,1,1}, x_{4,4,8}, x_{5,5,1}\}, \qquad \text{㉜} = \{x_{1,3,1}, x_{2,2,3}, x_{3,1,1}, x_{4,4,8}, x_{5,5,2}\},$$

$$\text{㉝} = \{x_{1,3,1}, x_{2,2,4}, x_{3,1,1}, x_{4,4,6}, x_{5,5,1}\}, \qquad \text{㉞} = \{x_{1,3,1}, x_{2,2,4}, x_{3,1,1}, x_{4,4,6}, x_{5,5,2}\},$$

$$\text{㉟} = \{x_{1,3,1}, x_{2,2,4}, x_{3,1,2}, x_{4,4,9}, x_{5,5,1}\}, \qquad \text{㊱} = \{x_{1,3,1}, x_{2,2,4}, x_{3,1,2}, x_{4,4,9}, x_{5,5,2}\},$$

$$\text{㊲} = \{x_{1,3,1}, x_{2,2,5}, x_{3,1,1}, x_{4,4,4}, x_{5,5,1}\}, \qquad \text{㊳} = \{x_{1,3,1}, x_{2,2,5}, x_{3,1,1}, x_{4,4,4}, x_{5,5,2}\},$$

$$\text{㊴} = \{x_{1,3,1}, x_{2,2,5}, x_{3,1,2}, x_{4,4,7}, x_{5,5,1}\}, \qquad \text{㊵} = \{x_{1,3,1}, x_{2,2,5}, x_{3,1,2}, x_{4,4,7}, x_{5,5,2}\},$$

$$\text{㊶} = \{x_{1,3,2}, x_{2,2,1}, x_{3,1,1}, x_{4,4,10}, x_{5,5,1}\}, \qquad \text{㊷} = \{x_{1,3,2}, x_{2,2,1}, x_{3,1,1}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{㊸} = \{x_{1,3,2}, x_{2,2,4}, x_{3,1,1}, x_{4,4,9}, x_{5,5,1}\}, \quad \text{㊹} = \{x_{1,3,2}, x_{2,2,4}, x_{3,1,1}, x_{4,4,9}, x_{5,5,2}\},$$

$$\text{㊺} = \{x_{1,3,2}, x_{2,2,5}, x_{3,1,1}, x_{4,4,7}, x_{5,5,1}\}, \quad \text{㊻} = \{x_{1,3,2}, x_{2,2,5}, x_{3,1,1}, x_{4,4,7}, x_{5,5,2}\},$$

$$\text{㊼} = \{x_{1,3,2}, x_{2,2,5}, x_{3,1,2}, x_{4,4,10}, x_{5,5,1}\}, \quad \text{㊽} = \{x_{1,3,2}, x_{2,2,5}, x_{3,1,2}, x_{4,4,10}, x_{5,5,2}\},$$

$$\text{㊾} = \{x_{1,3,1}, x_{2,2,3}, x_{3,4,2}, x_{4,1,1}, x_{5,5,1}\}, \quad \text{㊿} = \{x_{1,3,1}, x_{2,2,3}, x_{3,4,2}, x_{4,1,1}, x_{5,5,2}\},$$

$$\text{�51} = \{x_{1,3,1}, x_{2,2,3}, x_{3,5,2}, x_{4,1,1}, x_{5,4,2}\}, \quad \text{�52} = \{x_{1,3,1}, x_{2,2,5}, x_{3,5,1}, x_{4,1,2}, x_{5,4,1}\},$$

$$\text{�53} = \{x_{1,3,2}, x_{2,2,4}, x_{3,5,1}, x_{4,1,1}, x_{5,4,2}\}, \quad \text{�54} = \{x_{1,4,1}, x_{2,1,2}, x_{3,2,2}, x_{4,5,1}, x_{5,3,1}\},$$

$$\text{�55} = \{x_{1,4,3}, x_{2,1,1}, x_{3,2,1}, x_{4,5,1}, x_{5,3,2}\}, \quad \text{�56} = \{x_{1,4,4}, x_{2,1,1}, x_{3,2,2}, x_{4,5,1}, x_{5,3,2}\},$$

$$\text{�57} = \{x_{1,4,1}, x_{2,1,2}, x_{3,5,2}, x_{4,2,1}, x_{5,3,1}\}, \quad \text{�58} = \{x_{1,4,2}, x_{2,1,1}, x_{3,5,1}, x_{4,2,2}, x_{5,3,2}\},$$

$$\text{㊾9} = \{x_{1,4,4}, x_{2,1,1}, x_{3,5,1}, x_{4,2,1}, x_{5,3,1}\}, \quad \text{�60} = \{x_{1,4,4}, x_{2,1,1}, x_{3,5,2}, x_{4,2,1}, x_{5,3,2}\},$$

$$\text{�61} = \{x_{1,4,2}, x_{2,2,3}, x_{3,1,1}, x_{4,5,1}, x_{5,3,1}\}, \quad \text{�62} = \{x_{1,4,2}, x_{2,2,4}, x_{3,1,2}, x_{4,5,1}, x_{5,3,2}\},$$

$$\text{�63} = \{x_{1,4,3}, x_{2,2,1}, x_{3,1,1}, x_{4,5,1}, x_{5,3,2}\}, \quad \text{�64} = \{x_{1,4,3}, x_{2,2,4}, x_{3,1,1}, x_{4,5,1}, x_{5,3,1}\},$$

$$\text{�65} = \{x_{1,4,3}, x_{2,2,5}, x_{3,1,2}, x_{4,5,1}, x_{5,3,2}\}, \quad \text{�66} = \{x_{1,4,4}, x_{2,2,5}, x_{3,1,1}, x_{4,5,1}, x_{5,3,1}\},$$

$$\text{�67} = \{x_{1,4,1}, x_{2,2,5}, x_{3,5,1}, x_{4,1,2}, x_{5,3,2}\}, \quad \text{�68} = \{x_{1,4,2}, x_{2,2,1}, x_{3,5,1}, x_{4,1,1}, x_{5,3,2}\},$$

$$\text{�69} = \{x_{1,4,2}, x_{2,2,4}, x_{3,5,1}, x_{4,1,1}, x_{5,3,1}\}, \quad \text{ㄱ0} = \{x_{1,4,2}, x_{2,2,4}, x_{3,5,2}, x_{4,1,1}, x_{5,3,2}\},$$

$$\text{ᄀ1} = \{x_{1,4,3}, x_{2,2,5}, x_{3,5,1}, x_{4,1,1}, x_{5,3,1}\}, \quad \text{ᄀ2} = \{x_{1,4,3}, x_{2,2,5}, x_{3,5,2}, x_{4,1,1}, x_{5,3,2}\},$$

Let z_1 consist of a marked point in all regions of the Heegaard diagram except for D_i with

$$i = 4, 16, 17, 21, 22, 23, 25, 26, 27, 36, 37, 38, 58, 59, 60, 61, 62, 63, 64, 68.$$

A neighborhood of these latter domains is illustrated in Figure 9, where the aforementioned domains are colored green.

The differential corresponding to the Heegaard diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1)$ is illustrated in Figure 10. In fact, most of the positive Whitney disks of Maslov index 1 for $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1)$, which connect two of the aforementioned 72 generators, have polygonal domains, and their contribution to the differential is thus equal to 1. There are precisely 12 disks ϕ_i for $i = 1, \ldots, 7$, and ϕ_j' for $j = 1, 2, 5, 6, 7$, with nonpolygonal domains, where we have

$$\phi_1 \in \pi_2(\text{㉑}, \text{㊺}), \quad \phi_1' \in \pi_2(\text{㉒}, \text{㊻}), \quad \phi_2 \in \pi_2(\text{⑬}, \text{㊲}), \quad \phi_2' \in \pi_2(\text{⑭}, \text{㊳}),$$

$$\phi_3 \in \pi_2(\text{③}, \text{①}), \quad \phi_4 \in \pi_2(\text{ᄀ1}, \text{㉔}), \quad \phi_5 \in \pi_2(\text{㊼}, \text{㊸}), \quad \phi_5' \in \pi_2(\text{㊽}, \text{㊹}),$$

$$\phi_6 \in \pi_2(\text{㊴}, \text{㉝}), \quad \phi_6' \in \pi_2(\text{㊵}, \text{㉞}), \quad \phi_7 \in \pi_2(\text{㉟}, \text{㉛}), \quad \phi_7' \in \pi_2(\text{㊱}, \text{㉜}).$$

The domains associated with these disks are

$$D(\phi_1) = D(\phi_1') = D_4 + D_{16} + D_{58} + D_{59} + D_{60} + D_{61}, \quad D(\phi_2) = D(\phi_2') = D(\phi_1) + D_{62} + D_{63} + D_{64},$$
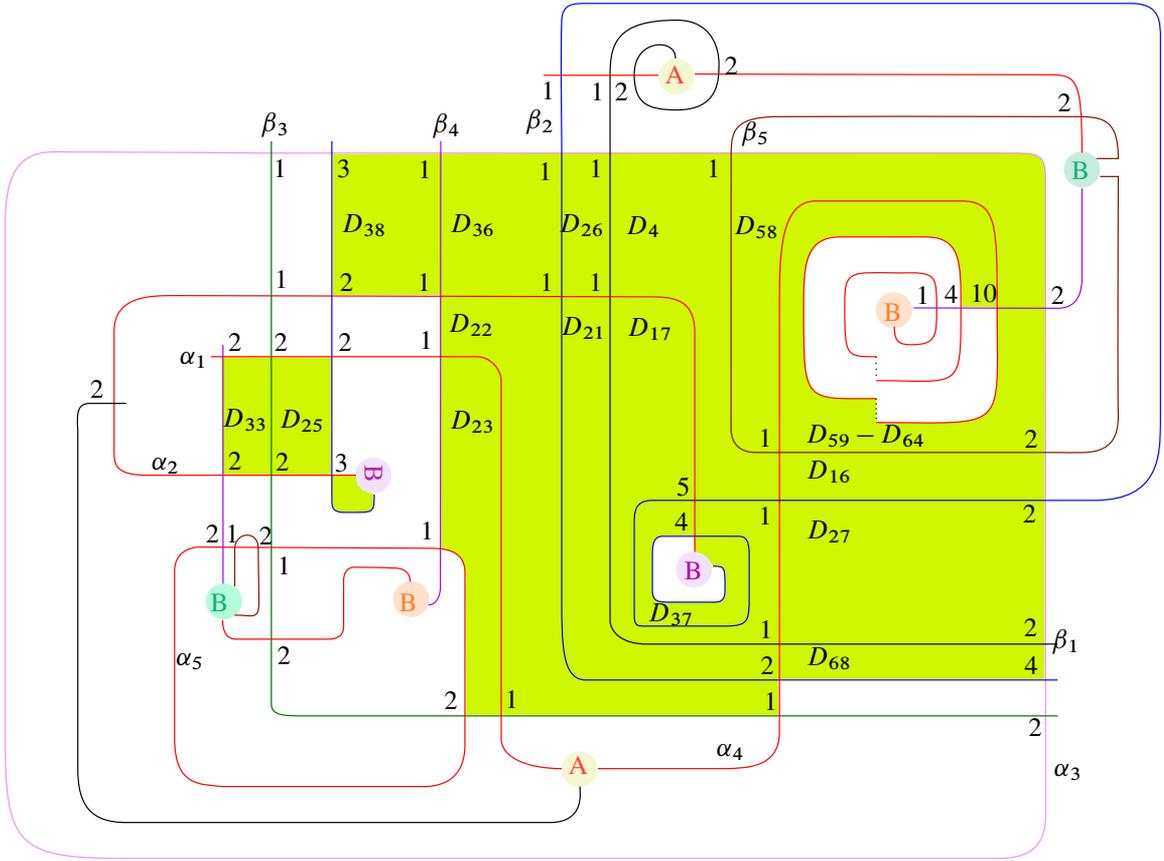
Figure 9: Part of the Heegaard diagram, where the marked points z_1 are in the domains other than those in green.

$$D(\phi_3) = D_4 + D_{17} + D_{21} + D_{22} + D_{23} + D_{26} + D_{36}, \qquad D(\phi_4) = D_4 + D_{17},$$

$$D(\phi_5) = D(\phi_5') = D_4 + D_{16} + D_{17} + D_{27} + D_{58} + D_{59}, \quad D(\phi_6) = D(\phi_6') = D(\phi_5) + D_{60} + D_{61} + D_{62},$$

$$D(\phi_7) = D(\phi_7') = D(\phi_5) + D_{60}.$$

By Lemma 4.1, $\#\widehat{M}(\phi_i) = \#\widehat{M}(\phi_i') = 1$, for $i = 1, \ldots, 4$ and $j = 1, 2$. Moreover, by Lemma 4.2, $\#\widehat{M}(\phi_i) = \#\widehat{M}(\phi_i') = 1$ for $i = 5, 6, 7$. Thus, the differential is as illustrated in Figure 10, and $H_*(\widehat{CF}(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1))$ is generated by

$$C = \{C_1 = \textcircled{49}, C_2 = \textcircled{50}, C_3 = \textcircled{53}, C_4 = \textcircled{69}\}.$$

To compute $\widehat{HF}(M, \mathfrak{s}_2)$, we need to determine the matrices l, n, p and k in Lemma 3.1. All possible positive disks with Maslov index 1 between the generators in C are

$$\psi_1', \psi_1 \in \pi_2(\textcircled{49}, \textcircled{50}), \quad \psi_2', \psi_2 \in \pi_2(\textcircled{53}, \textcircled{69}),$$

$$\psi_3 \in \pi_2(\textcircled{49}, \textcircled{53}), \qquad \psi_4 \in \pi_2(\textcircled{50}, \textcircled{69}).$$
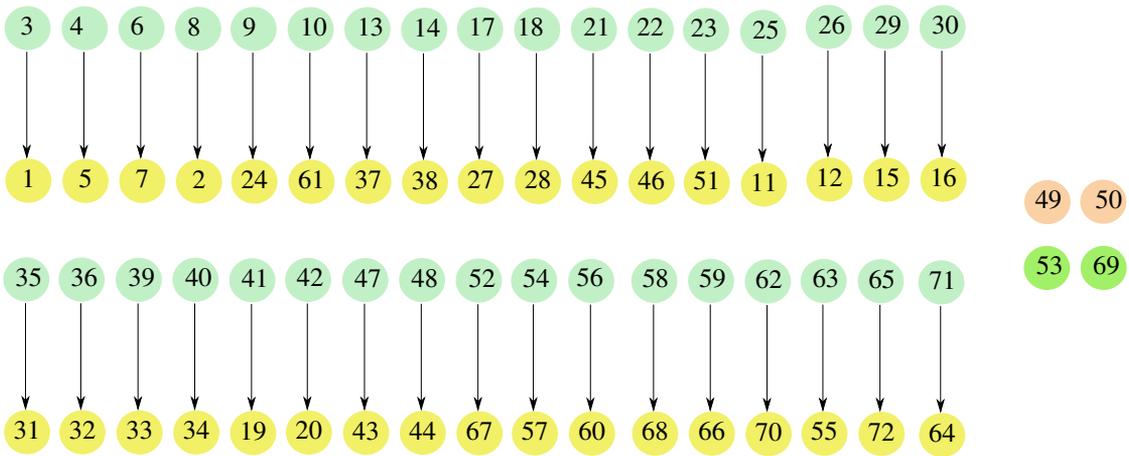
Figure 10: The differential corresponding to the diagram $(\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z_1)$.

For $i = 1, 2$, the domain associated with ψ'_i is a polygon. The domain associated with ψ_1 is shown in Figure 11 as the union of yellow, blue and pink regions. By Lemma 4.3, $\#\widehat{M}(\psi_1) = 1$. The domain associated with ψ_3 is shown in Figure 11 as the union of blue, brown and pink regions. Setting $V = \#\widehat{M}(\psi_3)$, we find

$$l = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 + e^{P_1} & 0 & 0 & 0 \\ V & 0 & 0 & 0 \\ 0 & * & * & 0 \end{pmatrix}.$$

Let A and B denote the chain complexes generated by all the generators colored in light green and yellow in Figure 10, respectively. Denote the generators of A and B by A_i and B_i, respectively. We may choose the labeling of the aforementioned generators of A and B such that k is a lower triangular matrix. Therefore,

$$p(I + k)^{-1}n = pn + pkn + pk^2n + \cdots.$$

For $j \geq 0$, since the coefficients are in $\mathbb{Z}_2$, each nonzero entry in pk^jn is of the form

$$(pk^jn)_{wv} = \sum_{r_1,\ldots,r_{j+1}} p_{wr_1}k_{r_1r_2} \cdots k_{r_jr_{j+1}}n_{r_{j+1}v},$$

and implies the existence of positive disks λ_t of Maslov index 1 for $t = 1, \ldots, j + 1$, where

$$\lambda_1 \in \pi_2(C_v, B_{r_{j+1}}), \qquad \lambda_{j+2} \in \pi_2(A_{r_1}, C_w), \qquad \lambda_t \in \pi_2(A_{r_t}, B_{r_{t-1}}) \quad t = 2, \ldots, j + 1.$$

and $\#\widehat{M}(\lambda_t) = 1$. In particular, $D(\lambda_t) > 0$ for all t and

$$(10) \qquad D(\lambda_t) \subset \sum_{t=1}^{j+2} D(\lambda_t) = \sum_{t=1}^{j+1} D(\lambda'_t) + D(\lambda) \pm D(P_1), \quad \mu(\lambda_t) = 1, \quad D(\lambda_t) > 0,$$

for some positive Whitney disks $\lambda'_t \in \pi_2(A_{r_t}, B_{r_t})$ and $\lambda \in \pi_2(C_v, C_w)$ of Maslov index 1. Potentially, there are only two such sequences satisfying (10), which are shown in Figure 12. Here $\psi_5 \in \pi_2(\textcircled{49}, \textcircled{51})$
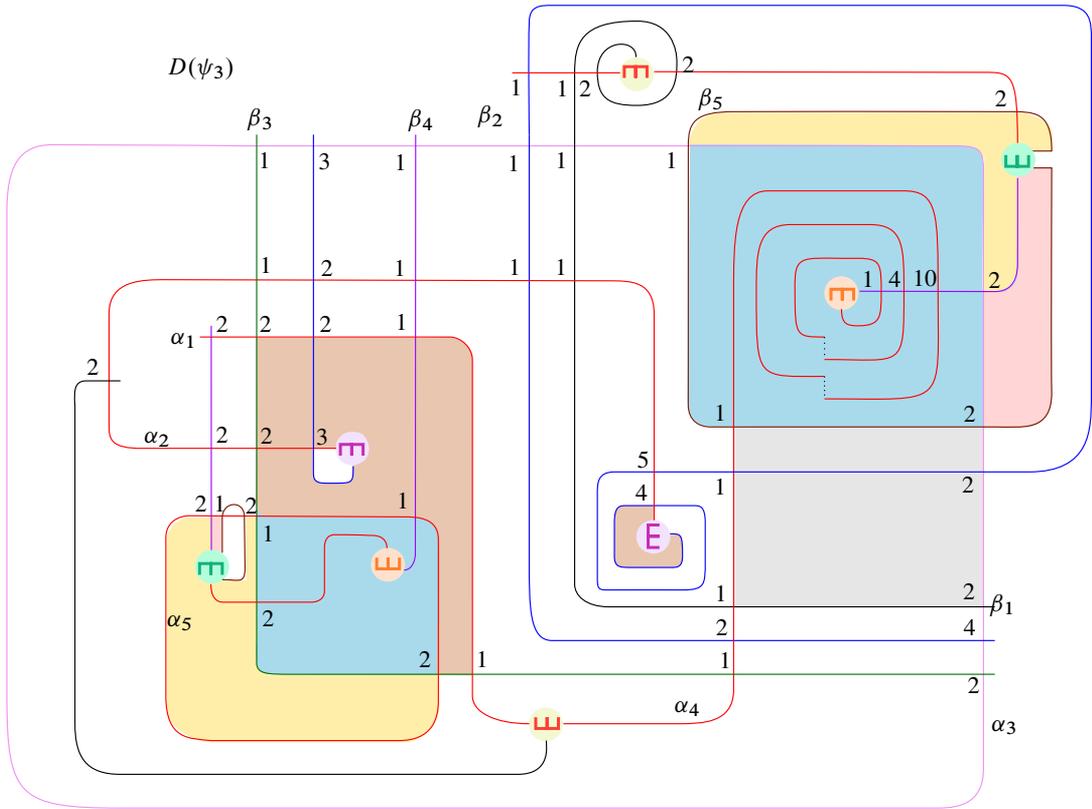
Figure 11: If I_1, I_2, I_3, I_4 and I_5 denote the set of indices of domains colored yellow, blue, pink, gray and brown respectively, the domains associated with ψ_1, ψ_3 and ψ_6 are given by $D(\psi_1) = \sum_{i \in I_1 \cup I_2 \cup I_3} D_i$, $D(\psi_3) = \sum_{i \in I_2 \cup I_3 \cup I_5} D_i$ and $D(\psi_6) = \sum_{i \in I_2 \cup I_4 \cup I_5} D_i$.

is a disk with a polygonal domain. The domain associated with the disk $\psi_6 \in \pi_2(\textcircled{23}, \textcircled{53})$ is shown in Figure 11 as the union of gray, brown and blue regions.

Lemma 6.1 *With the above notation in place, we have $\#\widehat{M}(\psi_3) = \#\widehat{M}(\psi_6)$.*

Proof The domains associated with ψ_3 and ψ_6 are extended in two ways in Figure 13. The domains $D_i^\bullet$ for $\bullet = a, b, c, d$ denote the components of the regions colored pink, gray, yellow and green,
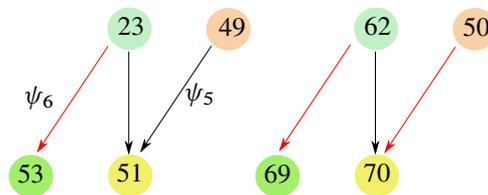
Figure 12: Potential sequences corresponding to nonzero summands $p_{wr_1} k_{r_1 r_2} \cdots k_{r_j r_{j+1}} n_{r_{j+1} v}$ in $(pk^j n)_{wv}$.
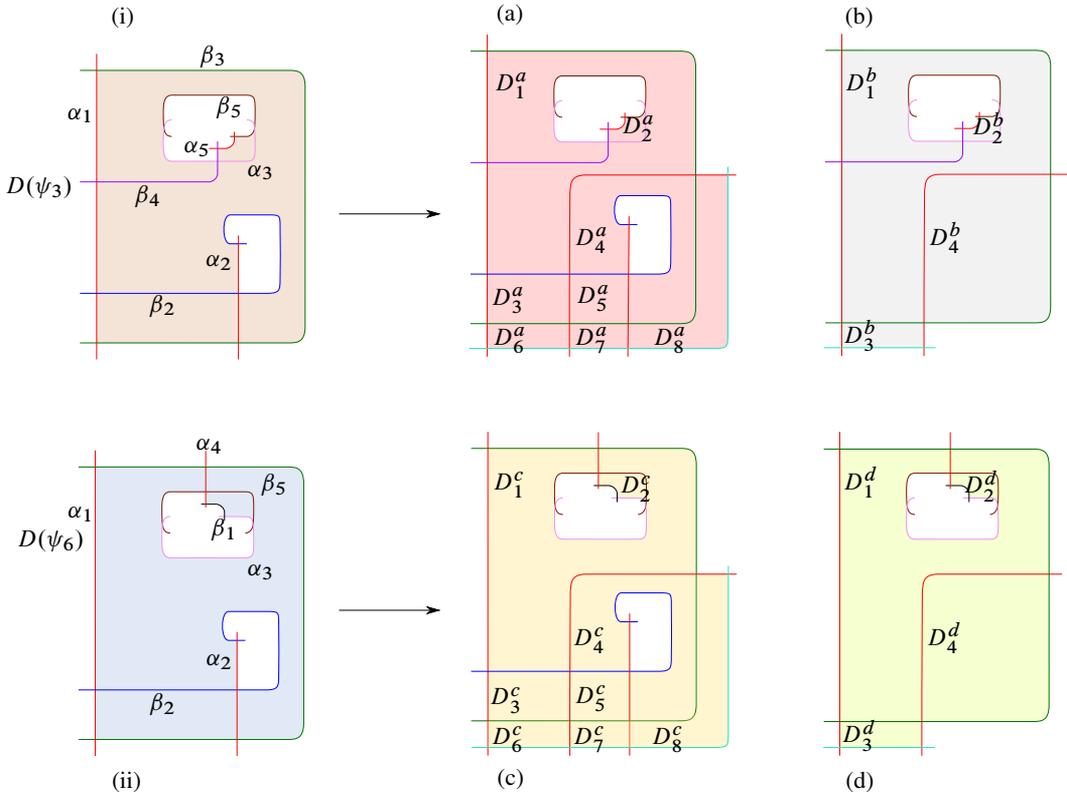
Figure 13: Domains associated with disks η_a, η_b, η_c and η_d, in (a), (b), (c) and (d), respectively.

respectively. They determine the domains of Whitney disks η_a, η_b, η_c and η_d, respectively, with domains $D(\eta_\bullet) = \sum_i D_i^\bullet$. Note that for $\bullet = a, c$ the values of i are in $\{1, \ldots, 8\}$, while for $\bullet = b, d$ the values of i are in $\{1, \ldots, 4\}$. The possible degenerations for the disks η_a, η_b, η_c and η_d are given by

$$\eta_\bullet = \phi_1^\bullet * \sigma_1^\bullet = \sigma_2^\bullet * \phi_2^\bullet = \sigma_3^\bullet * \phi_3^\bullet, \quad \bullet = a, c,$$

$$\eta_\bullet = \phi_1^\bullet * \sigma_1^\bullet = \sigma_2^\bullet * \phi_2^\bullet = \phi_3^\bullet * \sigma_3^\bullet, \quad \bullet = b, d,$$

where the corresponding domains are given by

$$D(\phi_1^\bullet) = \sum_{j=1}^{5} D_j^\bullet, \qquad D(\phi_2^\bullet) = D_1^\bullet + \sum_{j=3}^{8} D_j^\bullet, \qquad D(\phi_3^\bullet) = D_1^\bullet + D_2^\bullet + D_3^\bullet + D_6^\bullet,$$

$$D(\phi_1^\star) = D_1^\star + D_2^\star + D_3^\star, \quad D(\phi_2^\star) = D_1^\star + D_3^\star + D_4^\star, \qquad D(\phi_3^\star) = D_1^\star + D_2^\star + D_4^\star,$$

$$D(\sigma_3^\bullet) = D_4^\bullet + D_5^\bullet + D_7^\bullet + D_8^\bullet,$$

for $\bullet = a, c$ and $\star = b, d$, while the domains associated with σ_j^a, σ_j^c, σ_i^b and σ_i^d are polygons for $j = 1, 2$ and $i = 1, 2, 3$. By Lemma 4.1, we also have $\#\hat{M}(\sigma_3^\bullet) = 1$, $\bullet = a, c$. Therefore,

$$(11) \qquad \sum_{i=1}^{3} \#\hat{M}(\phi_i^\bullet) = 0 \quad \text{for } \bullet = a, b, c, d.$$

On the other hand,

$$D(\phi_1^a) = D(\psi_3), \quad D(\phi_1^c) = D(\psi_6), \quad D(\phi_2^a) = D(\phi_2^c), \quad D(\phi_2^b) = D(\phi_2^d),$$
$$D(\phi_3^a) = D(\phi_1^b), \quad D(\phi_3^c) = D(\phi_1^d), \quad D(\phi_3^b) = D(\phi_3^d).$$

Thus by (11), we have $V = \#\hat{M}(\psi_3) = \#\hat{M}(\psi_6)$. □

Having established Lemma 6.1, we conclude that

$$p(I+k)^{-1}n = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & * & * & * \\ V & * & * & * \\ 0 & * & * & * \end{pmatrix} \implies l + p(I+k)^{-1}n = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1+e^{P_1} & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

This means that ⑨ survives in $\widehat{HF}(M, \mathfrak{s}_2)$, and the latter group is thus nontrivial.

References

[Alishahi and Eftekhary 2015]   **A S Alishahi**, **E Eftekhary**, *A refinement of sutured Floer homology*, J. Symplectic Geom. 13 (2015) 609–743   MR Zbl

[Andrews and Curtis 1965]   **J J Andrews**, **M L Curtis**, *Free groups and handlebodies*, Proc. Amer. Math. Soc. 16 (1965) 192–195   MR Zbl

[Bagherifard 2021]   **N Bagherifard**, *Three-manifolds with boundary and the Andrews–Curtis transformations*, preprint (2021)   arXiv 2109.13844

[Brown 1984]   **R Brown**, *Coproducts of crossed P–modules: applications to second homotopy groups and to the homology of groups*, Topology 23 (1984) 337–345   MR Zbl

[Burns and Macedońska 1993]   **R G Burns**, **O Macedońska**, *Balanced presentations of the trivial group*, Bull. Lond. Math. Soc. 25 (1993) 513–526   MR Zbl

[Eftekhary 2015]   **E Eftekhary**, *Floer homology and splicing knot complements*, Algebr. Geom. Topol. 15 (2015) 3155–3213   MR Zbl

[Eftekhary 2018]   **E Eftekhary**, *Bordered Floer homology and existence of incompressible tori in homology spheres*, Compos. Math. 154 (2018) 1222–1268   MR Zbl

[Hanselman et al. 2024]   **J Hanselman**, **J Rasmussen**, **L Watson**, *Bordered Floer homology for manifolds with torus boundary via immersed curves*, J. Amer. Math. Soc. 37 (2024) 391–498   MR Zbl

[Hog-Angeloni and Metzler 1993]   **C Hog-Angeloni**, **W Metzler**, *The Andrews–Curtis conjecture and its generalizations*, from "Two-dimensional homotopy and combinatorial group theory" (C Hog-Angeloni, W Metzler, A J Sieradski, editors), Lond. Math. Soc. Lect. Note Ser. 197, Cambridge Univ. Press (1993) 365–380   MR Zbl

[Juhász 2006]   **A Juhász**, *Holomorphic discs and sutured manifolds*, Algebr. Geom. Topol. 6 (2006) 1429–1457   MR Zbl

[Lekili 2013]   **Y Lekili**, *Heegaard–Floer homology of broken fibrations over the circle*, Adv. Math. 244 (2013) 268–302   MR Zbl

[Miller and Schupp 1999] **C F Miller, III**, **P E Schupp**, *Some presentations of the trivial group*, from "Groups, languages and geometry" (R H Gilman, editor), Contemp. Math. 250, Amer. Math. Soc., Providence, RI (1999) 113–115 MR Zbl

[Myasnikov et al. 2002] **A D Myasnikov**, **A G Myasnikov**, **V Shpilrain**, *On the Andrews–Curtis equivalence*, from "Combinatorial and geometric group theory" (S Cleary, R Gilman, A G Myasnikov, V Shpilrain, editors), Contemp. Math. 296, Amer. Math. Soc., Providence, RI (2002) 183–198 MR Zbl

[Ozsváth and Szabó 2004a] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and genus bounds*, Geom. Topol. 8 (2004) 311–334 MR Zbl

[Ozsváth and Szabó 2004b] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. 159 (2004) 1159–1245 MR Zbl

[Ozsváth and Szabó 2004c] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR Zbl

[Sarkar and Wang 2010] **S Sarkar**, **J Wang**, *An algorithm for computing some Heegaard Floer homologies*, Ann. of Math. 171 (2010) 1213–1236 MR Zbl

[Thurston 1986] **W P Thurston**, *A norm for the homology of* 3*–manifolds*, Mem. Amer. Math. Soc. 339, Amer. Math. Soc., Providence, RI (1986) MR Zbl

[Wright 1975] **P Wright**, *Group presentations and formal deformations*, Trans. Amer. Math. Soc. 208 (1975) 161–169 MR Zbl

School of Mathematics, Institute for Research in Fundamental Sciences (IPM)
Tehran, Iran

School of Mathematics, Institute for Research in Fundamental Sciences (IPM)
Tehran, Iran

neda.bagherifard@gmail.com, eaman@ipm.ir

Morse elements in Garside groups are strongly contracting

MATTHIEU CALVEZ

BERT WIEST

We prove that in the Cayley graph of any braid group modulo its center $B_n/Z(B_n)$, equipped with Garside's generating set, the axes of all pseudo-Anosov braids are strongly contracting. More generally, we consider a Garside group G of finite type with cyclic center. We prove that in the Cayley graph of $G/Z(G)$, equipped with the Garside generators, the axis of any Morse element is strongly contracting. As a consequence, we prove that Morse elements act loxodromically on the additional length graph of G.

20F36; 20F65

1 Introduction

For a finitely generated group G, equipped with some fixed finite generating set, and an element $g \in G$ of infinite order, one can study the axis $\mathrm{axis}(g) = \langle g \rangle$, seen as a set of vertices in the Cayley graph $\Gamma(G)$. There are many different ways of formalizing the idea that this axis might "look like a geodesic in a hyperbolic space".

A particularly weak notion is that $\langle g \rangle$ is quasi-isometrically embedded in G. A particularly strong condition is that the axis is *strongly contracting*, which is equivalent to being *strongly constricting* [Arzhantseva et al. 2015]. There are many intermediate notions — for instance the axis could be hyperbolically embedded [Dahmani et al. 2017; Osin 2016; 2018], it could be rank one [Hamenstädt 2009; Sisto 2018], it could be Morse [Dahmani et al. 2017; Sisto 2016], it could be contracting in the sense of [Abbott et al. 2021; Arzhantseva et al. 2015], it could have various other contraction and divergence properties [Arzhantseva et al. 2017], or constriction properties [Arzhantseva et al. 2015].

In this paper we will be interested in two of these properties, namely the Morse property and the strong contraction property. Precise definitions can be found in Sections 3 and 5.

An element g (or its axis in the Cayley graph $\Gamma(G)$) is said to be *Morse* if this axis is quasi-isometrically embedded in $\Gamma(G)$, and if for each pair of constants (K, L), there exists a constant $M_g^{(K,L)}$ such that every (K, L)–quasigeodesic between two points of the axis travels in an $M_g^{(K,L)}$–neighborhood of the axis. A remarkable example of the Morse property is the result of Behrstock [2006] that pseudo-Anosov elements in mapping class groups are Morse — see also [Sisto 2016; Dahmani et al. 2017]. (In fact, [Duchin and Rafi 2009, Theorem 4.2] implies that their axes satisfy the stronger condition of being contracting.)

Saying that the axis of an element g (or indeed any other subset A of the Cayley graph $\Gamma(G)$) has the *strong contraction* property means, roughly speaking, that there is a constant C such that taking any ball in $\Gamma(G)$ disjoint from A, and projecting this ball to A via a closest-point projection, yields a subset of A of diameter at most C.

One first crucial observation about these two properties is that the Morse property is invariant under quasi-isometry (eg when looking at the axis of an element in the Cayley graphs of G with respect to two different generating sets), whereas the strong contraction/constriction property is not. One reason for the failure of quasi-isometry invariance is that these strong properties make reference to actual distances and geodesics, *not* to quasigeodesics.

There is one well-known family of groups with a very natural family of geodesics between any pair of points in the Cayley graphs, namely the Garside groups. The notions of Garside theory needed in this paper will be recalled in Section 2. We will be interested specifically in Δ–*pure Garside groups of finite type*.

A Garside group of finite type G is generated by a finite lattice $\mathscr{D}$ with a top element called Δ. Garside groups are bi-automatic — in particular, every element $g \in G$ is represented by a unique word in a certain normal form, with letters in $\mathscr{D}^{\pm 1}$; these normal form words represent *geodesics* in the Cayley graph of G with respect to $\mathscr{D}$. We will also require that our Garside groups are Δ–pure, or equivalently Zappa–Szép indecomposable — this condition means in particular that the center of G is infinite cyclic and is generated by some power Δ^e.

The most famous examples of Garside groups of finite type are the braid groups, and more generally the Artin–Tits groups of spherical type [Brieskorn and Saito 1972; Charney 1992; Deligne 1972]. (In this setting, the Δ–pureness condition is equivalent to the defining Coxeter graph being connected.)

Since the infinite subgroup $\langle \Delta^e \rangle$ of G is central, there cannot be any elements whose axes in the Cayley graph $\Gamma(G, \mathscr{D})$ are Morse or strongly contracting. Instead, we will study the axes of elements in the Cayley graph of G modulo its center. We will let $\overline{\Gamma} = \Gamma(G/Z(G), \mathscr{D})$ and we will say that an element of a Garside group G is Morse if its axis in $\overline{\Gamma}$ is Morse.

Actually, for our proof it will be technically convenient not to work with the model space $\overline{\Gamma} = \Gamma(G, \mathscr{D})/\langle \Delta^e \rangle$, but with the following quasi-isometric variation. We define $\mathscr{X}$ to be the quotient of the Cayley graph $\Gamma(G, \mathscr{D})$ under the right $\langle \Delta \rangle$–action: $\mathscr{X} = \Gamma(G, \mathscr{D})/\langle \Delta \rangle$. We note that this graph $\mathscr{X}$ is the 1–skeleton of the simplicial complex previously considered in [Bestvina 1999; Charney et al. 2004]. The graph $\mathscr{X}$ was also studied in [Dehornoy et al. 2015, VIII.3.2], under the name $\mathscr{G}^0$, and in [Calvez and Wiest 2017a; 2017b].

Our main result is as follows (see Theorem 5.5 for a precise version):

Theorem 1.1 *Suppose G is a Δ–pure Garside group of finite type. Suppose g is an element of G whose axis in $\mathscr{X}$ (or, equivalently, in $\overline{\Gamma}$) is Morse. Then this axis is strongly contracting, both as a subset of $\mathscr{X}$ and as a subset of $\overline{\Gamma}$.*

Corollary 1.2 *Consider the braid group B_n, equipped with the generating set $\mathscr{D}_{\mathrm{classical}}$ or $\mathscr{D}_{\mathrm{dual}}$ coming from its classical or dual Garside structure. Then in the Cayley graphs $\Gamma(B_n/Z(B_n), \mathscr{D}_{\mathrm{classical}})$ and $\Gamma(B_n/Z(B_n), \mathscr{D}_{\mathrm{dual}})$, the axis of any pseudo-Anosov braid is strongly contracting.*

Remark 1.3 Here is some context for these results:

(i) Rafi and Verberne [2021] have constructed a pseudo-Anosov element of the mapping class group of the five-punctured sphere (which contains $B_4/\langle \Delta^2 \rangle$ as a subgroup of index 5), and a generating set for this mapping class group, such that the axis of this pseudo-Anosov element in the corresponding Cayley graph is not strongly contracting. Our Corollary 1.2 supports the idea that examples such as that of [Rafi and Verberne 2021] can only exist under a pathological choice of generating set.

(ii) Let S be a surface of finite type, and $\mathscr{T}(S)$ its Teichmüller space, equipped with the Teichmüller metric and the $\mathrm{Mod}(S)$–action. Minsky [1996] proved that the axis of every pseudo-Anosov element of $\mathrm{Mod}(S)$ has the strong contraction property in $\mathscr{T}(S)$.

(iii) There is a hierarchy of contraction properties: strongly contracting implies contracting which in turn implies sublinearly contracting, which is equivalent to being Morse [Arzhantseva et al. 2017]; neither of the implications is an equivalence [Brady and Tran 2022; Rafi and Verberne 2021]. However, under a CAT(0)–hypothesis there *is* a strong converse: Morse geodesics in CAT(0)–spaces are strongly contracting [Cashen 2020; Sultan 2014] (as are axes of hyperbolic isometries not bounding a half-flat [Bestvina and Fujiwara 2009, Theorem 5.4]). Thus Theorem 1.1 is an indicator that Garside groups have CAT(0)–like behavior; this gives further evidence that the answer to the following question may be affirmative.

Question 1.4 Are all Garside groups CAT(0)? (Note that it is not even known whether all braid groups are CAT(0) [Brady and McCammond 2010; Haettel et al. 2016; Jeong 2023].)

Our Theorem 1.1 begs the question which elements of well-known Garside groups *are* Morse. For braid groups, we know the answer from [Behrstock 2006]: it's the pseudo-Anosov elements. For other irreducible Artin–Tits groups of spherical type, however, the question is open. The authors [Calvez and Wiest 2017b] hand-constructed some elements in each such group whose axes are strongly contracting, and in particular Morse. We believe, however, that all plausible candidates for being Morse really are Morse:

Conjecture 1.5 *An element a of an irreducible Artin–Tits groups of spherical type A is Morse if and only if its image in $A/Z(A)$ has virtually cyclic centralizer.*

An application of our results (indeed, the authors' original motivation for this research) concerns the *additional length graph* $\mathscr{C}_{\mathrm{AL}}(G)$ [Calvez and Wiest 2017a; 2017b; 2021]. To any Garside group G

one can associate a δ–hyperbolic graph $\mathscr{C}_{\mathrm{AL}}(G)$ whose relation to G is loosely analogous to the curve graph's relation to the mapping class group. Indeed, if B_n is the n–strand braid group, then $\mathscr{C}_{\mathrm{AL}}(B_n)$ is conjectured to be quasi-isometric to the curve graph of the $(n+1)$–punctured sphere.

Using Theorem 1.1 we can prove (see Theorem 6.5 and Corollary 6.7)

Proposition 1.6 *Suppose G is a Δ–pure Garside group of finite type. If g is a Morse element of G, then the action of g on $\mathscr{C}_{\mathrm{AL}}(G)$ is loxodromic and weakly properly discontinuous. In particular, if G contains a Morse element then $\mathscr{C}_{\mathrm{AL}}(G)$ has infinite diameter.*

The plan of this paper is as follows. In Section 2 we review some, mostly standard, elements from Garside theory which we will need. In Section 3 we recall the definition of the Morse property and prove a first contraction property for the axes of Morse elements in Garside groups. In Section 4 we define, in a Garside-theoretical fashion, a projection to the axis of any element satisfying a Garside-theoretical rigidity condition. For elements satisfying both the rigidity and the Morse condition, we strengthen our previous contraction result, and deduce that our projection is uniformly close to any closest-point projection. Section 5 contains a precise definition of the strong contraction property and the proof of Theorem 1.1. In Section 6 we present the applications of the main result to the additional length graph.

2 Garside groups

The notion of a Garside group stems from Garside's approach to solving the conjugacy problem in the braid groups [Garside 1969]. Soon generalized to Artin–Tits groups of spherical type [Brieskorn and Saito 1972; Deligne 1972], this approach was first axiomatized in [Dehornoy 2002; Dehornoy and Paris 1999] and thoroughly studied over the first decade of the 2000s. The book [Dehornoy et al. 2015] provides a comprehensive account of what is now called "Garside theory".

Definition 2.1 Let G be a group; G is a *Garside group* with *Garside structure* (G^+, Δ) if G^+ is a submonoid of G such that $G^+ \cap G^{+^{-1}} = \{1\}$ and there exists an element $\Delta \in G^+$ with the following properties:

(1) The partial order relations $\preccurlyeq$ and $\succcurlyeq$ on G defined by
  - $x \preccurlyeq y$ (x is a *prefix* of y) if and only if $x^{-1} y \in G^+$,
  - $x \succcurlyeq y$ (y is a *suffix* of x) if and only if $xy^{-1} \in G^+$

  are lattice orders on G; that is, all $x, y \in G$ admit a unique greatest common prefix $x \wedge y$, a unique greatest common suffix $x \wedge^{\!\shortmid} y$, a unique least common right multiple $x \vee y$ and a unique least common left multiple $x \vee^{\!\shortmid} y$.

(2) The set $\mathscr{D} = \{x \in G^+ \mid x \preccurlyeq \Delta\} = \{x \in G^+ \mid \Delta \succcurlyeq x\}$ generates G^+ as a monoid and G as a group.

(3) For all $x \in G^+ \setminus \{1\}$,
$$\|x\| = \sup\{k \mid \exists\, a_1, \ldots, a_k \in G^+ \setminus \{1\} \text{ such that } x = a_1 \cdots a_k\} < \infty.$$

The elements of G^+ are called *positive*, Δ is the *Garside element* and the elements of $\mathscr{D}$ are called *simple*. The elements x of G^+ such that $\|x\| = 1$ are called *atoms* and they form a subset of $\mathscr{D}$.

Given a simple element s, its *right-complement* $\partial(s)$ is defined by $\partial(s) = s^{-1}\Delta$ and its *left-complement* is defined by $\partial^{-1}(s) = \Delta s^{-1}$; both $\partial(s)$ and $\partial^{-1}(s)$ belong to $\mathscr{D}$. Conjugation by Δ will be denoted by τ (that is, for $g \in G$, $\tau(g) = \Delta^{-1}g\Delta$); notice that for every simple element s, $\partial^2(s) = \partial(\partial(s)) = \tau(s)$.

We shall make the additional assumption that G is *of finite type*, ie that the set of simple elements $\mathscr{D}$ is finite. In this case, it follows that τ has finite order, and we will denote this order by e. The element Δ^e is then central in G. We shall further assume that G is Δ–*pure*. This property was defined in [Picantin 2001], and shown in [Gebhardt and Tawn 2016, Theorem 39] to be equivalent to indecomposability as a Zappa–Szép product. All the reader needs to know about Δ–pure Garside groups of finite type is that their center is cyclic, generated by Δ^e, ie $Z(G) = \langle \Delta^e \rangle$ [Picantin 2001]. For instance, Artin–Tits groups of spherical type are Garside groups, and they are Δ–pure if and only if the defining Coxeter graph is connected [Picantin 2001, Proposition 4.7].

Notation 2.2 Throughout this paper, G denotes a Δ–pure Garside group of finite type. We denote by e the positive integer such that $Z(G) = \langle \Delta^e \rangle$. When we talk about Cayley graphs, it is always understood that the generating set is obtained from the set of simple elements $\mathscr{D}$. We will use the notation Γ for the Cayley graph $\Gamma(G, \mathscr{D})$. Also, we define $\overline{\Gamma} = \Gamma(G/Z(G))$, the Cayley graph of G modulo its center, with respect to the generators which are the images of $\mathscr{D}$ in $G/Z(G)$. The corresponding graph metrics will be denoted by d_Γ and $d_{\overline{\Gamma}}$ respectively.

To each element of G we associate three integer numbers as follows.

Definition 2.3 [El-Rifai and Morton 1994, Section 1] Let $g \in G$. The *infimum* of g is defined by $\inf(g) = \max\{r \in \mathbb{Z} \mid \Delta^r \preccurlyeq g\}$, the *supremum* of g is defined by $\sup(g) = \min\{s \in \mathbb{Z} \mid g \preccurlyeq \Delta^s\}$ and the *canonical length* of g is defined by $\ell(g) = \sup(g) - \inf(g)$.

It is well known that each element of G can be written uniquely as an irreducible fraction involving elements of G^+ — the letters D and N in the following stand for "denominator" and "numerator" respectively.

Lemma 2.4 [Charney 1999, Lemma 4.4] *Let $g \in G$.*

(i) **Left-fraction** *There is a unique pair of positive elements $(D_l(g), N_l(g))$ such that*
$$D_l(g) \wedge N_l(g) = 1 \quad and \quad g = D_l(g)^{-1} N_l(g).$$

In particular, if c is any positive element such that cg is positive, then $c \succcurlyeq D_l(g)$.

(ii) **Right-fraction** *There is a unique pair of positive elements $(D_r(g), N_r(g))$ such that*
$$D_r(g) \wedge^{\shortuparrow} N_r(g) = 1 \quad and \quad g = N_r(g)D_r(g)^{-1}.$$

In particular, if c is any positive element such that gc is positive, then $D_r(g) \preccurlyeq c$.

(iii)   *We have the equalities*

$$\inf(D_l(g)) = \inf(D_r(g)), \quad \sup(D_l(g)) = \sup(D_r(g)),$$
$$\inf(N_l(g)) = \inf(N_r(g)), \quad \sup(N_l(g)) = \sup(N_r(g)).$$

Moreover, every element of G can be decomposed as follows. Recall that a pair of simple elements $(s, t) \in \mathcal{D}^2$ is *left-weighted* if $\partial(s) \wedge t = 1$ and *right-weighted* if $\partial^{-1}(t) \wedge^{\urcorner} s = 1$.

Proposition 2.5 [Adyan 1984; Dehornoy 2002, Section 3]   *Let $g \in G$. Let $p = \inf(g)$ and $r = \ell(g)$.*

(i)   *There exists a unique decomposition $g = \Delta^p s_1 \cdots s_r$, where $s_1, \ldots, s_r \in \mathcal{D} \setminus \{1, \Delta\}$ and for every $1 \leqslant i \leqslant r - 1$, (s_i, s_{i+1}) is left-weighted.*

(ii)   *Similarly, there exists a unique decomposition $g = s'_r \cdots s'_1 \Delta^p$, where $s'_1, \ldots, s'_r \in \mathcal{D} \setminus \{1, \Delta\}$ and for every $1 \leqslant i \leqslant r - 1$, (s'_{i+1}, s'_i) is right-weighted.*

*These decompositions are called **left normal form** and **right normal form** of g, respectively.*

Considering the latter normal forms together with the fractional decompositions of Lemma 2.4 yields a slightly different notion of normal form.

Definition 2.6 [Dehornoy 2002, Proposition 3.9]   Let $g \in G$. If $\inf(g) < 0 < \sup(g)$ and $D_l(g) = a_1 \cdots a_r$ and $N_l(g) = b_1 \cdots b_s$ are left normal forms, then the *left mixed normal form* of g is the representation $g = a_r^{-1} \cdots a_1^{-1} b_1 \cdots b_s$; similarly, the *right mixed normal form* is the representation $g = b'_s \cdots b'_1 a_1'^{-1} \cdots a_r'^{-1}$, where $a'_r \cdots a'_1$ and $b'_s \cdots b'_1$ are the respective right normal forms of $D_r(g)$ and $N_r(g)$. If $\inf(g) \geqslant 0$, then the left (resp. right) mixed normal form of g coincides with the left (resp. right) normal form of g given by Proposition 2.5. If $\sup(g) \leqslant 0$, then the left (resp. right) mixed normal form of g is the formal inverse of the left (resp. right) normal form of g^{-1} given by Proposition 2.5.

These mixed normal forms have an important geometric meaning:

Lemma 2.7 [Charney and Meier 2004, Lemma 3.1]   *In Γ, the Cayley graph of G with respect to $\mathcal{D}$, mixed normal forms are geodesics.*

Finally, we shall need one more Garside-theoretical definition:

Definition 2.8   Let $x \in G$ with right normal form $x = x_r \cdots x_1 \Delta^p$. We say that x is *right-rigid* if its preferred simple suffix $\mathfrak{p}^{\urcorner}(x) := \tau^p(x_1) \wedge^{\urcorner} \partial^{-1}(x_r)$ is trivial. In particular, if $\inf(x) = 0$, then for $k \geqslant 1$, the right normal form of x^k consists of the concatenation of k copies of the right normal form of x.

Our aim is, of course, to study the geometry of G modulo its center $Z(G) = \langle \Delta^e \rangle$. However, it is technically far more convenient not to study the Cayley graph $\overline{\Gamma} = \Gamma(G/Z(G))$ directly, but a very closely related graph, which we will denote by $\mathscr{X}$. This graph $\mathscr{X}$ is the 1–skeleton of "Bestvina's normal

form complex" considered in [Bestvina 1999; Charney et al. 2004], and it has been described previously in [Calvez and Wiest 2017a; Dehornoy et al. 2015, Chapter VIII, Section 3.2]. We recall the definition:

Notation 2.9 We denote by $\mathcal{X}$ the quotient of $\Gamma = \Gamma(G)$ by the right-action of $\langle \Delta \rangle$,

$$\mathcal{X} = \Gamma(G, \mathcal{D})/\langle \Delta \rangle.$$

- The vertices of $\mathcal{X}$ are the left-cosets of G modulo $\langle \Delta \rangle$, $\{g\langle \Delta \rangle \mid g \in G\}$. Each vertex $g\langle \Delta \rangle$ of $\mathcal{X}$ possesses a unique *distinguished representative $\underline{g}$*, which is by definition the representative satisfying $\inf(\underline{g}) = 0$: given $g \in G$, we have $\underline{g} = g\Delta^{-\inf(g)}$. We denote by $*$ the vertex $\langle \Delta \rangle$, whose distinguished representative is the trivial element of G.

- Two vertices $g\langle \Delta \rangle$ and $h\langle \Delta \rangle$ of $\mathcal{X}$ are connected by an edge if there is a proper simple element $s \in \mathcal{D}$ such that $\underline{g}s \in h\langle \Delta \rangle$; this is equivalent to the existence of a proper simple element t such that $\underline{h}t \in g\langle \Delta \rangle$.

The following provides more precise information about adjacent vertices of $\mathcal{X}$:

Lemma 2.10 [Bestvina 1999, Lemma 3.4] *Suppose that $g\langle \Delta \rangle$ and $h\langle \Delta \rangle$ are adjacent vertices of $\mathcal{X}$. Then there exists a proper simple element $s \in \mathcal{D} \setminus \{1, \Delta\}$ such that one of the following holds:*

- $\underline{g}s = \underline{h}$ *(and in this case we have $\underline{h}\,\partial s = \underline{g}\,\Delta$), or*
- $\underline{h}s = \underline{g}$ *(and in this case we have $\underline{g}\,\partial s = \underline{h}\,\Delta$).*

Notation 2.11 We denote by $d_{\mathcal{X}}$ the graph metric on the graph $\mathcal{X}$; for $g, h \in G$, we sometimes write $d_{\mathcal{X}}(g, h)$ for $d_{\mathcal{X}}(g\langle \Delta \rangle, h\langle \Delta \rangle)$. Note that the groups G and $G/Z(G)$ act isometrically by left-translations on $\mathcal{X}$ by $g \cdot (g'\langle \Delta \rangle) = (gg')\langle \Delta \rangle$.

The spaces $\mathcal{X}$ and $\overline{\Gamma}$ are very closely related:

Proposition 2.12 *There is an **isometric** embedding $\iota : \mathcal{X} \hookrightarrow \overline{\Gamma}$ with $\lfloor \frac{1}{2}e \rfloor$–dense image (ie every vertex of $\overline{\Gamma}$ is at distance at most $\lfloor \frac{1}{2}e \rfloor$ from a vertex belonging to $\iota(\mathcal{X})$). In particular, ι is a quasi-isometry.*

Proof If $g, h \in G$ represent adjacent vertices of $\mathcal{X}$, then we know from Lemma 2.10 that $\underline{g}$ and $\underline{h}$ represent adjacent vertices in the Cayley graph Γ of G, and thus also in the Cayley graph $\overline{\Gamma}$ of $G/\langle \Delta^e \rangle$. This means that the map

$$\iota : \{\text{vertices of } \mathcal{X}\} \hookrightarrow \{\text{vertices of } \overline{\Gamma}\}, \quad g\langle \Delta \rangle \mapsto \underline{g}\langle \Delta^e \rangle$$

sends adjacent vertices to adjacent vertices, and thus induces a well-defined and 1–Lipschitz map of graphs $\iota : \mathcal{X} \hookrightarrow \overline{\Gamma}$.

In the other direction, there is a natural projection map

$$p : \{\text{vertices of } \overline{\Gamma}\} \to \{\text{vertices of } \mathcal{X}\}, \quad g\langle \Delta^e \rangle \mapsto g\langle \Delta \rangle,$$

which induces a well-defined map of graphs $p: \overline{\Gamma} \to \mathcal{X}$. Both ι and p are 1–Lipschitz, and $p \circ \iota = \mathrm{id}_{\mathcal{X}}$. This implies that ι is an isometric embedding.

Now we look at the opposite composition

$$\iota \circ p: \overline{\Gamma} \to \overline{\Gamma}, \quad g\langle \Delta^e \rangle \mapsto \underline{g}\langle \Delta^e \rangle.$$

We observe that $d_{\overline{\Gamma}}(g\langle \Delta^e \rangle, \underline{g}\langle \Delta^e \rangle) < \lfloor \tfrac{1}{2}e \rfloor$, which means that $\iota \circ p$ is at distance $\lfloor \tfrac{1}{2}e \rfloor$ from $\mathrm{id}_{\overline{\Gamma}}$; thus the image of ι is $\lfloor \tfrac{1}{2}e \rfloor$–dense. $\qquad \square$

We now recall from [Calvez and Wiest 2017a] the notion of a preferred path between two vertices in $\mathcal{X}$.

Definition 2.13 Given any vertex $g\langle \Delta \rangle$ of $\mathcal{X}$, let $s_1 \cdots s_r$ be the left normal form of $\underline{g}$; the *preferred path* between $*$ and $g\langle \Delta \rangle$ is the path

$$*, s_1\langle \Delta \rangle, \ldots, (s_1 \cdots s_r)\langle \Delta \rangle = g\langle \Delta \rangle.$$

Then, for $g, h \in G$, we denote by $A(g, h)$ the g–left translate of the preferred path between $*$ and $g^{-1}h\langle \Delta \rangle$ — so $A(g, h)$ is a path in $\mathcal{X}$ with starting point $g\langle \Delta \rangle$ and end point $h\langle \Delta \rangle$.

Remark 2.14 If $g' \in g\langle \Delta \rangle$ and $h' \in h\langle \Delta \rangle$, then the paths $A(g, h)$ and $A(g', h')$ coincide, so we have a preferred path $A(g\langle \Delta \rangle, h\langle \Delta \rangle)$ between any pair of points $g\langle \Delta \rangle, h\langle \Delta \rangle$ in $\mathcal{X}$.

Here is a proof of this coincidence of the two paths: Suppose the left normal form of $\underline{g^{-1}h}$ is $s_1 \cdots s_n$. Then by definition, the ith vertex on the path $A(g, h)$ in $\mathcal{X}$ is represented by $g \cdot s_1 \cdots s_i$. Let us calculate, for comparison, the ith vertex on the path $A(g\Delta^k, h\Delta^\ell)$. We have

$$\underline{(g\Delta^k)^{-1}h\Delta^\ell} = \underline{\Delta^{-k} \cdot g^{-1}h} = \underline{\tau^k(g^{-1}h)} = \tau^k(s_1) \cdots \tau^k(s_n).$$

The ith vertex of the $g\Delta^k$–translate of $A(*, \underline{(g\Delta^k)^{-1}h\Delta^\ell})$ is thus represented by

$$g\Delta^k \cdot \tau^k(s_1) \cdots \tau^k(s_i) = g \cdot s_1 \cdots s_i \cdot \Delta^k,$$

which represents the same point of $\mathcal{X}$.

We record some basic properties of the preferred paths (recall that for $g, h \in G$, $g \wedge h$ is the unique greatest common prefix of g and h):

Proposition 2.15 (i) *Let $g, h \in G$; let $p = \underline{g} \wedge \underline{h}$. The preferred path $A(g, h)$ is the concatenation of the preferred paths $A(g, p)$ and $A(p, h)$.*

(ii) *Preferred paths are symmetric: for all $g, h \in G$, $A(g, h)$ is the reverse of $A(h, g)$.*

(iii) *Preferred paths are preserved by left-translation: for all $g, h, k \in G$, $A(kg, kh) = kA(g, h)$.*

(iv) *Preferred paths are geodesics in $\mathcal{X}$ and for all $g, h \in G$, $d_{\mathcal{X}}(g, h) = d_{\Gamma}(\underline{g}, \underline{h})$.*

(v) *Balls in $\mathcal{X}$ are convex: if $g, h \in G$ and $k\langle \Delta \rangle \in A(g, h)$, then*

$$d_{\mathcal{X}}(*, k\langle \Delta \rangle) \leqslant \max(d_{\mathcal{X}}(*, g\langle \Delta \rangle), d_{\mathcal{X}}(*, h\langle \Delta \rangle)).$$

Proof Parts (i) and (ii) correspond to [Calvez and Wiest 2017a, Lemma 4] and [Calvez and Wiest 2017a, Lemma 5], respectively. For (iii), write the left normal form of $\underline{g^{-1}h}$ as $z_1 \cdots z_r$, so that $A(g, h)$ is by definition the path

$$g\langle\Delta\rangle = \underline{g}\langle\Delta\rangle, \underline{g}z_1\langle\Delta\rangle, \ldots, \underline{g}z_1 \cdots z_r\langle\Delta\rangle = h\langle\Delta\rangle.$$

Note that $\underline{kg} = \underline{k}\underline{g}\Delta^j$ (with $j = \inf(g) - \inf(kg)$); therefore the left normal form of $\underline{(kg)^{-1}kh}$ is $\tau^j(z_1) \cdots \tau^j(z_r)$. So by definition, $A(kg, kh)$ is the path

$$kg\langle\Delta\rangle = \underline{kg}\langle\Delta\rangle, \underline{kg}\tau^j(z_1)\langle\Delta\rangle, \ldots, \underline{kg}\tau^j(z_1 \cdots z_r)\langle\Delta\rangle,$$

but for all $1 \leqslant i \leqslant r$, we have

$$\underline{kg}\tau^j(z_1 \cdots z_i) = \underline{k}\underline{g}\Delta^j\Delta^{-j}z_1 \cdots z_i\Delta^j = \underline{k}\underline{g}z_1 \cdots z_i\Delta^j.$$

Thus $A(kg, kh)$ is the path

$$kg\langle\Delta\rangle, k\underline{g}z_1\langle\Delta\rangle, \ldots, k\underline{g}z_1 \cdots z_r\langle\Delta\rangle,$$

the k–left translate of $A(g, h)$, as claimed.

To see (iv), recall first that mixed normal forms are geodesics in Γ (Lemma 2.7). Let $g, h \in G$. It is shown in the proof of [Calvez and Wiest 2017a, Lemma 4] that the path $A(g, h)$ in $\mathcal{X}$ has the exact length of the mixed normal form of $g^{-1}h$, say r. Suppose that there was a shorter path in $\mathcal{X}$ between $g\langle\Delta\rangle$ and $h\langle\Delta\rangle$, that is a sequence of vertices $g\langle\Delta\rangle = g_0\langle\Delta\rangle, g_1\langle\Delta\rangle, \ldots, g_k\langle\Delta\rangle = h\langle\Delta\rangle$, with $k < r$. Then by Lemma 2.10 there are simple elements $s_1, \ldots, s_k$ such that for $1 \leqslant i \leqslant k$, we have $\underline{g_i} = \underline{g_{i-1}}s_i$ or $\underline{g_i} = \underline{g_{i-1}}s_i^{-1}$; setting t_i to be s_i or s_i^{-1} accordingly, we obtain that $\underline{h} = \underline{g_k} = \underline{g_0}t_1 \cdots t_k = \underline{g}t_1 \cdots t_k$, where each t_i is either a simple element or the inverse of a simple element. This contradicts the fact that mixed normal forms are geodesics in Γ.

Now, let us prove (v). In view of (iv), the distances involved are the respective lengths of the left normal forms of $\underline{k}$, $\underline{g}$ and $\underline{h}$. Let $p = \underline{g} \wedge \underline{h}$; write $\underline{g} = pa$ and $\underline{h} = pb$ (with $a, b \in G^+$), and let $a = a_1 \cdots a_r$ and $b = b_1 \cdots b_s$ be the respective left normal forms. By the proof of [Calvez and Wiest 2017a, Lemma 4], the distinguished representatives of the vertices along the path $A(g, h)$ are (in this order)

$$\underline{g} = pa_1 \cdots a_r, \ldots, pa_1, p, pb_1, \ldots, pb_1 \cdots b_s = \underline{h}.$$

Any of these is a prefix of $\underline{g}$ or $\underline{h}$; thus its left normal form is at most as long as that of $\underline{g}$ or $\underline{h}$ and the claim is proved. $\qquad\square$

Lemma 2.16 (fellow traveler property) *Suppose that $g, g', h, h' \in G$ are such that $d_{\mathcal{X}}(g, g') = 1$ and $d_{\mathcal{X}}(h, h') = 1$. Then the set of vertices along the path $A(g, h)$ and the set of vertices along the path $A(g, h')$ in $\mathcal{X}$ are at Hausdorff distance 1. Also, the analogous statement holds for the paths $A(g, h)$ and $A(g', h)$.*

Proof By symmetry of preferred paths (Proposition 2.15(ii)), the second statement follows from the first. After a left translation, we may assume $g = 1$, so we must show the claim for $A(*, h\langle\Delta\rangle)$ and $A(*, h'\langle\Delta\rangle)$.
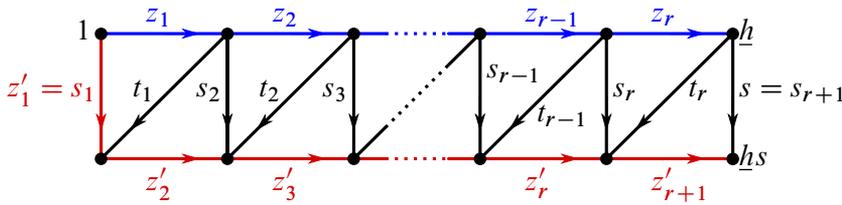
Figure 1: The normal form paths of $\underline{h}$ (in blue) and of $\underline{h}s$ (in red) stay at Hausdorff distance 1 in the Cayley graph of G.

We may assume from the hypothesis (and Lemma 2.10) that there exists a simple element s such that $\underline{h}s = \underline{h}'$. Let $z_1 \cdots z_r$ be the left normal form of $\underline{h}$. The vertices along $A(*, h\langle\Delta\rangle)$ are $z_1 \cdots z_i\langle\Delta\rangle$, for $0 \leqslant i \leqslant r$.

Now, the left normal form of $\underline{h}' = \underline{h}s$ may be obtained, for instance, from the algorithm given in [Gebhardt and González-Meneses 2010, Proposition 1.2], which is illustrated in Figure 1. It goes as follows. Set $s_{r+1} = s$; for $i = r, \ldots, 1$, define recursively the simple elements $t_i = \partial(z_i) \wedge s_{i+1}$, $s_i = z_i t_i$ and $z'_{i+1} = t_i^{-1} s_{i+1}$. Finally define $z' = s_1$. Then $z'_1 \cdots z'_r z'_{r+1}$ (or $z'_1 \cdots z'_r$, if $z'_{r+1} = 1$) is the left normal form of $\underline{h}$.

We then see that for $i = 1, \ldots, r$, we have $z'_1 \cdots z'_i = z_1 \cdots z_i t_i$; hence the $(i+1)^{\text{st}}$ vertex along $A(*, h'\langle\Delta\rangle)$ is at distance 1 from the $(i+1)^{\text{st}}$ vertex along $A(*, h\langle\Delta\rangle)$. $\qquad\square$

Lemma 2.17 (concatenation of normal form paths) *Let $g, h, k \in G$ be such that $\underline{g} \preccurlyeq \underline{h} \preccurlyeq \underline{k}$. Then:*

(i) *The concatenation of the paths $A(g, h)$ and $A(h, k)$ is a $(2, 0)$–quasigeodesic connecting $g\langle\Delta\rangle$ to $k\langle\Delta\rangle$.*

(ii) *Suppose in addition that the paths $A(g, h)$ and $A(g, k)$ have the same length; then the concatenation of $A(h, k)$ and $A(k, g)$ is a $(2, 0)$–quasigeodesic connecting $h\langle\Delta\rangle$ to $g\langle\Delta\rangle$.*

Proof Let $a = \underline{g}^{-1}\underline{h}$, $b = \underline{h}^{-1}\underline{k}$ and $c = \underline{g}^{-1}\underline{k}$. Under our hypothesis, $\inf(a) = \inf(b) = \inf(c) = 0$ and $c = ab$. Denote by $a = a_1 \cdots a_r$ and $b = b_1 \cdots b_s$ the respective left normal forms.

(i) Denote by α the concatenation of paths under consideration. Let $p\langle\Delta\rangle$ be the kth vertex on $A(g, h)$, for some $1 \leqslant k \leqslant r$; let $q\langle\Delta\rangle$ be the $(l+1)^{\text{st}}$ vertex on $A(h, k)$, for some $1 \leqslant l \leqslant s$. The distance $d_{\mathscr{H}}(p, q)$ is the length of $A(p, q)$, that is $\sup(\underline{p}^{-1}q) = \sup(a_k \cdots a_r b_1 \cdots b_l)$. As the subwords $a_k \cdots a_r$ and $b_1 \cdots b_l$ are in left normal form, we obtain $\overline{d_{\mathscr{H}}(p, q)} \geqslant \max\{r - k + 1, l\}$. Note that the portion of α between $p\langle\Delta\rangle$ and $q\langle\Delta\rangle$ has length $r - k + 1 + l$. We then obtain

$$\tfrac{1}{2}(r - k + 1 + l) \leqslant \max\{r - k + 1, l\} \leqslant d_{\mathscr{H}}(p, q) \leqslant r - k + 1 + l \leqslant 2(r - k + 1 + l);$$

hence α is a $(2, 0)$–quasigeodesic.

(ii) The extra hypothesis is saying that $\sup(a) = \sup(c)$. Consider the vertices $g'\langle\Delta\rangle = \underline{h}^{-1}g\langle\Delta\rangle$, $h'\langle\Delta\rangle = \underline{h}^{-1}h\langle\Delta\rangle = *$ and $k'\langle\Delta\rangle = \underline{h}^{-1}k\langle\Delta\rangle$; our goal is to show that the concatenation of $A(h', k')$

and $A(k', g')$ is a $(2, 0)$–quasigeodesic connecting $h'\langle\Delta\rangle$ to $g'\langle\Delta\rangle$. Let us compute the respective distinguished representatives: we have $\underline{h}' = 1$, $\underline{k}' = b$ and

$$\underline{g}' = \underline{h}^{-1}\underline{g} = \underline{a}^{-1} = a^{-1}\Delta^{\sup(a)}.$$

Moreover, $a^{-1}\Delta^{\sup(a)} = b(c^{-1}\Delta^{\sup(a)}) = b(c^{-1}\Delta^{\sup(c)})$ and $c^{-1}\Delta^{\sup(c)}$ is positive, so $b \preccurlyeq \underline{g}'$. We thus have $\underline{h}' \preccurlyeq \underline{k}' \preccurlyeq \underline{g}'$ and we are in a position to apply Lemma 2.17(i) to obtain the desired claim. $\qquad\square$

3 Morse elements in Garside groups

In this section we will define Morse elements in Garside groups, and prove some preliminary results on them. First, we set some notation and we recall the definition. We recall that throughout, G denotes a Δ–pure Garside group of finite type.

Notation 3.1 Let $g \in G$.

- The *axis* of g in $\mathcal{X}$ is the set of vertices $\{g^t\langle\Delta\rangle \mid t \in \mathbb{Z}\}$.
- The *axis* of g in $\overline{\Gamma}$ is the set of vertices $\{g^t\langle\Delta\rangle \mid t \in \mathbb{Z}\}$.

Remark 3.2 We caution the reader that two elements g and $g\Delta$ may have completely different axes in $G/Z(G)$, and thus in $\mathcal{X}$. For instance in braid groups, it may very well happen that some element g is pseudo-Anosov while $g\Delta$ is reducible.

Definition 3.3 (a) Let (X, d) be a metric space, and let $\gamma \colon \mathbb{Z} \to X$ be a map. We say that γ (or equivalently its image) is *Morse* if
  (i) γ is a quasi-isometric embedding, and
  (ii) for every pair (Λ, K) with $\Lambda \geqslant 1$ and $K \geqslant 0$, there is a number $M^{(\Lambda, K)}$ (the Morse constant) such that every (Λ, K)–quasigeodesic in X connecting two points, $\gamma(i)$ and $\gamma(j)$ (for $i, j \in \mathbb{Z}$), of the image of γ remains in the $M^{(\Lambda, K)}$–neighborhood of the image of γ in X.

(b) Let H be a group generated by a finite set S. An infinite order element $h \in H$ is *Morse* if the map $\gamma \colon \mathbb{Z} \to \mathrm{Cay}(H, S)$, $t \mapsto h^t$, is Morse in the sense of (a). In this situation, given $\Lambda \geq 1$ and $K \geq 0$ we denote by $M_h^{(\Lambda, K)}$ the associated Morse constant.

Notation 3.4 Since many quasigeodesics in this paper will be $(2, 0)$–quasigeodesics, we will use the simplified notation $M_h = M_h^{(2,0)}$.

Since a Δ–pure Garside group of finite type has an infinite-cyclic center, it cannot contain any Morse elements. We adapt the definition as follows, keeping in mind Proposition 2.12 which says that the projection $\overline{\Gamma} \to \mathcal{X}$, $g\langle\Delta^e\rangle \mapsto g\langle\Delta\rangle$, is a quasi-isometry and the fact that the Morse property is invariant under quasi-isometry.

Definition 3.5 We say that an element g of a Δ–pure Garside group of finite type G is *Morse* if any of the following equivalent condition holds:

  (i) the image of g in $G/Z(G)$ is Morse in the sense of Definition 3.3(b);

  (ii) the axis of g in $\overline{\Gamma}$ is Morse in the sense of Definition 3.3(a);

  (iii) the axis of g in $\mathscr{X}$ is Morse in the sense of Definition 3.3(a).

Example 3.6 Pseudo-Anosov braids are Morse; indeed, their projections to the group $B_n/Z(B_n)$, which is a finite-index subgroup of the mapping class group of an $(n{+}1)$–times punctured sphere, are pseudo-Anosov mapping classes in this group. By [Behrstock 2006], these are Morse.

The following is a key technical result; recall that the notion of right-rigidity is introduced in Definition 2.8.

Proposition 3.7 *Every Morse element of G has a power which is conjugate to a right-rigid element. Moreover, this right-rigid element can be required to be of the form $\Delta^{e \cdot m} x$, with $m \in \mathbb{Z}$, $\inf(x) = 0$, and x right-rigid.*

In order to prepare the proof of Proposition 3.7, we need the following lemma. It is well known to experts (see eg [Tran 2019, Corollary 4.16]), but for the sake of completeness we will give an elementary proof.

Lemma 3.8 *Let H be a group generated by a finite set S. If $h \in H$ is a Morse element in the sense of Definition 3.3(b), then $\langle h \rangle$ has finite index in the centralizer $Z(h)$ in H.*

Proof Denote by d the word distance associated to S in H. Let $A \geqslant 1$ and $B \geqslant 0$ such that $t \mapsto h^t$ is an (A, B)–quasi-isometric embedding of $\mathbb{Z}$ in $\mathrm{Cay}(H, S)$. Let $M = M_h^{(4A, B)}$.

We claim that if $zh = hz$, then $d(z, \langle h \rangle) \leqslant M$. Granted this claim, we then see that for all $z \in Z(h)$, there is some $m \in \mathbb{Z}$ such that $d(z, h^m) = d(1, z^{-1} h^m) \leqslant M$, that is the coset $z^{-1} \langle h \rangle$ has a representative of word length at most M. It follows that each coset of $Z(h)$ modulo $\langle h \rangle$ has a representative of word length at most M, so there are only finitely many cosets.

To prove the claim, assume on the contrary that $d(z, \langle h \rangle) = K > M$, for some $z \in Z(h)$. Let r be such that $d(z, h^r) = K$; up to replacing z by $h^{-r} z$, we may assume that $K = d(z, \langle h \rangle) = d(z, 1)$. Choose N big enough that $d = d(h^{-N}, h^N) > 4K$. Then we piecewise define a map $\gamma \colon \{0, \ldots, 2K{+}2N\} \to \mathrm{Cay}(H, S)$ as follows:

  • For $0 \leqslant t \leqslant K$, $\gamma(t) = \gamma_1(t)$ follows a geodesic from h^{-N} to zh^{-N}.

  • For $K \leqslant t \leqslant K + 2N$, $\gamma(t) = \gamma_2(t) = zh^{t-K-N}$.

  • For $K + 2N \leqslant t \leqslant 2K + 2N$, $\gamma(t) = \gamma_3(t)$ follows a geodesic from zh^N to h^N.

To conclude, we shall observe that γ defines a $(4A, B)$–quasigeodesic with endpoints h^{-N} and h^N on the axis of h, hence contradicting the hypothesis that h is Morse (as the vertex $z = \gamma(K + N)$ does not lie in the M–neighborhood of $\langle h^n, n \in \mathbb{Z} \rangle$). $\quad\square$

Let $0 \leqslant s, t \leqslant 2K + 2N$. The upper bound $d(\gamma(s), \gamma(t)) \leqslant A|s - t| + B$ follows immediately from the fact that γ is a juxtaposition of geodesics and an (A, B)–quasigeodesic. To obtain the lower bound, we consider two cases.

First, let $0 \leqslant s \leqslant K$ and $K \leqslant t \leqslant K + 2N$ (the situation is symmetric for $K + 2N \leqslant s \leqslant 2K + 2N$). Because $zh^{-N} = \gamma(K)$ realizes the minimum possible distance between h^{-N} and $\{zh^n, n \in \mathbb{Z}\}$, we have $d(\gamma(s), \gamma(t)) \geqslant d(\gamma(s), \gamma(K))$. Then we have

$$d(\gamma(s), \gamma(K)) + d(\gamma(K), \gamma(t)) \leqslant 2d(\gamma(s), \gamma(K)) + d(\gamma(s), \gamma(t)) \leqslant 3d(\gamma(s), \gamma(t))$$

and the lower bound $d(\gamma(s), \gamma(t)) \geqslant \frac{1}{3A}|s - t| - \frac{1}{3}B$ follows.

Second, let $0 \leqslant s \leqslant K$ and $K + 2N \leqslant t \leqslant 2K + 2N$. By our choice of N, we have

$$d(\gamma(s), \gamma(t)) \geqslant \tfrac{1}{2}d \geqslant \tfrac{1}{4}d + K \geqslant \tfrac{1}{4}d + \tfrac{1}{2}K \geqslant \tfrac{1}{4}d + \tfrac{1}{4}d(\gamma(s), zh^{-N}) + \tfrac{1}{4}d(\gamma(t), zh^N)$$

and we obtain the lower bound $d(\gamma(s), \gamma(t)) \geqslant \frac{1}{4A}|t - s| - \frac{1}{4}B$. □

Proof of Proposition 3.7 For the proof of the first sentence, we recall the following result, proven (although not stated in this form) in [Birman et al. 2007, Theorem 3.23 and Corollary 3.24]: Suppose that a certain element of a Garside group is not a root of a central element (infinite order condition), and that every element commuting with it has a common power with it, up to multiplication by a central power of Δ (small centralizer condition). Then the element has a power which is conjugate to a right-rigid element. This result from [Birman et al. 2007] implies the first sentence, because any Morse element satisfies the infinite order condition, and it satisfies the small centralizer condition by Lemma 3.8.

For the proof of the second sentence, we simply remark that for a *rigid* element g and an integer k,

$$\inf(g^k) = k \cdot \inf(g).$$

Thus by taking a further power, we can achieve that the infimum of the rigid conjugate is a multiple of e. □

Remark 3.9 There is an alternative proof of Proposition 3.7, which does not use Lemma 3.8 or the paper [Birman et al. 2007], but which is rather reminiscent of the "pumping lemma" [Epstein et al. 1992, Theorem 1.2.13]. The idea is that the right normal form of g^n has to stay close to the axis of the Morse element g. Now right normal forms belong to a language recognized by a finite state automaton; thus for large enough n, this right normal form has a middle segment with periodic behavior. This periodic segment is a right-rigid conjugate of a power of g. We leave the details as an exercise.

Notation 3.10 From here on, all diagrams in this paper will take place in $\mathcal{X}$ (not in the Cayley graph $\Gamma(G)$ or in $\overline{\Gamma} = \Gamma(G/Z(G))$). Also, in the diagrams, we simplify the notation, labeling a vertex g if it is represented by a group element g — strictly speaking, it should be labeled $g\langle\Delta\rangle$.
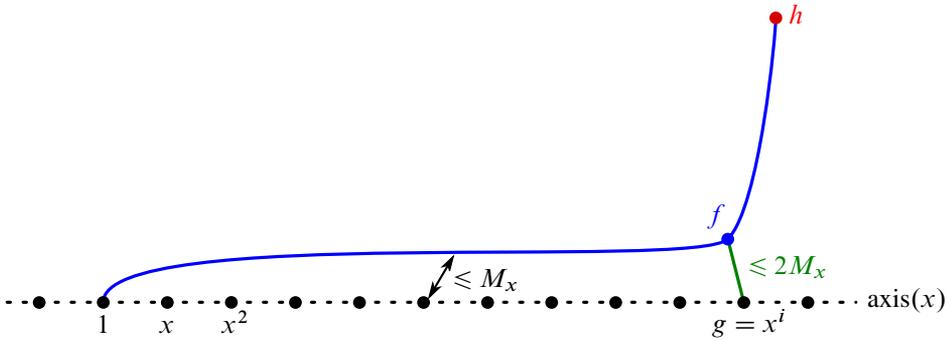
Figure 2: The statement of Proposition 3.11. Here the segment from $*$ to $f\langle\Delta\rangle$, with $f = \underline{h} \wedge \Delta^{\ell_i}$, has the same length as the segment from $*$ to $g\langle\Delta\rangle$, namely ℓ_i.

Proposition 3.11 (preferred paths stay close to the axis) *Suppose x is a Morse element of G satisfying* $\inf(x^k) = 0$ *for every positive integer k. For $i \in \mathbb{N}$, define $\ell_i = \ell(x^i)$. Let $h \in G$ and suppose that there exists an integer $i \geqslant 0$ such that $x^i \preccurlyeq \underline{h}$. Consider the initial segment of the path $A(1, h)$ in $\mathscr{X}$ of length ℓ_i (the same length as x^i). Then this segment stays within distance M_x from the axis of x:*

$$\text{if } 0 \leqslant k \leqslant \ell_i \text{ then } d_{\mathscr{X}}((\underline{h} \wedge \Delta^k)\langle\Delta\rangle, \text{axis}(x)) \leqslant M_x.$$

Moreover,

$$d_{\mathscr{X}}((\underline{h} \wedge \Delta^{\ell_i})\langle\Delta\rangle, x^i\langle\Delta\rangle) \leqslant 2M_x.$$

Proof Let $f = \underline{h} \wedge \Delta^{\ell_i}$ and $g = x^i$. Of course, $\underline{f} = f$, and our first hypothesis on x says that $\underline{g} = g$. Because $x^i \preccurlyeq \underline{h}$, we have $1 \preccurlyeq \underline{g} \preccurlyeq \underline{f}$ and by construction of f, the paths $A(1, f)$ and $A(1, g)$ have the same length. By Lemma 2.17(ii), the concatenation of paths $A(1, f)$, followed by $A(f, g)$ is a $(2, 0)$–quasigeodesic. This yields the first statement, recalling that $M_x = M_x^{(2,0)}$ is the Morse constant.

For the second statement, we notice that in particular, the vertex $f\langle\Delta\rangle$ is at distance at most M_x from some point $x^k\langle\Delta\rangle$ on the axis. By the triangle inequality,

$$\ell_i - M_x \leqslant d_{\mathscr{X}}(*, x^k\langle\Delta\rangle) \leqslant \ell_i + M_x,$$

and therefore $x^k\langle\Delta\rangle$ lies in an M_x–neighborhood of $g\langle\Delta\rangle$. We conclude that

$$d_{\mathscr{X}}(f\langle\Delta\rangle, g\langle\Delta\rangle) \leqslant d_{\mathscr{X}}(f\langle\Delta\rangle, x^k\langle\Delta\rangle) + d_{\mathscr{X}}(x^k\langle\Delta\rangle, g\langle\Delta\rangle) \leqslant 2M_x. \qquad \square$$

4 Projection to the axis

In this section, we will define, in Garside-theoretical terms, a projection from $\mathscr{X}$ to the axis of any element x of G, provided that $\inf(x) = 0$ and that x satisfies the additional hypothesis of being *right-rigid* (Definition 2.8). If, moreover, x is Morse, then this projection satisfies a contraction property (Proposition 4.8) which extends Proposition 3.11. This will be sufficient for deducing that our projection coincides, up to a bounded error, with any closest point projection (Corollary 4.9).

Remark 4.1 We remark that the occurrence of the condition of right-rigidity in our context is quite surprising, as this is a condition on the *right* normal form of x, whereas otherwise, we are generally using *left* normal forms throughout this paper.

Lemma 4.2 *Let* $x \in G \setminus \{1_G\}$ *be such that* $\inf(x^k) = 0$ *for all* $k \in \mathbb{N}$. *Let* $h \in G$. *Then the set* $\{k \in \mathbb{Z} \mid x \npreceq \underline{x^k h}\}$ *is nonempty and bounded above.*

In other words, we are claiming that there are indices k with $x \npreceq \underline{x^k h}$, but if k is sufficiently large then $x \preceq \underline{x^k h}$.

Proof We will study the bi-infinite sequence $(\inf(x^k h))_{k \in \mathbb{Z}}$. This sequence is nondecreasing: indeed, by [El-Rifai and Morton 1994, Section 1], for any group elements g and h, we have $\inf(g \cdot h) \geq \inf(g) + \inf(h)$; in particular, $\inf(x^{k+1} h) \geq \inf(x) + \inf(x^k h) \geq \inf(x^k h)$.

The same argument yields the inequality $\inf(x^k) \geq \inf(x^k h) + \inf(h^{-1}) = \inf(x^k h) - \sup(h)$, which we will be using in the equivalent form

$$\inf(x^k h) \leq \inf(x^k) + \sup(h).$$

Let us now prove that the set $\{k \in \mathbb{Z} \mid x \npreceq \underline{x^k h}\}$ is nonempty. The preceding inequality implies that $\lim_{k \to -\infty} \inf(x^k h) = -\infty$, and in particular that there exists some index k with $\inf(x^{k-1} h) < \inf(x^k h)$.

Now if an index k has the property that $x \preceq \underline{x^k h}$ then $\inf(x^{-1} \underline{x^{-k} h}) = 0$. However,

$$x^{-1} \underline{x^{-k} h} = x^{-1} \cdot x^{-k} h \cdot \Delta^{-\inf(x^{-k} h)} = x^{-k-1} h \cdot \Delta^{-\inf(x^{-k} h)}$$

and thus

$$\inf(x^{-k-1} h) = \inf(x^{-1} \underline{x^{-k} h}) + \inf(x^{-k} h) = \inf(x^{-k} h).$$

As seen above, this cannot be true for all indices k, so there must exist an index k with $x \npreceq \underline{x^k h}$.

We now turn to the proof that, by contrast, $x \preceq \underline{x^k h}$ for sufficiently large values of k. In fact, we will prove the stronger result that there exists $k_0 \in \mathbb{N}$ such that for $k \geq k_0$, $x^{k-k_0} \preceq \underline{x^k h}$.

We look at positive indices k, for which, we recall, $\inf(x^k) = 0$. As seen above, we have

$$\inf(x^k h) \leq \inf(x^k) + \sup(h) = \sup(h) \text{ for } k \geq 0.$$

Thus the increasing integer sequence $(\inf(x^k h))_{k \in \mathbb{N}}$ is bounded above by $\sup(h)$, and hence it is eventually constant: there exist $k_0 \in \mathbb{N}$ and $n_0 \in \mathbb{Z}$, with $n_0 \leq \sup(h)$, such that for all $k \geq k_0$, $\inf(x^k h) = n_0$. Now, for $k \geq k_0$,

$$\underline{x^k h} = x^k h \Delta^{-n_0} = x^{k-k_0} x^{k_0} h \Delta^{-n_0} = x^{k-k_0} \underline{x^{k_0} h}$$

which finishes the proof. $\square$

Definition 4.3 (Garside-theoretical projection to the axis of x) Let $x \in G$ be such that $\inf(x^k) = 0$ for all $k \in \mathbb{N}$.

(a)   Let $h \in G$. We define the integer

$$\lambda(h) = -\max\{k \in \mathbb{Z}, x \not\preccurlyeq \underline{x^k h}\}$$

and we note that for any $t \in \mathbb{Z}$, $\lambda(h\Delta^t) = \lambda(h)$.

(b)   We define a map $\pi \colon G \to G$ as follows: for any $h \in G$, we set $\pi(h) = x^{\lambda(h)}$. For any $t \in \mathbb{Z}$, we have $\pi(h\Delta^t) = \pi(h)$ so we can define the *Garside-theoretical projection to the axis of* x to be the map (which we denote by the same letter)

$$\pi \colon \mathscr{X} \to \mathscr{X}, \quad h\langle\Delta\rangle \mapsto x^{\lambda(h)}\langle\Delta\rangle.$$

From now on, for the rest of the paper, we make the stronger hypothesis that x satisfies $\inf(x) = 0$ and is *right-rigid* (see Definition 2.8). This implies in particular that $\inf(x^k) = 0$ for all $k \in \mathbb{N}$.

Lemma 4.4 *Let $x \in G$ with $\inf(x) = 0$ be a **right-rigid** element. Let $h \in G$. The following are equivalent*:

   (i) $x \preccurlyeq \underline{h}$;

  (ii)   *for all $k \geqslant 0$, $x^{k+1} \preccurlyeq \underline{x^k h}$.*

Proof   Only (i) $\Longrightarrow$ (ii) needs a proof. Suppose that $x \preccurlyeq \underline{h}$, and let $k > 0$. We claim that $\underline{x^k h} = x^k \underline{h}$ (ie that $\inf(x^k \underline{h}) = 0$). In order to prove this claim, we write $\underline{h} = xa$ with a positive. Since x is right-rigid, the condition $\inf(xa) = 0$ implies $\inf(x^{k+1}a) = 0$ (see [Calvez and Wiest 2017b, Lemma 1]). This means that $\inf(x^k \underline{h}) = 0$, proving the claim. Now (ii) is an immediate consequence.   □

When x is right-rigid, Lemma 4.4 yields a clean interpretation of the Garside-theoretical projection to the axis of x (see Figure 3):

Lemma 4.5   *Let $x \in G$ with $\inf(x) = 0$ be a right-rigid element. Let $h \in G$. For $m \in \mathbb{Z}$, the following conditions are equivalent*:

   (i) $m = -\lambda(h)$;

  (ii)   (a) $x \not\preccurlyeq \underline{x^m h}$, *and*
             (b) $x \preccurlyeq \underline{x^{m+1}h}$;

 (iii)   *for every $k \geqslant 0$*,
             (a) $x^{k+1} \not\preccurlyeq \underline{x^{m+k}h}$, *and*
             (b) $x^k \preccurlyeq \underline{x^{m+k}h}$.

In particular, whenever $\lambda(h) > 0$, we have $x^{\lambda(h)} \preccurlyeq \underline{h}$. Also, for every $k \in \mathbb{Z}$, $\lambda(x^k) = k$.

Proof   Implications (i) $\Longrightarrow$ (ii) and (iii) $\Longrightarrow$ (i) are clear by definition of λ. Assume (ii); this yields immediately statement (iii) for $k = 0$. Suppose $k > 0$. For (iii)(a), suppose on the contrary that
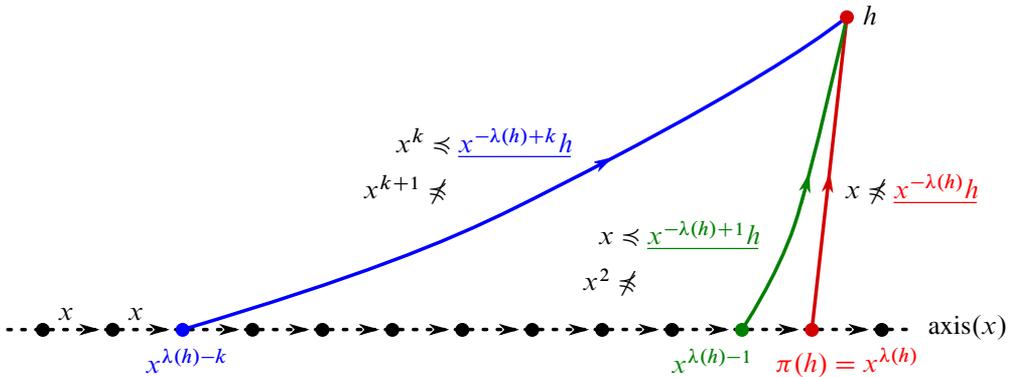
Figure 3: The definition of the projection π from $\mathscr{X}$ to the axis, for $h \in G$, and its properties (supposing that x is right-rigid).

$x^{k+1} \preccurlyeq \underline{x^{m+k}h}$. Then $\underline{x^{m+k}h} = x^{k+1}a$ for some positive a and $\underline{x^m h} = xa$, contradicting (ii)(a). For (iii)(b), we use the hypothesis (ii)(b) that $x \preccurlyeq \underline{x^{m+1}h}$; the conclusion then follows immediately from Lemma 4.4. □

Lemma 4.6 *Let $x \in G$ with $\inf(x) = 0$ be a right-rigid element. Let $z \in G$ with $\inf(z) = 0$, and let s be a simple element. Suppose that $x \npreccurlyeq z$. Then $x^2 \npreccurlyeq zs$.*

Proof Let $x_r \cdots x_2 x_1$ be the right normal form of x. Assume, contrary to the claim, that $x^{-2}zs$ is positive. Then by Lemma 2.4(ii), $D_r(x^{-2}z) \preccurlyeq s$. We deduce (Lemma 2.4(iii)) that $D_l(x^{-2}z)$ is also simple. On the other hand, by Lemma 2.4(i), $x^2 \succcurlyeq D_l(x^{-2}z)$ and, as x is right-rigid, $x_1 \succcurlyeq D_l(x^{-2}z)$. It then follows that $x x_r \cdots x_2 \preccurlyeq x^2 D_l(x^{-2}z)^{-1} \preccurlyeq z$, a contradiction. □

Proposition 4.7 *(π is $\ell(x)$–Lipschitz) Let $x \in G$ be a right-rigid element with $\inf(x) = 0$ and canonical length $\ell = \ell(x)$. If $g, h \in G$ satisfy $d_{\mathscr{X}}(g, h) = 1$, then $|\lambda(g) - \lambda(h)| \leqslant 1$ and $d_{\mathscr{X}}(\pi(g), \pi(h)) \leqslant \ell$.*

Proof Let $k_g = -\lambda(g)$ and $k_h = -\lambda(h)$. Apply Lemma 2.10 to the adjacent vertices $x^{k_g} g \langle \Delta \rangle$ and $x^{k_g} h \langle \Delta \rangle$: for some simple element s, either $(x^{k_g} g)s = \underline{x^{k_g} h}$ or $(x^{k_g} h)s = \underline{x^{k_g} g}$ holds. Let us consider the first case. By definition of k_g, $x \npreccurlyeq \underline{x^{k_g} g}$ and by Lemma 4.6, $x^2 \npreccurlyeq (x^{k_g} g)s = \underline{x^{k_g} h}$. From Lemma 4.4, we deduce that $x \npreccurlyeq \underline{x^{k_g-1}h}$. This means that $k_g - 1 \leqslant k_h$. In the second case, we obtain that $x \npreccurlyeq \underline{x^{k_g} h}$, whence $k_g \leqslant k_h$. A similar reasoning applied to the adjacent vertices $x^{k_h} g \langle \Delta \rangle$ and $x^{k_h} h \langle \Delta \rangle$ shows that either $k_h - 1 \leqslant k_g$ or $k_h \leqslant k_g$. In either case, we obtain the desired claim for $\lambda(g) = -k_g$ and $\lambda(h) = -k_h$.

Finally, the inequality $d_{\mathscr{X}}(\pi(g), \pi(h)) \leqslant \ell$ is an immediate consequence. □

Let us now combine the rigidity and the Morse hypothesis on x:

Proposition 4.8 *Let $x \in G$ be a right-rigid Morse element with $\inf(x) = 0$ and canonical length ℓ. Let π be the Garside-theoretical projection to $\mathrm{axis}(x)$. Then there exists a $D \in \mathbb{N}$ such that for any $h \in G$ and*
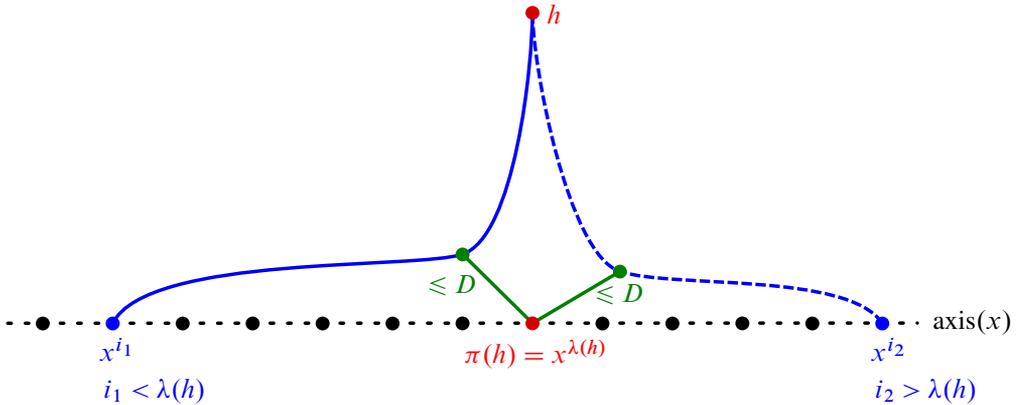
Figure 4: The statement of Proposition 4.8.

for any $i \in \mathbb{Z}$, the preferred geodesic $A(x^i, h)$ in $\mathscr{X}$ passes at distance at most D from $\pi(h\langle\Delta\rangle)$: there exists some $h' \in G$ with $h'\langle\Delta\rangle$ belonging to $A(x^i, h)$ such that

$$d_{\mathscr{X}}(h', \pi(h)) \leqslant D.$$

Specifically, we can take $D = (\ell + 1) \cdot M_x$.

Proof We treat the cases $i < \lambda(h)$ and $i > \lambda(h)$ separately.

If $i < \lambda(h)$, then after the action of x^{-i}, we can assume without loss of generality that $i = 0$. Noting that $x^{\lambda(h)} \preccurlyeq \underline{h}$, we are then precisely in the situation of Proposition 3.11. Note that by the rigidity hypothesis, $\ell(x^{\lambda(h)}) = \lambda(h) \cdot \ell$. Thus, if we define $h' = \underline{h} \wedge \Delta^{\lambda(h) \cdot \ell}$, we have

$$d_{\mathscr{X}}(h', \pi(h)) \leqslant 2 \cdot M_x.$$

If $i > \lambda(h)$, then this time we will assume, again without loss of generality (after the action of $x^{-\lambda(h)+1}$), that $\lambda(h) = 1$. Thus $\underline{\pi(h)} = x$ and $x \preccurlyeq \underline{h}$ but $x^2 \npreccurlyeq \underline{h}$ (Lemma 4.5).

Let $h' = x^i \wedge \underline{h}$ and note that $\underline{h}' = h'$. We know from Proposition 2.15(i) that the vertex $h'\langle\Delta\rangle$ lies on the preferred geodesic $A(h, x^i)$, and our aim now is to bound its distance from $\pi(h\langle\Delta\rangle)$.

We make two observations about the vertex $h'\langle\Delta\rangle$. The first observation is that it lies in the M_x–neighborhood of axis(x), where we recall that M_x is the Morse constant for $(2, 0)$–quasigeodesics with endpoints on axis(x). This follows from Lemma 2.17(i) and the fact that $x \preccurlyeq \underline{h}' \preccurlyeq x^i$. (For later reference — see Remark 5.6 — we observe that we even have $x \preccurlyeq \underline{h}' \preccurlyeq x^i \wedge \Delta^{\sup(\underline{h}')}$.)

The second observation about the vertex $h'\langle\Delta\rangle$ is that it has the same projection to the axis as $h\langle\Delta\rangle$,

$$\pi(h') = \pi(h).$$

Here is a proof of this fact. We have to prove that $\lambda(x^i \wedge \underline{h}) = \lambda(h) = 1$. By Lemma 4.5, it suffices to prove that

$$x \preccurlyeq x^i \wedge \underline{h} \quad \text{but} \quad x^2 \npreccurlyeq x^i \wedge \underline{h}.$$
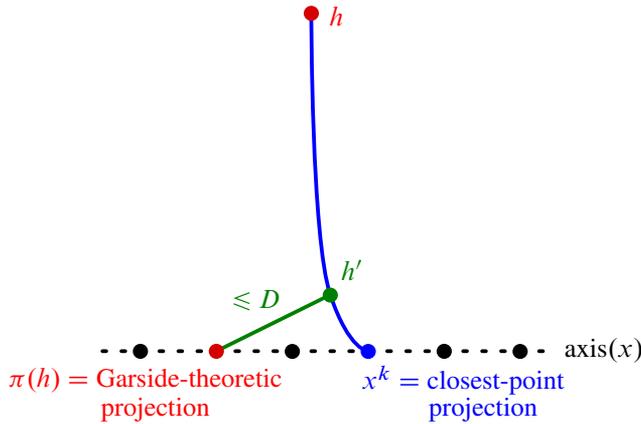
Figure 5: The projection π is uniformly close to any closest-point projection.

Keeping in mind the hypothesis that $i \geqslant 2$, this follows immediately from the analogous condition on $\underline{h}$. This completes the proof of the second observation about $h'\langle\Delta\rangle$.

By the first observation, there exists an integer k with

$$d_{\mathscr{X}}(h', x^k) \leqslant M_x.$$

Since the projection π is ℓ–Lipschitz (Proposition 4.7), we have

$$d_{\mathscr{X}}(\pi(h'), x^k) = d_{\mathscr{X}}(\pi(h'), \pi(x^k)) \leqslant \ell \cdot M_x.$$

Applying the triangle-inequality we obtain

$$d_{\mathscr{X}}(h', \pi(h')) \leqslant \ell \cdot M_x + M_x = (\ell + 1) \cdot M_x.$$

Also, by the second observation above we have $\pi(h') = \pi(h)$, so

$$d_{\mathscr{X}}(h', \pi(h)) \leqslant (\ell + 1) \cdot M_x.$$

The proof of Proposition 4.8 is complete, with

$$D = \max(2 \cdot M_x, (\ell + 1) \cdot M_x) = (\ell + 1) \cdot M_x. \qquad \square$$

We deduce that π is uniformly close to the closest point projection:

Corollary 4.9 *Let $x \in G$ with $\inf(x) = 0$ be a right-rigid Morse element, and π be the Garside-theoretical projection to $\mathrm{axis}(x)$. Let $h \in G$ and let $x^k\langle\Delta\rangle$ be any point of the axis such that*

$$d_{\mathscr{X}}(h, x^k) = \min_{i \in \mathbb{Z}} d_{\mathscr{X}}(h, x^i).$$

Then

$$d_{\mathscr{X}}(\pi(h), x^k) \leqslant 2D,$$

where D is the constant promised by Proposition 4.8.

Proof By Proposition 4.8, there is a point $h'\langle\Delta\rangle$ on the geodesic $A(h, x^k)$ such that $d_{\mathcal{X}}(\pi(h), h') \leqslant D$. Since $x^k\langle\Delta\rangle$ is a point on the axis as close as possible to $h\langle\Delta\rangle$, we must have

$$d_{\mathcal{X}}(x^k, h') \leqslant d_{\mathcal{X}}(\pi(h), h') \leqslant D.$$

By the triangle inequality, we conclude

$$d_{\mathcal{X}}(\pi(h), x^k) \leqslant 2D. \qquad \square$$

5 The strong contraction property

In this section we recall the definition of the strong contraction property and the strong constriction property. Then we prove the main result of this paper: in a Δ–pure Garside group of finite type G, the axis of any Morse element is strongly contracting.

The following two definitions and a proof of their equivalence can be found in [Arzhantseva et al. 2015].

Definition 5.1 Let (X, d) be a metric space and let $\mathcal{A}$ be any subset of X. A map $\rho : X \to \mathcal{A}$ is C–*strongly contracting* for $C \geqslant 0$ if the following hold:

  (i) ρ is coarsely equivalent to $i\,d_{\mathcal{A}}$ on $\mathcal{A}$: for every $a \in \mathcal{A}$, $d(\rho(a), a) \leqslant C$;

 (ii) for all $x \in X$, $d(x, \rho(x)) - d(x, \mathcal{A}) \leqslant C$;

(iii) for all $u, v \in X$, $d(u, v) < d(v, \mathcal{A}) - C$ implies that $d(\rho(u), \rho(v)) \leqslant C$ (ie if a ball in $\mathcal{X}$ is disjoint from a C–neighborhood of $\mathcal{A}$, then its image under ρ is contained in a ball of radius C).

The map ρ is *strongly contracting* if there exists a nonnegative integer C such that ρ is C–strongly contracting. The subset $\mathcal{A} \subset X$ is *strongly contracting* if there exists a strongly contracting map $X \to \mathcal{A}$.

Note [Arzhantseva et al. 2015, Lemma 2.8] that a strongly contracting map $\rho : X \to \mathcal{A}$ satisfies in fact a strengthened version of clause (ii) in Definition 5.1; namely, ρ is *coarsely a closest point-projection* to $\mathcal{A}$, meaning that for all $x \in X$, there exists an $a \in \mathcal{A}$ with $d(x, \mathcal{A}) = d(x, a)$ such that $d(\rho(x), a)$ is uniformly bounded.

As proven in [Arzhantseva et al. 2015, Proposition 2.9], a map ρ is strongly contracting if and only if it is *strongly constricting*; this alternative characterization will be useful in Lemma 5.3 and Section 6:

Definition 5.2 Let (X, d) be a metric space and let $\mathcal{A}$ be any subset of X. A map $\rho : X \to \mathcal{A}$ is C–*strongly constricting* for $C \geqslant 0$ if the following hold:

  (i) ρ is coarsely equivalent to $\mathrm{id}_{\mathcal{A}}$ on $\mathcal{A}$: for every $a \in \mathcal{A}$, $d(\rho(a), a) \leqslant C$;

 (ii) for every geodesic γ in X with endpoints x_0 and x_1, if $d(\rho(x_0), \rho(x_1)) > C$, then $d(\rho(x_i), \gamma) < C$ for $i = 0, 1$.

The map ρ is *strongly constricting* if there exists a nonnegative integer C such that ρ is C–strongly constricting.

Lemma 5.3 Let (X, d) be a metric space and let $\mathscr{A}$ be a subset of X. Let $\rho \colon X \to \mathscr{A}$ be a strongly contracting map.

(i) Suppose that $\mathscr{B} \subset X$ is another subset of X with $d_{\mathrm{Hausdorff}}(\mathscr{A}, \mathscr{B}) < \infty$. Then there is a strongly contracting map $\rho' \colon X \to \mathscr{B}$.

(ii) Let (X', d') be another metric space. Suppose there is an isometric and quasisurjective embedding $\iota \colon X \hookrightarrow X'$. Then there is a strongly contracting map $\rho' \colon X' \to \iota(\mathscr{A})$.

Proof For (i), let $\delta = d_{\mathrm{Hausdorff}}(\mathscr{A}, \mathscr{B})$. We construct ρ' by choosing, for any $x \in X$, a point $b \in \mathscr{B}$ with $d(b, \rho(x)) \leqslant \delta$, and declaring that $\rho'(x) = b$. Thus ρ and ρ' are δ–coarsely equivalent. Now it is an easy exercise to show that if ρ is C–strongly constricting then ρ' is $(C + 2\delta)$–strongly constricting.

For (ii), let ε be such that the ε–neighborhood of $\iota(X)$ in X' is all of X'. We define ρ' by choosing, for every $x' \in X'$, a point $x \in X$ with $d'(\iota(x), x') \leqslant \varepsilon$, and declaring that $\rho'(x') = \iota(\rho(x))$. We have to prove that ρ' is strongly contracting. More precisely, supposing that $\rho \colon X \to \mathscr{A}$ is C–strongly contracting, our aim is to prove that ρ' is $(C + 3\varepsilon)$–strongly contracting. We shall prove only part (iii) of Definition 5.1; the other two clauses can be checked easily.

For any point $v' \in X'$, consider the ball B' centered in v' and of radius $d'(v', \mathscr{A}) - C - 3\varepsilon$. If we choose a point of $\iota(X)$ at distance at most ε from each point of B', we obtain a subset of X which is contained in a ball in X centered at some point v (with $d'(\iota(v), v') \leqslant \varepsilon$) and of radius $d'(v', \mathscr{A}) - C - \varepsilon$. Since $d'(v', \mathscr{A}) - C - \varepsilon \leqslant d(v, \mathscr{A}) - C$, the projection $\rho'(B')$ is contained in $\iota(\rho(B))$, where B is the ball in X centered in v and of radius $d(v, \mathscr{A}) - C$. By hypothesis, $\mathrm{diam}(\rho(B)) \leqslant C < C + 3\varepsilon$, which is what we wanted to prove. $\qquad\square$

Proposition 5.4 Let G be a Δ–pure Garside group of finite type. Let $x \in G$ with $\inf(x) = 0$ be a right-rigid Morse element. The Garside-theoretical projection π to $\mathrm{axis}(x)$ is $5D$–strongly contracting, where D is the constant promised by Proposition 4.8.

Proof The first and second conditions of Definition 5.1 follow respectively from the last statement of Lemma 4.5 and Corollary 4.9, which asserts the stronger condition that π is coarsely a closest-point projection. Let us prove that condition (iii) is satisfied. Let $h, g \in G$ be such that $d_{\mathscr{X}}(h, g) \leqslant d_{\mathscr{X}}(h\langle\Delta\rangle, \mathrm{axis}(x))$ (that is, $g\langle\Delta\rangle$ lies in a ball in $\mathscr{X}$ centered at $h\langle\Delta\rangle$ and disjoint from the axis of x).

Let us write $r_* = d_{\mathscr{X}}(h, \pi(h))$. Now,

$$d_{\mathscr{X}}(h, g) \leqslant d_{\mathscr{X}}(h\langle\Delta\rangle, \mathrm{axis}(x)) \leqslant r_*.$$

By Proposition 4.8, the preferred geodesic $A(g, \pi(h))$ contains a point $g'\langle\Delta\rangle$ at distance at most D from $\pi(g\langle\Delta\rangle)$. By Proposition 2.15(v) (convexity of balls), we have $d_{\mathscr{X}}(h, g') \leqslant r_*$.

Let us now study the preferred geodesic $A(h, g')$. We have just seen that it is of length at most r_*. Moreover, by Lemma 2.16, it is at Hausdorff distance at most D from $A(h, \pi(g))$, which in turn passes
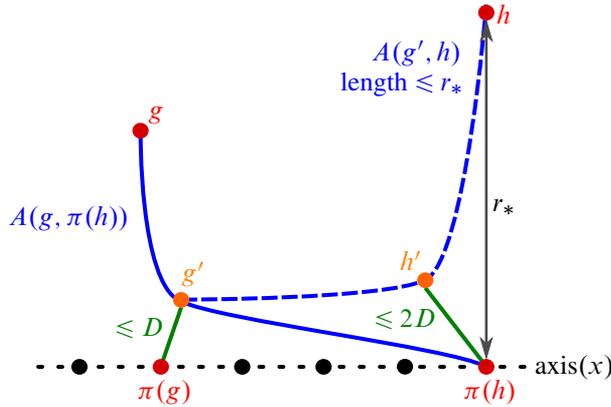
Figure 6: The proof of Proposition 5.4.

at distance at most D from $\pi(h\langle\Delta\rangle)$ (by Proposition 4.8 again). Thus $A(h, g')$ contains a point $h'\langle\Delta\rangle$ at distance at most $2D$ from $\pi(h\langle\Delta\rangle)$.

Now $d_{\mathscr{X}}(h', h) \geqslant r_* - 2D$ by the triangle inequality. Therefore,

$$d_{\mathscr{X}}(g', h') = d_{\mathscr{X}}(g', h) - d_{\mathscr{X}}(h', h) \leqslant 2D,$$

and we obtain the desired conclusion

$$d_{\mathscr{X}}(\pi(g), \pi(h)) \leqslant d_{\mathscr{X}}(\pi(g), g') + d_{\mathscr{X}}(g', h') + d_{\mathscr{X}}(h', \pi(h)) \leqslant D + 2D + 2D = 5D. \qquad \square$$

The following is the main result of this paper:

Theorem 5.5 (strong contraction property of axes) *Let G be a Δ–pure Garside group of finite type. Let $g \in G$ be a Morse element. Then*:

(i) *In $\mathscr{X} = \Gamma(G)/\langle\Delta\rangle$, the axis $\{g^k\langle\Delta\rangle \mid k \in \mathbb{Z}\} \subset \mathscr{X}$ is strongly contracting.*

(ii) *In $\overline{\Gamma} = \Gamma(G/Z(G))$, the axis $\{g^k Z(G) \mid k \in \mathbb{Z}\} \subset \overline{\Gamma}$ is strongly contracting.*

Proof of Theorem 5.5   First we recall that the axis of g being Morse in $\mathscr{X}$ or in $\overline{\Gamma}$ are equivalent properties, because the property of being Morse is invariant under quasi-isometry.

Now, by Proposition 3.7, there is an element $x \in G$ with $\inf(x) = 0$ which is right-rigid, and which is obtained from g by taking a power, conjugating by some element $a \in G$, and multiplying by a central element. Thus in both spaces, $\mathscr{X}$ and $\overline{\Gamma}$, taking the axis of x and translating it by the action of a yields a subset which is at finite Hausdorff distance from the axis of g. By Proposition 5.4, the axis of x is strongly contracting in $\mathscr{X}$, and so is its image under the a–action; by Lemma 5.3(i), the axis of g is strongly contracting in $\mathscr{X}$.

For Theorem 5.5(ii) we recall from Proposition 2.12 that there is an isometric embedding $\iota \colon \mathscr{X} \hookrightarrow \overline{\Gamma}$ with $\lfloor \frac{1}{2}e \rfloor$–dense image. The vertices of the image are those which are represented by elements g with

$\inf(g) \equiv 0 \mod e$. In particular, the axis of x in $\overline{\Gamma}$ is the image under ι of the axis of x in $\mathscr{X}$. By Proposition 5.4 and Lemma 5.3(ii), the axis of x in $\overline{\Gamma}$ is strongly contracting (and so is its a–translate). By Lemma 5.3(i), the axis of g in $\overline{\Gamma}$ is also strongly contracting. □

Remark 5.6 The proof of our main theorem (Theorem 5.5) did not use the full strength of the Morse hypothesis. Let us state carefully what is required in order to prove that the axis of an element g of a Δ–pure Garside group of finite type G is strongly contracting.

First, we need g to have some power which is conjugate to a rigid element, which we will call x, with $\inf(x)$ divisible by e. By the result of [Birman et al. 2007] cited in the proof of Proposition 3.7, this is the case if the centralizer of g in $G/Z(G)$ is virtually cyclic. (Alternatively, as seen in Proposition 3.7, it follows if g is Morse.) We note that the axis of g is strongly contracting if and only if the axis of the rigid element x is, as seen in the proof of Theorem 5.5.

Second, we need x to satisfy a Morse hypothesis: let us say that a rigid element with $\inf(x) = 0$ is *weakly Garside Morse* if there exists an integer M such that, for any integer i,

(a) if $h \in G$ with $x^i \preccurlyeq h$, $\inf(h) = \inf(x^i) = 0$ and $\sup(h) = \sup(x^i)$, then $\sup(g^{-1}h) \leqslant M$, and similarly

(b) if $h \in G$ with $h \preccurlyeq x^i$, $\inf(h) = \inf(x^i) = 0$ and $\sup(h) = \sup(x^i)$, then $\sup(h^{-1}g) \leqslant M$.

Roughly speaking, the weak Garside version of the Morse condition requires that in any isosceles triangle in $\mathscr{X}$ made up of three Garside normal form paths, and where one of the two equal sides is part of the axis of x, the third side must be of universally bounded length.

Our proof of Theorem 5.5 shows that for any element g of any Δ–pure Garside group of finite type, the above two conditions imply strong contractibility of the axis of g, and in particular the full Morse property.

Indeed, there are three places in the proof where the Morse property was used: first, in Proposition 3.11, but this proposition is an immediate consequence of part (a) of the weak Garside Morse property. Second, the Morse property was used in the proof of Proposition 4.8, which has two cases; we observe that the first case follows again from part (a) of the weak Garside Morse property, and the second case only uses part (b) of this property, not the full Morse property. Third, the proof of Proposition 5.4 uses the Morse property, but only indirectly, by citing Proposition 4.8.

Remark 5.7 We conjecture that arbitrary bi-infinite axes in $\mathscr{X}$ satisfying the Morse property (not just the periodic ones) are strongly contracting.

6 Consequences for the additional length graph

In this section, we record a consequence of Theorem 5.5 for the study of the *additional length graph* $\mathscr{C}_{\mathrm{AL}}(G)$ of a Garside group G. This graph was introduced in [Calvez and Wiest 2017a], and further

studied in [Calvez and Wiest 2017b; 2021] (see also [Calvez 2022]). We briefly recall the definition and the main results from [Calvez and Wiest 2017a]:

Definition 6.1 Let G be a Garside group of finite type.

(a) An element $h \in G$ is *absorbable* if it satisfies two conditions:
   (i) $\inf(h) = 0$ or $\sup(h) = 0$;
   (ii) there exists an element $g \in G$ which "absorbs" h, meaning that $\inf(g) = \inf(gh)$ and $\sup(g) = \sup(gh)$.

(b) The *additional length graph* $\mathscr{C}_{\mathrm{AL}}(G)$ of G is the (usually locally infinite) graph with the same set of vertices and edges as $\mathscr{X}$, but with, additionally, a new edge between vertices $g\langle\Delta\rangle$ and $h\langle\Delta\rangle$ whenever there is an absorbable element $s \in G$ so that $\underline{g}s \in h\langle\Delta\rangle$. The graph metric of $\mathscr{C}_{\mathrm{AL}}(G)$ is denoted by d_{AL}: for vertices $g\langle\Delta\rangle$, $h\langle\Delta\rangle$ of $\mathscr{C}_{\mathrm{AL}}$, we sometimes write $d_{\mathrm{AL}}(g, h) = d_{\mathrm{AL}}(g\langle\Delta\rangle, h\langle\Delta\rangle)$.

Proposition 6.2 [Calvez and Wiest 2017a, Lemmas 1–3] (i) *An element $h \in G$ is absorbable if and only if h^{-1} is.*

(ii) *If $h = h_1 \cdot h_2 \cdot h_3$, with $\inf(h) = \inf(h_1) = \inf(h_2) = \inf(h_3) = 0$ is absorbable, then h_1, h_2 and h_3 are also absorbable.*

(iii) *Suppose that $h \in G$ is absorbable; then there exists an absorbing element g with $\inf(g) = 0$ and $\sup(g) = \ell(h)$.*

Since there is a natural inclusion $\mathscr{X} \hookrightarrow \mathscr{C}_{\mathrm{AL}}(G)$, we can interpret the family of paths $A(g, h)$ from Definition 2.13 as a family of paths in $\mathscr{C}_{\mathrm{AL}}(G)$.

Proposition 6.3 (properties of $\mathscr{C}_{\mathrm{AL}}$ [Calvez and Wiest 2017a, Theorem 1]) *Let G be a Garside group of finite type.*

(i) *The additional length graph $\mathscr{C}_{\mathrm{AL}}(G)$ is 60–hyperbolic.*

(ii) *The paths $A(g, h)$ form a uniform family of unparametrized quasigeodesics in $\mathscr{C}_{\mathrm{AL}}(G)$.*

Remark 6.4 (a) If G is the braid group B_n, equipped with the classical or dual Garside structure, then $\mathscr{C}_{\mathrm{AL}}(G)$ is conjectured to be quasi-isometric to the curve graph of the $(n+1)$–times punctured sphere.

(b) Note that Proposition 6.3 is *not* claiming that $\mathrm{diam}(\mathscr{C}_{\mathrm{AL}}(G)) = \infty$. For instance, the group $G = \mathbb{Z}^3$ carries a Garside structure with $\Delta = (1, 1, 1)$ (see [Dehornoy et al. 2015, Chapter 1,1.1]), for which all elements $(k, 0, 0)$, $(0, k, 0)$ and $(0, 0, k)$ (with $k \in \mathbb{Z}$) are absorbable, so that $\mathrm{diam}(\mathscr{C}_{\mathrm{AL}}(\mathbb{Z}^3)) = 3$. By contrast, for any Artin group of spherical type A we do have $\mathrm{diam}(\mathscr{C}_{\mathrm{AL}}(A)) = \infty$; the proof of this fact in [Calvez and Wiest 2017b] involved an explicit Garside-theoretical construction of elements with (very) strongly constricting axes.

Theorem 6.5 *For a Δ–pure Garside group of finite type G we consider the additional length graph $\mathscr{C}_{AL}(G)$, equipped with the $G/Z(G)$–action. For any Morse element g, the action of g on $\mathscr{C}_{AL}(G)$ is loxodromic and WPD.*

Here WPD is the weak proper discontinuity condition of [Bestvina and Fujiwara 2002]:

Definition 6.6 The action of $g \in G/Z(G)$ on $\mathscr{C}_{AL}(G)$ is *weakly properly discontinuous* (WPD) if for every (equivalently, for any) $k \in G$, and for every $\kappa > 0$, there exists $N > 0$ such that for all $n \geqslant N$, the set

$$\{h \in G/Z(G) \mid d_{AL}(k, hk) \leqslant \kappa, \ d_{AL}(g^N k, hg^n k) \leqslant \kappa\}$$

is finite.

Corollary 6.7 *If G contains a Morse element then $\mathrm{diam}(\mathscr{C}_{AL}(G)) = \infty$.*

Corollary 6.8 *Pseudo-Anosov braids act loxodromically and WPD on $\mathscr{C}_{AL}(B_n)$.*

Proof of Theorem 6.5 Let g a Morse element of G; by Proposition 3.7, g has a power which is conjugate to a right-rigid element of the form $\Delta^{em} x$ with x right-rigid and $\inf(x) = 0$. Thus it suffices to prove the theorem for a right-rigid Morse element x with $\inf(x) = 0$.

We know from Proposition 5.4 and from [Arzhantseva et al. 2015, Proposition 2.9] that there is a constant $C \in \mathbb{N}$ such that the Garside-theoretical projection $\pi \colon \mathscr{X} \to \mathrm{axis}(x)$ is C–strongly constricting.

Lemma 6.9 *There is a constant $F \in \mathbb{N}$ with the following property: suppose that we have $h_1, h_2 \in G$ and an absorbable element $s \in G$ such that $\underline{h}_1 s \in h_2 \langle \Delta \rangle$. Then*

$$d_{\mathscr{X}}(\pi(h_1), \pi(h_2)) < F.$$

Proof After exchanging the roles of h_1 and h_2, if necessary, we can suppose that $\inf(s) = 0$ (rather than $\sup(s) = 0$). We are going to prove that the bound $F = 2M_x^{(2,C)} + 6C$ works, where $M_x^{(2,C)}$ is the Morse constant for $(2, C)$–quasigeodesics with endpoints on $\mathrm{axis}(x)$.

If $d_{\mathscr{X}}(\pi(h_1), \pi(h_2)) \leqslant C$, then we are done. If $d_{\mathscr{X}}(\pi(h_1), \pi(h_2)) > C$, then the C–strong constriction property implies that the Garside normal form of s (as a word in the letters $\mathscr{D}$) can be cut into three pieces, yielding a factorization $s = s_1 \cdot s_2 \cdot s_3$ with $\inf(s_1) = \inf(s_2) = \inf(s_3) = 0$, and such that (see Figure 7)

$$d_{\mathscr{X}}(\underline{h}_1 s_1, \pi(h_1)) \leqslant C \quad \text{and} \quad d_{\mathscr{X}}(\underline{h}_1 s_1 s_2), \pi(h_2)) \leqslant C.$$

By Proposition 6.2(ii), all three factors s_1, s_2 and s_3 are absorbable. In particular, s_2 is. Let us denote by ℓ_2 the Garside length of s_2 — thus $\inf(s_2) = 0$ and $\sup(s_2) = \ell_2$.

As seen in [Calvez and Wiest 2021], absorbability of s_2 means that there is a geodesic triangle in $\mathscr{X}$ which is equilateral of side length ℓ_2, and one of whose sides is the geodesic $A(\underline{h}_1 s_1, \underline{h}_1 s_1 s_2)$. Moreover, for any two points in two different sides of this triangle, with distances d_1 and d_2 from the shared
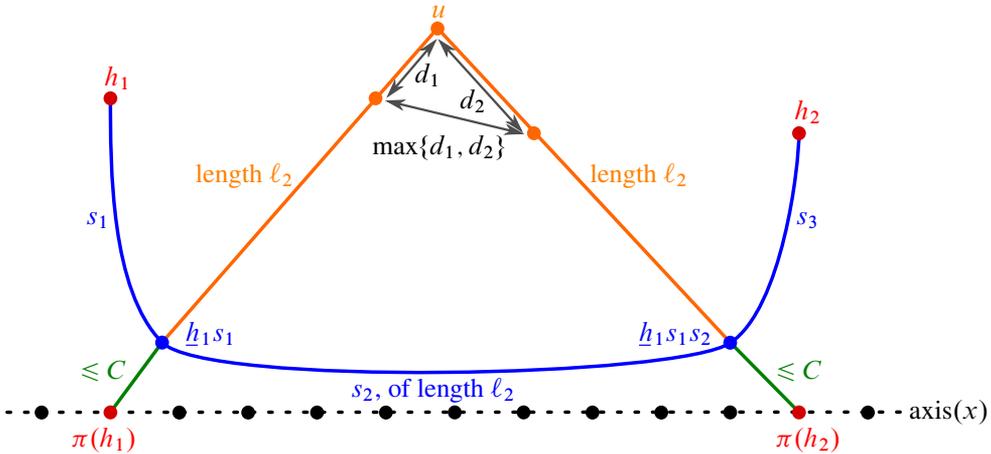
Figure 7: The proof of Lemma 6.9.

corner of the triangle, the distance of the two points in $\mathscr{X}$ is $\max(d_1, d_2)$. In particular, the triangle is $(2, 0)$–quasi-isometrically embedded in $\mathscr{X}$ (compare Lemma 2.17).

Let $u \in G$ be such that $u\langle\Delta\rangle$ is the corner of the triangle furthest from the axis of x. We claim that the distance of $u\langle\Delta\rangle$ from the axis is at least $\frac{1}{2}l_2 - 2C$. Indeed,

$$\ell_2 - C \leq d_{\mathscr{X}}(u, \pi(h_i)) \leq \ell_2 + C$$

for $i = 1, 2$. Moreover, for any $k \in \mathbb{Z}$ (not necessarily between $\lambda(h_1)$ and $\lambda(h_2)$) we have by the triangle inequality

$$d_X(u, x^k)$$
$$\geq \max\big(d_{\mathscr{X}}(x^k, \pi(h_2)) - \ell_2 - C, \ell_2 - C - d_{\mathscr{X}}(x^k, \pi(h_1)), \ell_2 - C - d_{\mathscr{X}}(x^k, \pi(h_2)), d_{\mathscr{X}}(x^k, \pi(h_1)) - \ell_2 - C\big).$$

Using the fact that $\ell_2 - 2C \leq d_{\mathscr{X}}(\pi(h_1), \pi(h_2)) \leq \ell_2 + 2C$, one can calculate that, depending on k, one of these four values is always at least $\frac{1}{2}\ell_2 - 2C$. This completes the proof of the claim.
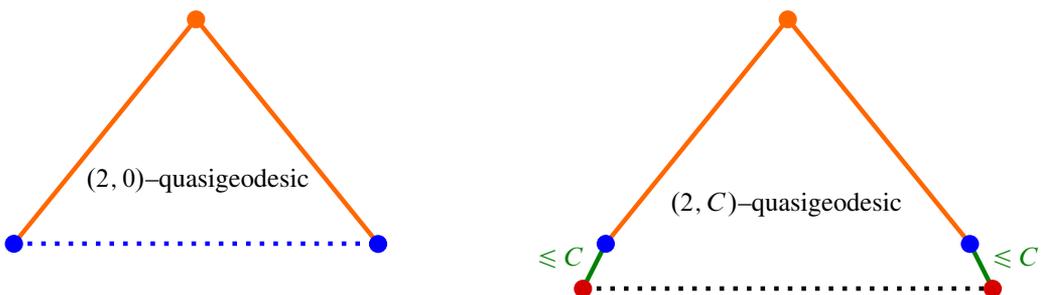
Figure 8: Left: the unit-speed parametrization of this path is a $(2, 0)$–quasigeodesic. Right: γ, which coincides with the previous path except for jumps of size at most C at the starting and endpoint, is a $(2, C)$–quasigeodesic.

Now consider the path $\gamma\colon[0, 2\ell_2]\to\mathscr{X}$

- with $\gamma(0)=\pi(h_1\langle\Delta\rangle)$,
- which for $t\in\,]0,\ell_2]$ follows a unit speed parametrization of $A(\underline{h}_1 s_1, u)$,
- which for $t\in[\ell_2, 2\ell_2[$ follows a unit speed parametrization of $A(u, \underline{h}_1 s_1 s_2)$, and
- with $\gamma(2\ell_2)=\pi(h_2\langle\Delta\rangle)$.

We see that γ is a $(2, C)$–quasigeodesic (because, apart from the jumps of size at most C at the starting and end points it is a $(2, 0)$–quasigeodesic).

Thus, the Morse condition for the axis of x implies that $\frac{1}{2}\ell_2 - 2C\leqslant M_x^{(2,C)}$, or equivalently,

$$\ell_2\leqslant 2M_x^{(2,C)}+4C.$$

Therefore,

$$d_{\mathscr{X}}(\pi(h_1), \pi(h_2))\leqslant\ell_2+2C\leqslant 2M_x^{(2,C)}+6C. \qquad\square$$

Coming back to the proof of Theorem 6.5, suppose that the action of x is not loxodromic. This means that $t_n := d_{\mathrm{AL}}(1, x^n)$ grows sublinearly with n. Consider elements $1 = h_0, h_1, \ldots, h_{t_n} = x^n$ of G and a geodesic in $\mathscr{C}_{\mathrm{AL}}(G)$ between $*$ and $x^n\langle\Delta\rangle$ through the vertices $h_i\langle\Delta\rangle$. The sublinear growth means that for sufficiently large values of n, there must be an $i\in\{1, 2, \ldots, t_n\}$ such that

$$d_{\mathscr{X}}(h_{i-1}, h_i)\geqslant\max\{F, \ell(x)\}.$$

This contradicts either Lemma 6.9 or the Lipschitz property of π (Proposition 4.7). This completes the proof that the x–action on $\mathscr{C}_{\mathrm{AL}}(G)$ is loxodromic.

We now turn to the proof that the x–action on $\mathscr{C}_{\mathrm{AL}}(G)$ is WPD. Fix $\kappa > 0$. Define

$$S_x^{(\kappa,n)} = \{h\in G/Z(G) \mid d_{\mathrm{AL}}(1, h) < \kappa \text{ and } d_{\mathrm{AL}}(x^n, hx^n) < \kappa\}.$$

We look at the situation in $\mathscr{X}$: letting $E = \max\{F, \ell(x)\}$, Lemma 6.9 and Proposition 4.7 tell us that for $h\in S_g^{(\kappa,n)}$,

$$d_{\mathscr{X}}(1, \pi(h))\leqslant E\cdot\kappa \quad\text{and}\quad d_{\mathscr{X}}(x^n, \pi(hx^n))\leqslant E\cdot\kappa.$$

We now choose N sufficiently large that $d_{\mathscr{X}}(1, x^N) > C + 2\cdot E\cdot\kappa$ — then for all integers n with $n\geqslant N$ we also have $d_{\mathscr{X}}(1, x^n) > C + 2\cdot E\cdot\kappa$, and by the triangle inequality

$$d_{\mathscr{X}}(\pi(h), \pi(hx^n)) > C.$$

The strong constriction property of π then guarantees that the geodesic $A(h, hx^n)$ passes through points $a\langle\Delta\rangle$ and $b\langle\Delta\rangle$ at distance at most C from $\pi(h\langle\Delta\rangle)$ and $\pi(hx^n\langle\Delta\rangle)$ respectively, and hence at distance at most $C + E\cdot\kappa$ from $*$ and $x^n\langle\Delta\rangle$ respectively. Therefore, $d_{\mathscr{X}}(a, b)\geqslant d_{\mathscr{X}}(1, x^n) - 2\cdot C - 2\cdot E\cdot\kappa$.

On the other hand, the geodesic $A(h, hx^n)$ has the same length as the segment of the axis $A(1, x^n)$ (as it is its image under left-translation by h). Thus,

$$d_{\mathscr{X}}(h, a) = d_{\mathscr{X}}(1, x^n) - d_{\mathscr{X}}(a, b) - d_{\mathscr{X}}(b, hx^n)\leqslant d_{\mathscr{X}}(1, x^n) - d_{\mathscr{X}}(a, b)\leqslant 2\cdot C + 2\cdot E\cdot\kappa.$$
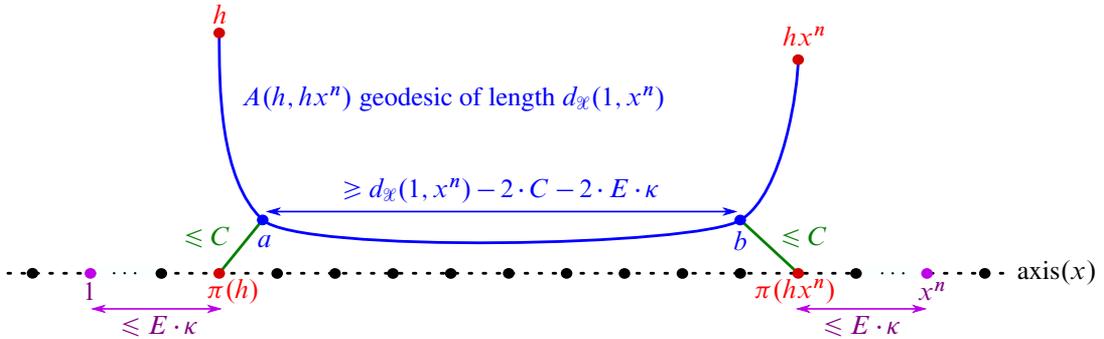
Figure 9: The proof that the action of x is WPD.

We conclude that $d_{\mathcal{X}}(1, h) \leqslant 3 \cdot (C + E \cdot \kappa)$. There are only finitely many elements $h \in G/Z(G)$ with this property. This completes the proof that the action of g is WPD. In fact we have proven something slightly stronger than what was required: we found a bound on the size of the set $S_g^{(r,n)}$ which does not depend on n, as long as $n \geqslant N$. $\qquad\square$

References

[Abbott et al. 2021] **C Abbott**, **J Behrstock**, **M G Durham**, *Largest acylindrical actions and stability in hierarchically hyperbolic groups*, Trans. Amer. Math. Soc. Ser. B 8 (2021) 66–104 MR Zbl

[Adyan 1984] **S I Adyan**, *Fragments of the word Δ in the braid group*, Mat. Zametki 36 (1984) 25–34 MR Zbl In Russian; translated in Math. Notes 36 (1984) 505–510

[Arzhantseva et al. 2015] **G N Arzhantseva**, **C H Cashen**, **J Tao**, *Growth tight actions*, Pacific J. Math. 278 (2015) 1–49 MR Zbl

[Arzhantseva et al. 2017] **G N Arzhantseva**, **C H Cashen**, **D Gruber**, **D Hume**, *Characterizations of Morse quasi-geodesics via superlinear divergence and sublinear contraction*, Doc. Math. 22 (2017) 1193–1224 MR Zbl

[Behrstock 2006] **J A Behrstock**, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. 10 (2006) 1523–1578 MR Zbl

[Bestvina 1999] **M Bestvina**, *Non-positively curved aspects of Artin groups of finite type*, Geom. Topol. 3 (1999) 269–302 MR Zbl

[Bestvina and Fujiwara 2002] **M Bestvina**, **K Fujiwara**, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. 6 (2002) 69–89 MR Zbl

[Bestvina and Fujiwara 2009] **M Bestvina**, **K Fujiwara**, *A characterization of higher rank symmetric spaces via bounded cohomology*, Geom. Funct. Anal. 19 (2009) 11–40 MR Zbl

[Birman et al. 2007] **J S Birman**, **V Gebhardt**, **J González-Meneses**, *Conjugacy in Garside groups, I: Cyclings, powers and rigidity*, Groups Geom. Dyn. 1 (2007) 221–279 MR Zbl

[Brady and McCammond 2010]   **T Brady**, **J McCammond**, *Braids, posets and orthoschemes*, Algebr. Geom. Topol. 10 (2010) 2277–2314   MR   Zbl

[Brady and Tran 2022]   **N Brady**, **H C Tran**, *Divergence of finitely presented subgroups of* CAT(0) *groups*, Pacific J. Math. 316 (2022) 1–52   MR   Zbl

[Brieskorn and Saito 1972]   **E Brieskorn**, **K Saito**, *Artin–Gruppen und Coxeter–Gruppen*, Invent. Math. 17 (1972) 245–271   MR   Zbl

[Calvez 2022]   **M Calvez**, *Euclidean Artin–Tits groups are acylindrically hyperbolic*, Groups Geom. Dyn. 16 (2022) 963–983   MR   Zbl

[Calvez and Wiest 2017a]   **M Calvez**, **B Wiest**, *Curve graphs and Garside groups*, Geom. Dedicata 188 (2017) 195–213   MR   Zbl

[Calvez and Wiest 2017b]   **M Calvez**, **B Wiest**, *Acylindrical hyperbolicity and Artin–Tits groups of spherical type*, Geom. Dedicata 191 (2017) 199–215   MR   Zbl

[Calvez and Wiest 2021]   **M Calvez**, **B Wiest**, *Hyperbolic structures for Artin–Tits groups of spherical type*, from "Geometry at the frontier: symmetries and moduli spaces of algebraic varieties" (P Comparin, E Esteves, H Lange, S Reyes-Carocca, editors), Contemp. Math. 766, Amer. Math. Soc., Providence, RI (2021) 83–98   MR   Zbl

[Cashen 2020]   **C H Cashen**, *Morse subsets of* CAT(0) *spaces are strongly contracting*, Geom. Dedicata 204 (2020) 311–314   MR   Zbl

[Charney 1992]   **R Charney**, *Artin groups of finite type are biautomatic*, Math. Ann. 292 (1992) 671–683   MR   Zbl

[Charney 1999]   **R Charney**, *Injectivity of the positive monoid for some infinite type Artin groups*, from "Geometric group theory down under" (J Cossey, W D Neumann, M Shapiro, editors), de Gruyter, Berlin (1999) 103–118   MR   Zbl

[Charney and Meier 2004]   **R Charney**, **J Meier**, *The language of geodesics for Garside groups*, Math. Z. 248 (2004) 495–509   MR   Zbl

[Charney et al. 2004]   **R Charney**, **J Meier**, **K Whittlesey**, *Bestvina's normal form complex and the homology of Garside groups*, Geom. Dedicata 105 (2004) 171–188   MR   Zbl

[Dahmani et al. 2017]   **F Dahmani**, **V Guirardel**, **D Osin**, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. 1156, Amer. Math. Soc., Providence, RI (2017)   MR   Zbl

[Dehornoy 2002]   **P Dehornoy**, *Groupes de Garside*, Ann. Sci. École Norm. Sup. 35 (2002) 267–306   MR   Zbl

[Dehornoy and Paris 1999]   **P Dehornoy**, **L Paris**, *Gaussian groups and Garside groups, two generalisations of Artin groups*, Proc. Lond. Math. Soc. 79 (1999) 569–604   MR   Zbl

[Dehornoy et al. 2015]   **P Dehornoy**, **F Digne**, **E Godelle**, **D Krammer**, **J Michel**, *Foundations of Garside theory*, EMS Tracts in Math. 22, Eur. Math. Soc., Zürich (2015)   MR   Zbl

[Deligne 1972]   **P Deligne**, *Les immeubles des groupes de tresses généralisés*, Invent. Math. 17 (1972) 273–302   MR   Zbl

[Duchin and Rafi 2009]   **M Duchin**, **K Rafi**, *Divergence of geodesics in Teichmüller space and the mapping class group*, Geom. Funct. Anal. 19 (2009) 722–742   MR   Zbl

[El-Rifai and Morton 1994]   **E A El-Rifai**, **H R Morton**, *Algorithms for positive braids*, Q. J. Math. 45 (1994) 479–497   MR   Zbl

[Epstein et al. 1992] **D B A Epstein**, **J W Cannon**, **D F Holt**, **S V F Levy**, **M S Paterson**, **W P Thurston**, *Word processing in groups*, Jones & Bartlett, Boston, MA (1992) MR Zbl

[Garside 1969] **F A Garside**, *The braid group and other groups*, Q. J. Math. 20 (1969) 235–254 MR Zbl

[Gebhardt and González-Meneses 2010] **V Gebhardt**, **J González-Meneses**, *The cyclic sliding operation in Garside groups*, Math. Z. 265 (2010) 85–114 MR Zbl

[Gebhardt and Tawn 2016] **V Gebhardt**, **S Tawn**, *Zappa–Szép products of Garside monoids*, Math. Z. 282 (2016) 341–369 MR Zbl

[Haettel et al. 2016] **T Haettel**, **D Kielak**, **P Schwer**, *The 6–strand braid group is* CAT(0), Geom. Dedicata 182 (2016) 263–286 MR Zbl

[Hamenstädt 2009] **U Hamenstädt**, *Rank-one isometries of proper* CAT(0)*–spaces*, from "Discrete groups and geometric structures" (K Dekimpe, P Igodt, A Valette, editors), Contemp. Math. 501, Amer. Math. Soc., Providence, RI (2009) 43–59 MR Zbl

[Jeong 2023] **S G Jeong**, *The seven-strand braid group is* CAT(0), Manuscripta Math. 171 (2023) 563–581 MR Zbl

[Minsky 1996] **Y N Minsky**, *Quasi-projections in Teichmüller space*, J. Reine Angew. Math. 473 (1996) 121–136 MR Zbl

[Osin 2016] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR Zbl

[Osin 2018] **D V Osin**, *Groups acting acylindrically on hyperbolic spaces*, from "Proceedings of the International Congress of Mathematicians, II" (B Sirakov, P N de Souza, M Viana, editors), World Sci., Hackensack, NJ (2018) 919–939 MR Zbl

[Picantin 2001] **M Picantin**, *The center of thin Gaussian groups*, J. Algebra 245 (2001) 92–122 MR Zbl

[Rafi and Verberne 2021] **K Rafi**, **Y Verberne**, *Geodesics in the mapping class group*, Algebr. Geom. Topol. 21 (2021) 2995–3017 MR Zbl

[Sisto 2016] **A Sisto**, *Quasi-convexity of hyperbolically embedded subgroups*, Math. Z. 283 (2016) 649–658 MR Zbl

[Sisto 2018] **A Sisto**, *Contracting elements and random walks*, J. Reine Angew. Math. 742 (2018) 79–114 MR Zbl

[Sultan 2014] **H Sultan**, *Hyperbolic quasi-geodesics in* CAT(0) *spaces*, Geom. Dedicata 169 (2014) 209–224 MR Zbl

[Tran 2019] **H C Tran**, *On strongly quasiconvex subgroups*, Geom. Topol. 23 (2019) 1173–1235 MR Zbl

IRISIB
Brussels, Belgium
IRMAR - UMR CNRS 6625, Université de Rennes
Rennes, France

calvez.matthieu@gmail.com, bertold.wiest@univ-rennes1.fr

Homotopy ribbon discs with a fixed group

ANTHONY CONWAY

In the topological category, the classification of homotopy ribbon discs is known when the fundamental group G of the exterior is $\mathbb{Z}$ and the Baumslag–Solitar group $BS(1, 2)$. We prove that if a group G is geometrically 2–dimensional and satisfies the Farrell–Jones conjecture, then a condition involving the fundamental group ensures that exteriors of aspherical homotopy ribbon discs with fundamental group G are s–cobordant rel boundary. When G is good, this leads to the classification of such discs. As an application, for any knot $J \subset S^3$ whose knot group $G(J)$ is good, we classify the homotopy ribbon discs for $J \,\#\, {-}J$ whose complement has group $G(J)$. A similar application is obtained for $BS(m, n)$ when $|m - n| = 1$.

[57K10, 57N35, 57N70, 57R67](57K10)

1 Introduction

Given a knot $K \subset S^3$, we consider the problem of classifying locally flat discs $D \subset D^4$ with boundary K, up to topological ambient isotopy rel boundary. Naturally K need not bound such a disc (ie K need not be *slice*), but if it does, then it is conjectured that it necessarily bounds one for which the inclusion induced map $\pi_1(S^3 \setminus K) \to \pi_1(D^4 \setminus D)$ is surjective; such discs are called *homotopy ribbon*. For this reason, and for technical purposes, we restrict our attention to homotopy ribbon discs with boundary K. Additionally, observe that if D_1 and D_2 are two ambiently isotopic slice discs with boundary K, then their groups must be isomorphic: $\pi_1(D^4 \setminus D_1) \cong \pi_1(D^4 \setminus D_2)$. Our goal here is to study the following question:

Question 1.1 *Given a knot $K \subset S^3$ and a ribbon group G, can one describe the set of homotopy ribbon discs for K with group G, considered up to topological ambient isotopy rel boundary?*

Here a group is called *ribbon* if it arises as $\pi_1(D^4 \setminus D)$ for some (smoothly embedded) ribbon disc $D \subset D^4$ ($D \subset D^4$ is *ribbon* if the restriction of the radial function $D^4 \to \mathbb{R}$ to D is Morse and admits no local maxima). We work with ribbon groups instead of fundamental groups of locally flat disc exteriors for convenience: the former admit an algebraic characterisation [Friedl and Teichner 2005, Theorem 2.1], while no such description appears to be known for the latter [loc. cit., Question 1.7]. Examples of ribbon groups include $G = \mathbb{Z}$ and the Baumslag–Solitar group $G = BS(1, 2)$, and in those cases Question 1.1 has been fully resolved [Friedl and Teichner 2005; Conway and Powell 2021]. The answers, which will be partially recalled in Remark 1.11, both rely on Freedman's 5–dimensional s–cobordism theorem [1982]

and therefore make use of the fact that $\mathbb{Z}$ and $\mathrm{BS}(1, 2)$ are *good* groups. We refer to [Powell and Ray 2021, Definition 12.12] for the precise definition of a good group and to [Kim et al. 2021, Chapter 19] for a survey, but note that the class of good groups contains all groups of subexponential growth as well as all elementary amenable groups (eg solvable groups). At the time of writing, it is unknown whether all groups are good; this is equivalent to the question of whether the free group F_2 is good [loc. cit., Proposition 19.7].

Remark 1.2 The only elementary amenable ribbon groups are $\mathbb{Z}$ and $\mathrm{BS}(1, 2)$, as can be seen by combining [Hillman 2002, Corollary 2.6.1] with the fact that ribbon groups have deficiency one and abelianise to $\mathbb{Z}$. As a consequence, if the class of good ribbon groups were eventually shown to coincide with the class of elementary amenable ribbon groups, then the current article would contain no new classification result. On the other hand, Theorem 1.7 contains criteria for certain disc exteriors to be s–cobordant rel boundary and holds regardless of the state of the art on the class of good groups. We also hope that the approach taken here will be of interest given the recent surge of activity around the topic of 2–discs in the 4–ball, both in the smooth and topological category [Juhász and Zemke 2020; Conway and Powell 2021; Hayden 2020; 2021; Sundberg and Swann 2022; Hayden et al. 2021; Hayden and Sundberg 2021; Lipshitz and Sarkar 2022; Dai et al. 2023].

In order to give a flavour of our results without listing technical assumptions this early on, we mention a corollary of our main theorems (Theorems 1.7 and 1.10). To state this result succinctly, we introduce some terminology. A *G–ribbon disc* refers to a homotopy ribbon disc $D \subset D^4$ with $\pi_1(D^4 \setminus D) \cong G$, and given a knot K, we write $\mathcal{D}_G(K)$ for the set of rel boundary topological ambient isotopy classes of G–ribbon discs with boundary K. We also write M_K for the result of 0–surgery on K and use $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)$ to denote the set of epimorphisms $\pi_1(M_K) \twoheadrightarrow G$ that satisfy (FT) below. While this definition will be discussed in greater detail in the next couple of sections, for the moment we simply note that $\mathrm{Aut}(G)$ acts on $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)$ by postcomposition, allowing us to consider the orbit set $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$. Mapping a G–ribbon disc $D \in \mathcal{D}_G(K)$ with aspherical complement to the inclusion induced homomorphism $\pi_1(M_K) \twoheadrightarrow \pi_1(D^4 \setminus D)$ determines an element $\Phi(D)$ in this orbit set.

Theorem *Fix a knot $K \subset S^3$.*

(i) *If G is a knot group (ie $G = \pi_1(S^3 \setminus J)$ for some knot J), then exteriors of G–ribbon discs $D_1, D_2 \in \mathcal{D}_G(K)$ are s–cobordant rel boundary if $\Phi(D_1) = \Phi(D_2)$. If G is good, then Φ induces a bijection $\mathcal{D}_G(K) \approx \mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$.*

(ii) *If $m, n \in \mathbb{Z}$ are such that $|m - n| = 1$ and G is the Baumslag–Solitar group $\mathrm{BS}(m, n) = \langle a, b \mid ab^m a^{-1} = b^n \rangle$, then exteriors of aspherical G–ribbon discs $D_1, D_2 \in \mathcal{D}_G(K)$ are s–cobordant rel boundary if $\Phi(D_1) = \Phi(D_2)$. If G is good, then Φ induces a bijection $\mathcal{D}_G^a(K) \approx \mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$, where $\mathcal{D}_G^a(K) \subset \mathcal{D}_G(K)$ denotes the subset of G–ribbon discs with aspherical exterior.*

Examples 1.12 and 1.13 describe how this follows from Theorems 1.7 and 1.10. Additionally, as we explain in more detail in Remark 1.11, this theorem recovers the previously known classifications for $BS(0, 1) = \mathbb{Z}$ and $BS(1, 2)$ since, for these groups, homotopy ribbon disc exteriors are known to be aspherical.

1.1 Existence

We recall and motivate a sufficient condition for the existence of a G–ribbon disc with boundary K, which is due to Friedl and Teichner [2005, Theorem 1.9]. First, if K bounds a locally flat disc $D \subset D^4$, then $\partial N_D = M_K$, where $N_D := D^4 \setminus \nu D$ is the exterior of D and M_K denotes the 3–manifold obtained by 0–framed surgery on K. Next, if $D \subset D^4$ is a G–ribbon disc for a knot K, then there is an epimorphism $\pi_1(M_K) \twoheadrightarrow \pi_1(N_D) \cong G$ and (N_D, M_K) satisfies Poincaré duality or, using surgery theory jargon, is a (4–dimensional) *Poincaré pair*. If, additionally, the disc exterior $N_D = D^4 \setminus \nu D$ is aspherical, then we have a homotopy equivalence $N_D \simeq K(G, 1)$ and we deduce that $(K(G, 1), M_K)$ is a Poincaré pair.

Remark 1.3 It is expected that ribbon disc exteriors are aspherical [Gordon 1981, Conjecture 6.5]; see also [Howie 1985]. As noted in [Friedl and Teichner 2005, Section 2], this would imply the *ribbon group conjecture*: ribbon groups are geometrically 2–dimensional.[1] Here recall that a group G is called *geometrically 2–dimensional* if $K(G, 1)$ is (homotopy equivalent to) a 2–complex. Both statements are in fact particular cases of the *Whitehead conjecture*, which states that every connected subcomplex of a 2–dimensional aspherical CW–complex is itself aspherical [Whitehead 1941]; see [Rosebrock 2007] for a nice overview. Howie [1982, Theorem 5.2] proved that locally indicable ribbon groups are geometrically 2–dimensional. On the other hand, to the best of our knowledge, the Whitehead conjecture is not known to imply that exteriors of homotopy ribbon discs are aspherical; see also Remark 1.11.

We argued that if D is a G–ribbon disc with aspherical exterior and boundary a knot K, then $\pi_1(M_K) \twoheadrightarrow \pi_1(N_D) \cong G$ is an epimorphism and $(K(G, 1), M_K)$ is a Poincaré pair. On the other hand, if we start with an epimorphism $\pi_1(M_K) \twoheadrightarrow G$ onto a group G, then there is an embedding $\varphi \colon M_K \hookrightarrow K(G, 1) = BG$ that induces the given surjection on fundamental groups and, if G is geometrically 2–dimensional, then [Friedl and Teichner 2005, Lemma 3.2] shows that $(K(G, 1), M_K)$ is a Poincaré pair if and only if the induced map satisfies

(FT) $\qquad\qquad \varphi^* \colon H^i(BG; \mathbb{Z}[G]) \to H^i(M_K; \mathbb{Z}[G]_\varphi)$ is an isomorphism for $i = 1, 2$.

Under an additional condition on the group G, Friedl and Teichner [2005, Theorem 1.9 and Lemma 3.2] prove that this leads to a sufficient condition for K to bound a G–ribbon disc.

Theorem 1.4 (Friedl and Teichner) *Let $K \subset S^3$ be a knot and let G be a good geometrically 2–dimensional ribbon group such that $\widetilde{L}_4^h(\mathbb{Z}[G]) = 0$. If $\varphi \colon \pi_1(M_K) \twoheadrightarrow G$ is an epimorphism that satisfies (FT), then there exists a G–ribbon disc $D \subset D^4$ with aspherical exterior and boundary K such that the composition $\pi_1(M_K) \twoheadrightarrow \pi_1(N_D) \cong G$ agrees with φ.*

---

[1] Friedl and Teichner refer to geometrically 2–dimensional groups as *aspherical* groups.

Remark 1.5 • Friedl and Teichner actually prove a stronger result. Instead of asking for G to be geometrically 2–dimensional, they merely demand that $H_3(G) = 0$ and $H^i(G; \mathbb{Z}[G]) = 0$ for $i > 2$, and instead of assuming that G is ribbon, they only require that G be finitely presented and satisfy $H_1(G) = \mathbb{Z}$ and $H_2(G) = 0$. Finally, they do not require G to be good, only that the surgery sequence (with h–decorations) be exact for all 4–dimensional Poincaré pairs (X, M) with $\pi_1(X) = G$.

• The fact that the disc exterior is aspherical is implicit in [loc. cit., proof of Theorem 1.9]: their surgery-theoretic argument yields a disc D whose exterior $N_D = D^4 \setminus \nu D$ is homotopy equivalent to $K(G, 1)$, which is aspherical.

• The groups $\mathbb{Z}$ and $\mathrm{BS}(1, 2)$ satisfy all the assumptions of Theorem 1.4. Additionally, for those groups, condition (FT) simplifies considerably. Indeed if G is poly-(torsion-free abelian) (or PTFA for short), then (FT) reduces to

(Ext) $$\mathrm{Ext}^1_{\mathbb{Z}[G]}(H_1(M_K; \mathbb{Z}[G]_\varphi), \mathbb{Z}[G]) = 0,$$

and for $G = \mathbb{Z}$ it reduces further to the condition $\Delta_K = 1$; all of this is explained in [loc. cit., Sections 1 and 4, and Lemma 3.3].

1.2 Uniqueness and classification

We now return to the set $\mathcal{D}_G(K)$ of rel boundary topological ambient isotopy classes of G–ribbon discs with boundary K. In fact, we will mostly be concerned with the subset $\mathcal{D}_G^a(K) \subset \mathcal{D}_G(K)$ of discs with aspherical exteriors. To that effect, inspired by [Hambleton et al. 2009, Definition 1.2], we describe some assumptions on the group G that we will require:

Definition 1.6 A group G *satisfies properties* W–AA if

(W) the Whitehead group $\mathrm{Wh}(G)$ vanishes,

(A4) the assembly map $A_4 \colon H_4(BG; \boldsymbol{L}_\bullet) \to L_4(\mathbb{Z}[G])$ is an isomorphism,[2] and

(A5) the assembly map $A_5 \colon H_5(BG; \boldsymbol{L}_\bullet) \to L_5(\mathbb{Z}[G])$ is surjective.

We will mostly use these conditions as a blackbox, but note that thanks to extensive work on the Farrell–Jones conjecture (see [Lück 2021] for a survey) they should not be thought of as insurmountable restrictions. We discuss all of this in more detail in Remark 1.11 and refer to [Ranicki 1992; Chang and Weinberger 2021; Lück 2020; 2021] for background on assembly maps in L–theory. Returning to our aim of describing $\mathcal{D}_G(K)$, we consider the set

(Epi) $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G) := \{\varphi \colon \pi_1(M_K) \to G \mid \varphi \text{ is an epimorphism that satisfies (FT)}\},$

---

[2]In the work of Hambleton, Kreck and Teichner [Hambleton et al. 2009] W–AA only requires A_4 to be injective.

and observe that it is acted upon (by postcomposition) by the group $\mathrm{Aut}(G)$ of automorphisms of G. Thanks to the discussion leading up to Theorem 1.4, note that sending a G–ribbon disc with aspherical exterior to an epimorphism $\pi_1(M_K) \twoheadrightarrow \pi_1(N_D) \cong G$ defines a map

$$\Phi: \mathcal{D}_G^a(K) \to \mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$$

which does not depend on the choice of the isomorphism $\pi_1(N_D) \cong G$. If G is a good geometrically 2–dimensional ribbon group such that $\widetilde{L}_4(\mathbb{Z}[G]) = 0$, then Theorem 1.4 ensures that Φ is surjective. Our main technical result gives conditions on G for Φ to be injective and, in the absence of the goodness condition on G, for exteriors of G–ribbon discs to be s–cobordant rel boundary.

Theorem 1.7 *Let K be a knot and let G be a geometrically 2–dimensional group that satisfies* (W) *and* (A5). *If D_1 and D_2 are two G–ribbon discs with aspherical exteriors and boundary K such that $\Phi(D_1) = \Phi(D_2)$, then the disc exteriors N_{D_1} and N_{D_2} are s–cobordant rel boundary.*

If in addition to these conditions the group G is good, then the discs D_1 and D_2 are ambiently isotopic rel boundary.

We note that this result can alternatively be stated with normal subgroups instead of epimorphisms, as this is easier to verify in practice. To state this concisely, given a slice disc D for a knot K, we use $\iota_D: \pi_1(M_K) \to \pi_1(N_D)$ to denote the inclusion-induced map.

Corollary 1.8 *Let K be a knot and let G be a geometrically 2–dimensional group that satisfies* (W) *and* (A5). *If D_1 and D_2 are two G–ribbon discs with aspherical exteriors and boundary K such that $\ker(\iota_{D_1}) = \ker(\iota_{D_2})$, then the disc exteriors N_{D_1} and N_{D_2} are s–cobordant rel boundary.*

If in addition to these conditions the group G is good, then the discs D_1 and D_2 are ambiently isotopic rel boundary.

For smoothly embedded discs, the hypotheses of these results can be relaxed:

Remark 1.9 *If D_1 and D_2 are ribbon discs with aspherical exteriors and $\pi_1(N_{D_i}) \cong G$ for $i = 1, 2$, then the assumption that G be geometrically 2–dimensional can be omitted in both Theorem 1.7 and Corollary 1.8: in this case $K(G, 1) \simeq N_{D_i}$ has the homotopy type of a 2–complex.*

Combining Theorems 1.4 and 1.7 we obtain an answer to Question 1.1, provided we make some restrictions on the ribbon group G and require the ribbon disc exteriors to be aspherical.

Theorem 1.10 *Let $K \subset S^3$ be a knot and let G be a geometrically 2–dimensional good ribbon group that satisfies properties W–AA. Mapping a G–ribbon disc D to the epimorphism $\pi_1(M_K) \twoheadrightarrow \pi_1(N_D) \cong G$ defines a bijection Φ between*

  (i) *the set $\mathcal{D}_G^a(K)$ of G–ribbon discs with aspherical exterior and boundary K, considered up to ambient isotopy rel boundary, and*

  (ii) *the set $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$ defined in* (Epi).

Proof We argue in Remark 2.1 that since G is a geometrically 2–dimensional ribbon group with $\mathrm{Wh}(G) = 0$, requiring G to satisfy condition (A4) is equivalent to asking for $\widetilde{L}_4(\mathbb{Z}[G]) = 0$. Thus the hypotheses of Theorem 1.4 are satisfied and so Φ is surjective. The injectivity of Φ follows from Theorem 1.7, which we can apply since G satisfies properties W–AA. ☐

Remark 1.11 • If the ribbon group conjecture (or more optimistically the Whitehead conjecture) were true, then requiring G to be geometrically 2–dimensional would be superfluous; recall Remark 1.3. It is also tempting to conjecture that exteriors of G–ribbon discs are aspherical, and in this case we would have $\mathcal{D}_G^a(K) = \mathcal{D}_G(K)$. This latter conjecture holds when G is PTFA [Conway and Powell 2021, Lemma 2.1] (eg when $G = \mathbb{Z}$ and $G = \mathrm{BS}(1, 2)$) and is a consequence of the Whitehead conjecture if the disc exterior is homotopy equivalent to a 2–complex.

• The groups $\mathbb{Z}$ and $\mathrm{BS}(1, 2)$ satisfy the hypotheses of Theorem 1.10, and in this case unpacking the definition of $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$ recovers [loc. cit., Theorems 1.5 and 1.6]. Instead of repeating those statements, we note that for $G = \mathbb{Z}$, $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$ has at most one element, while for $G = \mathrm{BS}(1, 2)$ it has at most two [loc. cit., Section 4]. Estimating the cardinality of this set in general appears to be more challenging. Naturally, the set $\mathcal{D}_G(K)$ is often empty: for example, we refer to [Friedl and Teichner 2005, Corollary 3.4] for an obstruction (based on the Alexander polynomial) to a knot K bounding a G–ribbon disc.

• As we alluded to in Corollary 1.8, the classification result of Theorem 1.10 can be stated in terms of normal subgroups of $\pi_1(M_K)$ instead of epimorphisms originating from $\pi_1(M_K)$: to a G–ribbon disc D, one associates the normal subgroup $\ker(\pi_1(M_K) \twoheadrightarrow \pi_1(N_D))$ of $\pi_1(M_K)$. This was the perspective taken in [Conway and Powell 2021] where, using that $\mathrm{BS}(1, 2)$ is metabelian, the results were then formulated using submodules of the Alexander module $H_1(M_K; \mathbb{Z}[t^{\pm 1}])$; the details are in [loc. cit., Section 3].

• The requirement that the group be good is hard to verify in practice. On the other hand, G satisfies properties W–AA if it is geometrically 2–dimensional and satisfies the Farrell–Jones conjecture: if a group G is geometrically 2–dimensional, then $K(G, 1)$ is a 2–complex and the claim now follows as in [Kasprowski and Land 2022, Lemma 2.3]. (The core of the argument will be recalled both in the proof of Theorem 1.7 and in Remark 2.1.) We treat the Farrell–Jones conjecture as a blackbox, but refer the interested reader to [Lück 2021] for a survey and to [Lück 2021, Chapter 15] for a list of groups for which the conjecture is known to hold.

Example 1.12 We argue that the group $G(J) = \pi_1(S^3 \setminus J)$ of a classical knot $J \subset S^3$ is a geometrically 2–dimensional ribbon group that satisfies properties W–AA. Thus Theorem 1.7 provides a criterion for exteriors of $G(J)$–ribbon discs to be s–cobordant rel boundary and, if $G(J)$ is additionally assumed to be good, then Theorem 1.10 classifies $G(J)$–ribbon discs for $J \# -J$.

The group of $J \subset S^3$ is ribbon (the ribbon knot $J \# -J$ bounds a smoothly embedded ribbon disc with group $G(J)$ as explained in [Friedl and Teichner 2005, page 2135]). The sphere theorem ensures that
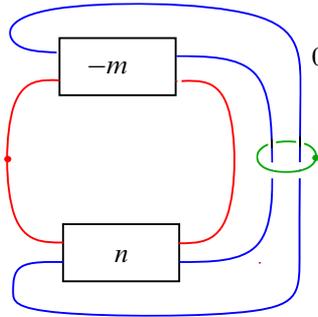
Figure 1: Assuming that $|m-n| = 1$, this figure depicts a handle diagram of a ribbon disc exterior with fundamental for $BS(m, n)$. Indeed, since $|m - n| = 1$, the red and blue knots form a handle diagram for D^4 in which the green knot is sliced by a ribbon disc D with $\pi_1(N_D) = BS(m, n)$.

$G(J)$ is geometrically 2–dimensional (the knot exterior is aspherical and has the homotopy type of a 2–complex; see eg [Lickorish 1997, Theorem 11.7]). The Farrell–Jones conjecture holds for $G(J)$ because it holds for the fundamental group of any 3–manifold with boundary [Lück 2021, Theorem 15.1(e)].

Since knot groups are PTFA by work of Strebel [1974], $G(J)$–ribbon discs are aspherical by [Conway and Powell 2021, Lemma 2.1] and thus $\mathcal{D}_{G(J)}(J \# -J) = \mathcal{D}_{G(J)}^a(J \# -J)$. Finally, as we noted in Remark 1.5, since $G(J)$ is PTFA we can use condition (Ext) instead of condition (FT).

Example 1.13 We argue that for $m, n \in \mathbb{Z}$ with $|m - n| = 1$, the Baumslag–Solitar group $BS(m, n)$ is a geometrically 2–dimensional ribbon group that satisfies properties W–AA. Thus Theorem 1.7 provides a criterion for exteriors of aspherical $BS(m, n)$–ribbon discs to be s–cobordant rel boundary and, if $BS(m, n)$ is additionally assumed to be good, then Theorem 1.10 classifies $BS(m, n)$–ribbon discs with aspherical exteriors.

The fact that $BS(m, n)$ is ribbon when $|m - n| = 1$ can be seen by looking at the handle diagram depicted in Figure 1. Baumslag–Solitar groups are geometrically 2–dimensional: the universal cover of the presentation 2–complex for $\langle a, b \mid ba^m b^{-1} = b^n \rangle$ is homeomorphic to the product of $\mathbb{R}$ with a tree; see eg [Freden et al. 2011, Section 2]. Additionally, every Baumslag–Solitar group $BS(m, n)$ satisfies the Farrell–Jones conjecture [Farrell and Wu 2015; Gandini et al. 2015].

We conclude with a brief final remark concerning asphericity. Our methods rely heavily on G–ribbon disc exteriors (conjecturally) being aspherical. Currently, nonaspherical 4–manifolds with boundary M_K and fundamental group G are poorly understood beyond the group $G = \mathbb{Z}$ [Conway and Powell 2023]. This is the reason why we only work in D^4 instead of in other 4–manifolds.

Conventions Throughout this article, we work in the topological category. Manifolds are assumed to be compact and oriented. Homeomorphisms, homotopy equivalences and isotopies are *rel boundary* if they fix the boundary pointwise. If M_1 and M_2 are two n–manifolds with boundary Y, a cobordism between M_1 and M_2 is *relative* Y if, when restricted to Y, it is the product $Y \times [0, 1]$.

2 Proof of the main technical result

We recall the statement of Theorem 1.7 and prove it. Let K be a knot and let G be a geometrically 2–dimensional group that satisfies (W) and (A5). The aim is to prove that if D_1 and D_2 are two G–ribbon discs with aspherical exteriors and boundary K such that $\Phi(D_1) = \Phi(D_2) \in \mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$, then the disc exteriors N_{D_1} and N_{D_2} are s–cobordant rel boundary and, if G is additionally assumed to be good, then D_1 and D_2 are ambiently isotopic rel boundary.

Proof of Theorem 1.7 Assume that D_1 and D_2 are two G–ribbon discs with aspherical exteriors and boundary K and that their epimorphisms agree in $\mathrm{Epi}^{\mathrm{FT}}(\pi_1(M_K), G)/\mathrm{Aut}(G)$. We must show that the exteriors N_{D_1} and N_{D_2} are s–cobordant rel boundary. If we additionally assume that G is good, then Freedman's 5–dimensional relative s–cobordism theorem will then ensure that N_{D_1} and N_{D_2} are in fact homeomorphic rel boundary. That D_1 and D_2 are ambiently isotopic rel boundary follows by applying Alexander's trick, as noted in [Conway and Powell 2021, Lemma 2.5]. Our strategy decomposes into two steps. The first uses the conditions on the epimorphisms to show that id_{M_K} extends to a homotopy equivalence $N_{D_1} \simeq N_{D_2}$. The second uses surgery theory to improve this homotopy equivalence to an s–cobordism rel boundary; here is where we rely on properties (W) and (A5) as well as on the fact that G is good.

We start with the first step. Since the epimorphisms of D_1 and D_2 agree, there exists an automorphism Ψ of G that makes the following diagram commute:

$$\begin{array}{ccc} \pi_1(M_K) & \xrightarrow{\;=\;} & \pi_1(M_K) \\ {\scriptstyle \iota_{D_1}}\big\downarrow & & \big\downarrow{\scriptstyle \iota_{D_2}} \\ \pi_1(N_{D_1}) & & \pi_1(N_{D_2}) \\ {\scriptstyle \cong}\big\downarrow & & \big\downarrow{\scriptstyle \cong} \\ G & \xrightarrow{\;\Psi,\cong\;} & G. \end{array}$$

Since the bottom vertical maps in this diagram are isomorphisms, we deduce that there exists an isomorphism $g \colon \pi_1(N_{D_1}) \cong \pi_1(N_{D_2})$ such that $g \circ \iota_{D_1} = \iota_{D_2}$; such isomorphisms were called *compatible* in [loc. cit., Section 2]. As the D_i have aspherical exteriors, the obstruction theory argument from [loc. cit., end of proof of Lemma 2.1] shows that the identity $\mathrm{id}_{M_K} \colon M_K \to M_K$ extends to a homotopy equivalence $f \colon N_{D_1} \to N_{D_2}$ which induces g on fundamental groups.

We now move on to the second step: we use surgery theory to improve the homotopy equivalence f to an s–cobordism $N_{D_1} \cong_{s\text{–cob}} N_{D_2}$ rel boundary. We describe the argument very briefly for readers that are familiar with surgery theory before giving some more details. Consider the surgery sequence, where we can ignore decorations thanks to condition (W),

$$\mathcal{N}(N_{D_2} \times [0,1], \partial(N_{D_2} \times [0,1])) \xrightarrow{\sigma_5}{}_5 (\mathbb{Z}[G]) \to \mathcal{S}(N_{D_2}, \partial N_{D_2}) \xrightarrow{\eta} N(N_{D_2}, \partial N_{D_2}) \xrightarrow{\sigma_4}{}_4 (\mathbb{Z}[G]).$$

We use that disc exteriors have trivial H_2 to deduce that η is the zero map. More concretely, we obtain a degree-1 normal map

(1) $$(F', f, \mathrm{id}_{N_{D_2}}) \colon (W', N_{D_1}, N_{D_2}) \to (N_{D_2} \times [0,1], N_{D_2}, N_{D_2})$$

that we can assume to be 2–connected by surgery below the middle dimension. We then use property (A5) and the fact that G is geometrically 2–dimensional to deduce that σ_5 is surjective. We infer that N_{D_1} and N_{D_2} are s–cobordant either by appealing to the exactness of the surgery sequence (which requires G to be good) or by using the surjectivity of σ_5 to replace F' by another degree-1 normal map with vanishing surgery obstruction (despite being slightly longer, this argument has the advantage of not requiring G to be good). Thus the fact that N_{D_1} and N_{D_2} are s–cobordant rel boundary can be proved without using that G is good. The homeomorphism classification result then follows from Freedman's 5–dimensional relative s–cobordism theorem, which we can apply if G is good.

We give more details. The set $\mathcal{N}(N_{D_2}, \partial N_{D_2})$ consists of equivalences classes of degree-1 normal maps $M \to N_{D_2}$ that restrict to a homeomorphism on the boundary. Two such degree-1 normal maps $f_i \colon M_i \to N_{D_2}$ for $i = 1, 2$ are equivalent if there exists a rel boundary cobordism (W, M_1, M_2) and a degree-1 normal map

$$(W, M_1, M_2) \to (N_{D_2} \times [0,1], N_{D_2}, N_{D_2})$$

that restricts to f_i on M_i for $i = 1, 2$. A homotopy equivalence $h \colon M \to N_{D_2}$ rel boundary is in particular a degree-1 normal map that we denote by $\eta(h) \in \mathcal{N}(N_{D_2}, \partial N_{D_2})$.

We claim that η is the zero map. Under the isomorphism

(2) $$\mathcal{N}(N_{D_2}, \partial N_{D_2}) \cong H^4(N_{D_2}, \partial N_{D_2}) \oplus H^2(N_{D_2}, \partial N_{D_2}; \mathbb{Z}_2) = H^4(N_{D_2}, \partial N_{D_2}) \cong \mathbb{Z}$$

we have $\eta(h) = \frac{1}{8}(\sigma(M) - \sigma(N_{D_2}))$; this fact is well known to surgeons, but we refer to [Conway and Powell 2021, Proposition 2.2] in case the reader is curious about the details. Since the signature of a disc exterior vanishes and h is a homotopy equivalence, we deduce that $\eta(h) = 0$, as claimed.

We assert that the map σ_5 from the surgery sequence is surjective. This relies on surgery spectra and the algebraic theory of surgery. We treat this largely as a blackbox, but note that this part of surgery theory was developed by Quinn [1970; 1971] and Ranicki [1979; 1981]; we also refer to [Chang and Weinberger 2021, Section 4.4] for a nice overview of these topics and to [Cencelj et al. 2009, Section 4] for a helpful account of the rel boundary case. Using the relation between the assembly map and the

surgery obstruction (as mentioned for example in [Chang and Weinberger 2021, page 158]) and the fact that N_{D_2} is a $K(G, 1)$, the following diagram commutes:

$$
\begin{array}{ccc}
\mathcal{N}(N_{D_2} \times [0,1], \partial(N_{D_2} \times [0,1])) & \xrightarrow{\sigma_5} & L_5(\mathbb{Z}[G]) \\
\downarrow{\cong} & & \downarrow{=} \\
H_5(N_{D_2}; \boldsymbol{L}\langle 1\rangle_\bullet) \xrightarrow{\cong} H_5(N_{D_2}; \boldsymbol{L}_\bullet) \longrightarrow & & L_5(\mathbb{Z}[G]) \\
\downarrow{\cong} & \downarrow{\cong} & \downarrow{=} \\
H_5(\mathrm{B}G; \boldsymbol{L}\langle 1\rangle_\bullet) \xrightarrow{\cong} H_5(\mathrm{B}G; \boldsymbol{L}_\bullet) \xrightarrow{A_5} & & L_5(\mathbb{Z}[G]).
\end{array}
$$

Here $\boldsymbol{L}_\bullet$ denotes the L–theory spectrum of the integers and $\boldsymbol{L}\langle 1\rangle_\bullet$ denotes its 1–connective cover. The fact that $H_5(\mathrm{B}G; \boldsymbol{L}\langle 1\rangle_\bullet) \to H_5(\mathrm{B}G; \boldsymbol{L}_\bullet)$ is an isomorphism follows because $K(G, 1)$ admits a 2–dimensional CW–model (the Atiyah–Hirzebruch spectral sequence argument is the same as in [Kasprowski and Land 2022, proof of Lemma 2.3]), and the fact that the top left vertical map is an isomorphism is a fact from algebraic surgery theory; see eg [Cencelj et al. 2009, (27)]. Using this commutative diagram and property (A5) (which stipulates that the assembly map A_5 is surjective), one deduces that σ_5 is surjective.

There are now two closely related ways to conclude that N_{D_1} and N_{D_2} are s–cobordant rel boundary. The first way is shorter, but uses that the group G is good: since η is the zero map, σ_5 is surjective and the surgery sequence is exact (because G is good), the structure set $\mathcal{S}(N_{D_2}, \partial N_{D_2})$ — to which f belongs — is trivial. The second argument (inspired by [Kasprowski and Land 2022]) is slightly longer but does not require that the group G be good: Since $\eta \equiv 0$, there is a rel boundary cobordism (W, N_{D_1}, N_{D_2}) and a degree-1 normal map

$$
(F, f, \mathrm{id}_{N_{D_2}}) : (W, N_{D_1}, N_{D_2}) \to (N_{D_2} \times [0,1], N_{D_2}, N_{D_2}).
$$

Perform surgery below the middle dimension on the interior of W to obtain the 2–connected degree-1 normal map F' with surgery obstruction $x := \sigma(F') \in L_5(\mathbb{Z}[G])$ that we alluded to in (1). Using the surjectivity of σ_5, one can find a degree-1 normal map

$$
\Psi : (V, N_{D_2}, N_{D_2}) \to (N_{D_2} \times [0,1], N_{D_2}, N_{D_2})
$$

that restricts to the identity on both boundary components and with $-x$ as its surgery obstruction; stacking Ψ on top of F' leads to a degree-1 normal map F'' with vanishing surgery obstruction $\sigma(F'') \in L_5(\mathbb{Z}[G])$, and it follows that F'' is normal bordant rel $M_K \times [0,1]$ to a homotopy equivalence. Thus, we have two arguments for why N_{D_1} and N_{D_2} are s–cobordant rel boundary.

If G is good, we can apply Freedman's 5–dimensional relative s–cobordism theorem [Freedman and Quinn 1990, Theorem 7.1A] and it follows that N_{D_1} and N_{D_2} are homeomorphic rel boundary. As we already mentioned, [Conway and Powell 2021, Lemma 2.5] implies that the discs are ambiently isotopic rel boundary. $\square$

We conclude by proving a statement that was used in the proof of Theorem 1.10:

Remark 2.1 Assume that G is a geometrically 2–dimensional ribbon group with vanishing Whitehead torsion (condition (W)). We claim that G satisfies $\widetilde{L}_4(\mathbb{Z}[G]) = 0$ if and only if it satisfies (A4), which stipulates that the assembly map $A_4 \colon H_4(BG; \boldsymbol{L}_\bullet) \to L_4(\mathbb{Z}[G])$ is an isomorphism. Since G is a ribbon group, there is a (smoothly embedded) ribbon disc $D \subset D^4$ with $\pi_1(N_D) \cong G$. This time N_D might not be aspherical, but it is still a 2–complex with vanishing H_2. An Atiyah–Hirzebruch spectral sequence argument therefore shows that $H_4(N_D; \boldsymbol{L}\langle 1\rangle_\bullet) \to H_4(BG; \boldsymbol{L}\langle 1\rangle_\bullet)$ is an isomorphism. Here it is helpful to note that $H_2(G) = 0$: use $H_2(N_D) = 0$ together with the exact sequence $\pi_2(N_D) \to H_2(N_D) \to H_2(\pi_1(N_D)) \to 0$; see eg [Brown 1982, (0.1)]. The same argument as above then produces the following commutative diagram:

$$
\begin{array}{ccc}
\mathcal{N}(N_D, M_K) & \xrightarrow{\quad \sigma_4 \quad} & L_4(\mathbb{Z}[G]) \\
\downarrow{\scriptstyle\cong} & & \downarrow{\scriptstyle=} \\
H_4(N_D; \boldsymbol{L}\langle 1\rangle_\bullet) \xrightarrow{\ \cong\ } H_4(N_D; \boldsymbol{L}_\bullet) & \longrightarrow & L_4(\mathbb{Z}[G]) \\
\downarrow{\scriptstyle\cong} \qquad\qquad \downarrow{\scriptstyle\cong} & & \downarrow{\scriptstyle=} \\
H_4(BG; \boldsymbol{L}\langle 1\rangle_\bullet) \xrightarrow{\ \cong\ } H_4(BG; \boldsymbol{L}_\bullet) & \xrightarrow{\ A_4\ } & L_4(\mathbb{Z}[G]).
\end{array}
$$

As explained in (2) and [Freedman and Quinn 1990, Section 11.3B], the surgery obstruction σ_4 maps the set of normal invariants $\mathcal{N}(N_D, \partial N_D) \cong \mathbb{Z}$ isomorphically onto the $L_4(\mathbb{Z}) \cong \mathbb{Z}$–summand of $L_4(\mathbb{Z}[G]) = L_4(\mathbb{Z}) \oplus \widetilde{L}_4(\mathbb{Z}[G])$. The claim now follows by combining this fact with the commutativity of the diagram.

References

[Brown 1982] **K S Brown**, *Cohomology of groups*, Graduate Texts in Math. 87, Springer (1982) MR Zbl

[Cencelj et al. 2009] **M Cencelj**, **Y V Muranov**, **D Repovš**, *On structure sets of manifold pairs*, Homology Homotopy Appl. 11 (2009) 195–222 MR Zbl

[Chang and Weinberger 2021] **S Chang**, **S Weinberger**, *A course on surgery theory*, Ann. of Math. Stud. 211, Princeton Univ. Press (2021) MR Zbl

[Conway and Powell 2021] **A Conway**, **M Powell**, *Characterisation of homotopy ribbon discs*, Adv. Math. 391 (2021) art. id. 107960 MR Zbl

[Conway and Powell 2023] **A Conway**, **M Powell**, *Embedded surfaces with infinite cyclic knot group*, Geom. Topol. 27 (2023) 739–821 MR Zbl

[Dai et al. 2023] **I Dai**, **A Mallick**, **M Stoffregen**, *Equivariant knots and knot Floer homology*, J. Topol. 16 (2023) 1167–1236 MR Zbl

[Farrell and Wu 2015] **F T Farrell**, **X Wu**, *Isomorphism conjecture for Baumslag–Solitar groups*, Proc. Amer. Math. Soc. 143 (2015) 3401–3406 MR Zbl

[Freden et al. 2011] **E M Freden**, **T Knudson**, **J Schofield**, *Growth in Baumslag–Solitar groups, I: Subgroups and rationality*, LMS J. Comput. Math. 14 (2011) 34–71 MR Zbl

[Freedman 1982] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR Zbl

[Freedman and Quinn 1990] **M H Freedman**, **F Quinn**, *Topology of 4–manifolds*, Princeton Math. Ser. 39, Princeton Univ. Press (1990) MR Zbl

[Friedl and Teichner 2005] **S Friedl**, **P Teichner**, *New topologically slice knots*, Geom. Topol. 9 (2005) 2129–2158 MR Zbl

[Gandini et al. 2015] **G Gandini**, **S Meinert**, **H Rüping**, *The Farrell–Jones conjecture for fundamental groups of graphs of abelian groups*, Groups Geom. Dyn. 9 (2015) 783–792 MR Zbl

[Gordon 1981] **C M Gordon**, *Ribbon concordance of knots in the 3–sphere*, Math. Ann. 257 (1981) 157–170 MR Zbl

[Hambleton et al. 2009] **I Hambleton**, **M Kreck**, **P Teichner**, *Topological 4–manifolds with geometrically two-dimensional fundamental groups*, J. Topol. Anal. 1 (2009) 123–151 MR Zbl

[Hayden 2020] **K Hayden**, *Exotically knotted disks and complex curves*, preprint (2020) arXiv 2003.13681

[Hayden 2021] **K Hayden**, *Corks, covers, and complex curves*, preprint (2021) arXiv 2107.06856

[Hayden and Sundberg 2021] **K Hayden**, **I Sundberg**, *Khovanov homology and exotic surfaces in the 4–ball*, preprint (2021) arXiv 2108.04810

[Hayden et al. 2021] **K Hayden**, **A Kjuchukova**, **S Krishna**, **M Miller**, **M Powell**, **N Sunukjian**, *Brunnian exotic surface links in the 4–ball* (2021) arXiv 2106.13776 To appear in Michigan Math. J.

[Hillman 2002] **J A Hillman**, *Four-manifolds, geometries and knots*, Geom. Topol. Monogr. 5, Geom. Topol. Publ., Coventry (2002) MR Zbl

[Howie 1982] **J Howie**, *On locally indicable groups*, Math. Z. 180 (1982) 445–461 MR Zbl

[Howie 1985] **J Howie**, *On the asphericity of ribbon disc complements*, Trans. Amer. Math. Soc. 289 (1985) 281–302 MR Zbl

[Juhász and Zemke 2020] **A Juhász**, **I Zemke**, *Distinguishing slice disks using knot Floer homology*, Selecta Math. 26 (2020) art. id. 5 MR Zbl

[Kasprowski and Land 2022] **D Kasprowski**, **M Land**, *Topological 4–manifolds with 4–dimensional fundamental group*, Glasg. Math. J. 64 (2022) 454–461 MR Zbl

[Kim et al. 2021] **M H Kim**, **P Orson**, **J Park**, **A Ray**, *Good groups*, from "The disc embedding theorem" (S Behrens, B Kalmár, M H Kim, M Powell, A Ray, editors), Oxford Univ. Press (2021) 273–282 MR Zbl

[Lickorish 1997] **W B R Lickorish**, *An introduction to knot theory*, Graduate Texts in Math. 175, Springer (1997) MR Zbl

[Lipshitz and Sarkar 2022] **R Lipshitz**, **S Sarkar**, *A mixed invariant of nonorientable surfaces in equivariant Khovanov homology*, Trans. Amer. Math. Soc. 375 (2022) 8807–8849 MR Zbl

[Lück 2020] **W Lück**, *Assembly maps*, from "Handbook of homotopy theory" (H Miller, editor), CRC, Boca Raton, FL (2020) 851–890 MR Zbl

[Lück 2021] **W Lück**, *Isomorphism conjectures in K– and L–theory*, preprint (2021) Available at https://him-lueck.uni-bonn.de/data/ic.pdf

[Powell and Ray 2021] **M Powell**, **A Ray**, *Gropes, towers, and skyscrapers*, from "The disc embedding theorem" (S Behrens, B Kalmár, M H Kim, M Powell, A Ray, editors), Oxford Univ. Press (2021) 171–184 MR Zbl

[Quinn 1970]   **F Quinn**, *A geometric formulation of surgery*, from "Topology of manifolds" (J C Cantrell, C H Edwards, Jr, editors), Markham, Chicago (1970) 500–511   MR   Zbl

[Quinn 1971]   **F Quinn**, *$B_{(\mathrm{TOP}_n)}\sim$ and the surgery obstruction*, Bull. Amer. Math. Soc. 77 (1971) 596–600   MR   Zbl

[Ranicki 1979]   **A Ranicki**, *The total surgery obstruction*, from "Algebraic topology" (J L Dupont, I H Madsen, editors), Lecture Notes in Math. 763, Springer (1979) 275–316   MR   Zbl

[Ranicki 1981]   **A Ranicki**, *Exact sequences in the algebraic theory of surgery*, Math. Notes 26, Princeton Univ. Press (1981)   MR   Zbl

[Ranicki 1992]   **A A Ranicki**, *Algebraic L–theory and topological manifolds*, Cambridge Tracts in Math. 102, Cambridge Univ. Press (1992)   MR   Zbl

[Rosebrock 2007]   **S Rosebrock**, *The Whitehead conjecture: an overview*, Sib. Èlektron. Mat. Izv. 4 (2007) 440–449   MR   Zbl

[Strebel 1974]   **R Strebel**, *Homological methods applied to the derived series of groups*, Comment. Math. Helv. 49 (1974) 302–332   MR   Zbl

[Sundberg and Swann 2022]   **I Sundberg**, **J Swann**, *Relative Khovanov–Jacobsson classes*, Algebr. Geom. Topol. 22 (2022) 3983–4008   MR   Zbl

[Whitehead 1941]   **J H C Whitehead**, *On adding relations to homotopy groups*, Ann. of Math. 42 (1941) 409–428   MR   Zbl

Massachusetts Institute of Technology
Cambridge, MA, United States

Current address: *Department of Mathematics, University of Texas at Austin*
Austin, TX, United States

anthony.conway@austin.utexas.edu

Tame and relatively elliptic $\mathbb{CP}^1$–structures on the thrice-punctured sphere

Samuel A Ballas
Philip L Bowers
Alex Casella
Lorenzo Ruffoni

Suppose a relatively elliptic representation ρ of the fundamental group of the thrice-punctured sphere S is given. We prove that all projective structures on S with holonomy ρ and satisfying a tameness condition at the punctures can be obtained by grafting certain circular triangles. The specific collection of triangles is determined by a natural framing of ρ. In the process, we show that (on a general surface Σ of negative Euler characteristics) structures satisfying these conditions can be characterized in terms of their Möbius completion, and in terms of certain meromorphic quadratic differentials.

30F30, 57M50

1 Introduction

This paper deals with the geometry of surfaces which are locally modeled on the geometry of the Riemann sphere $\mathbb{CP}^1$, and their grafting deformations. Throughout the paper, Σ denotes an orientable surface with finitely many punctures (and no boundary) and $\overline{\Sigma}$ denotes the closed orientable surface where the punctures have been filled in. While the main technical core of the paper holds for a general Σ with negative Euler characteristic (see Sections 3 and 4), Section 5 deals specifically with the case of a thrice-punctured sphere, which we denote by S.

The structures under consideration here are known as complex projective structures or $(\mathrm{PSL}_2\mathbb{C}, \mathbb{CP}^1)$–structures. We denote respectively by $\mathcal{T}(\Sigma)$ and $\mathcal{P}(\Sigma)$ the deformation spaces of complex and complex projective structures on Σ. We also denote by $\mathcal{R}(\Sigma)$ the space of representations of $\pi_1(\Sigma)$ into $\mathrm{PSL}_2\mathbb{C}$, up to conjugation by $\mathrm{PSL}_2\mathbb{C}$. We have natural forgetful maps

$$\pi : \mathcal{P}(\Sigma) \to \mathcal{T}(\Sigma) \quad \text{and} \quad \mathrm{Hol} : \mathcal{P}(\Sigma) \to \mathcal{R}(\Sigma),$$

respectively recording the underlying complex structure and holonomy representation. We refer the reader to Section 3 for precise definitions, and to [Dumas 2009] for a general survey about $\mathbb{CP}^1$–structures. For more on the geometry of the deformation space, see [Faraco 2020].

Classic examples of complex projective structures are given by hyperbolic metrics (seen as $(\mathrm{PSL}_2\mathbb{R}, \mathbb{H}^2)$–structures), but a general projective structure is not defined by a Riemannian metric, nor is it completely determined by its holonomy (not even in the Fuchsian case, see for instance [Calsamiglia et al. 2014a; Goldman 1987]). However, under some additional conditions Hol is known to be a local homeomorphism (see [Gupta and Mj 2021; Hejhal 1975; Luo 1993]), ie a structure is at least locally determined by its holonomy. A major question in the field is the description in geometric terms of all structures having the same holonomy.

Grafting Conjecture [Gallo et al. 2000, Problems 12.1.1–2] *Two complex projective structures have the same holonomy if and only if it is possible to obtain one from the other by some sequence of graftings and degraftings.*

Here *grafting* refers to a geometric surgery on Σ which consists in cutting Σ open along a curve and inserting a domain from $\mathbb{CP}^1$, and *degrafting* is the inverse operation. For the reader familiar with grafting deformations: by grafting we will always mean *projective 2π–grafting*. This construction allows one to change a structure without changing its holonomy, and iterating this construction shows that Hol has infinite fibers. The Grafting Conjecture has been verified for closed surfaces: the case of (quasi-)Fuchsian representations is due to Goldman [1987], and Baba [2010; 2012; 2015; 2017] has addressed the case of generic (ie totally loxodromic) representations in a series of papers.

Inspired by a specific question about punctured spheres in [Gallo et al. 2000, Problem 12.2.1], we propose a study of certain structures on the thrice-punctured sphere, and we prove the Grafting Conjecture in this setting (see Section 1.2 of this introduction for a comparison with related results available in the literature). It is worth noticing that the complex projective geometry around a puncture is much more interesting than the underlying complex geometry. As an example, consider the two structures on the thrice-punctured sphere given by the complete hyperbolic metric of finite area and by the inclusion $\mathbb{CP}^1 \setminus \{0, 1, \infty\} \subseteq \mathbb{CP}^1$; they are not isomorphic as complex projective structures, but they have the same underlying complex structure.

The study of holonomy fibers also has an analytic motivation coming from the classical monodromy problems for ODEs, ie generalization of Hilbert's XXI problem. Since the work of Poincaré [1908],

projective structures have been known as a geometric counterpart to second-order linear ODEs. In more recent years, some monodromy problems for such ODEs have successfully been approached in terms of holonomy problems for projective structures (see [Calsamiglia et al. 2019; Chenakkod et al. 2022; Gallo et al. 2000; Gupta 2021; Gupta and Mj 2020; Kapovich 2020]).

We consider structures satisfying some regularity conditions at the punctures, which can be roughly stated as follows (see Section 3.1 for precise definitions):

- **Tameness**   Each local chart has a limit along arcs going off into a puncture.
- **Relative ellipticity**   Each peripheral holonomy (ie the holonomy around each puncture) is a nontrivial elliptic element in $\mathrm{PSL}_2\mathbb{C}$.
- **Nondegeneracy**   There is no pair of points $p_\pm \in \mathbb{CP}^1$ such that the entire holonomy preserves the set $\{p_\pm\}$.

Motivating examples of tame structures arise from the study of triangle groups and automorphism groups (as in [Faraco and Ruffoni 2019, Remark 2.13]), and more generally from metrics of constant curvature with cones or cusps. Tameness is not a generic condition in the space of all complex projective structures, but is a natural case to consider. Indeed, it corresponds to the condition that the associated second-order linear ODE has *regular singular points* (see Theorem E below). It turns out that the peripheral holonomy of a tame structure can only be trivial, parabolic, or elliptic (see Lemma 3.1.3), so the second condition is a generic condition within the space of holonomies of tame structures. In particular, it implies that there are no *apparent singularities* (ie no puncture has trivial holonomy).

For an arbitrary surface Σ, we denote by $\mathcal{P}^{\ominus}(\Sigma)$ the subspace of $\mathcal{P}(\Sigma)$ consisting of nondegenerate tame and relatively elliptic structures; the white disk in the superscript represents the local invariance under a rotation, and the black dot the possibility to extend the charts to the puncture. The tameness condition provides a natural choice of a fix point for each peripheral holonomy, ie a *framing* for the holonomy representation (see Corollary 3.1.5). We observe that grafting preserves this natural framing, which suggests a more precise formulation of the Grafting Conjecture in the noncompact case. Our main result in the case of the thrice-punctured sphere S is the following, which confirms the conjecture, in the spirit of [Gallo et al. 2000, Problem 12.2.1].

Theorem A   *Two structures in $\mathcal{P}^{\ominus}(S)$ have the same framed holonomy if and only if it is possible to obtain one from the other by some combination of graftings and degraftings along ideal arcs.*

Here an arc is *ideal* if it starts and ends at a puncture. To the best of our knowledge this is the first result in this direction for the case of noncompact surfaces with nontrivial holonomy around the punctures.

The representations involved here are representations of the free group $\mathbb{F}_2 = \pi_1(S)$ generated by elliptic elements. Representations satisfying certain rationality conditions correspond to the classical triangle groups, but the general ones are nondiscrete. In all cases we construct an explicit list of triangular membranes (ie immersions of a triangle in $\mathbb{CP}^1$) realizing these representations, and identify the ones

that are *atomic*; these can be taken as basic building blocks that can be grafted to reconstruct all the projective structures in $\mathscr{P}^{\odot}(S)$. Theorem A is a consequence of the following theorem.

Theorem B *Every $\sigma \in \mathscr{P}^{\odot}(S)$ is obtained by grafting on an atomic triangular structure with the same framed holonomy.*

Another consequence of Theorem B is a handy description of the moduli space $\mathscr{P}^{\odot}(S)$ with positive real coordinates, which we plan to address in a future work.

When a representation $\rho\colon \pi_1(S) \to \mathrm{PSL}_2\mathbb{C}$ is unitary (ie is conjugate into $\mathrm{PSU}(2)$), it preserves a spherical metric, and a structure $\sigma \in \mathscr{P}^{\odot}(S)$ is given by a spherical metrics with cone points. This special case of Theorem B is implicit in the proof of [Mondello and Panov 2016, Theorem 3.8], which constructs such spherical metrics by gluing together spherical triangles and bigons. Grafting a spherical metric results in a spherical metric, with increased angles at the cones. However in general this is not always the case; for example the structure obtained by grafting a hyperbolic structure is not defined by any Riemannian metric.

While our results about the Grafting Conjecture are for the case of the thrice-punctured sphere S, the main technical core of the paper applies to any noncompact surface Σ of negative Euler characteristic, and is of independent interest. It consists of a characterization of structures from $\mathscr{P}^{\odot}(\Sigma)$ in terms of their Möbius completion (see Section 3 and [Kulkarni and Pinkall 1994]) and in terms of meromorphic projective structures (see Section 4 and [Allegretti and Bridgeland 2020]). The easy case of structures on a twice-punctured sphere can be worked out concretely; see Remark 3.3.8. In the remaining part of the introduction we present our main results in the general case (see Section 1.1), as well as a comparison with other work in the literature about the Grafting Conjecture (see Section 1.2).

1.1 Results for general surfaces

The universal cover $\widetilde{\Sigma}$ of Σ is a topological disk. It admits a natural decoration obtained by adding ideal points at infinity "above" the punctures. We call these ideal points *ends*. This gives rise to a natural enlargement of $\widetilde{\Sigma}$ that we call the *end-extension*, and denote by $\widetilde{\Sigma}^{\#}$. Part of the paper is concerned with understanding the behavior of the developing map in the limit to an end.

Möbius completion Any complex projective structure σ on Σ can be used to define another natural extension of $\widetilde{\Sigma}$, known as the *Möbius completion* $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$, which comes with a (noncanonical) structure of a complete metric space (see [Kulkarni and Pinkall 1994]). For instance, when σ is induced by a spherical metric with cone points, $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ coincides with $\widetilde{\Sigma}^{\#}$, while when σ is induced by a complete hyperbolic metric of finite area $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ identifies with the closed disk model for the hyperbolic plane $\mathbb{H}^2 \cup \mathbb{RP}^1$ (see Examples 3.2.3 and 3.2.4).

The topologies on $\widetilde{\Sigma}^{\#}$ and on $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ are not in general compatible. One of the main technical contributions of this paper is a study of the geometry of the Möbius completion $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ for $\sigma \in \mathscr{P}^{\odot}(\Sigma)$, and of its relation

with the end-extension $\widetilde{\Sigma}^{\#}$ (see Section 3). Tameness of a structure σ implies that its developing map admits natural continuous extensions $\mathrm{dev}_{\#}$ to the end-extension $\widetilde{\Sigma}^{\#}$ and dev_{σ} to the Möbius completion $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$. We study the local properties of $\mathrm{dev}_{\#}$ and dev_{σ} around the ends.

Theorem C *Let* $\sigma \in \mathscr{P}(\Sigma)$ *be nondegenerate and without apparent singularities. Let* $j^{\#} \colon \widetilde{\Sigma} \to \widetilde{\Sigma}^{\#}$ *and* $j_{\sigma} \colon \widetilde{\Sigma} \to \mathrm{M}^{\sigma}(\widetilde{\Sigma})$ *be the natural embeddings. Then* $\sigma \in \mathscr{P}^{\odot}(\Sigma)$ *if and only if there exists a continuous open* $\pi_1(\Sigma)$*–equivariant embedding* $j_{\sigma}^{\#} \colon \widetilde{\Sigma}^{\#} \to \mathrm{M}^{\sigma}(\widetilde{\Sigma})$ *that makes the following diagram commute:*

$$
\begin{array}{ccc}
 & \widetilde{\Sigma}^{\#} & \\
 \stackrel{j^{\#}}{\nearrow} & \Big\downarrow j_{\sigma}^{\#} & \stackrel{\mathrm{dev}_{\#}}{\searrow} \\
\widetilde{\Sigma} & & \mathbb{CP}^1 \\
 \stackrel{j_{\sigma}}{\searrow} & \mathrm{M}^{\sigma}(\widetilde{\Sigma}) & \stackrel{\mathrm{dev}_{\sigma}}{\nearrow}
\end{array}
$$

In this statement, continuity is a consequence of tameness of σ, and openness is a consequence of relative ellipticity.

In general, the developing map for a projective structure is a surjection onto $\mathbb{CP}^1$, in which case it fails to be a global covering map. However, under certain circumstances it is known to be a covering map onto a component of the domain of discontinuity in $\mathbb{CP}^1$ for its holonomy representation (see for instance [Kra 1971a, Theorem 1]). But in general the holonomy group is not discrete, so it has no domain of discontinuity. The following statement shows that in our context some local covering behavior can be guaranteed around ends.

Theorem D *Let* $\sigma \in \mathscr{P}^{\odot}(\Sigma)$*, and let* E *be an end. Then there is a neighborhood* $\hat{N}_E$ *of* E *in* $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ *onto which the developing map for* σ *restricts to a branched covering map*, *branching only at* E*, and with image a round disk in* $\mathbb{CP}^1$.

These neighborhoods should be regarded as an analogue of the round balls considered in [Kulkarni and Pinkall 1994], but "centered" at ideal points in the Möbius completion. While Theorem D is stated as a local fact, we actually show that such a neighborhood can be chosen to be so large as to have another ideal point on its boundary. We use the existence of these neighborhoods to define a local geometric invariant, which we call the *index* (see Section 3.4). This number measures the angle described by the developing map at a puncture, and provides a notion of complexity for an inductive proof of Theorem B.

Meromorphic projective structures A second major ingredient (once again valid for an arbitrary noncompact surface Σ) consists of an analytic description of structures in $\mathscr{P}^{\odot}(\Sigma)$ as meromorphic projective structures in the sense of [Allegretti and Bridgeland 2020]. These are projective structures whose developing map is defined by solving certain differential equations with coefficients given by meromorphic quadratic differentials on the closed surface $\overline{\Sigma}$ (with poles corresponding to the punctures of Σ; see Section 4.1 for precise definitions). The local control from Theorem D allows us to obtain the following result.

Theorem E *Let $\sigma \in \mathscr{P}(\Sigma)$ and let $X \in \mathscr{T}(\Sigma)$ be the underlying complex structure. Then $\sigma \in \mathscr{P}^{\odot}(\Sigma)$ if and only if X is a punctured Riemann surface and σ is represented by a meromorphic quadratic differential on X with double poles and reduced exponents in $\mathbb{R} \setminus \mathbb{Z}$.*

Here the parametrization of projective structures by quadratic differentials is the classical one in terms of the Schwarz derivative, which here is taken with respect to any compatible holomorphic structure on the closed Riemann surface obtained by filling the punctures (eg the constant curvature uniformization). From this point of view, the index of a structure at a puncture corresponds to the absolute value of the exponents of the quadratic differential, so it can be computed in terms of its residues.

It should also be noted that work of Luo [1993] guarantees that Hol is a local homeomorphism for this class of meromorphic projective structures, as there are no apparent singularities. Therefore fibers of Hol in $\mathscr{P}^{\odot}(\Sigma)$ are discrete, and in particular it makes sense to seek a description of them in terms of a discrete geometric surgery such as the type of grafting that we consider in this paper.

Outline of the proof of Theorem B Let S be the thrice-punctured sphere, and let $\sigma \in \mathscr{P}^{\odot}(S)$, with developing map dev and holonomy ρ. By Theorems C and D, dev extends continuously and equivariantly to the ends, and restricts to a branched covering map on a suitable neighborhood of each end. This allows us to define the index of σ at each puncture. Then we construct a circular triangle such that the pillowcase obtained by doubling it provides a structure $\sigma_0 \in \mathscr{P}^{\odot}(S)$ with holonomy ρ. Note that such a triangle is not unique in general. A careful analysis of the framing of ρ defined by σ shows that such a triangle can be found with the same framing for ρ. On such a triangle, we find a suitable combination of disjoint ideal arcs that are graftable, and we show that if sufficiently many grafting regions are inserted, the resulting structure $\sigma' \in \mathscr{P}^{\odot}(S)$ has the same indices as σ. By Theorem E, σ and σ' can be represented by two meromorphic differentials on the Riemann sphere $\mathbb{CP}^1$ with double poles at 0, 1 and ∞. Two such differentials on $\mathbb{CP}^1$ are completely determined by their residues, and in this case residues can be computed directly from the indices, hence are the same. So we conclude that $\sigma = \sigma'$. □

1.2 Relation to other work about the Grafting Conjecture

Following seminal work of Thurston (see [Baba 2020; Dumas 2009; Kamishima and Tan 1992]), grafting (in its general version) has been successfully used as a tool to explore the deformation space of $\mathbb{CP}^1$–structures. The grafting we consider here preserves the holonomy representation, hence can be used to explore holonomy fibers. The classical case is that of structures on a closed surface with Fuchsian holonomy, which was considered by Goldman [1987]. Our work displays some technical differences, that we summarize here for the expert reader.

Framing The main results for closed surfaces in [Baba 2012; 2015; 2017; Calsamiglia et al. 2014b; Goldman 1987] confirm the Grafting Conjecture, ie that two structures with the same holonomy differ by grafting. In our noncompact case there is a natural framing for the holonomy which needs to be taken into

consideration, as it is preserved by grafting (see Lemma 3.1.7). We prove that having the same framed holonomy is not only necessary, but also sufficient, for two structures on the thrice-punctured sphere to differ by grafting.

Basepoints for holonomy fibers   When $\rho: \pi_1(\Sigma) \to \mathrm{PSL}_2\mathbb{C}$ is Fuchsian, the holonomy fiber $\mathrm{Hol}^{-1}(\rho)$ contains a preferred structure, namely the hyperbolic structure $\mathbb{H}^2/\rho(\pi_1(\Sigma))$. This structure serves as a basepoint, ie any other structure in $\mathrm{Hol}^{-1}(\rho)$ can be obtained by grafting it (see [Goldman 1987]). In this paper, we show that every representation coming from $\mathscr{P}^{\ominus}(S)$ is generated by reflections in the sides of a circular triangle in $\mathbb{CP}^1$. Even when such a representation ρ is nondiscrete, the pillowcase obtained by doubling the triangle provides a basepoint in the holonomy fiber $\mathrm{Hol}^{-1}(\rho)$. A first guess is that every structure in $\mathrm{Hol}^{-1}(\rho)$ is obtained by grafting this pillowcase. However, this is not the case, because of the aforementioned framing, which is given by the vertices of the triangle. In Section 2.3 we identify the list of the structures that can be taken as basepoints in the above sense, which we call atomic. Interestingly, they are not all embedded geodesic triangles for some invariant metric.

Type of grafting curves   In the classical Fuchsian case it is enough to perform grafting along simple closed geodesics on the hyperbolic basepoint (see [Goldman 1987]). Here we consider grafting along ideal arcs, ie arcs that start and end at punctures. Grafting along open arcs is also known as bubbling in the literature (see [Calsamiglia et al. 2014a; Francaviglia and Ruffoni 2021; Gallo et al. 2000; Ruffoni 2019; 2021]). Most structures considered here are not metric, but they still have a well-defined notion of circular arc. We show that in most cases grafting arcs can be chosen to be circular.

Uniqueness of grafting curves   In the classical Fuchsian case grafting curves are homotopically nontrivial, and are uniquely determined by the structure itself (see [Goldman 1987]). Here grafting regions do not carry any topology (they are disks), hence they should not be expected to be canonically associated with the structure. Indeed it is quite common for a structure to arise from different graftings on different atomic structures.

Outline of the paper

Section 2 contains background material about the geometry of circles and circular triangles in $\mathbb{CP}^1$ (see Sections 2.1 and 2.2). In Section 2.3 we provide a classification of certain triangular immersions that will serve as the atomic structures for our main grafting results. This classification is referred to in different parts of the paper, and it is summarized in Tables 1, 2 and 3 of the appendix.

Section 3 introduces the main geometric definitions, ie that of tameness and relative ellipticity. In Section 3.2 we study the geometry of the Möbius completion for a general surface and address Theorem C. The proof of Theorem D is in Section 3.3, where we show that the developing map restricts to a nice branched cover around each end. This is used in Section 3.4 to define the index of a puncture, and in Section 4 to obtain a characterization of tame and relatively elliptic structures in terms of quadratic

differentials on a general Riemann surface. In particular we show that the geometric notion of index can be also defined and computed analytically. Theorem E is contained in Section 4.2.

Finally, in Section 5 we restrict our attention to the case of the thrice-punctured sphere S. In Section 5.1 we define the class of triangular structures on S, based on Section 2.3, and in Section 5.2 we prove the main grafting results of Theorems A and B.

Acknowledgements

2 Basics on complex projective geometry

In this chapter we collect some background about the geometry of the Riemann sphere, on which our geometric structures will be modeled, mainly to fix notation and terminology. Let $\mathbb{CP}^1$ denote the set of complex lines through the origin in $\mathbb{C}^2$, ie the quotient of $\mathbb{C}^2 \setminus \{0\}$ by scalar complex multiplication. We fix identifications of $\mathbb{CP}^1$ with the extended complex plane $\mathbb{C} \cup \{\infty\}$ and the unit sphere $\mathbb{S}^2$. Through them, $\mathbb{CP}^1$ inherits a natural complex structure, an orientation, and a spherical metric. A *circle* in $\mathbb{CP}^1$ is a circle or a line in $\mathbb{C} \cup \{\infty\}$. Every circle divides $\mathbb{CP}^1$ into two disks, each of which has a standard identification with the hyperbolic plane which respects the underlying complex structure. We denote by $\mathrm{PSL}_2\mathbb{C}$ the group of projective classes of 2-by-2 complex matrices of determinant 1. This group acts on $\mathbb{CP}^1$ by *Möbius transformations*

$$\mathrm{PSL}_2\mathbb{C} \times \mathbb{CP}^1 \to \mathbb{CP}^1, \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad z \mapsto \frac{az+b}{cz+d}.$$

For elements in $\mathrm{PSL}_2\mathbb{C}$, traces and determinants are not well defined. However there is a two-to-one map $\mathrm{SL}_2\mathbb{C} \to \mathrm{PSL}_2\mathbb{C}$ such that $\pm A \mapsto [A]$. Therefore, given an element $G \in \mathrm{PSL}_2\mathbb{C}$, we can always assume it to be in $\mathrm{SL}_2\mathbb{C}$ modulo a sign. It follows that $\det(G)$, $|\mathrm{tr}(G)|$ and $\mathrm{tr}(G)^2$ are well-defined quantities. The action of $\mathrm{PSL}_2\mathbb{C}$ on $\mathbb{CP}^1$ is faithful, and simply transitive on triples of pairwise distinct points. In particular, we can always map three distinct points (p_1, p_2, p_3) to $(0, 1, \infty)$. Möbius transformations are conformal, preserve cross ratios and preserve circles. Three distinct points in $\mathbb{CP}^1$ determine a unique circle through them. Great circles are geodesic circles in the underlying spherical metric. However, elements of $\mathrm{PSL}_2\mathbb{C}$ are generally not isometries, and so the set of great circles is not $\mathrm{PSL}_2\mathbb{C}$–invariant.

A nontrivial element $G \in \mathrm{PSL}_2\mathbb{C}$ is classified as

- *parabolic* if $\mathrm{tr}(G)^2 = 4$,
- *elliptic* if $\mathrm{tr}(G)^2$ is real and $\mathrm{tr}(G)^2 < 4$,
- *loxodromic* otherwise.
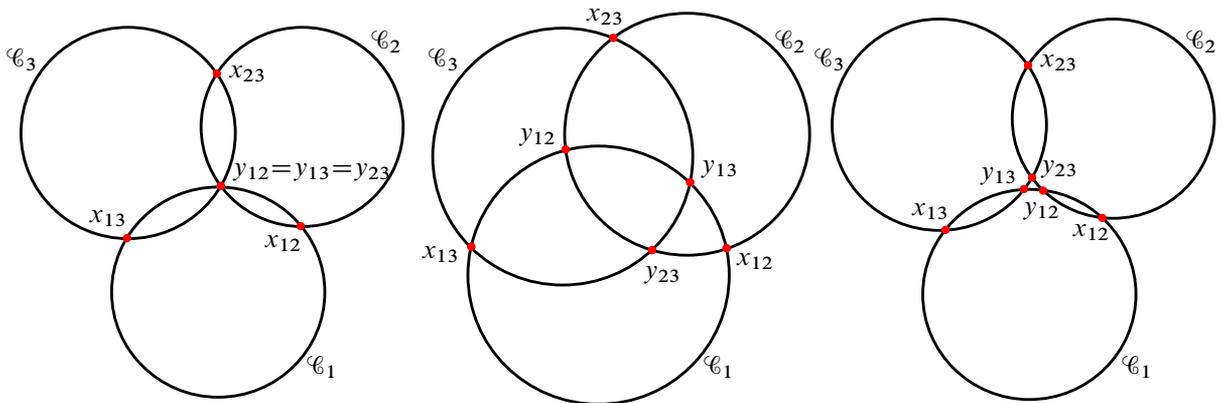
Figure 1: From left to right: a Euclidean, spherical and hyperbolic configuration of circles.

2.1 Configurations of circles

Let $\mathfrak{C} = (\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3)$ be an (*ordered*) *configuration* of three distinct circles in $\mathbb{CP}^1$. The configuration $\mathfrak{C}$ is *nondegenerate* if every pair $\mathscr{C}_i, \mathscr{C}_j$ intersects in exactly two points $\{x_{ij}, y_{ij}\}$, and the set of pairwise intersection points has at least four elements. Henceforth, all configurations will be assumed to be nondegenerate. Also notice that by definition $\mathfrak{C}$ is an ordered triple.

A configuration of circles is *Euclidean* if the circles have a common intersection point. In this case there are exactly four intersection points. If the configuration is not Euclidean, since every circle divides $\mathbb{CP}^1$ into two disjoint regions, then $\mathscr{C}_1$ separates $\{x_{23}, y_{23}\}$ if and only if $\mathscr{C}_2$ separates $\{x_{13}, y_{13}\}$ if and only if $\mathscr{C}_3$ separates $\{x_{12}, y_{12}\}$. In that case, we say that the configuration $\mathfrak{C}$ is *spherical*. Otherwise, it is *hyperbolic* (see Figure 1).

Remark 2.1.1 A configuration of circles induces a CW–structure on $\mathbb{CP}^1$, in which the 2–cells are either bigons, triangles or quadrilaterals; in the spherical case the structure is simplicial and isomorphic to an octahedron. Given two configurations of circles $\mathfrak{C}^i = (\mathscr{C}_1^i, \mathscr{C}_2^i, \mathscr{C}_3^i)$ of the same kind (Euclidean, spherical or hyperbolic), there is always (at least) one CW–isomorphism of $\mathbb{CP}^1$ mapping $\mathscr{C}_k^1$ to $\mathscr{C}_k^2$. For spherical and hyperbolic configurations, it is enough to consider orientation preserving CW–isomorphisms. On the other hand, if $\mathfrak{C} = (\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3)$ is a Euclidean configuration of circles, there is no orientation preserving CW–isomorphism mapping $(\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3)$ to $(\mathscr{C}_1, \mathscr{C}_3, \mathscr{C}_2)$; the obstruction being the cyclic order of the circles at the common intersection point.

The connection between a configuration of circles and the corresponding geometries is well known. We recall it in the next result (cf Figure 2).

Lemma 2.1.2 *Let $\mathfrak{C}$ be a configuration of three circles.*

- *If $\mathfrak{C}$ is Euclidean, let y be the common intersection point. Then $\mathbb{CP}^1 \setminus \{y\}$ admits a Euclidean metric for which the circles in $\mathfrak{C}$ are geodesics.*
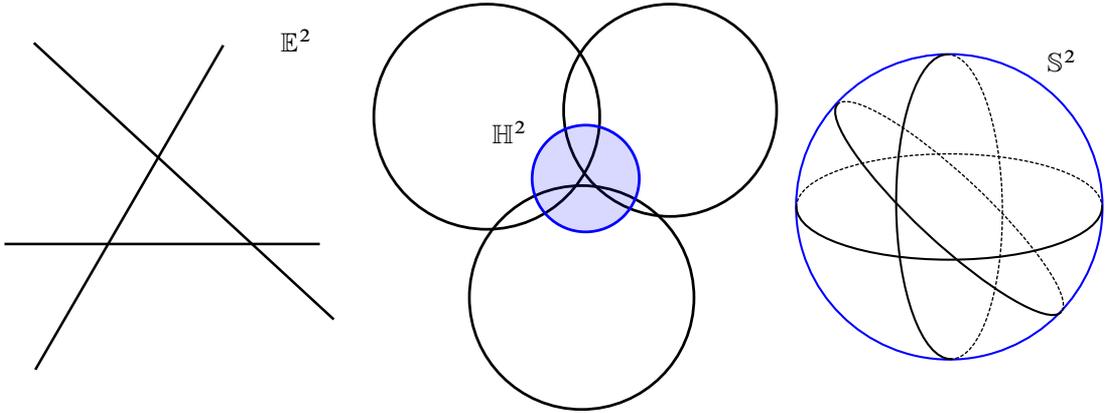
Figure 2: Euclidean, hyperbolic and spherical configurations are related to the corresponding geometries.

- If $\mathfrak{C}$ is *spherical*, *then there is a Möbius transformation* $G \in \mathrm{PSL}_2\mathbb{C}$ *such that* $G \cdot \mathfrak{C}$ *are great circles for the underlying spherical metric.*

- If $\mathfrak{C}$ is *hyperbolic*, *then there is a unique circle* $\mathscr{C}_{\mathbb{H}}$ *orthogonal to every circle in* $\mathfrak{C}$. *In particular, each connected component* $D_{\mathbb{H}}$ *of* $\mathbb{CP}^1 \setminus \mathscr{C}_{\mathbb{H}}$ *admits a hyperbolic metric for which the intersections of* $\mathfrak{C}$ *with* $D_{\mathbb{H}}$ *are geodesics.*

Any two distinct circles $\mathscr{C}_1$ and $\mathscr{C}_2$ in a configuration $\mathfrak{C}$ intersect in two points. If x is a point of intersection, then we can use the orientation of $\mathbb{CP}^1$ to determine the *anticlockwise angle* $\angle_x \mathscr{C}_1 \mathscr{C}_2$ from $\mathscr{C}_1$ to $\mathscr{C}_2$ at x (see Figure 3). We have that

$$\angle_x \mathscr{C}_2 \mathscr{C}_1 = \pi - \angle_x \mathscr{C}_1 \mathscr{C}_2 = \angle_y \mathscr{C}_1 \mathscr{C}_2,$$

where y is the other point of intersection of $\mathscr{C}_1$ and $\mathscr{C}_2$. It is a simple exercise in complex projective geometry to show that a configuration of circles is uniquely determined (up to Möbius transformations) by the ordered triple of angles at three points.

Lemma 2.1.3 *For* $i \in \{1, 2\}$, *let* $\mathfrak{C}^i = (\mathscr{C}_1^i, \mathscr{C}_2^i, \mathscr{C}_3^i)$ *be a configuration of circles. For every pair of circles in* $\mathfrak{C}^i$ *let* $x_{jk}^i \in \mathscr{C}_j^i \cap \mathscr{C}_k^i$ *be an intersection point such that*

$$\angle_{x_{12}^1} \mathscr{C}_1^1 \mathscr{C}_2^1 = \angle_{x_{12}^2} \mathscr{C}_1^2 \mathscr{C}_2^2, \quad \angle_{x_{23}^1} \mathscr{C}_2^1 \mathscr{C}_3^1 = \angle_{x_{23}^2} \mathscr{C}_2^2 \mathscr{C}_3^2, \quad \angle_{x_{13}^1} \mathscr{C}_1^1 \mathscr{C}_3^1 = \angle_{x_{13}^2} \mathscr{C}_1^2 \mathscr{C}_3^2.$$

Then there is a Möbius transformation $M \in \mathrm{PSL}_2\mathbb{C}$ *such that* $M \cdot \mathfrak{C}^1 = \mathfrak{C}^2$ *with* $M \cdot x_{jk}^1 = x_{jk}^2$.
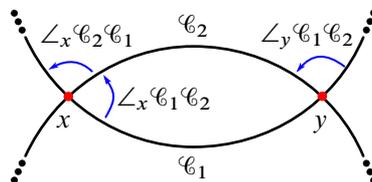

Figure 3: The anticlockwise angle between two circles at a point of intersection.

2.2 Elliptic Möbius transformations

In this section we prove a correspondence between configurations of circles and certain triples of elliptic Möbius transformations (Corollary 2.2.7).

As defined above, a nontrivial Möbius transformation $G \in \mathrm{PSL}_2\mathbb{C}$ is said to be elliptic if $\mathrm{tr}(G)^2$ is real and $\mathrm{tr}(G)^2 < 4$. An elliptic transformation fixes exactly two points of $\mathbb{CP}^1$. Let $G \in \mathrm{PSL}_2\mathbb{C}$ be elliptic. The *rotation angle* $\mathrm{Rot}(G, x) \in (0, 2\pi)$ *of* G *at a fixed point* x is the angle of anticlockwise rotation of G at x (more precisely of dG_x on $T_x\mathbb{CP}^1$). If x and y are the fixed points of G, a Möbius transformation mapping x, y to $0, \infty$ conjugates G to the element of $\mathrm{PSL}_2\mathbb{C}$

$$(2.2.1) \qquad \begin{bmatrix} e^{i\frac{1}{2}\mathrm{Rot}(G,x)} & 0 \\ 0 & e^{-i\frac{1}{2}\mathrm{Rot}(G,x)} \end{bmatrix}.$$

The definition of rotation angle implies the following result.

Lemma 2.2.1 *Let* $G \in \mathrm{PSL}_2\mathbb{C}$ *be elliptic with fixed points* $\{x, y\}$. *Then*

$$\mathrm{Rot}(G, y) = 2\pi - \mathrm{Rot}(G, x) = \mathrm{Rot}(G^{-1}, x).$$

The *rotation invariant* of an elliptic transformation G is the unordered pair

$$\mathrm{Rot}(G) := \{\mathrm{Rot}(G, x), \mathrm{Rot}(G, y)\}.$$

Lemma 2.2.2 *Let* $G \in \mathrm{PSL}_2\mathbb{C}$ *be elliptic, and let* $\theta \in (0, 2\pi)$. *Then* $\theta \in \mathrm{Rot}(G)$ *if and only if* $4\cos^2\left(\frac{1}{2}\theta\right) = \mathrm{tr}^2(G)$.

Proof Both the rotation angle and the trace operator are invariant under conjugation; thus we may assume that G is normalized as in (2.2.1). The equation $4\cos^2\left(\frac{1}{2}\theta\right) = \mathrm{tr}^2(G)$ has precisely two solutions in $(0, 2\pi)$, of the form

$$\theta_1 = 2\arccos\left(\tfrac{1}{2}|\mathrm{tr}(G)|\right) \quad \text{and} \quad \theta_2 = 2\pi - 2\arccos\left(\tfrac{1}{2}|\mathrm{tr}(G)|\right),$$

where we fix a determination of arccos in $[0, \pi]$. A direct computation shows that $\mathrm{Rot}(G) = \{\theta_1, \theta_2\}$, concluding the proof. $\qquad\square$

Given the fixed points of G, the rotation invariant is enough to determine G up to inversion, while the rotation angle is a complete invariant.

Lemma 2.2.3 *Let* $G, H \in \mathrm{PSL}_2\mathbb{C}$ *be two elliptic transformations. Then*

(1) $\mathrm{Rot}(G) = \mathrm{Rot}(H) \iff \mathrm{tr}^2(G) = \mathrm{tr}^2(H) \iff G, H$ *are conjugate.*

(2) *If* G *and* H *have the same fixed points* $\{x, y\}$, *then*

$$\mathrm{Rot}(G) = \mathrm{Rot}(H) \iff G = H^{\pm 1},$$

  and in particular

$$\mathrm{Rot}(G, x) = \mathrm{Rot}(H, x) \iff G = H.$$

Proof (1) Two elliptics with the same rotation invariants must have the same trace squared by the previous Lemma 2.2.2. But this is a complete invariant of conjugacy classes for semisimple elements of $\mathrm{PSL}_2\mathbb{C}$.

(2) Since G and H share the same fixed points, we can simultaneously normalize them as in (2.2.1). Both statements follow from comparing the two normal forms.   □

Next we analyze the connection between elliptic transformations, whose product is elliptic, and configurations of circles in $\mathbb{CP}^1$. First, we recall the following result from [Gallo et al. 2000, Lemma 3.4.1].

Lemma 2.2.4 *Let* $G, H \in \mathrm{PSL}_2\mathbb{C}$ *be elliptic transformations with at most one common fixed point, and such that the product* GH *is elliptic. Then the fixed points of* G *and* H *are contained in a unique circle* $\mathcal{C}_{G,H}$.

We recall that given any two distinct circles $\mathcal{C}_1$ and $\mathcal{C}_2$ intersecting at a point x, the (anticlockwise) angle from $\mathcal{C}_1$ to $\mathcal{C}_2$ at x is denoted by $\angle_x\mathcal{C}_1\mathcal{C}_2$ (see Section 2.1).

Lemma 2.2.5 *Let* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *be distinct circles in* $\mathbb{CP}^1$ *meeting exactly at two points,* x *and* y. *Let* J_i *denote the reflection in* $\mathcal{C}_i$. *Then the product* $G = J_2 J_1$ *is an elliptic transformation fixing* x *and* y *with*

$$\mathrm{Rot}(G, x) = 2\angle_x\mathcal{C}_1\mathcal{C}_2 \quad \text{and} \quad \mathrm{Rot}(G, y) = 2\angle_y\mathcal{C}_1\mathcal{C}_2.$$

Proof Since Möbius transformations are conformal, we can normalize so that $x = 0$ and $y = \infty$. Under the standard identification $\mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$, we can further normalize so that $\mathcal{C}_1 = \mathbb{R} \cup \{\infty\}$. Then $\mathcal{C}_2$ is a Euclidean line through 0 and ∞. In this setting

$$J_1(z) = \bar{z} \quad \text{and} \quad J_2(z) = e^{i2(\angle_x\mathcal{C}_1\mathcal{C}_2)}\bar{z},$$

and the statement follows from a direct computation.   □

Henceforth we fix the following notation. Given G and H, distinct elliptic transformations whose product GH is elliptic, we denote by $\{p_G, q_G\}$ (resp. $\{p_H, q_H\}$) the fixed points of G (resp. H), by $\mathcal{C}_{G,H}$ the unique circle through $\{p_G, q_G, p_H, q_H\}$ (see Lemma 2.2.4), and by $J_{G,H}$ the reflection about $\mathcal{C}_{G,H}$.

Lemma 2.2.6 *Let* (A, B, C) *be an ordered triple of elliptic transformations with at most one common fixed point, and such that* $ABC = 1$. *Then*

(1) $\mathcal{C}_{A,C} \cap \mathcal{C}_{A,B} = \{p_A, q_A\}$;

(2) $2\angle_{p_A}\mathcal{C}_{A,B}\mathcal{C}_{A,C} = \mathrm{Rot}(A, p_A)$ *and* $2\angle_{q_A}\mathcal{C}_{A,B}\mathcal{C}_{A,C} = \mathrm{Rot}(A, q_A)$;

(3) $A = J_{A,C} J_{A,B}$.

Proof We begin by noticing that two of the three elliptic transformations share a common fixed point p if and only if p is fixed by all three of them. Hence there are either four or six distinct fixed points. Then statement (1) follows from Lemma 2.2.4.

Next, we recall that Möbius transformations are conformal; thus without loss of generality we can simultaneously normalize (A, B, C) so that $(p_A, q_A, p_B) = (0, \infty, 1)$. It follows that $\mathscr{C}_{A,B} = \mathbb{R} \cup \{\infty\}$. If we let $\theta := \frac{1}{2} \operatorname{Rot}(A, 0)$, then the three elliptic transformations take the forms

$$A = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}, \quad B = \begin{bmatrix} a & b \\ c & \bar{a} \end{bmatrix}, \quad C^{-1} = AB = \begin{bmatrix} ae^{i\theta} & be^{i\theta} \\ ce^{-i\theta} & \bar{a}e^{-i\theta} \end{bmatrix},$$

where $|\operatorname{tr} B| = 2|\operatorname{Re}(a)| < 2$ (the relation between the diagonal elements of B is implied by the fact that C is elliptic). We remind the reader that we are always taking representatives in $\mathrm{SL}_2\mathbb{C}$ modulo a sign. Using that $\det(B) = 1$ and that B fixes 1, it follows that b and c are purely imaginary. In particular, there are choices of signs for which

$$b = -i \operatorname{Im}(a) \pm \sqrt{\operatorname{Re}^2(a) - 1} \quad \text{and} \quad c = i \operatorname{Im}(a) \pm \sqrt{\operatorname{Re}^2(a) - 1}.$$

We claim that C^{-1} has fixed points of the form $te^{i\theta}$ for $t \in \mathbb{R} \setminus \{0\}$. Since C and C^{-1} have the same fixed points, this will imply that $\angle_0 \mathscr{C}_{A,B} \mathscr{C}_{A,C} = \theta = \frac{1}{2} \operatorname{Rot}(A, 0)$. To this end, we look for real solutions of the equation

$$te^{i\theta} = AB \cdot te^{i\theta} = \frac{ate^{2i\theta} + be^{i\theta}}{ct + \bar{a}e^{-i\theta}} \iff ct^2 - 2i \operatorname{Im}(ae^{i\theta})t - b = 0.$$

Since b and c are purely imaginary, this polynomial has real roots if and only if its discriminant $-4 \operatorname{Im}(ae^{i\theta})^2 + 4bc$ is negative. But that follows from

$$1 = \det(AB) = \|a\|^2 - bc = \|ae^{i\theta}\|^2 - bc = \operatorname{Re}(ae^{i\alpha})^2 + \operatorname{Im}(ae^{i\theta})^2 - bc,$$

and

$$2 > |\operatorname{tr}(AB)| = |2 \operatorname{Re}(ae^{i\theta})|.$$

This concludes the proof of the first part of (2), while the rest follows from the definition of the anticlockwise angle between two circles and Lemma 2.2.1.

For the last statement of the lemma, recall that $G := J_{A,C} J_{A,B}$ is an elliptic Möbius transformation with fixed points $\{p_A, q_A\}$ (Lemma 2.2.5). Then G has the same fixed points and rotation angles as A; thus $G = A$ by Lemma 2.2.3. □

Lemmas 2.2.4, 2.2.5 and 2.2.6 have the following straightforward consequence.

Corollary 2.2.7 *There is a bijection*

$$\left\{ \begin{array}{c} \text{configurations} \\ \text{of three circles} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{ordered triples of elliptic transformations with} \\ \text{at most one common fixed point and product 1} \end{array} \right\}$$

where $(\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3) \mapsto (J_3 J_1, J_1 J_2, J_2 J_3)$ *and* $(A, B, C) \mapsto (\mathscr{C}_{A,B}, \mathscr{C}_{B,C}, \mathscr{C}_{A,C})$.

2.3 Triangular immersions

In this section we define certain immersions of the standard 2–simplex in $\mathbb{CP}^1$. Lemmas 2.3.1, 2.3.3 and 2.3.4 prove the existence of immersions with certain requirements on the angles at the vertices. These

are the ones we call atomic, and are listed in Tables 1, 2 and 3 of the appendix. Then we study some invariants of such immersions, and conclude in Corollary 2.3.7 that they are essentially determined by the image of the vertices, up to a minor ambiguity.

Let $\triangle := \{(x_1, x_2, x_3) \in \mathbb{R}^3_{\geq 0} \mid x_1 + x_2 + x_3 = 1\}$ be the *standard 2–dimensional simplex*. Let $\{V_1, V_2, V_3\} \subset \triangle$ be its set of vertices $V_1 = (1, 0, 0)$, $V_2 = (0, 1, 0)$ and $V_3 = (0, 0, 1)$, and let $e_{ij} \subset \triangle$ be the edge between V_i and V_j. We endow $\triangle$ with the orientation induced form the ordering (V_1, V_2, V_3) of its vertices.

A *triangular immersion* is an orientation preserving immersion $\tau \colon \triangle \to \mathbb{CP}^1$ such that each $\tau(e_{ij})$ is contained in a circle. In particular, we require τ to be locally injective everywhere except at the vertices. When every $\tau(e_{ij})$ is contained in a great circle, the triangle $\triangle$ inherits a spherical metric with geodesic boundary and cone angles at the vertices. This is usually referred to as a *spherical triangular membrane* in the literature [Eremenko 2004; Mondello and Panov 2016]. Triangular immersions are relevant to this paper as they produce natural examples of $\mathbb{CP}^1$–structures on the thrice-punctured sphere (see Section 5.1 for details.)

Henceforth, we will often make the abuse of notation of referring to both the triangular immersion and its image in $\mathbb{CP}^1$ by τ, when it is not necessary to make a distinction. The image of the vertices (resp. edges) of $\triangle$ are the *vertices* (resp. *edges*) of τ. Since edges of τ are arcs of circles, τ has well-defined *angles* at the vertices. When τ is not locally injective at a vertex, the angle is larger than 2π, and τ should be thought as "spreading over" $\mathbb{CP}^1$. The orientation of $\triangle$ and the ordering of its vertices induce an orientation on τ, and an ordering of its vertices and of its angles (which agree with the orientation and ordering induced by the orientation of $\mathbb{CP}^1$).

Configurations of circles and triangular immersions are related to one another. If τ is a triangular immersion, each one of its edges extends to a unique circle giving a (possibly degenerate) configuration $\mathfrak{C}_\tau$ of three circles. In this case we say that $\mathfrak{C}_\tau$ *supports* τ. When $\mathfrak{C}_\tau$ is nondegenerate, we say that τ is *nondegenerate*. When the interior of the image of τ is disjoint from $\mathfrak{C}_\tau$, we say that τ is *enclosed in* $\mathfrak{C}_\tau$. These are exactly those triangular immersions whose (interior of the) images are the connected components of $\mathbb{CP}^1 \setminus \mathfrak{C}_\tau$. Necessary and sufficient conditions on the angles of τ for it to be enclosed in $\mathfrak{C}_\tau$ are well known, but we provide a short proof as we could not find a direct reference.

Lemma 2.3.1 *Let (a, b, c) be an ordered triple of angles in $(0, \pi)^3$.*

(1) **Euclidean triangles** *There is a Euclidean configuration of circles $\mathfrak{C}$ and a triangular immersion τ enclosed in $\mathfrak{C}$ with angles (a, b, c) if and only if one of the following conditions are satisfied:*

(2.3.1) $a + b + c = \pi, \quad -a + b + c = \pi, \quad a - b + c = \pi, \quad a + b - c = \pi.$

(2) **Hyperbolic triangles** *There is a hyperbolic configuration of circles $\mathfrak{C}$ and a triangular immersion τ enclosed in $\mathfrak{C}$ with angles (a, b, c) if and only if*

(2.3.2) $a + b + c < \pi.$
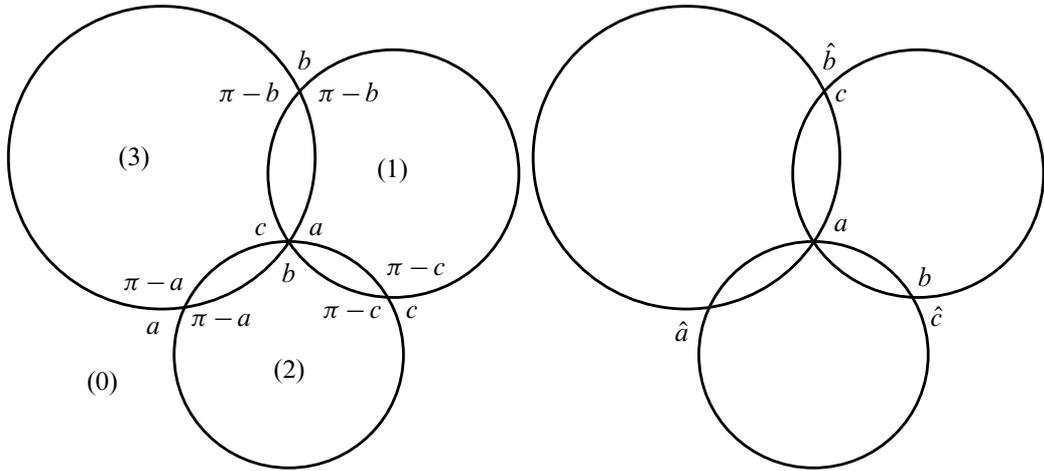
Figure 4: Two Euclidean configurations. Both support an enclosed triangular immersion with angles $(a, b, c) \in (0, \pi)^3$, such that either $a + b + c = \pi$ (left) or $-a + b + c = \pi$ (right).

(3) **Spherical triangles** There is a spherical configuration of circles $\mathfrak{C}$ and a triangular immersion T enclosed in $\mathfrak{C}$ with angles (a, b, c) if and only if (a, b, c) satisfies

$$(2.3.3) \qquad a + b + c > \pi, \quad a + \pi > b + c, \quad b + \pi > c + a, \quad c + \pi > a + b.$$

Proof (1) Let τ be a triangular immersion enclosed in a Euclidean configuration of circles $\mathfrak{C}_\tau$. Then there is a common intersection point y, and $\mathbb{CP}^1 \setminus \{y\}$ admits a Euclidean metric for which the circles in $\mathfrak{C}_\tau$ are geodesics (see Lemma 2.1.2). In this setting, it is easy to check that each one of the four triangular immersions that are enclosed in $\mathfrak{C}_\tau$ have angles

- (0) (a, b, c),
- (1) $(a, \pi - c, \pi - b)$,
- (2) $(\pi - c, b, \pi - a)$,
- (3) $(\pi - b, \pi - a, c)$,

each one satisfying exactly one of the equalities in (2.3.1) (see Figure 4).

The converse implication is well known for $a + b + c = \pi$. If $-a + b + c = \pi$, we consider the angles $\hat{a} = a$, $\hat{b} = \pi - c$ and $\hat{c} = \pi - b$. Clearly $(\hat{a}, \hat{b}, \hat{c}) \in (0, \pi)^3$ and $\hat{a} + \hat{b} + \hat{c} = \pi$; therefore there is a Euclidean triangle with angles $(\hat{a}, \hat{b}, \hat{c})$ supported by some configuration of circles. One of the other enclosed triangular immersions has angles (a, b, c) (see Figure 4). The same strategy applies to the other cases.

(2) Let τ be a triangular immersion enclosed in a hyperbolic configuration of circles $\mathfrak{C}_\tau$. Let $\mathscr{C}_{\mathbb{H}}$ be the circle that is orthogonal to the family $\mathfrak{C}_\tau$ (Lemma 2.1.2). In this case there are precisely two triangular immersions that are enclosed in $\mathfrak{C}_\tau$, and they are both disjoint from $\mathscr{C}_{\mathbb{H}}$. It follows that τ is a hyperbolic triangle in one of the two connected components of $\mathbb{CP}^1 \setminus \mathscr{C}_{\mathbb{H}}$; thus inequality (2.3.2) is a consequence
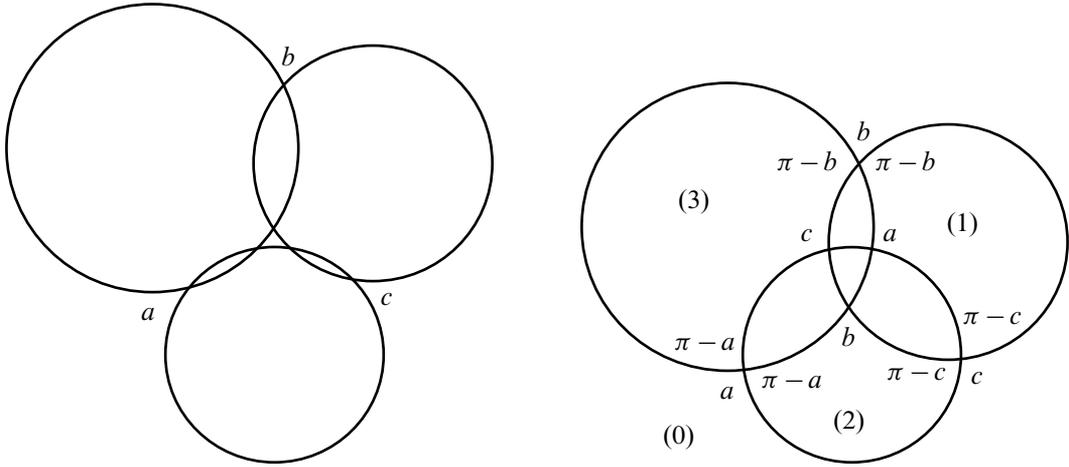
Figure 5: A hyperbolic configuration and a spherical configuration. They both support an enclosed triangular immersion with angles $(a, b, c) \in (0, \pi)^3$, such that either $a + b + c = \pi$ (left) or (2.3.3) is satisfied (right).

of the formula for hyperbolic area of triangles (see Figure 5). The converse implication is [Ratcliffe 2006, Theorem 3.5.9].

(3) Finally, let τ be a triangular immersion enclosed in a spherical configuration of circles $\mathfrak{C}_\tau$. By Lemma 2.1.2, we can realize this configuration of circles by great circles. So every triangular region τ enclosed in $\mathfrak{C}_\tau$ is a geodesic triangle for the standard spherical metric. By the area formula for spherical triangles, we have that

$$a + b + c = \pi + \text{Area}(\tau) > \pi.$$

The other inequalities (2.3.3) are obtained by applying Gauss–Bonnet to the enclosed triangular regions adjacent to τ (see Figure 5), whose angles are

(1) $(a, \pi - c, \pi - b)$,

(2) $(\pi - c, b, \pi - a)$,

(3) $(\pi - b, \pi - a, c)$.

The converse implication is a simple adaptation of [Ratcliffe 2006, Theorem 3.5.9] using the law of cosines in spherical geometry (see [Ratcliffe 2006, Exercise 2.5.8]). $\qquad\square$

Remark 2.3.2 For convenience, Lemma 2.3.1 is stated just in terms of the existence of a triangular immersion τ. Although we will not need it, we remark that it is a simple consequence of Lemma 2.1.3 that τ is also unique up to Möbius transformations. The same is true for the following results.

Given an enclosed triangular immersion τ, there are two simple operations that one can perform to construct new triangular immersions supported by the same configuration of circles. The first one consists in extending τ by a full disk, by "pushing" an edge of τ to its complement in its supporting circle
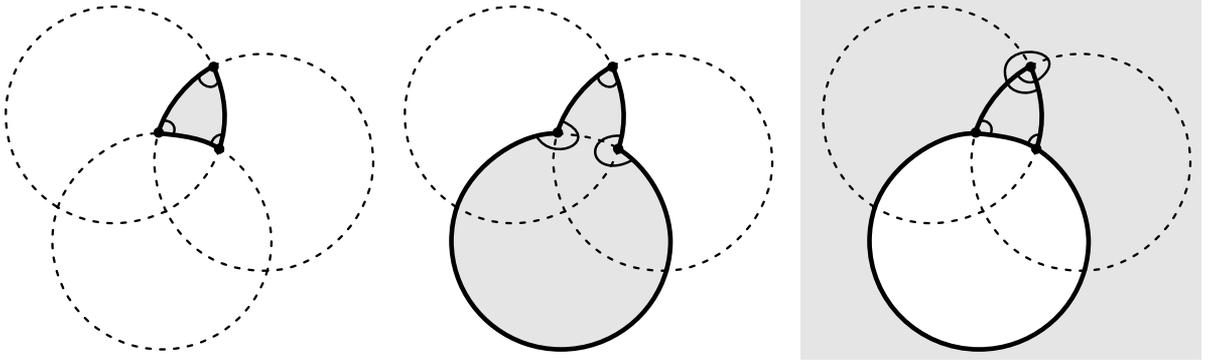
Figure 6: A triangular immersion can be manipulated to new triangular immersions by adding an entire disk, or by taking a full turn around a vertex.

(Figure 6). This operation increases the two angles adjacent to the pushed edge by π. The second manipulation involves making a full turn around a vertex, by extending the opposite edge to cover its entire supporting circle (Figure 6). This operation increases the angle at the highlighted vertex by 2π. It will be remarked later on how these operations are related to grafting the associated triangular structure (see Example 5.1.2).

On the other hand, there are triangular immersions that do not arise from these operations, whose existence we prove now.

Lemma 2.3.3 *Let* (a, b, c) *be an ordered triple of angles such that*

$$a \in (0, \pi) \cup (\pi, 2\pi) \quad \text{and} \quad b, c \in (0, \pi).$$

Then there is a configuration of circles $\mathfrak{C}$ *and a triangular immersion* τ *supported by* $\mathfrak{C}$ *with angles* (a, b, c).

Proof First suppose $a \in (0, \pi)$. Those cases where (a, b, c) satisfies one of the conditions (2.3.1), (2.3.2) or (2.3.3) from Lemma 2.3.1 are covered by that lemma. Hence suppose $a + b + c > \pi$, but at least one of the other inequalities in (2.3.3) is not satisfied. Up to permuting a, b, c we may assume that $a + \pi < b + c$. Let

$$\hat{a} = a, \quad \hat{b} = \pi - b, \quad \hat{c} = \pi - c.$$

Then $\hat{a} + \hat{b} + \hat{c} = a + (\pi - b) + (\pi - c) < \pi$ by assumption; therefore by Lemma 2.3.1 there is a hyperbolic configuration of circles $\mathfrak{C}$ and a triangular immersion $\hat{\tau}$ enclosed in $\mathfrak{C}$ with angles $(\hat{a}, \hat{b}, \hat{c})$. Figure 7 (on the left) shows that the same configuration of circles supports a triangular immersion with angles (a, b, c).

Now suppose $a \in (\pi, 2\pi)$. Consider the relations

(1) (i) $a + b + c > 3\pi$,
 (ii) $a + b + c = 3\pi$,
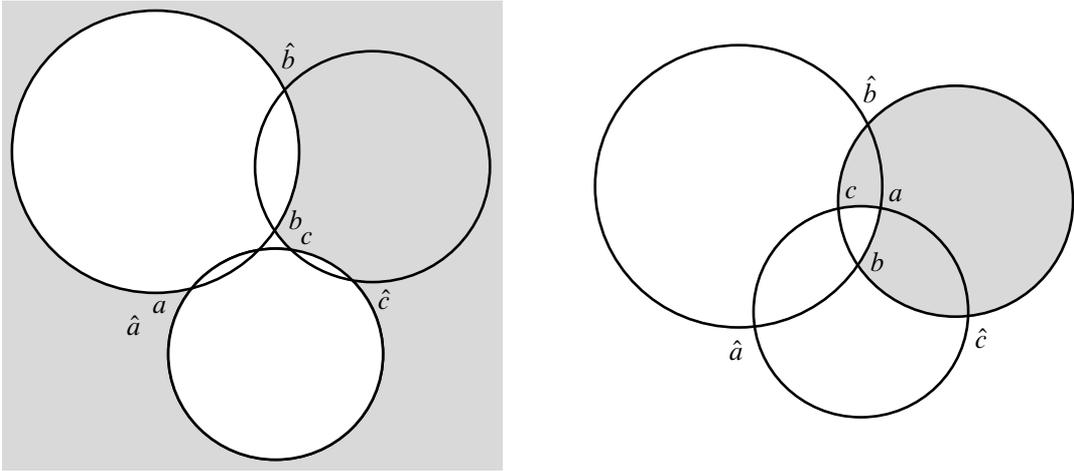
Figure 7: Left: a triangular immersion on a hyperbolic configuration with angles $(a, b, c) \in (0, \pi)^3$ such that $a + b + c > \pi$ but $a + \pi < b + c$. Right: a triangular immersion supported by a spherical configuration.

(2) (i) $a - b - c > \pi$,

    (ii) $a - b - c = \pi$,

(3) (i) $a - b + c \le \pi$,

    (ii) $a + b - c \le \pi$.

We observe that these three groups of inequalities are mutually exclusive, as any two of them imply the following contradictions:

$$(1) + (2) \implies a \ge 2\pi, \quad (1) + (3)(i) \implies b \ge \pi, \quad (2) + (3)(i) \implies c \le 0,$$

$$(3)(i) + (3)(ii) \implies a \le \pi, \quad (1) + (3)(ii) \implies c \ge \pi, \quad (2) + (3)(ii) \implies b \le 0.$$

If one of those inequalities is satisfied, we define

$$\hat{a} = 2\pi - a, \quad \hat{b} = \pi - b, \quad \hat{c} = \pi - c \quad \text{if (1)(i) is satisfied,}$$
$$\hat{a} = 2\pi - a, \quad \hat{b} = \pi - c, \quad \hat{c} = \pi - b \quad \text{if (1)(ii) is satisfied,}$$
$$\hat{a} = 2\pi - a, \quad \hat{b} = b, \quad\quad\ \hat{c} = c \quad\quad\ \text{if (2)(i) is satisfied,}$$
$$\hat{a} = 2\pi - a, \quad \hat{b} = c, \quad\quad\ \hat{c} = b \quad\quad\ \text{if (2)(ii) is satisfied,}$$
$$\hat{a} = a - \pi, \quad\ \ \hat{b} = \pi - b, \quad \hat{c} = c \quad\quad\ \text{if (3)(i) is satisfied,}$$
$$\hat{a} = a - \pi, \quad\ \ \hat{b} = b, \quad\quad\ \hat{c} = \pi - c \quad \text{if (3)(ii) is satisfied.}$$

In each case, the assumption implies that $\hat{a} + \hat{b} + \hat{c} \le \pi$; therefore Lemma 2.3.1 applies to give a Euclidean or hyperbolic configuration of circles $\mathfrak{C}$ and a triangular immersion $\hat{\tau}$ enclosed in $\mathfrak{C}$ with angles $(\hat{a}, \hat{b}, \hat{c})$. Figures 8 and 9 show that the same configuration of circles supports a triangular immersion with angles (a, b, c).
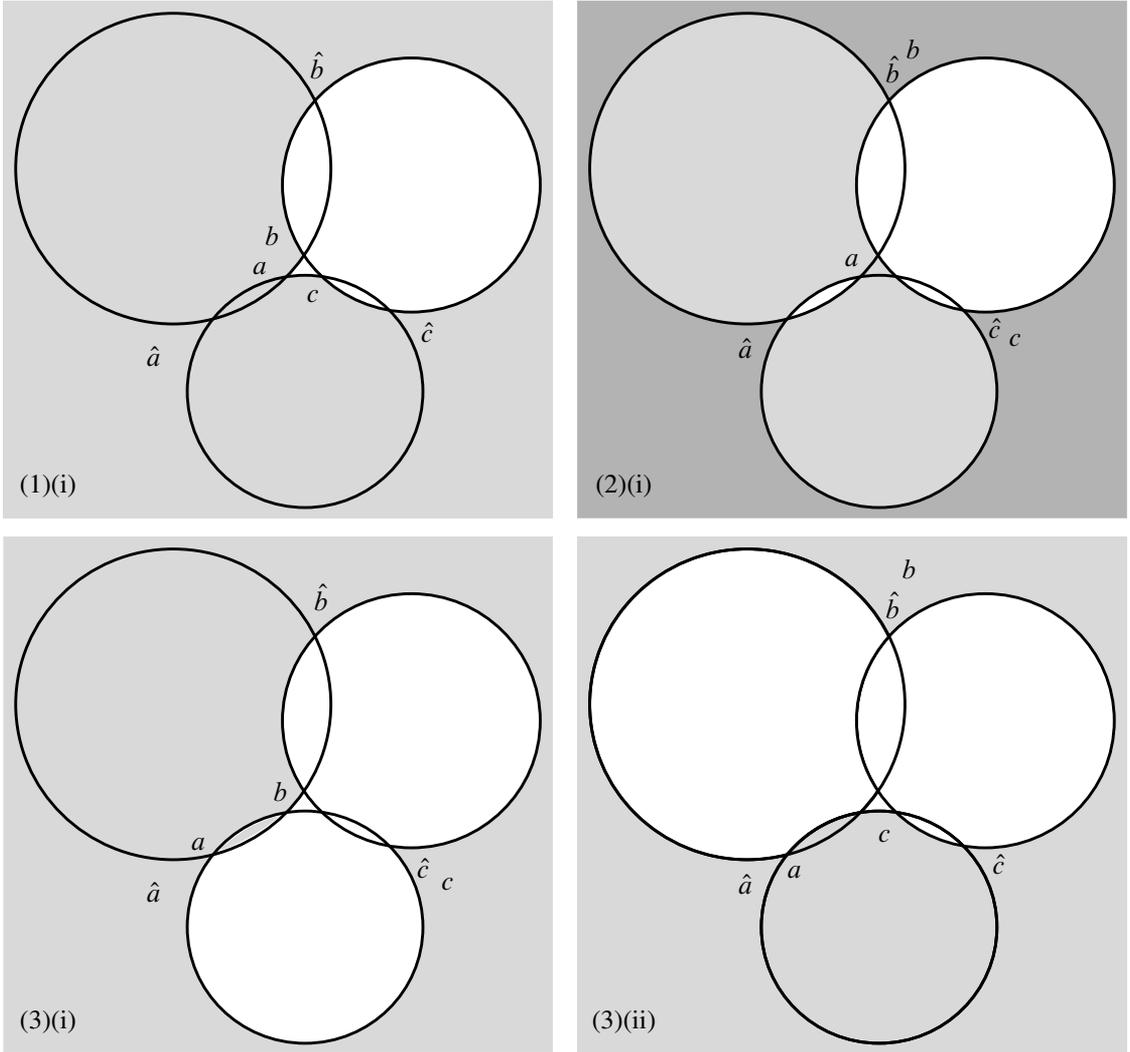
Figure 8: Different triangular immersions with angles (a, b, c), supported by hyperbolic configurations. We remark that the one depicted in (2) covers the darker triangle twice.

Finally, let $(\neg 1)$, $(\neg 2)$, $(\neg 3)(i)$ and $(\neg 3)(ii)$ be the opposite of the inequalities (1), (2), $(3)(i)$ and $(3)(ii)$, and suppose (a, b, c) satisfies all of $(\neg 1)$, $(\neg 2)$, $(\neg 3)(i)$ and $(\neg 3)(ii)$. We define

$$\hat{a} = 2\pi - a, \quad \hat{b} = \pi - b, \quad \hat{c} = \pi - c.$$

This time $\hat{a} + \hat{b} + \hat{c} = 4\pi - a - b - c > \pi$ because of $(\neg 1)$. Moreover,

$$\hat{a} + \pi = 3\pi - a > 2\pi - b - c = \hat{b} + \hat{c} \quad \text{by } (\neg 2),$$
$$\hat{b} + \pi = 2\pi - b > 3\pi - a - c = \hat{a} + \hat{c} \quad \text{by } (\neg 3)(i),$$
$$\hat{c} + \pi = 2\pi - c > 3\pi - a - b = \hat{a} + \hat{b} \quad \text{by } (\neg 3)(ii).$$
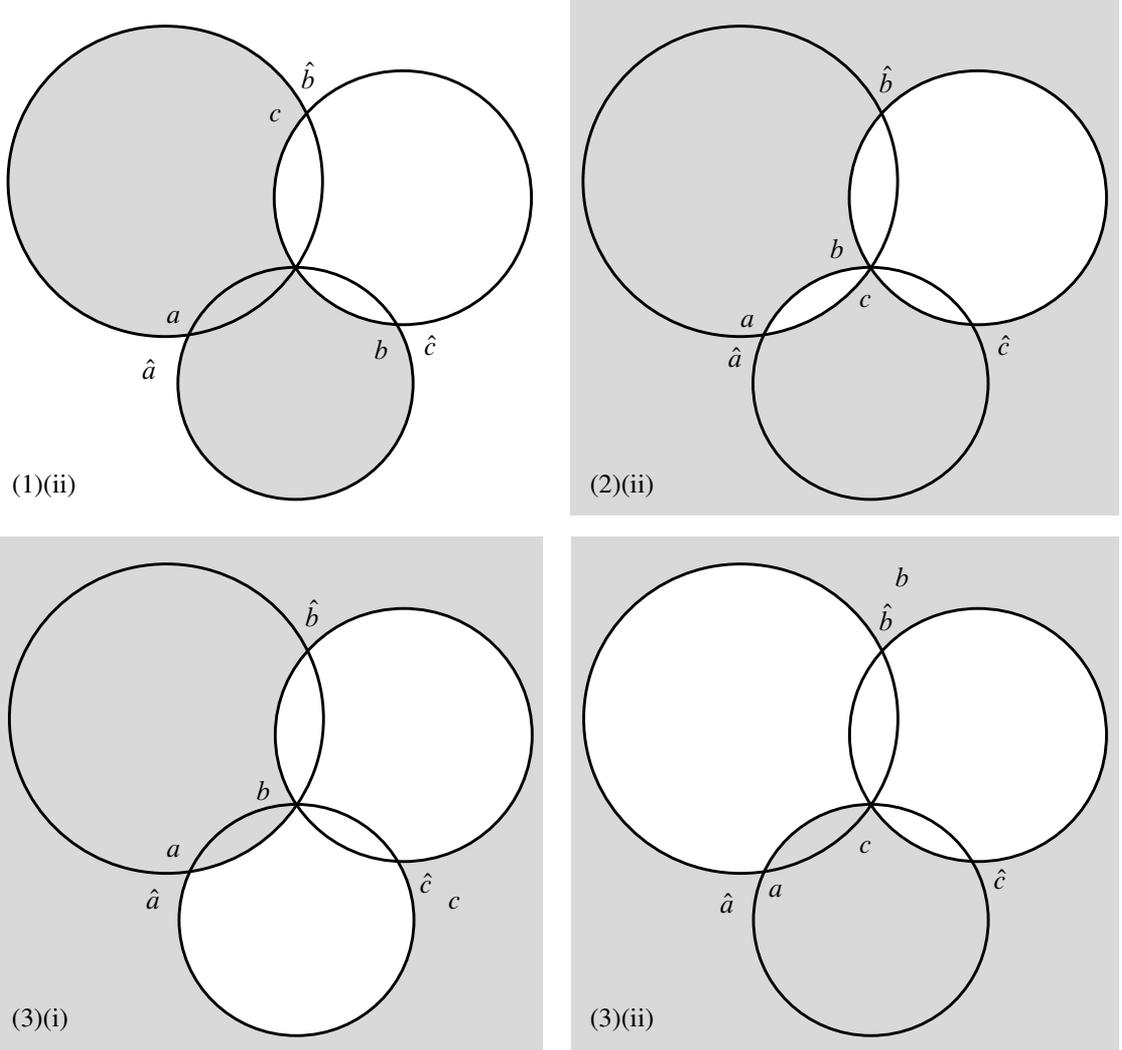
Figure 9: Different triangular immersions with angles (a, b, c), supported by Euclidean configurations.

By Lemma 2.3.1, there is a spherical configuration of circles $\mathfrak{C}$ and a triangular immersion $\hat{\tau}$ enclosed in $\mathfrak{C}$ with angles $(\hat{a}, \hat{b}, \hat{c})$. See Figure 7 for a triangular immersion with angles (a, b, c) supported by the same configuration $\mathfrak{C}$. $\square$

Due to the degenerate nature of Euclidean configurations, there is one additional case that needs to be considered, which we address in the next lemma.

Lemma 2.3.4 *Let* (a, b, c) *be an ordered triple of angles such that*

$$a \in (2\pi, 3\pi), \quad b, c \in (0, \pi), \quad a - b - c = \pi.$$

Then there is a configuration of circles $\mathfrak{C}$ *and a triangular immersion* τ *supported by* $\mathfrak{C}$ *with angles* (a, b, c).
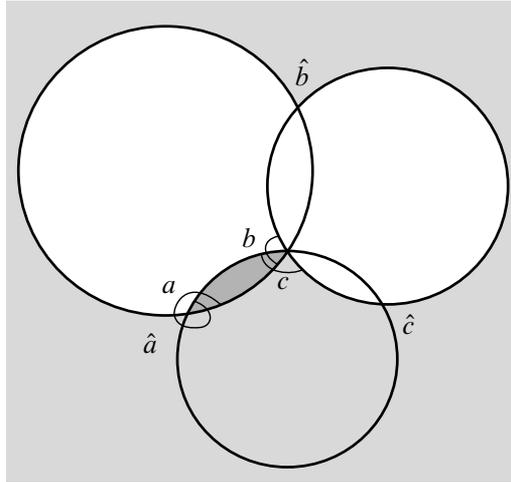
Figure 10: The additional triangular immersion mentioned in Lemma 2.3.4. Notice that $a > 2\pi$, hence the darker bigon is covered twice.

Proof Let
$$\hat{a} = a - 2\pi, \quad \hat{b} = \pi - b, \quad \hat{c} = \pi - c.$$

Then $\hat{a}, \hat{b}, \hat{c} \in (0, \pi)$ and $\hat{a} + \hat{b} + \hat{c} = a - 2\pi + \pi - b + \pi - c = \pi$; therefore by Lemma 2.3.1 there is a Euclidean configuration of circles $\mathfrak{C}$ and a triangular immersion $\hat{\tau}$ enclosed in $\mathfrak{C}$ with angles $(\hat{a}, \hat{b}, \hat{c})$. Figure 10 shows that the same configuration of circles supports a triangular immersion with angles (a, b, c). □

The triangular immersions constructed in the proofs of Lemmas 2.3.1, 2.3.3 and 2.3.4, which are depicted in Figures 7, 8, 9 and 10, are the starting point to construct all complex projective structures of interest in this paper. For this reason, we will refer to them as the *atomic* triangular immersions. They are *Euclidean/hyperbolic/spherical* depending on the type of the underlying configuration of circles. In Lemmas 2.3.3 and 2.3.4 exactly one angle is allowed to be larger than π, and we have assumed that to be the first one for simplicity. This normalization is inessential, and the same statements and proofs hold if one chooses a different angle to be the large one. This should be regarded as a change of marking (ie a permutation of the vertices of the simplex on which the triangular immersions are defined), and we call *atomic triangular immersion* any triangular immersion obtained in this way. Theorem B and Corollary 5.2.3 will show that, in a precise sense, this is indeed the minimal collection of triangular immersions to be considered.

We remark that the proofs of these lemmas are explicit, and construct a concrete collection of triangular immersions. Notice that for every triple of real numbers (a, b, c), two of which are in $(0, \pi)$ and one is in $(0, \pi) \cup (\pi, 2\pi) \cup (2\pi, 3\pi)$, there is a unique atomic triangular immersion with those angles. This allows us to organize the atomic triangular immersions in Tables 1, 2 and 3. We now define the other features listed in those tables.

Let $\tau\colon \triangle \to \mathbb{CP}^1$ be an atomic triangular immersion, and let $\mathfrak{C}_\tau$ be the configuration of circles that supports it. The configuration $\mathfrak{C}_\tau$ is either of spherical, Euclidean or hyperbolic *type*. The *target angles* of τ are the numbers $(\hat{a}, \hat{b}, \hat{c})$ defined as follows.

- If (a, b, c) satisfies the hypothesis of Lemma 2.3.1 then $(\hat{a}, \hat{b}, \hat{c}) = (a, b, c)$.

- If (a, b, c) does not satisfy the hypothesis of Lemma 2.3.1 then $(\hat{a}, \hat{b}, \hat{c})$ is defined as in the proofs of Lemmas 2.3.3 and 2.3.4, depending on what conditions are satisfied, and up to permuting the angles as appropriate.

The target angles of τ satisfy the hypothesis of Lemma 2.3.1. Therefore there is a triangular immersion $\hat{\tau}$ with angles $(\hat{a}, \hat{b}, \hat{c})$, which we call the *target triangular immersion*. If $\mathfrak{C}_\tau := (\mathscr{C}_{12}, \mathscr{C}_{23}, \mathscr{C}_{13})$, it follows from the construction that $\hat{\tau}$ is supported either by $\mathfrak{C}_\tau$ or by $\mathfrak{C}_\tau^* := (\mathscr{C}_{12}, \mathscr{C}_{13}, \mathscr{C}_{23})$, but the latter only happens in the Euclidean cases of Figure 4 (right) and Figure 9(1)–(2). In addition, $\hat{\tau}$ is always enclosed (while τ may not be). All the above pictures representing the atomic triangular immersions have been normalized so that $\hat{\tau}(\triangle)$ contains the point at infinity in its interior.

For pairwise distinct $i, j, k \in \{1, 2, 3\}$, consider the circle $\mathscr{C}_{ij} \in \mathfrak{C}_\tau$ supporting $\hat{\tau}(e_{ij})$; the intersection $\mathscr{C}_{ij} \cap \mathscr{C}_{jk}$ consists of two points: one is $\hat{\tau}(V_j)$, and we define $\hat{\tau}(V_j)'$ to be the other one. The collection $\{\hat{\tau}(V_j), \hat{\tau}(V_j)' \mid j = 1, 2, 3\}$ accounts for all the points of intersection of the circles in $\mathfrak{C}_\tau$, which are the possible vertices for τ. Note that by construction we always have $\{\tau(V_1), \hat{\tau}(V_1)\} \subseteq \mathscr{C}_{12} \cap \mathscr{C}_{13}$. We say a vertex $\tau(V_j)$ of τ is *positive* if there exists k such that $\tau(V_j) = \hat{\tau}(V_k)$, ie if it coincides with a vertex of $\hat{\tau}$, and we say it is *negative* otherwise. This defines a triple of *signs* $(s_1(\tau), s_2(\tau), s_3(\tau)) \in \{\pm\}^3$ associated to τ. In the Euclidean case, we additionally decorate this triple; we define it to be $(s_1(\tau), s_2(\tau), s_3(\tau))$ when $\hat{\tau}$ is supported by $\mathfrak{C}_\tau$, and to be $(s_1(\tau), s_2(\tau), s_3(\tau))^*$ when $\hat{\tau}$ is supported by $\mathfrak{C}_\tau^*$.

Remark 2.3.5 The Euclidean case (see Table 3) displays all possible cases for the triple of signs, including the extra $*$ decoration, with the only exception of the cases in which all vertices are negative. This cannot happen as it would mean that τ maps all vertices to the common intersection point of the configuration of circles, but this never happens for an atomic triangular immersion. The extra $*$ decoration is not needed for the hyperbolic and spherical cases as they are less degenerate than the Euclidean ones, in the sense that circles in $\mathfrak{C}$ have six distinct intersection points, which allows for more flexibility in the definition of the atomic immersions. See Tables 1 and 2. In the hyperbolic case we find all possible cases for the signs. In the spherical case we only see the triples $(\pm, \pm, \pm)$. This is because a spherical configuration of circles has only triangular complementary regions (while the complement of a hyperbolic configuration has different shapes, with only two triangles). As a result it is much easier for a spherical atomic triangular immersion to be enclosed, and equal to its own target triangular immersion.

Lemma 2.3.6 *Let τ be an atomic triangular immersions supported by a configuration of circles $\mathfrak{C}$. Then $\hat{\tau}$ is uniquely determined by $\mathfrak{C}$ and the vertices of τ.*
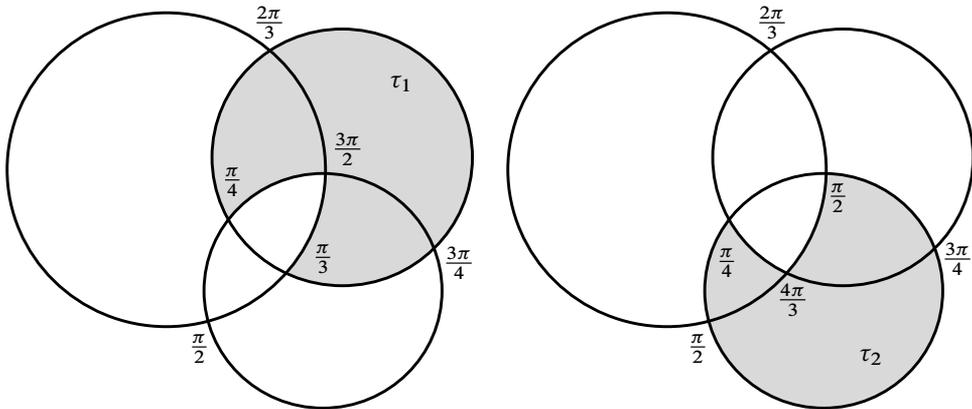
Figure 11: Two atomic triangular immersions supported by the same (spherical) configuration of circles, and with the same signs $(-, -, -)$.

Proof Let $\mathfrak{C} = (\mathscr{C}_{12}, \mathscr{C}_{23}, \mathscr{C}_{13})$ and recall that we have $\tau(e_{jk}) \subseteq \mathscr{C}_{jk}$ for $j, k = 1, 2, 3$, by definition of what it means for a triangular immersion to be supported by a configuration of circles. Moreover, by construction $\{\tau(V_1), \hat{\tau}(V_1)\} \subseteq \mathscr{C}_{12} \cap \mathscr{C}_{13}$.

Suppose that $\mathfrak{C}$ is Euclidean. Then $\hat{\tau}$ is the unique enclosed triangular immersion mapping to the Euclidean triangle such that $\hat{\tau}(V_1) = (\mathscr{C}_{12} \cap \mathscr{C}_{13}) \setminus \{\infty\}$.

Next, if $\mathfrak{C}$ is hyperbolic, then let $\mathscr{C} = \mathscr{C}_{\mathbb{H}}$ be the dual circle from Lemma 2.1.2. If $\mathfrak{C}$ is spherical, then let $\mathscr{C}$ be a circle which separates the vertices of τ from the other intersection points of circles in $\mathfrak{C}$. In either case $\hat{\tau}$ is the unique enclosed triangular immersion which has image disjoint from $\mathscr{C}$, is supported by $\mathfrak{C}$, and such that $\{\tau(V_1), \hat{\tau}(V_1)\} \subseteq \mathscr{C}_{12} \cap \mathscr{C}_{13}$. We additionally remark that $\hat{\tau}$ is always on the left of $\mathscr{C}$ with respect to the orientation induced by $\mathfrak{C}$. $\square$

Corollary 2.3.7 *Let $\mathfrak{C}$ be a configuration of circles. Let τ_1 and τ_2 be two atomic triangular immersions supported by $\mathfrak{C}$, such that $\tau_1(V_j) = \tau_2(V_j)$ for all $j \in \{1, 2, 3\}$. Then $\hat{\tau}_1 = \hat{\tau}_2$. Moreover, if (a_i, b_i, c_i) are the angles of τ_i, then exactly one of the following happens:*

(1) *$(a_1, b_1, c_1) = (a_2, b_2, c_2)$ and $\tau_1 = \tau_2$;*

(2) *$(a_1 - a_2, b_1 - b_2, c_1 - c_2) = (\pi, -\pi, 0)$ up to permutation.*

Proof The first assertion follows directly from Lemma 2.3.6. As a direct consequence, τ_1 and τ_2 have the same target angles and the same triple of signs. A direct inspection of Tables 1, 2 and 3 proves the desired relations between the angles, just by imposing equalities of the respective target angles. In particular, recall that atomic triangular immersions are uniquely determined by their angles; hence $(a_1, b_1, c_1) = (a_2, b_2, c_2)$ implies $\tau_1 = \tau_2$. $\square$

Example 2.3.8 Let τ_1 and τ_2 be two atomic triangular immersions with angles

$$(a_1, b_1, c_1) = \left(\tfrac{3\pi}{2}, \tfrac{\pi}{3}, \tfrac{\pi}{4}\right) \quad \text{and} \quad (a_2, b_2, c_2) = \left(\tfrac{\pi}{2}, \tfrac{4\pi}{3}, \tfrac{\pi}{4}\right).$$

These immersions correspond to the second and third row of Table 2, respectively. They are supported by the same spherical configuration of circles $\mathfrak{C}$, with target angles $(\hat{a}, \hat{b}, \hat{c}) = \left(\frac{\pi}{2}, \frac{2\pi}{3}, \frac{3\pi}{4} \right)$, and share the same signs $(-, -, -)$. In particular, $\hat{\tau}_1 = \hat{\tau}_2$. Furthermore, τ_1 can be transformed into τ_2 by first adding a disk and then removing another disk (see Figure 11).

Remark 2.3.9 Some of the sign invariants in each of Tables 1, 2 and 3 occur exactly once. If two triangular immersions have same such signs, then they are equal by Corollary 2.3.7 case (1). This applies for instance to atomic triangular immersions arising from Lemma 2.3.4, depicted in Figure 10.

3 Tame and relatively elliptic $\mathbb{CP}^1$–structures

In this chapter we define the geometric structures of interest in this paper, and study the geometry they induce on the universal cover. The reader can find the proofs of Theorems C and D in Sections 3.2 and 3.3 respectively.

Let $\overline{\Sigma}$ be a closed oriented surface and let $\{x_1, \ldots, x_n\} \subset \overline{\Sigma}$ be n distinct points such that the *punctured surface* $\Sigma := \overline{\Sigma} \setminus \{x_1, \ldots, x_n\}$ has negative Euler characteristic. If g is the genus of $\overline{\Sigma}$, this is equivalent to $2g + n > 2$, and it implies that Σ admits a complete hyperbolic metric of finite area. The points $\{x_1, \ldots, x_n\}$ are the *punctures* of Σ.

A *complex projective structure* ($\mathbb{CP}^1$–structure in short) on Σ is a maximal atlas of charts into $\mathbb{CP}^1$ with transition maps in $\mathrm{PSL}_2\mathbb{C}$ (see [Dumas 2009; Gunning 1967]). A $\mathbb{CP}^1$–structure can be described by a *developing pair* (dev, ρ) consisting of a *developing map* and a *holonomy representation*

$$\mathrm{dev} \colon \widetilde{\Sigma} \to \mathbb{CP}^1, \quad \rho \colon \pi_1(\Sigma) \to \mathrm{PSL}_2\mathbb{C},$$

satisfying the equivariance condition

$$\mathrm{dev}(\gamma \cdot x) = \rho(x) \cdot \mathrm{dev}(x) \quad \text{for all } x \in \widetilde{\Sigma}, \ \gamma \in \pi_1(\Sigma).$$

There is a natural equivalence relation on the set of complex projective structures on a surfaces for which two pairs (dev, ρ) and (dev', ρ') are equivalent if there is $A \in \mathrm{PSL}_2\mathbb{C}$ so that $\mathrm{dev}' = A \circ \mathrm{dev}$ and $\rho' = A\rho A^{-1}$ (up to isotopy of Σ). The *deformation space of marked $\mathbb{CP}^1$–structures on* Σ is the space of equivalence classes of complex projective structures and it is denoted by $\mathscr{P}(\Sigma)$. We denote by $\mathscr{R}(\Sigma)$ the *space of conjugacy classes of representations* of $\pi_1(\Sigma)$ into $\mathrm{PSL}_2\mathbb{C}$. We prefer not to use the GIT quotient because some of the representations of interest in this paper are reducible. The *holonomy map* is the forgetful map

$$\mathrm{Hol} \colon \mathscr{P}(\Sigma) \to \mathscr{R}(\Sigma), \quad [(\mathrm{dev}, \rho)] \mapsto [\rho].$$

Every $\mathbb{CP}^1$–structure has a natural underlying complex structure (or equivalently a conformal structure). We define $\mathscr{P}^\bullet(\Sigma) \subset \mathscr{P}(\Sigma)$ to be the subset of $\mathbb{CP}^1$–structures on Σ whose underlying conformal structure around every puncture is the complex punctured disk $\mathbb{D}^* := \{z \in \mathbb{C} \mid 0 < |z| < 1\}$.

The space of interest in this paper is the subspace $\mathcal{P}^{\circ}(\Sigma)$ of $\mathcal{P}^{\bullet}(\Sigma)$ of those structures whose developing map is *tame* and whose holonomy representation is *relatively elliptic*. We will define these terms in [Section 3.1](#).

3.1 Ends, framing, and grafting

Let $\widetilde{\Sigma}$ be the topological universal cover of Σ, and choose an identification $\widetilde{\Sigma} \cong \mathbb{H}^2$ coming from a uniformization of Σ as a complete hyperbolic surface of finite area. An *end* E of $\widetilde{\Sigma}$ is defined to be the fixed point of a parabolic deck transformation in the boundary of $\mathbb{H}^2$ in the closed disk model. For every puncture x of Σ, we denote by $\mathcal{E}_x(\widetilde{\Sigma})$ the *set of ends covering* x (see [Remark 3.1.1](#) for more details), and by

$$\mathcal{E}(\widetilde{\Sigma}) := \bigcup_x \mathcal{E}_x(\widetilde{\Sigma})$$

the *set of all ends*. The *end-extension* of $\widetilde{\Sigma}$ is the topological space $\widetilde{\Sigma}^{\#} = \widetilde{\Sigma} \cup \mathcal{E}(\widetilde{\Sigma})$, equipped with the topology generated by all open sets of $\widetilde{\Sigma}$ together with the *horocyclic neighborhoods* of the ends, ie sets of the form $N = N_0 \cup \{E\}$ where N_0 is an open disk in the closed disk model for $\mathbb{H}^2$ which is tangent to the boundary at E. The action of $\pi_1(\Sigma)$ on $\widetilde{\Sigma}$ naturally extends to a continuous (neither free nor proper) action on $\widetilde{\Sigma}^{\#}$. The quotient of $\mathcal{E}(\widetilde{\Sigma})$ by this action is precisely the set of punctures of Σ.

Remark 3.1.1 Ends cover the punctures of Σ in the sense that the universal cover projection $\widetilde{\Sigma} \to \Sigma$ admits a continuous extension to a map $\widetilde{\Sigma}^{\#} \to \overline{\Sigma}$. In particular, a sequence of points $x_n \in \widetilde{\Sigma}$ converges to an end $E \in \mathcal{E}(\widetilde{\Sigma})$ if and only if its projection to Σ is a sequence of points converging (in $\overline{\Sigma}$) to the puncture covered by E. This happens if and only if x_n eventually enters every horocyclic neighborhood of E.

Remark 3.1.2 Notice that $\widetilde{\Sigma}$ is open and dense in $\widetilde{\Sigma}^{\#}$, but this is not the same topology as the one induced from the closed disk model for $\mathbb{H}^2$. Indeed the topology of $\widetilde{\Sigma}^{\#}$ is strictly finer; the natural inclusion of $\widetilde{\Sigma}^{\#}$ into the closed disk is continuous but not open. Furthermore, the topology induced on the collection of ends is discrete, so $\widetilde{\Sigma}^{\#}$ is not compact. Actually not even locally compact, as ends do not have compact neighborhoods.

Recall that a *peripheral element* $\delta_x \in \pi_1(\Sigma)$ is the homotopy class of a *peripheral loop* (also denoted by δ_x) around the puncture x. If E_x is an end covering x, then δ_x is a generator of the stabilizer of E_x in $\pi_1(\Sigma)$. We make the convention that δ_x is the *positive* peripheral element if the corresponding peripheral loop is *positively* oriented, namely it turns anticlockwise around x (with respect to the orientation of Σ). This convention is chosen to match the convention that the angle between two circles is also taken in the anticlockwise direction.

Let $\sigma \in \mathcal{P}(\Sigma)$ be represented by a developing pair (dev, ρ). We say that σ is

- *tame at a puncture x* if dev admits a continuous extension

$$(\mathrm{dev}_{\#})_x \colon \widetilde{\Sigma} \cup \mathcal{E}_x(\widetilde{\Sigma}) \to \mathbb{CP}^1;$$

- *tame* if dev admits a continuous extension

$$\mathrm{dev}_\#\colon \widetilde{\Sigma} \cup \mathscr{E}(\widetilde{\Sigma}) \to \mathbb{CP}^1$$

  (note that this is equivalent to σ being tame at each puncture);

- *relatively elliptic* if the holonomy representation is relatively elliptic, ie the holonomy of every peripheral element is an elliptic Möbius transformation;

- *degenerate* if the holonomy representation is degenerate in the sense of [Gupta 2021, Definition 2.4], ie if either one of the following happens:

  – there are two points $p_\pm \in \mathbb{CP}^1$ such that the entire holonomy preserves the set $\{p_\pm\}$ and the holonomy of every peripheral element fixes $p_\pm$ individually;

  – there exists a point $p \in \mathbb{CP}^1$ such that the entire holonomy fixes p and the holonomy of every peripheral element is parabolic or identity.

The property of being degenerate is related (but not equivalent) to the more classical notions of *reducible* or *elementary* representations. In the case of punctured spheres, a degenerate representation is always reducible; on the other hand a representation generated by rotations of the Euclidean plane around different points is reducible but nondegenerate (see [Gupta 2021, Section 2.4] for a discussion).

The above notions are invariant under conjugation of representations in $\mathrm{PSL}_2\mathbb{C}$ and postcomposition of developing maps by Möbius transformations, thus they do not depend on the choice of representative pair (dev, ρ). The *deformation space of $\mathbb{CP}^1$–structures on Σ which are tame*, *relatively elliptic and nondegenerate* is $\mathscr{P}^\odot(\Sigma)$. The image of $\mathscr{P}^\odot(\Sigma)$ under the holonomy map is $\mathscr{R}^\odot(\Sigma) := \mathrm{Hol}(\mathscr{P}^\odot(\Sigma))$.

Lemma 3.1.3 *Let $\sigma \in \mathscr{P}(\Sigma)$ and let (dev, ρ) be a developing pair. Let x be a puncture and suppose that σ is tame at x. Let E_x be an end covering x and let $\delta_x \in \pi_1(S)$ be a peripheral element fixing it. Then*

 (1) *the map $(\mathrm{dev}_\#)_x$ is ρ–equivariant. In particular, the transformation $\rho(\delta_x)$ fixes $\mathrm{dev}_\#(E_x)$;*

 (2) *the transformation $\rho(\delta_x)$ is either trivial, parabolic or elliptic.*

Proof   (1) Follows by equivariance of dev and continuity of the extension $\mathrm{dev}_\#$.

 (2) Let $p := (\mathrm{dev}_\#)_x(E_x)$ be one of the fixed points of $\rho(\delta_x)$, and assume by contradiction that $\rho(\delta_x)$ is hyperbolic or loxodromic. Then it has another fixed point q and there is a $\rho(\delta_x)$–invariant simple arc ℓ joining them. Let η be an initial segment of ℓ starting at p and ending at some other point y on ℓ, and lift it to an arc $\tilde{\eta}$ starting at E_x. Consider the family of arcs $\tilde{\eta}_n := \delta_x^n \cdot \tilde{\eta}$, for $n \in \mathbb{Z}$. Up to replacing δ_x with its inverse, the sequence $\{(\mathrm{dev}_\#)_x(\tilde{\eta}_n)\}$ converges to the whole curve ℓ as $n \to +\infty$, and shrinks to p as $n \to -\infty$. Hence for all $n \in \mathbb{Z}^+$ there is a point $x_n \in \tilde{\eta}_n$ developing to y. Then we have $x_n \to E_x$ in the topology of $\widetilde{\Sigma}^\#$, but also $(\mathrm{dev}_\#)_x(x_n) = y \neq p$, which contradicts the continuity of $(\mathrm{dev}_\#)_x$ at E_x. $\square$

We will see in Section 4 that if $\sigma \in \mathscr{P}^\odot(\Sigma)$ then $\sigma \in \mathscr{P}^\bullet(\Sigma)$, ie the underlying complex structure is that of a punctured Riemann surface. More precisely, σ can be defined by a suitable meromorphic

quadratic differential with double poles (Theorem E). However $\mathscr{P}^{\odot}(\Sigma)$ is strictly contained in $\mathscr{P}^{\bullet}(\Sigma)$, as the following examples show.

Example 3.1.4 We now collect examples of structures in $\mathscr{P}^{\bullet}(\Sigma)$ which are or are not in $\mathscr{P}^{\odot}(\Sigma)$. These examples show that being tame and having relatively elliptic holonomy are independent concepts.

• All structures induced by Euclidean or hyperbolic metrics with cone points of angles $2\pi\theta$ are in $\mathscr{P}^{\odot}(\Sigma)$, when $\theta \notin \mathbb{N}$. For spherical metrics one has to additionally require that they do not have coaxial holonomy (see [Mondello and Panov 2016]).

• The structure induced by a complete hyperbolic metric of finite area is tame, but its holonomy is not relatively elliptic because peripherals have parabolic holonomy. Hence it is in $\mathscr{P}^{\bullet}(\Sigma)$ but not in $\mathscr{P}^{\odot}(\Sigma)$.

• Let σ_0 be the structure induced by a constant curvature metric with cone points of angles $2\pi\theta$, for $\theta \notin \mathbb{N}$. Remove disks centered at the cones, turn them into crowns and perform infinitely many graftings along arcs joining the crown tips. The resulting structure is in $\mathscr{P}^{\bullet}(\Sigma)$ and has relative elliptic holonomy, but it is not tame, hence it is not in $\mathscr{P}^{\odot}(\Sigma)$. This construction is described in [Gupta and Mj 2021], where it is shown that these structures arise from meromorphic quadratic differentials with poles of order at least 3 on punctured Riemann surfaces. Compare Example 3.2.10.

• Let $\sigma_0 \in \mathscr{P}(\bar{\Sigma})$ be the complex projective structure induced by a hyperbolic metric on the closed surface $\bar{\Sigma}$. Pick a simple closed geodesic and let σ_n be the structure obtained by grafting along it n times. For $n \to \infty$ we obtain a punctured surface Σ with two punctures (possibly disconnected if the geodesic is separating) which is endowed with a complex projective structure in $\mathscr{P}^{\bullet}(\Sigma)$ (see [Hensel 2011]). However it is not tame, and peripherals have hyperbolic holonomy, so it is not in $\mathscr{P}^{\odot}(\Sigma)$. Compare Example 3.2.11.

We conclude this section by observing that structures in $\mathscr{P}^{\odot}(\Sigma)$ carry some additional piece of information which can be regarded as a decoration of the holonomy representation. A *framing* for a representation $\rho: \pi_1(\Sigma) \to \mathrm{PSL}_2\mathbb{C}$ consists of a choice of a fixed point in $\mathbb{CP}^1$ for the holonomy about each puncture (compare [Allegretti and Bridgeland 2020; Gupta 2021]). When considering representations up to conjugacy (as we do), a framing can equivalently be defined as a ρ–equivariant map $\mathscr{F}: \mathscr{E}(\tilde{\Sigma}) \to \mathbb{CP}^1$ from the space of ends to $\mathbb{CP}^1$. A framing is said to be *degenerate* if one of the following occurs (compare [Gupta 2021, Section 2.5]):

• $\mathscr{F}(\mathscr{E}(\tilde{\Sigma}))$ consists of two points, preserved as a set by every element, and fixed individually by the holonomy at every puncture;

• $\mathscr{F}(\mathscr{E}(\tilde{\Sigma}))$ consists of one point, fixed by every element, and the holonomy at every puncture is either parabolic or the identity.

Every framing of every nondegenerate representation is nondegenerate (see [Gupta 2021, Proposition 3.1]). In general, a $\mathbb{CP}^1$–structure can be framed in different ways, by arbitrarily picking the fixed point for each peripheral curve. However, tame structures can be canonically framed.
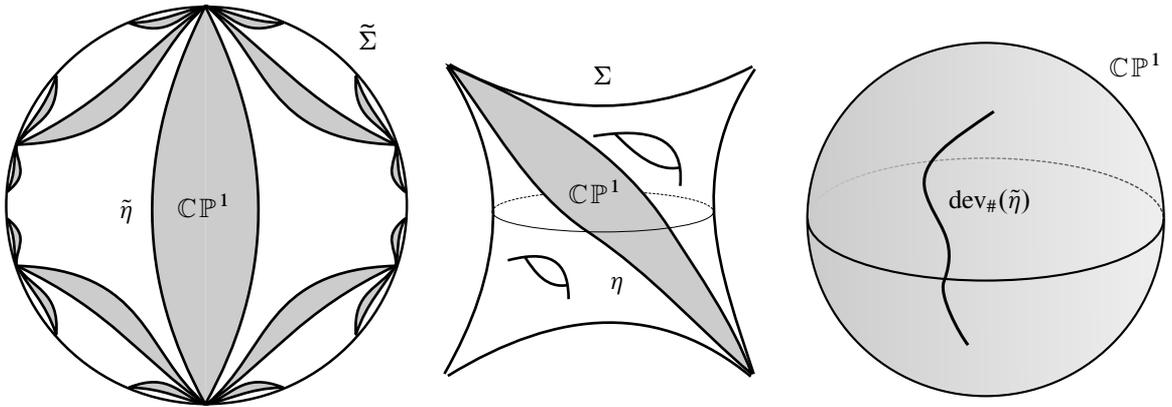
Figure 12: The structure on $\widetilde{\Sigma}$ induced by grafting a structure along a curve η.

Corollary 3.1.5 *Let $\sigma \in \mathscr{P}^{\odot}(\Sigma)$. Then the extension of a developing map provides a nondegenerate canonical framing for the holonomy.*

Proof Let (dev, ρ) be a developing pair defining σ. By Lemma 3.1.3 we know dev extends naturally to a map $\mathrm{dev}_{\#}$ on the space of ends. The restriction $\mathscr{F} = \mathrm{dev}_{\#}|_{\mathscr{E}(\widetilde{\Sigma})}$ provides the desired framing. The framing is nondegenerated because ρ itself is a nondegenerate representation. $\qquad\square$

In the following, whenever dealing with a structure $\sigma \in \mathscr{P}^{\odot}(\Sigma)$, we assume that this natural framing $\mathscr{F}$ has been chosen for its holonomy representation, and refer to the pair $(\rho, \mathscr{F})$ as its *framed holonomy*.

In this paper we are mostly interested in a surgery that can be used to deform $\mathbb{CP}^1$–structures and explore their moduli space. It was introduced by Maskit [1969] and later developed in unpublished work of Thurston (see [Baba 2020; Dumas 2009; Kamishima and Tan 1992] for some accounts). The specific version we are interested in is designed to create new structures from old ones without changing their holonomy. For convenience we define it just in the setting of $\mathbb{CP}^1$–structures in $\mathscr{P}^{\odot}(\Sigma)$.

Let $\sigma \in \mathscr{P}^{\odot}(\Sigma)$, and let $\eta\colon I \to \Sigma$ be an ideal arc (ie with endpoints in the set of punctures). We say η is *graftable* if it is simple and injectively developed, ie $\mathrm{dev}_{\#}$ is injective on some (every) lift of η to $\widetilde{\Sigma}^{\#}$, all the way to the ends. In particular, the two endpoints develop to two distinct points. When η is graftable, the developed image of any of its lifts $\mathrm{dev}_{\#}(\tilde{\eta})$ is a simple arc in $\mathbb{CP}^1$; hence $\mathbb{CP}^1 \setminus \mathrm{dev}_{\#}(\tilde{\eta})$ is a topological disk, endowed with a natural $\mathbb{CP}^1$–structure, which we call a *grafting region*.

Let $\sigma \in \mathscr{P}^{\odot}(\Sigma)$ and let $\eta\colon I \to \Sigma$ be a graftable arc. The *grafting* of σ along η is the $\mathbb{CP}^1$–structure $\mathrm{Gr}(\sigma, \eta)$ obtained by the following procedure: for each lift $\tilde{\eta}$ of η to the universal cover, cut $\widetilde{\Sigma}$ along $\tilde{\eta}$ and glue in a copy of the disk $\mathbb{CP}^1 \setminus \mathrm{dev}_{\#}(\tilde{\eta})$ using $\mathrm{dev}_{\#}$ as a gluing map. The obvious inverse operation is called *degrafting*. The structure on $\widetilde{\Sigma}$ induced by $\mathrm{Gr}(\sigma, \eta)$ looks like the union of the one induced by σ together with an equivariant collection of grafting regions, glued along all the possible lifts of η (see Figure 12).

Remark 3.1.6 If two graftable arcs η and η' have the same endpoints and are isotopic through graftable curves, then $\mathrm{Gr}(\sigma, \eta) = \mathrm{Gr}(\sigma, \eta')$, and $\mathrm{Gr}(\sigma, \eta)$ is graftable again along η (see [Calsamiglia et al. 2014a, Lemma 2.8] or [Ruffoni 2021, Section 2] for details). On the other hand, if η and η' are disjoint, then $\mathrm{Gr}(\sigma, \eta)$ (resp. $\mathrm{Gr}(\sigma, \eta')$) is graftable along η' (resp. η), and $\mathrm{Gr}(\mathrm{Gr}(\sigma, \eta), \eta') = \mathrm{Gr}(\mathrm{Gr}(\sigma, \eta'), \eta)$.

More generally, a grafting surgery can be defined along any graftable measured lamination on a $\mathbb{CP}^1$–structure, and the reader familiar with grafting deformations will identify the type of grafting introduced here as a type of projective 2π–grafting (see [Baba 2020; Dumas 2009; Kamishima and Tan 1992] for details). We record the following statement for future reference.

Lemma 3.1.7 *Let* $\sigma \in \mathscr{P}^{\odot}(\Sigma)$ *and* $\eta \colon I \to \Sigma$ *be a graftable arc. Then*

(1) $\mathrm{Hol}(\mathrm{Gr}(\sigma, \eta)) = \mathrm{Hol}(\sigma)$ (*ie grafting preserves the holonomy*),

(2) $\mathrm{Gr}(\sigma, \eta) \in \mathscr{P}^{\odot}(\Sigma)$,

(3) *grafting does not change the developed images of the punctures* (*ie grafting preserves the framed holonomy*).

Proof The first statement is well known in the literature for this type of grafting (see for instance [Baba 2020]). The statements about tameness and framing follow by pasting together the developing map for σ and the natural embedding of the grafting regions in $\mathbb{CP}^1$. □

3.2 The Möbius completion

In this section we prove Theorem C. Henceforth we fix a complex projective structure $\sigma \in \mathscr{P}(\Sigma)$ with developing pair (dev, ρ). First of all we recall the definition of a natural projective completion of $\widetilde{\Sigma}$ defined in terms of σ (see [Kulkarni and Pinkall 1994] for details). Let g_0 be a conformal Riemannian metric on $\mathbb{CP}^1$ (eg the standard spherical metric). Let $g := \mathrm{dev}^*(g_0)$ be the metric on $\widetilde{\Sigma}$ obtained by pullback, and let d be the associated distance function, ie

$$d(x, y) := \inf\{\ell_g(\eta) \mid \eta \colon [0, 1] \to \widetilde{\Sigma} \text{ is a rectifiable arc from } x \text{ to } y\}$$

where $\ell_g(\eta)$ denotes the length of γ with respect to the metric g. Notice that g is generally not invariant under deck transformations. By construction $(\widetilde{\Sigma}, d)$ is a path-connected length space. It is locally path-connected, but not necessarily geodesic. Moreover it is locally compact, but in general not proper, nor complete.

The *Möbius completion* $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ of $\widetilde{\Sigma}$ with respect to σ is defined to be the metric completion of $(\widetilde{\Sigma}, d)$. The subspace $\partial_{\infty}^{\sigma}(\widetilde{\Sigma}) := \mathrm{M}^{\sigma}(\widetilde{\Sigma}) \setminus \widetilde{\Sigma}$ is called the *ideal boundary* of $\widetilde{\Sigma}$ with respect to σ. We collect the following facts from [Kulkarni and Pinkall 1994, Section 2]:

(1) Different choices of the metric g_0 on $\mathbb{CP}^1$ or of the developing map for σ result in metrics on $\widetilde{\Sigma}$ having the same underlying uniform structure. So $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ does not depend (up to homeomorphism) on these choices.

(2) $\mathrm{dev}\colon \widetilde{\Sigma} \to \mathbb{CP}^1$ extends continuously to a map $\mathrm{dev}_\sigma\colon \mathrm{M}^\sigma(\widetilde{\Sigma}) \to \mathbb{CP}^1$.

(3) The action of $\pi_1(\Sigma)$ by deck transformations extends to an action by homeomorphisms on the Möbius completion.

Lemma 3.2.1 *The map dev_σ is ρ–equivariant.*

Proof Let $\xi \in \partial_\infty^\sigma(\widetilde{\Sigma})$ and let $x_n \in \widetilde{\Sigma}$ a Cauchy sequence converging to ξ. Then by continuity of dev_σ,

$$\mathrm{dev}_\sigma(\gamma \cdot \xi) = \mathrm{dev}_\sigma\big(\gamma \cdot \lim_{n\to\infty} x_n\big) = \lim_{n\to\infty} \mathrm{dev}_\sigma(\gamma \cdot x_n)$$
$$= \rho(\gamma) \cdot \lim_{n\to\infty} \mathrm{dev}_\sigma(x_n) = \rho(\gamma) \cdot \mathrm{dev}_\sigma(\xi). \qquad \square$$

Lemma 3.2.2 $\mathrm{M}^\sigma(\widetilde{\Sigma})$ *is a complete, path-connected and locally path-connected length space.*

Proof Completeness is trivial by construction. The completion of a length space is a length space (see for instance [Bridson and Haefliger 1999, I.3.6(3)]). Since $\widetilde{\Sigma}$ is path-connected and $\mathrm{M}^\sigma(\widetilde{\Sigma})$ is a length space, it follows that $\mathrm{M}^\sigma(\widetilde{\Sigma})$ is path-connected. Analogously one can obtain that $\mathrm{M}^\sigma(\widetilde{\Sigma})$ is locally path-connected. $\qquad \square$

The following examples describe more explicitly the Möbius completion for projective structures defined by certain constant curvature metrics. Notice they are both examples of hyperbolic Möbius structures with respect to the terminology introduced in [Kulkarni and Pinkall 1994, Section 2].

Example 3.2.3 Let $\sigma = (\mathrm{dev}, \rho)$ be defined by a complete hyperbolic metric of finite area on Σ. In this case $\mathrm{M}^\sigma(\widetilde{\Sigma})$ is homeomorphic to a closed disk, and $\partial_\infty^\sigma(\widetilde{\Sigma})$ to a circle. Ideal points are either ends, or limit points of complete lifts of closed geodesics. Indeed, $\rho\colon \pi_1(\Sigma) \to \mathrm{PSL}_2\mathbb{R}$ is an isomorphism onto Fuchsian group, and $\mathrm{dev}\colon \widetilde{\Sigma} \to \mathbb{CP}^1$ is a ρ–equivariant diffeomorphism with an open hemisphere.

Example 3.2.4 Let $\sigma = (\mathrm{dev}, \rho)$ be defined by a spherical metric on Σ, with cone singularities at the punctures. In this case $\mathrm{M}^\sigma(\widetilde{\Sigma}) = \widetilde{\Sigma}^\#$ is homeomorphic to the end-extension, and $\partial_\infty^\sigma(\widetilde{\Sigma}) = \mathscr{E}(\widetilde{\Sigma})$. Indeed, the action of $\pi_1(\Sigma)$ on $\widetilde{\Sigma}$ preserves a spherical metric and admits a fundamental domain D given by a geodesic spherical polygon having finite area A and all the vertices in the set of ends. Notice that each pair of nonintersecting edges of this polygon has positive finite distance, and let $L > 0$ be the minimum of such distances. Pick $\xi \in \partial_\infty^\sigma(\widetilde{\Sigma})$, and a rectifiable curve of finite length $\gamma\colon [0, 1) \to \mathrm{M}^\sigma(\widetilde{\Sigma})$ tending to ξ. If γ intersects finitely many fundamental domains, then it is eventually contained in a single one, hence ξ must be an end. If γ intersects infinitely many domains D_n, then the length of the arcs $\gamma \cap D_n$ converges to zero, so is eventually less than L. In particular, eventually all the domains D_n share a common vertex. By construction this vertex is an end and γ converges to it, which forces ξ to be an end.

Lemma 3.2.5 *For all $x \in \widetilde{\Sigma}$, $\xi \in \partial_\infty^\sigma(\widetilde{\Sigma})$ and $c > 0$ there is a continuous curve $\eta_c\colon [0, 1) \to \widetilde{\Sigma}$ such that $\eta_c(0) = x$, $\lim_{t\to 1} \eta_c(t) = \xi$ and $d(x, \xi) \leq \ell(\eta_c) \leq d(x, \xi) + c$.*

Proof By definition, $d(x, \xi) = \lim_{n \to \infty} d(x, y_n)$ for any Cauchy sequence $\{y_n\}$ converging to ξ. Let $\{y_n\}$ be a Cauchy sequence in $\widetilde{\Sigma}$ converging to ξ such that

$$d(x, y_i) \le d(x, \xi) + \frac{1}{i} \quad \text{and} \quad d(y_i, y_{i+1}) \le \frac{1}{i^2}.$$

Such sequence can be easily constructed from any Cauchy sequence by taking an appropriate subsequence. Since $\widetilde{\Sigma}$ is a length space, for all k there is a continuous curve $\gamma_k \colon [0, 1] \to \widetilde{\Sigma}$ such that $\gamma_k(0) = y_k$, $\gamma_k(1) = y_{k+1}$ and

$$\ell(\gamma_k) \le d(y_k, y_{k+1}) + \frac{1}{k^2} = \frac{2}{k^2}.$$

By concatenating these curves, for every i, we obtain a continuous curve $\gamma_i \colon [0, 1) \to \widetilde{\Sigma}$ such that $\gamma_i(0) = y_i$, $\lim_{t \to 1} \gamma_i(t) = \xi$, and

$$\ell(\gamma_i) \le \sum_{k=i}^{\infty} \frac{2}{k^2} =: T_i.$$

In particular, $\lim_{i \to \infty} \ell(\gamma_i) = \lim_{i \to \infty} T_i = 0$. Finally, let $\eta_i \colon [0, 1) \to \widetilde{\Sigma}$ be a continuous curve such that $\eta_i(0) = x$, $\eta_i(1) = y_i$, and

$$\ell(\eta_i) \le d(x, y_i) + \frac{1}{i^2}.$$

Let $f_i \colon [0, 1) \to \widetilde{\Sigma}$ be the continuous curve obtained by concatenating η_i with γ_i. Then f_i is a continuous curve such that $f_i(0) = x$, $\lim_{t \to 1} f_i(t) = \xi$ and

$$d(x, \xi) \le \ell(f_i) = \ell(\eta_i) + \ell(\gamma_i) \le d(x, y_i) + \frac{1}{i^2} + T_i \le d(x, \xi) + \frac{1}{i} + \frac{1}{i^2} + T_i.$$

Now let i such that $1/i + 1/i^2 + T_i < c$ and take $f_c := f_i$. $\square$

Lemma 3.2.6 *Let $\xi \in \partial_\infty^\sigma(\widetilde{\Sigma})$ and $\varepsilon > 0$. Then $B_\sigma(\xi, \varepsilon) \cap \widetilde{\Sigma}$ is path-connected.*

Proof First of all let us show that each path-component N of $B_\sigma(\xi, \varepsilon) \cap \widetilde{\Sigma}$ contains points arbitrarily close to ξ. Pick a base point $x \in N$, and let $R = d(x, \xi)$; notice $R < \varepsilon$. By Lemma 3.2.5 for all $c > 0$ we can pick a continuous curve $\eta_c \colon [0, 1) \to \widetilde{\Sigma}$ such that $\eta_c(0) = x$, $\lim_{t \to 1} \eta_c(t) = \xi$ and $R \le \ell(\eta_c) \le R + c$. For each $t \in [0, 1)$ we have

$$d(\eta_c(t), \xi) \le \ell(\eta_c([t, 1))) \le \ell(\eta_c([0, 1))) \le R + c.$$

In particular, for $c < \frac{1}{2}(\varepsilon - R)$ we get that $d(\eta_c(t), \xi) < \varepsilon$, ie η_c is entirely contained in $B_\sigma(\xi, \varepsilon) \cap \widetilde{\Sigma}$. Since it is a curve starting at x, it is then entirely contained in N; since it converges to ξ we get $\lim_{t \to 1} d(\eta_c(t), \xi) = 0$.

Suppose by contradiction that $B_\sigma(\xi, \varepsilon) \cap \widetilde{\Sigma}$ admits at least two different path-components N_1 and N_2. Let $x_k \in N_k$ be two points such that $d(x_k, \xi) < \frac{1}{4}\varepsilon$. In particular, $d(x_1, x_2) < \frac{1}{2}\varepsilon$. Since $(\widetilde{\Sigma}, d)$ is a length space, for every $\delta > 0$ we can find a continuous curve $\gamma_\delta \colon [0, 1] \to \widetilde{\Sigma}$ joining x_1 to x_2 of length at most $\frac{1}{2}\varepsilon + \delta$. Let now $z \in \gamma_\delta$. Without loss of generality let us assume that $d(z, x_1) \le d(z, x_2)$, so that by the triangle inequality we get

$$d(z, \xi) \le d(z, x_1) + d(x_1, \xi) \le \frac{1}{2}\left(\frac{1}{2}\varepsilon + \delta\right) + \frac{1}{4}\varepsilon = \frac{1}{2}\varepsilon + \frac{1}{4}\delta.$$

In particular, for each $\delta < \varepsilon$ we get that the curve η_δ is at distance at most ε from ξ. In particular, it is entirely contained in $B_\sigma(\xi, \varepsilon) \cap \widetilde{\Sigma}$, which contradicts the fact that x_1 and x_2 are in distinct path-components. $\qquad\square$

Our next goal is to define a cyclic order on $\partial_\infty^\sigma(\widetilde{\Sigma})$, which will induce a total order on $\partial_\infty^\sigma(\widetilde{\Sigma}) \setminus \{\xi\}$, for any $\xi \in \mathscr{E}(\widetilde{\Sigma})$.

Lemma 3.2.7 *For any pair of distinct points $(\xi_0, \xi_1) \in \partial_\infty^\sigma(\widetilde{\Sigma})$ there exists a simple continuous curve $\gamma \colon (0, 1) \to \widetilde{\Sigma}$ such that $\lim_{t \to 0} \gamma(t) = \xi_0$ and $\lim_{t \to 1} \gamma(t) = \xi_1$.*

Moreover, for any such curve γ, the space $M^\sigma(\widetilde{\Sigma}) \setminus Cl(\gamma)$ has exactly two path-components, which we call the left and right components, $C_L(\gamma)$ and $C_R(\gamma)$, with respect to the orientation of γ. The induced partition of $\partial_\infty^\sigma(\widetilde{\Sigma})$ as

$$\{\xi_0, \xi_1\} \cup (\partial_\infty^\sigma(\widetilde{\Sigma}) \cap C_L(\gamma)) \cup (\partial_\infty^\sigma(\widetilde{\Sigma}) \cap C_R(\gamma))$$

only depends on the ordered pair (ξ_0, ξ_1) and not on γ.

Proof Existence of γ is clear, for instance by Lemma 3.2.5. Let us show that its complement consists of exactly two path-components. $\widetilde{\Sigma} \setminus \gamma$ clearly has exactly two path components, so $M^\sigma(\widetilde{\Sigma}) \setminus Cl(\gamma)$ has at most two components (again by Lemma 3.2.5). We need to show that no ideal point can be joined by an arc to both components. This follows from Lemma 3.2.6.

To show that the induced decomposition of $\partial_\infty^\sigma(\widetilde{\Sigma})$ does not depend on the choice of γ, just notice that any two such curves are isotopic relatively to their endpoints in $\widetilde{\Sigma}$. $\qquad\square$

Hence we denote by $C_L(\xi_0, \xi_1) := \partial_\infty^\sigma(\widetilde{\Sigma}) \cap C_L(\gamma)$ and $C_R(\xi_0, \xi_1) = \partial_\infty^\sigma(\widetilde{\Sigma}) \cap C_R(\gamma)$ for any curve γ as in Lemma 3.2.7. We define the following ternary relation on $\partial_\infty^\sigma(\widetilde{\Sigma})$. If $\xi_0, \xi_1, \zeta \in \partial_\infty^\sigma(\widetilde{\Sigma})$ then we say they are in relation (denoted by $[\xi_0, \zeta, \xi_1]$) if $\zeta \in C_R(\xi_0, \xi_1)$, ie ζ is on the right of γ.

Remark 3.2.8 This relation defines a $\pi_1(\Sigma)$–invariant cyclic order on $\partial_\infty^\sigma(\widetilde{\Sigma})$.

The goal of the rest of this section is to explore the features of the Möbius completion and the ideal boundary in the case of structures from $\mathscr{P}^\odot(\Sigma)$.

Proposition 3.2.9 *A structure σ is tame if and only if the natural embedding $j_\sigma \colon \widetilde{\Sigma} \hookrightarrow M^\sigma(\widetilde{\Sigma})$ extends to a $\pi_1(\Sigma)$–equivariant continuous embedding $j_\sigma^\# \colon \widetilde{\Sigma}^\# \hookrightarrow M^\sigma(\widetilde{\Sigma})$. Moreover in this case $dev_\# = dev_\sigma \circ j_\sigma^\#$.*

Proof First assume the existence of a $\pi_1(\Sigma)$–equivariant continuous embedding $j_\sigma^\# \colon \widetilde{\Sigma}^\# \hookrightarrow M^\sigma(\widetilde{\Sigma})$. As remarked above there exists a continuous extension dev_σ of dev to $M^\sigma(\widetilde{\Sigma})$. Then $dev_\sigma \circ j^\sigma$ provides a continuous extension of dev to $\widetilde{\Sigma}^\#$, ie σ is tame.

Conversely let σ be tame, let $E \in \mathscr{E}(\widetilde{\Sigma})$ and $p_E = dev_\#(E)$. Since dev extends continuously to E, for all $\varepsilon > 0$ the set $N_\varepsilon = (dev_\#)^{-1}(B(p_E, \varepsilon))$ is an open neighborhood of E in $\widetilde{\Sigma}^\#$, containing points at distance

at most ε from E. Therefore we can construct a Cauchy sequence x_n in $\widetilde{\Sigma}$ converging to E (in $\widetilde{\Sigma}^\#$). We can associate to E the limit of x_n in the completion $\mathrm{M}^\sigma(\widetilde{\Sigma})$. Suppose y_n is another Cauchy sequence in $\widetilde{\Sigma}$ converging to E (in $\widetilde{\Sigma}^\#$). By definition of the topology on $\widetilde{\Sigma}$, continuity of $\mathrm{dev}_\#$ at E implies that $\mathrm{dev}_\#(x_n)$ and $\mathrm{dev}_\#(y_n)$ both converge to p_E. Hence y_n eventually enters each neighborhood N_ε. As a result we get $d(x_n, y_n) \leq 2\varepsilon$, which implies that the two sequences give rise to the same point in the completion. This defines the desired extension, which is (sequentially) continuous. Injectivity follows from the fact that any two ends are at a positive distance from each other. Moreover $\mathrm{dev}_\# = \mathrm{dev}_\sigma \circ j_\sigma^\#$ because they agree on the dense subset $\widetilde{\Sigma}$ and $\mathbb{CP}^1$ is Hausdorff. $\qquad\square$

In particular, tame structures have infinitely many ideal points, hence they are of hyperbolic type with respect to the classification in [Kulkarni and Pinkall 1994]. Moreover it should be noticed that ends do not have compact neighborhoods, so the completion fails to be locally compact or proper.

Example 3.2.10 Gupta and Mj [2021] considered structures obtained by grafting crowned hyperbolic surfaces, and showed that the local structure at the crown can be modeled by a meromorphic differential with a pole of sufficiently high order. For such a structure, every sequence going off to a puncture gives rise to an ideal point in the Möbius completion, but sequences converging in different Stokes sectors develop to sequences converging to different limit points in $\mathbb{CP}^1$, hence give rise to different ideal points in the Möbius completion. They are not tame structures (as observed in Example 3.1.4), and the space of ends does not embed continuously in their ideal boundary. Notice that Lemma 3.2.6 applies to each individual ideal point, while the intersection of $\widetilde{\Sigma}$ with the neighborhood of an end can fail to be connected.

Example 3.2.11 For a more extreme behavior, take a closed hyperbolic surface, and graft it along a geodesic pants decomposition infinitely many times. The underlying complex structure is being pinched along each pants curve, and in the limit the structure decomposes into a collection of thrice-punctured spheres (see [Hensel 2011, Section 6]). There, punctures do not give rise to well-defined ideal points; indeed, the structure has hyperbolic peripheral holonomy, hence it is not tame (by Lemma 3.1.3).

Remark 3.2.12 In general the embedding $j_\sigma^\#$ in Proposition 3.2.9 is not open. For instance consider the tame relatively parabolic structure induced by a complete finite area hyperbolic metric. In this case the completion is the closed disk, and we have already observed in Remark 3.1.2 that inclusion of the space of ends in it is not open. We will show below in Proposition 3.2.15 that having relatively parabolic holonomy is actually the only obstruction to the openness of $j_\sigma^\#$.

For a point $p \in \mathrm{M}^\sigma(\widetilde{\Sigma})$ we define the balls

$$\mathrm{B}(p, r) := \{z \in \widetilde{\Sigma} \mid d(p, z) < r\},$$
$$\mathrm{B}_\#(p, r) := \{z \in \widetilde{\Sigma}^\# \mid d(p, z) < r\},$$
$$\mathrm{B}_\sigma(p, r) := \{z \in \mathrm{M}^\sigma(\widetilde{\Sigma}) \mid d(p, z) < r\}.$$

By Proposition 3.2.9, $B(p, r) \subseteq B_\#(p, r) \subseteq B_\sigma(p, r)$ for any p and r, and these balls are open. For small values of r they also enjoy extra properties.

By Proposition 3.2.9 we know we can embed the space of ends in the ideal boundary $\partial_\infty^\sigma(\widetilde{\Sigma})$ of the Möbius completion $M^\sigma(\widetilde{\Sigma})$. So it makes sense for a given subset Z of $\widetilde{\Sigma}$ to consider its closure $\text{Cl}_\#(Z)$ in $\widetilde{\Sigma}^\#$ or $\text{Cl}_\sigma(Z)$ in $M^\sigma(\widetilde{\Sigma})$; by completeness of $M^\sigma(\widetilde{\Sigma})$, the latter is the same as the metric completion of Z with respect to some choice of metric as in the previous sections. In either case, the (*topological*) *boundary* of a subset Z is the difference between its *closure* and its *interior* $\partial Z := \text{Cl}(Z) \setminus \text{Int}(Z)$.

Lemma 3.2.13 *For each $p \in \widetilde{\Sigma}$ let $R = d(p, \partial_\infty^\sigma(\widetilde{\Sigma}))$. Then for all $r < R$,*

(1) $B(p, r) = B_\#(p, r) = B_\sigma(p, r)$,

(2) $\text{Cl}(B(p, r)) = \text{Cl}_\#(B_\#(p, r)) = \text{Cl}_\sigma(B_\sigma(p, r))$ *is complete.*

Proof Since the metric structure on $\widetilde{\Sigma}$ is induced by the Riemannian metric g_E, for sufficiently small radius, the metric balls are just balls for the Riemannian metric g_E. In particular, they are all disjoint from the ideal boundary, hence they coincide and their closure is complete and contained in $\widetilde{\Sigma}$. □

Lemma 3.2.14 *For each $p \in \widetilde{\Sigma}$ let $R = d(p, \partial_\infty^\sigma(\widetilde{\Sigma}))$; then for all $r \leq R$ the developing map induces an isometry between $\text{Cl}_\sigma(B_\sigma(p, r))$ and $\text{Cl}(B(\text{dev}(p), r))$.*

Proof Let I be the set of $r \in [0, R]$ such that the developing map induces an isometry between $\text{Cl}_\sigma(B_\sigma(p, r))$ and $\text{Cl}(B(\text{dev}(p), r))$. We are going to show that I is not empty, open on the right and closed on the right to conclude that $I = [0, R]$.

- $[0, \epsilon) \subset I$ for $\epsilon > 0$ small enough. This is because dev is a local isometry at p.

- If $[0, r) \subset I$ then $[0, r] \subset I$. Notice that the developing map induces an isometry between $\text{Cl}_\sigma\big(B_\sigma\big(p, r - \frac{1}{n}\big)\big)$ and $\text{Cl}\big(B\big(\text{dev}(p), r - \frac{1}{n}\big)\big)$ for all $n > 0$. This is enough to deduce that the developing map induces an isometry between $B_\sigma(p, r)$ and $B(\text{dev}(p), r)$. Since $\text{Cl}_\sigma(B_\sigma(p, r))$ is complete, and the metric completion is unique, the developing map induces an isometry between $\text{Cl}_\sigma(B_\sigma(p, r))$ and $\text{Cl}(B(\text{dev}(p), r))$.

- If $[0, r] \subset I$ for $r < R$ then $[0, r + \epsilon] \subset I$ for $\epsilon > 0$ small enough. Given that $r \in I$, the developing map induces an isometry between $\text{Cl}_\sigma(B_\sigma(p, r))$ and $\text{Cl}(B(\text{dev}(p), r))$. In particular, $\partial B_\sigma(p, r)$ is compact. Since $r < d(p, \partial_\infty^\sigma(\widetilde{\Sigma}))$, there is an ϵ–neighborhood of $\partial B_\sigma(p, r)$ on which dev is an isometry and $r + \epsilon \in I$. □

We call $\text{Cl}_\sigma(B_\sigma(p, R))$ the *maximal ball* centered at p. It is a maximal round ball containing p, in the sense of [Kulkarni and Pinkall 1994]. Our goal in Section 3.3 is to construct analogous "round neighborhoods" of all the ends, in the case of elliptic holonomy. We will need the following preliminary results.

Proposition 3.2.15 *Let $E \in \mathscr{E}(\widetilde{\Sigma})$, let σ be tame at E, and let N be an open horocyclic neighborhood of E. Then $j_\sigma^\#(N)$ is open if and only if E has nonparabolic holonomy.*

Proof Let δ_E be the peripheral element fixing E, let $R_E := \rho(\delta_E)$, and let $p_E = \mathrm{dev}_{\#}(E)$. By Lemma 3.2.13, every point in $j_\sigma^{\#}(N) \cap \widetilde{\Sigma}$ is in the interior of $j_\sigma^{\#}(N)$, so we only need to check whether E is in the interior of $j_\sigma^{\#}(N)$.

First, consider the case R_E is parabolic. Pick a point $x \in \partial j_\sigma^{\#}(N)$ that does not develop to p_E, eg on the image via $j_\sigma^{\#}(N)$ of a horocycle bounding N. Then $d(E, \delta_E^n(x)) \to 0$, ie $\delta_E^n(x) \to E$ in $\mathrm{M}^\sigma(\widetilde{\Sigma})$. So the sequence $\delta_E^n(x)$ must eventually enter in every open neighborhood of E in $\mathrm{M}^\sigma(\widetilde{\Sigma})$. However it clearly does not enter in $j_\sigma^{\#}(N)$ by construction, which shows $j_\sigma^{\#}(N)$ is not open.

So let us now assume R_E is nonparabolic; by Lemma 3.1.3 we know that since σ is tame at E, R_E is either the identity or elliptic. Since δ_E acts cocompactly on the boundary ∂N of N and $\mathrm{dev}_{\#}$ is a local diffeomorphism along ∂N, we have that $\mathrm{dev}_{\#}^{-1}(p_E) \cap \partial N$ is finite in any δ_E–fundamental domain. In particular, we can equivariantly modify N to a δ_E–invariant neighborhood $W \subset N$ of E, such that ∂W stays at finite distance from ∂N. By construction E is the only end in the closure of W.

When R_E is trivial or elliptic, the set $\mathrm{dev}_{\#}(\partial W)$ has compact closure in $\mathbb{CP}^1 \setminus \{p_E\}$. In particular, it sits in the annulus $\{z \in \mathbb{CP}^1 \mid R_1 \leq d_0(p_E, z) \leq R_2\}$, for some suitable radii $0 < R_1 \leq R_2$. For $r < R_1$ consider the open R_E–invariant ball $D_r \subseteq \mathbb{CP}^1$ of radius r around p_E, as well as the open ball $\mathrm{B}_\sigma(E, r)$. Observe that $\mathrm{dev}_\sigma(\mathrm{B}_\sigma(E, r))$ is contained in D_r, and so is disjoint from $\mathrm{dev}_{\#}(\partial W)$. We claim $\mathrm{B}_\sigma(E, r) \subseteq j_\sigma^{\#}(W) \subset j_\sigma^{\#}(N)$. By contradiction let $x \in \mathrm{B}_\sigma(E, r) \setminus j_\sigma^{\#}(W)$. Then connect x to E by a continuous arc γ contained in $\mathrm{B}_\sigma(E, r)$ (which is possible since we are in a length space). Then γ has to cross $\partial j_\sigma^{\#}(W)$, since ∂W separates E from the complement of W in $\widetilde{\Sigma}^{\#}$. Then $\mathrm{dev}_\sigma(\gamma)$ meets $\mathrm{dev}_{\#}(\partial W) = \mathrm{dev}_\sigma(\partial j_\sigma^{\#}(W))$, which leads to the desired contradiction. $\square$

We summarize the results of this section in the following statement.

Theorem C *Let $\sigma \in \mathscr{P}(\Sigma)$ be nondegenerate and without apparent singularities. Let $j^{\#} \colon \widetilde{\Sigma} \to \widetilde{\Sigma}^{\#}$ and $j_\sigma \colon \widetilde{\Sigma} \to \mathrm{M}^\sigma(\widetilde{\Sigma})$ be the natural embeddings. Then $\sigma \in \mathscr{P}^{\circ}(\Sigma)$ if and only if there exists a continuous open $\pi_1(\Sigma)$–equivariant embedding $j_\sigma^{\#} \colon \widetilde{\Sigma}^{\#} \to \mathrm{M}^\sigma(\widetilde{\Sigma})$ that makes the following diagram commute:*

$$
\begin{array}{ccc}
 & \widetilde{\Sigma}^{\#} & \\
{\scriptstyle j^{\#}} \nearrow & {\scriptstyle j_\sigma^{\#}} \downarrow & \searrow {\scriptstyle \mathrm{dev}_{\#}} \\
\widetilde{\Sigma} & & \mathbb{CP}^1 \\
{\scriptstyle j_\sigma} \searrow & & \nearrow {\scriptstyle \mathrm{dev}_\sigma} \\
 & \mathrm{M}^\sigma(\widetilde{\Sigma}) &
\end{array}
$$

Proof First assume $\sigma \in \mathscr{P}^{\circ}(\Sigma)$. Since σ is tame, by Proposition 3.2.9 we know that $j_\sigma \colon \widetilde{\Sigma} \hookrightarrow \mathrm{M}^\sigma(\widetilde{\Sigma})$ extends to a $\pi_1(\Sigma)$–equivariant continuous embedding $j_\sigma^{\#} \colon \widetilde{\Sigma}^{\#} \hookrightarrow \mathrm{M}^\sigma(\widetilde{\Sigma})$, and that $\mathrm{dev}_{\#} = \mathrm{dev}_\sigma \circ j_\sigma^{\#}$. To check that $j_\sigma^{\#}$ is open we argue as follows. Observe that the restriction of $j_\sigma^{\#}$ to $\widetilde{\Sigma}$ is just the natural embedding of $\widetilde{\Sigma}$ in its completion, which is open. So we only need to check the ends. Let E be an end; without loss of generality we can assume that an open neighborhood of E in $\widetilde{\Sigma}^{\#}$ is an open horocycle N. Since σ is relatively elliptic, Proposition 3.2.15 implies that $j_\sigma^{\#}(N)$ is an open neighborhood of $j_\sigma^{\#}(E)$ in $\mathrm{M}^\sigma(\widetilde{\Sigma})$.
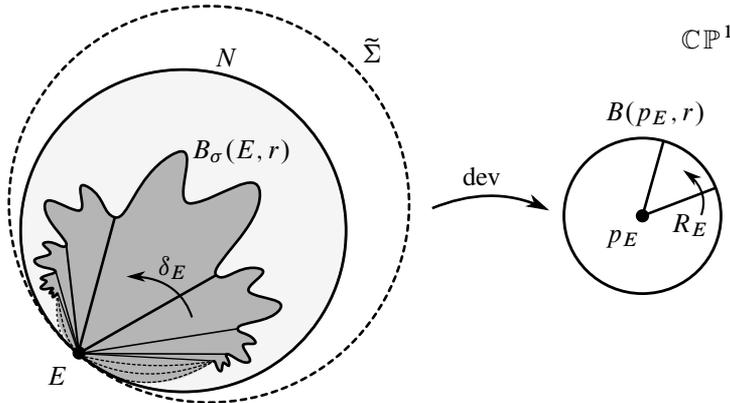
Figure 13: A horocycle containing a ball, in the elliptic case.

Conversely, assume the existence of the extension $j_\sigma^\#$ as in the statement. Its continuity implies tameness of σ by Proposition 3.2.9. Let E be an end. By Lemma 3.1.3 we know that the holonomy of σ at E is either trivial, parabolic or elliptic. The first case is excluded by the hypothesis that σ has no apparent singularities, and the second case by the hypothesis that $j_\sigma^\#$ is open, together with Proposition 3.2.15. Therefore σ is relatively elliptic. It is also assumed to be nondegenerate, hence we can conclude that $\sigma \in \mathscr{P}^\odot(\Sigma)$. □

Corollary 3.2.16 *If $\sigma \in \mathscr{P}(\Sigma)$, then $\mathscr{E}(\widetilde\Sigma)$ is a discrete subspace of $\partial_\infty^\sigma(\widetilde\Sigma)$.*

Proof Let E be an end. By Theorem C any horocyclic neighborhood N of E is open for the topology of $M^\sigma(\widetilde\Sigma)$, under the natural embedding $j_\sigma^\#$. By definition, N does not contain any other point of $\partial_\infty^\sigma(\widetilde\Sigma)$, hence E is an open point. □

Corollary 3.2.17 *Let σ be tame and relatively elliptic. For every end $E \in \mathscr{E}(\widetilde\Sigma)$, the action of the peripheral subgroup $\langle \delta_E \rangle$ on $M^\sigma(\widetilde\Sigma) \setminus \{E\}$ is proper and free.*

Proof The action on the Möbius completion extends the action by deck transformations, so the statement is trivial for points in $\widetilde\Sigma$. By Proposition 3.2.15, both metric balls and horocyclic neighborhoods provide fundamental systems of neighborhoods of the ends in the completion. So one can see that the action of δ_E on the subspace $\mathscr{E}(\widetilde\Sigma) \setminus E$ is proper and free. The case of a general ideal point follows from this fact together with the existence of a δ_E–invariant cyclic order on the ideal boundary (see Remark 3.2.8). □

3.3 Local properties of the developing map at an end

The main goal of this section is to prove Theorem D, about the behavior of developing maps around E for a structure $\sigma \in \mathscr{P}^\odot(\Sigma)$. If σ has developing pair (dev, ρ), and if $E \in \mathscr{E}(\widetilde\Sigma)$, then let $p_E := \mathrm{dev}_\#(E) \in \mathbb{CP}^1$ and let $\delta_E \in \pi_1(\Sigma)$ be a peripheral element fixing E. Then $R_E := \rho(\delta_E)$ is an elliptic Möbius transformation fixing p_E (Lemma 3.1.3); let q_E denote the other fixed point of R_E. We will construct a family

of δ_E–invariant neighborhoods of E which develop to R_E–invariant round disks in $\mathbb{CP}^1$, and on which $\mathrm{dev}_\#$ restricts to a branched covering (branching only at E).

While the results of the previous sections relied (but did not depend), on the choice of the background metric g_0 on $\mathbb{CP}^1$, we now want to exploit the fact that the peripheral holonomy is elliptic to pick a convenient metric. The topological structure of the Möbius completion is not affected by this (eg ideal points, etc), but finer metric statements (eg the shape and properties of individual metric balls) are. Let g_0 be the unique R_E–invariant spherical round metric on $\mathbb{CP}^1$ for which the fixed points p_E and q_E of R_E are antipodal points at distance 1. Let us denote by $g_E = \mathrm{dev}^*(g_0)$ the Riemannian metric and by d_E the distance function induced on $\widetilde{\Sigma}$. By construction, the Möbius completion is the metric completion of $(\widetilde{\Sigma}, d_E)$.

Lemma 3.3.1 *Let $U \subseteq \mathrm{M}^\sigma(\widetilde{\Sigma})$ be a δ_E–invariant neighborhood of E. Then the distance between δ_E–orbits defines a metric on $U/\langle \delta_E \rangle$ with respect to which the quotient map*

$$\pi_E : U \setminus \{E\} \to (U \setminus \{E\})/\langle \delta_E \rangle$$

is a locally isometric covering map.

Proof Let $\pi_E(u), \pi_E(v) \in U/\langle \delta_E \rangle$. Then their distance is defined to be

$$d(\pi_E(u), \pi_E(v)) := \inf\{d(\delta_E^n(u), \delta_E^m(v)) \mid m, n \in \mathbb{Z}\}.$$

Since the action on $U \setminus \{E\}$ is isometric, free and proper (Corollary 3.2.17), by [Bridson and Haefliger 1999, Proposition I.8.5] we get our statement in the complement of the end. To include the end it is enough to notice that it is an isolated fix point and that no orbit accumulates to it, since the holonomy is elliptic. $\qquad\square$

Lemma 3.3.2 *Let $U \subseteq \mathrm{M}^\sigma(\widetilde{\Sigma})$ be a δ_E–invariant neighborhood of E on which δ_E acts cocompactly. Then the following hold.*

(1) *U is complete.*

(2) *If $V \subseteq U$ is closed and δ_E–invariant, then V is complete and δ_E acts on V cocompactly.*

Proof (1) Let $x_n \in U$ be a Cauchy sequence. Let us denote by F_n a (coarse) compact fundamental domain for the action $\langle \delta_E \rangle \curvearrowright U$ containing x_n. If the sequence of F_n eventually stabilizes to some F, then eventually the sequence x_n lies entirely in F, hence converges in it by compactness. So let us assume that the sequence F_n does not stabilize. We claim that since x_n is a Cauchy sequence this forces $d_E(x_n, E)$ to decrease to zero, ie x_n converges to E. Indeed, since the holonomy is elliptic and the metric invariant, if $|n - m|$ is large enough then the shortest curve between a point in F_n and a point in F_m goes through E.

(2) If V is closed then it is complete by completeness of U. Let F be a (coarse) compact fundamental domain for the action $\langle \delta_E \rangle \curvearrowright U$. Since V is invariant we get $V/\langle \delta_E \rangle = (V \cap F)/\langle \delta_E \rangle$, and this is compact because $V \cap F$ is. $\qquad\square$

We have seen in Proposition 3.2.15 that, when the holonomy is elliptic, horocycles contain metric balls (see Figure 13). We now describe a sufficient condition on a metric ball to be fully contained in a horocycle. Notice that the following statement fails in the case of parabolic holonomy (see Remark 3.2.12).

Lemma 3.3.3 *For each $E \in \mathscr{E}(\widetilde{\Sigma})$ let $\rho_E := d_E(E, \partial^{\sigma}_{\infty}(\widetilde{\Sigma}) \setminus \{E\})$. Then $\rho_E > 0$ and for all $0 < r < \rho_E$, there is a proper horocyclic neighborhood of E containing $B_{\sigma}(E, r)$.*

Henceforth we call $\rho_E := d_E(E, \partial^{\sigma}_{\infty}(\widetilde{\Sigma}) \setminus \{E\})$ the *critical radius* of E.

Proof For the first part of the lemma, let V be a proper (ie $\mathrm{Cl}_{\#}(V) \subsetneq \Sigma \cup \{E\}$) horocycle based at E. By Proposition 3.2.15 V is open, so there is $r > 0$ such that $B_{\sigma}(E, r) \subseteq V$. We claim that $B_{\sigma}(E, r) \subset \Sigma \cup \{E\}$, from which it follows that $\rho_E \geq r > 0$. Recall that δ_E acts cocompactly on $\mathrm{Cl}_{\#}(V)$; therefore $\mathrm{Cl}_{\#}(V)$ is complete by Lemma 3.3.2. It follows that

$$\mathrm{Cl}_{\sigma}(B_{\sigma}(E, r)) \subseteq \mathrm{Cl}_{\sigma}(V) = \mathrm{Cl}_{\#}(V) \subsetneq \Sigma \cup \{E\}.$$

Next, let $r < \rho_E$. Suppose by contradiction that, for every proper horocyclic neighborhood N of E, there was a point $x \in B_{\sigma}(E, r) \setminus N$. Fix $\{N_k\}$ a sequence of proper horocyclic neighborhoods of E such that $N_k \subset N_{k+1}$ and $\bigcup N_k = \widetilde{\Sigma} \cup \{E\}$. Let $x_k \in B_{\sigma}(E, r) \setminus N_k$. For every k, let $r_k := d_E(E, \partial N_k)$. As δ_E acts cocompactly and by isometries on ∂N_k, there is some point on ∂N_k at distance r_k from E. The fact that $E \notin \partial N_k$ and the sequence $\{N_k\}$ is nested further implies that

$$r_k > 0, \quad r_k \leq r_{k+1}, \quad \lim_{k \to \infty} r_k = \rho_E.$$

Notice that the second inequality is due to the fact that ∂N_k separates E from ∂N_{k+1}. Similarly, ∂N_k separates E from x_k, therefore $r_k \leq d_E(E, x_k) < r$, hence in the limit we get $\rho_E = \lim_{k \to \infty} r_k \leq r$, in contradiction with the choice of r. $\square$

Corollary 3.3.4 *For each $E \in \mathscr{E}(\widetilde{\Sigma})$ and $0 < r < \rho_E$, $B_{\#}(E, r) = B_{\sigma}(E, r)$ and $\mathrm{Cl}_{\#}(B_{\#}(E, r))$ is complete. Moreover, $\mathrm{Cl}_{\#}(B_{\#}(E, r)) = \mathrm{Cl}_{\sigma}(B_{\sigma}(E, r))$.*

Proof By Lemma 3.3.3, the ball $B_{\#}(E, r)$ is contained in a proper horocyclic neighborhood V of E. It follows that $B_{\sigma}(E, r) \subset \widetilde{\Sigma} \cup \{E\}$ and so $B_{\#}(E, r) = B_{\sigma}(E, r)$.

By Lemma 3.3.2, the closed ball $\mathrm{Cl}_{\#}(B_{\#}(E, r))$ is complete.

Finally, since $\mathrm{Cl}_{\#}(B_{\#}(E, r))$ contains $B_{\#}(E, r)$ and is complete, it must contain the completion of $B_{\#}(E, r)$. Since $\mathrm{M}^{\sigma}(\widetilde{\Sigma})$ is complete we have that $\mathrm{Cl}_{\sigma}(B_{\sigma}(E, r))$ coincides with the completion of $B_{\sigma}(E, r)$. But we also know that $B_{\sigma}(E, r) = B_{\#}(E, r)$. So $\mathrm{Cl}_{\sigma}(B_{\sigma}(E, r))$ coincides with the completion of $B_{\#}(E, r)$, and it is therefore contained in $\mathrm{Cl}_{\#}(B_{\#}(E, r))$. $\square$

Recall that a metric space Z is *star-shaped at a point* $x \in Z$ if for every $y \in Z$ there is a geodesic in Z connecting x to y.

Lemma 3.3.5 *For each $E \in \mathscr{E}(\widetilde{\Sigma})$ and $0 < r < \rho_E$, the open ball $B_{\sigma}(E, r)$ is star-shaped at E.*

Proof Let $x \in \mathrm{B}_\sigma(E, r)$ and let $r' := d_E(x, E) < r$. By Lemma 3.2.5, for all $c > 0$ we can pick a continuous curve $\eta_c \colon [0, 1) \to \tilde{\Sigma}$ such that $\eta_c(0) = x$, $\lim_{t \to 1} \eta_c(t) = E$ and $r' \leq \ell(\eta_c) \leq r' + c$. For each $t \in [0, 1)$ we have

$$d(\eta_c(t), E) \leq \ell(\eta_c([t, 1))) \leq \ell(\eta_c([0, 1))) \leq r' + c.$$

In particular, for $c < r - r'$ we get that $d_E(\eta_c(t), E) < r$, ie η_c is entirely contained in $\mathrm{B}_\sigma(E, r) \cap \tilde{\Sigma}$. Let $\gamma_n \colon [0, 1) \to \mathrm{B}_\sigma(E, r)$ be the curve obtained for $c = \frac{1}{n}$.

Consider the quotient $\pi_E \colon \mathrm{Cl}_\sigma(\mathrm{B}_\sigma(E, r)) \to \mathrm{Cl}_\sigma(\mathrm{B}_\sigma(E, r))/\langle \delta_E \rangle =: Y$. It follows from Lemma 3.3.1 that π_E is a branched covering map onto a metric space, branching only at E; let us denote by d_Y the distance in Y. Moreover by Lemma 3.3.3 the ball $\mathrm{B}_\sigma(E, r)$ is properly contained in a horocycle. Since δ_E acts cocompactly on horocycles, it follows that Y is compact by Lemma 3.3.2. Notice that since δ_E acts by isometries and E is the only fixed point, we also have that $r' = d_E(x, E) = d_Y(\pi_E(x), \pi_E(E))$.

Projecting the curves γ_n to the quotient we obtain curves $\pi_E \circ \gamma_n \colon [0, 1) \to Y$ such that $\pi_E \circ \gamma_n(0) = \pi_E(x)$, $\lim_{t \to 1} \pi_E \circ \gamma_n(t) = \pi_E(E)$ and

$$d_Y(\pi_E(E), \pi_E(x)) = r' \leq \ell(\pi_E \circ \gamma_n) \leq r' + \tfrac{1}{n}.$$

In particular, by Arzelà–Ascoli we can extract a uniform limit $\bar{\gamma} \colon [0, 1] \to Y$. By the above length inequality we obtain

$$d_Y(\pi_E(E), \pi_E(x)) = r' = \ell(\bar{\gamma}) = \lim_{n \to \infty} \ell(\pi_E \circ \gamma_n),$$

ie $\bar{\gamma}$ is a geodesic from $\pi_E(x)$ to $\pi_E(E)$. Notice that it goes through $\pi_E(E)$ only at one endpoint; so we can lift it to a curve $\gamma \colon [0, 1) \to \mathrm{Cl}_\sigma(\mathrm{B}_\sigma(E, r))$ starting at x and limiting to E, of the same length r'. By the same argument as the beginning, γ is completely contained in the open ball $\mathrm{B}_\sigma(E, r)$, so this is the desired geodesic. $\qquad\square$

We now consider the restriction of dev_σ to a ball around an end $E \in \mathcal{E}(\tilde{\Sigma})$, ie

$$\mathrm{dev}_\sigma \colon \mathrm{B}_\sigma(E, r) \to \mathrm{B}(p_E, r),$$

and we find the values of r for which it is a covering map, branching only at E. The proof is reminiscent of (and based on) the classical fact that a local isometry from a complete Riemannian manifold to a connected one is a covering map. Notice that in our setting dev_σ is not locally isometric at E (not even locally injective), and on the other hand $\mathrm{Cl}_\sigma(\mathrm{B}_\sigma(E, r)) \setminus \{E\}$ is not complete. The proof shows how to deal with this, and also provides quantitative control on the critical radius.

Proposition 3.3.6 *For each $E \in \mathcal{E}(\tilde{\Sigma})$ we have that $\rho_E \leq 1$. Moreover,*

(1) *for each $0 < r \leq \rho_E$, dev_σ maps $\partial \mathrm{B}_\sigma(E, r)$ to $\partial \mathrm{B}(p_E, r)$;*

(2) *for each $0 < r \leq \rho_E$, $\mathrm{dev}_\sigma \colon \mathrm{B}_\sigma(E, r) \to \mathrm{B}(p_E, r)$ is a branched covering map, branching only at E.*

Proof We are first going to prove statements (1) and (2) for $r \leq \min\{\rho_E, 1\}$, and then we will show that $\rho_E \leq 1$.

We begin with the following observation. Suppose $r < \rho_E$ and let $x \in \partial B_\sigma(E, r)$. Then $d_E(E, x) = r$ and $d_0(p_E, \mathrm{dev}_\sigma(x)) \leq r$. Let $r' > 0$ be such that $r < r' < \rho_E$. Then $x \in B_\sigma(E, r')$ and by Lemma 3.3.5 there exists a geodesic γ_r from x to E contained in $B_\sigma(E, r')$. Observe that $r = \ell(\gamma_r) = \ell(\mathrm{dev}_\sigma(\gamma_r))$. Notice that dev is a local isometry on $\widetilde{\Sigma}$, so γ_r maps to a geodesic in $\mathbb{CP}^1$.

Next, additionally assume that $r < 1$, the diameter of $\mathbb{CP}^1$. Then the curve γ_r maps to a simple geodesic arc, starting from p_E and avoiding q_E, of length $r < 1$. Since the choice of x above was arbitrary, it follows that $\mathrm{dev}_\sigma(\partial B_\sigma(E, r)) \subseteq \partial B(p_E, r)$. In particular, it avoids q_E. This concludes the proof of (1) in the case where $r < \min\{\rho_E, 1\}$. The limiting case $r = \min\{\rho_E, 1\}$ follows by continuity of the developing map.

We now start the proof of (2). To begin with, we claim that when $r < \min\{\rho_E, 1\}$, each component of $\partial B_\sigma(E, r)$ is isometric to a complete line. Since $r < 1$, $\partial B(p_E, r)$ is a circle in $\mathbb{CP}^1$. Since $r < \rho_E$, we have that $\partial B_\sigma(E, r) \subset \widetilde{\Sigma}$ and dev_σ is a local homeomorphism on it. In particular, $\partial B_\sigma(E, r)$ is a 1–dimensional submanifold of $\widetilde{\Sigma}$; moreover it is closed in $\mathrm{Cl}_\sigma(B_\sigma(E, r))$, hence complete by Corollary 3.3.4. Then dev_σ induces a local isometry from the complete manifold $\partial B_\sigma(E, r)$ to the connected manifold $\partial B(p_E, r)$; it follows that it is a Riemannian covering map. Notice that $\langle \delta_E \rangle$ is an infinite cyclic group acting on $\partial B_\sigma(E, r)$ properly and freely by Corollary 3.2.17, hence each component of $\partial B_\sigma(E, r)$ must be isometric to a complete line.

Now we claim that, for all $0 < r \leq \min\{\rho_E, 1\}$, $\mathrm{dev}_\sigma \colon B_\sigma(E, r) \to B(p_E, r)$ is a branched covering map, branching only at E. First notice that

(3.3.1) $$B_\sigma(E, r) \setminus \{E\} = B_\sigma(E, r) \setminus \{\mathrm{dev}_\sigma^{-1}(p_E)\}.$$

Indeed suppose $z \in B_\sigma(E, r)$ is another point developing to p_E; then there is $r' < r \leq \rho_E$ such that $z \in B_\sigma(E, r')$, and a geodesic γ from z to E contained in $B_\sigma(E, r')$. Since $\mathrm{dev}_\sigma(E) = \mathrm{dev}_\sigma(z) = p_E$, this geodesic γ has to cover at least a great circle through p_E in $\mathbb{CP}^1$, hence $d_E(E, z) \geq 2$. But $r' < r \leq 1$ forbids this. In particular, we get a well-defined local homeomorphism

$$\varphi := \mathrm{dev}_\sigma \mid_{B_\sigma(E, r) \setminus \{E\}} \colon B_\sigma(E, r) \setminus \{E\} \to B(p_E, r) \setminus \{p_E\}.$$

It is enough to show that this is a covering map. We are going to show that every point in $B(p_E, r) \setminus \{p_E\}$ is evenly covered. Let $y \in B(p_E, r) \setminus \{p_E\}$ and let $r_y := d_0(p_E, y)$. Notice that $0 < r_y < r$. Since dev_σ is a covering map between $\partial B_\sigma(E, r_y)$ and $\partial B(p_E, r_y)$, there is $\epsilon_y > 0$ such that $B(y, \epsilon_y) \cap \partial B(p_E, r_y)$ is evenly covered. Let

$$\delta_y := \min\{\epsilon_y, r - r_y, r_y\}.$$

Notice that the ball $B(y, \delta_y)$ is entirely contained in $B(p_E, r) \setminus \{p_E\}$. Then we claim that $B(y, \delta_y)$ is evenly covered. Let $z \in \mathrm{dev}_\sigma^{-1}(y) \cap B_\sigma(E, r)$. By definition of δ_y, $B_\sigma(z, \delta_y)$ is entirely contained in

$B_\sigma(E, r) \setminus \{E\}$. In particular, it is smaller than the maximal ball centered at z, so it is isometrically mapped to $B(y, \delta_y)$ by dev_σ (Lemma 3.2.14). This implies that if $z' \in \mathrm{dev}_\sigma^{-1}(y) \cap B_\sigma(E, r)$ is different from z, then $B_\sigma(z, \delta_y) \cap B_\sigma(z', \delta_y) = \varnothing$. This concludes the proof of (2) in the case $r \le \min\{\rho_E, 1\}$.

Now suppose by contradiction that $\rho_E > 1$. Then there is r such that $1 < r < \rho_E$, and the open ball $B_\sigma(E, r)$ is star-shaped at E (Lemma 3.3.5). Moreover, the developing map maps $\partial B_\sigma(E, 1)$ to q_E, the only point at distance 1 from p_E in $\mathbb{CP}^1$. Since the open ball $B_\sigma(E, r)$ is entirely contained in $\widetilde{\Sigma}$, and contains $\partial B_\sigma(E, 1)$, the developing map is a local homeomorphism on $\partial B_\sigma(E, 1)$. In particular $\partial B_\sigma(E, 1)$ is discrete. On the other hand, for every $r' < 1 = \min\{\rho_E, 1\}$ we can apply the first part of the proof where we proved that dev_σ maps $\partial B_\sigma(E, r')$ to $\partial B(p_E, r')$, and

$$\lim_{r' \to 1^-} \partial B(p_E, r') = \{q_E\}.$$

This implies that, for radii $r' < 1 = \min\{\rho_E, 1\}$ sufficiently close to 1, $\partial B_\sigma(E, r')$ is a disjoint union of circles, contradicting that each connected component is isometric to a complete line. $\qquad \square$

Theorem D *Let $\sigma \in \mathscr{P}^{\odot}(\Sigma)$, and let E be an end. Then there is a neighborhood $\hat{N}_E$ of E in $\mathrm{M}^\sigma(\widetilde{\Sigma})$ onto which the developing map for σ restricts to a branched covering map, branching only at E, and with image a round disk in $\mathbb{CP}^1$.*

Proof We can just take $\hat{N}_E$ to be any ball $B_\sigma(E, r)$ satisfying the conditions of Proposition 3.3.6. $\qquad \square$

Let $E \in \mathscr{E}(\widetilde{\Sigma})$, and let ρ_E be its critical radius. The open metric ball $\hat{N}_E = B_\sigma(E, \rho_E)$ plays the role of a *canonical maximal neighborhood* of E, similar to the maximal round balls in [Kulkarni and Pinkall 1994]. Indeed, it develops to a round ball in $\mathbb{CP}^1$, and by definition of ρ_E, the boundary of $\hat{N}_E$ contains an ideal point. However, note that we have normalized things "locally" at E, by fixing the R_E–invariant round metric on $\mathbb{CP}^1$ for which the fixed points p_E and q_E of the holonomy at E are antipodal points of distance 1 (here $R_E = \rho(\delta_E)$ denotes the peripheral holonomy at E). Then $\hat{N}_E$ is defined as a metric ball for the induced metric g_E on $\mathrm{M}^\sigma(\widetilde{\Sigma})$. If E' is a different end, then the metric ball around E' (with respect to g_E) does not necessarily agree with $\hat{N}_{E'}$, which would be defined as a metric ball for the metric $g_{E'}$.

Moreover one can observe that when $r < \rho_E$ the ball $\partial B_\sigma(E, r)$ contains a horocycle and is contained in a horocycle. Therefore each component of its boundary is contained in the lune between two horocycles. Since $B_\sigma(E, r)$ is star-shaped at the end, and $\partial B_\sigma(E, r)$ is invariant under the action of the peripheral δ_E, we can see that $\partial B_\sigma(E, r)$ is actually connected and isometric to a complete line. In particular, $\partial B_\sigma(E, r)$ is the universal cover of $\partial B(p_E, r)$, and $B_\sigma(E, r) \setminus \{E\}$ is isometric to the universal cover of $B(p_E, r) \setminus \{p_E\}$.

Remark 3.3.7 If σ is the tame and relatively parabolic structure induced by complete hyperbolic metric of finite area, then dev is a global diffeomorphism, and horocycles develop to round disks. In particular, Theorem D holds for such a structure. However, there is no analogue of Theorem D in the general

parabolic case. For example, consider the structure obtained by grafting σ along an ideal arc, and let E be an end covering one of the endpoints of the grafting arc. If U is any δ_E–invariant neighborhood of E, then $V = \mathrm{dev}_\sigma(U)$ is invariant under a parabolic transformation and contains its fixed point $p_E = \mathrm{dev}_\#(E)$ in its interior. This forces $V = \mathbb{CP}^1$. In particular, we see that the local homeomorphism (analogous to the one considered in the proof of Proposition 3.3.6)

$$\varphi := \mathrm{dev}_\sigma\,|_{U\setminus\mathrm{dev}_\sigma^{-1}(p_E)} : U \setminus \mathrm{dev}_\sigma^{-1}(p_E) \to V \setminus \{p_E\} = \mathbb{C}$$

cannot be a covering map, because it is not injective and the image is simply connected.

Throughout this section we have worked under the normalization in which the fixed points $p_E, q_E \in \mathbb{CP}^1$ of the rotation R_E are antipodal points at distance 1. As established in Proposition 3.3.6, it follows that the critical radius of an end E satisfies $\rho_E \le 1$. We conclude this chapter by discussing what happens when a tame structure σ has an end E with elliptic holonomy and $\rho_E = 1$.

Remark 3.3.8 (structures on a twice-punctured sphere) Suppose σ is tame and has an end E with elliptic holonomy and $\rho_E = 1$. By (1) in Proposition 3.3.6 all the points on the boundary of $\widehat{N}_E$ must develop to q_E. It follows from the proof of Proposition 3.3.6 that in this case the boundary of $\widehat{N}_E$ cannot contain any isolated points in $\widetilde{\Sigma}$. As a result, $\widehat{N}_E = \widetilde{\Sigma}$. By tameness, this forces all the ends different from E to develop to q_E. We claim that in this case Σ must be a twice-punctured sphere, and σ is the structure associated to a power map $z \mapsto z^\alpha$ for some $\alpha \in \mathbb{R} \setminus \mathbb{Z}$. To see this, assume by contradiction that there is a peripheral element $\gamma \in \pi_1(\Sigma)$ distinct from any power of the peripheral element δ_E which fixes E. Then γ moves E to another end $\gamma E \ne E$. By equivariance and tameness of the developing map (see Lemma 3.1.3) we see that

$$q_E = \mathrm{dev}_\#(\gamma E) = \rho(\gamma)\,\mathrm{dev}_\#(E) = \rho(\gamma)p_E.$$

On the other hand, γ fixes an end $E' \ne E$. It follows that $\mathrm{dev}_\#(E') = q_E = \rho(\gamma)q_E$. We get $\rho(\gamma)p_E = \rho(\gamma)q_E$, which is absurd. Therefore all peripheral elements are powers of a fixed one. But the only orientable surface in which this happens is a sphere with two punctures. Notice that this surface has zero Euler characteristic, $\widetilde{\Sigma}$ identifies with $\mathbb{C}$, and we can normalize things so that $\mathrm{dev}(z) = e^{az}$, $p_E = 0$ and $q_E = \infty$, for some $a \in \mathbb{C}^*$. Deck transformations are generated by $z \mapsto z + 2\pi i$, and the holonomy by $w \mapsto e^{2\pi i a}w$; ellipticity of the holonomy means $a \in \mathbb{R} \setminus \mathbb{Z}$. The Möbius completion is obtained by adding just two ideal points, for $\mathrm{Re}(z) \to \pm\infty$, mapping to $q_E = \infty$ and $p_E = 0$ respectively. Structures of this type can be defined by a spherical metric with two cone points and coaxial holonomy.

Remark 3.3.9 Spherical metrics with cone points and coaxial holonomy exist also on a surface of negative Euler characteristic (see [Eremenko 2004; Mondello and Panov 2016]), and provide examples of structures with degenerate holonomy. However such a structure must have some apparent singularities (ie punctures with trivial holonomy, see [Gupta 2021]), whose presence forces the critical radius to be strictly less than 1 at every end with elliptic holonomy. Indeed, if E is an end with elliptic holonomy, then the

family of neighborhoods $B_\sigma(E, r)$ must hit another end (possibly one covering an apparent singularity) before $r = 1$. As an illustrative example, consider the structure obtained by puncturing an additional point on a sphere endowed with a spherical metric with two cone points.

3.4 The index of a puncture

Using the neighborhoods constructed in the previous section (namely Theorem D), we can define a numerical invariant of the complex projective structure for each puncture, which we call the *index*. This is essentially the angle that the developed image of a peripheral curve makes around the image of a corresponding end.

Let $\sigma \in \mathcal{P}^\odot(\Sigma)$ be a structure represented by a pair (dev, ρ). Let x be a puncture of Σ, and let η be a positive peripheral curve in Σ around x. This can be chosen so that for any end E covering x (in the sense of Remark 3.1.1), the lift of η which is asymptotic to E is entirely contained in a neighborhood V_E of E on which dev is a branched covering map, branching only at E (see Theorem D).

Let us fix an end E, and let $\delta_E \in \pi_1(\Sigma)$ be the positive peripheral deck transformation fixing E. We recall that $p_E := \text{dev}_\#(E)$ is one of the two fixed points for the elliptic transformation $\rho(\delta_E)$ (see Lemma 3.1.3). Let us normalize so that $\rho(\delta_E)$ fixes 0 and ∞. Let $\tilde\eta \subset V_E$ be the lift of η in V_E, and choose $\tilde\eta_0 \subset \tilde\eta$ to be a fundamental domain for the action $\langle\delta_E\rangle \curvearrowright \tilde\eta$. Let $\zeta := \text{dev}(\tilde\eta_0) \subset \text{dev}(V_E) \setminus \{0\}$. Notice that freely homotoping η deeper into the puncture results in a homotopy of ζ in the complement of $p_E = 0$, because there are no other preimages of p_E in V_E (see (3.3.1) in Proposition 3.3.6).

The *index of the structure σ at the puncture x* is defined to be the number

$$I_\sigma(x) := \text{Im}\left(\int_\zeta \frac{dz}{z}\right).$$

When clear from the context, we will usually drop the σ and write $I(x) = I_\sigma(x)$.

We remark explicitly that this definition does not depend on any of the choices involved. Indeed, let us choose a parametrization $\zeta \colon [0, 1] \to \mathbb{C} \setminus \{0\}$, $\zeta(s) = r(s)e^{i\theta(s)}$, where $\theta \colon [0, 1] \to \mathbb{R}$ is a determination of the argument function on $\mathbb{C} \setminus \{0\}$, and $r \colon [0, 1] \to \mathbb{R}$. A direct computation in local coordinates shows that

$$\int_\zeta \frac{dz}{z} = \log\left(\frac{r(1)}{r(0)}\right) + i(\theta(1) - \theta(0)).$$

Notice that since η is chosen to be a peripheral curve, its holonomy is elliptic. Therefore we get $r(1)e^{i\theta(1)} = e^{i\varphi}r(0)e^{i\theta(0)}$, where φ is such that $\rho(\delta_E)z = e^{i\varphi}z$. It follows that

$$I_\sigma(x) = 2\pi k + \varphi,$$

where $k \in \mathbb{Z}$ counts the number of times ζ turns around 0 anticlockwise. Notice that the index is always positive, since δ_E was chosen to be a positive peripheral.

Remark 3.4.1 Let $\sigma \in \mathcal{P}^{\odot}(\Sigma)$, and let x and y be punctures. If η is a graftable arc joining x to y, then

- if $x \neq y$ then $I_{Gr(\sigma,\eta)}(x) = I_\sigma(x) + 2\pi$ and $I_{Gr(\sigma,\eta)}(y) = I_\sigma(y) + 2\pi$;
- if $x = y$ then $I_{Gr(\sigma,\eta)}(x) = I_\sigma(x) + 4\pi$.

4 The complex analytic point of view

The theory of $\mathbb{CP}^1$–structures enjoys fundamental interactions with the study of second-order linear ODEs on complex domains, namely through the use of the Schwarzian derivative. The purpose of this chapter is to describe the complex analytic counterpart to the structures in $\mathcal{P}^{\odot}(\Sigma)$ (see Theorem E). These are described by meromorphic quadratic differentials satisfying certain conditions on their Laurent expansion around poles.

4.1 Local theory at regular singularities

We start by reviewing the classical theory for the convenience of the reader, with a particular focus to the behavior around singularities of the coefficients (see [Hille 1969; Ince 1944]). This will provide the local model for our structures around the punctures.

Let us consider a holomorphic function $q : \mathbb{D}^* \to \mathbb{C}$ on the punctured unit disk $\mathbb{D}^* = \{z \in \mathbb{C} \mid 0 < |z| < 1\}$ with a double pole at the origin with *leading coefficient* a, ie a function of the form $q(z) = a/z^2 + O(1/z)$. We will consider the second-order linear ODE

$$(4.1.1) \qquad\qquad u'' + \tfrac{1}{2}qu = 0 \quad \text{for } u : \mathbb{D}^* \to \mathbb{C},$$

as well as the *Schwarz equation*

$$(4.1.2) \qquad\qquad \mathcal{S}f = q \quad \text{for } f : \mathbb{D}^* \to \mathbb{CP}^1,$$

where the operator

$$\mathcal{S}f = \left(\frac{f''}{f'}\right)' - \frac{1}{2}\left(\frac{f''}{f'}\right)^2$$

is the *Schwarzian derivative*. The main properties of $\mathcal{S}$ are the following:

(1) **Invariance** $\mathcal{S}f = 0$ if and only if f is the restriction of some Möbius transformation.

(2) **Cocycle** If f and g are locally injective holomorphic functions for which the composition is defined, then $\mathcal{S}(f \circ g) = g^*(\mathcal{S}f) + \mathcal{S}g$.

The relationship between the two equations above is well known (see [Hille 1969, Appendix D]), and can be summarized as follows: if u_1 and u_2 are linearly independent solutions for (4.1.1), then $f = u_1/u_2$ is a solution for (4.1.2); conversely, any solution for (4.1.2) is obtained in this way. In both cases, since the domain of the equation is not simply connected, these equations can have nontrivial monodromy, ie solutions are to be considered as multivalued functions, or as single-valued functions on a suitable covering domain.

The classical theory of linear ODEs (see [Ince 1944, Section 15.3], or [Allegretti and Bridgeland 2020, Section 5] for a more recent treatment) provides an explicit description of the local solutions of (4.1.1). First, the *indicial equation* of (4.1.1) is given by

$$r(r-1) + \tfrac{1}{2}a = 0.$$

Let $r_1, r_2 \in \mathbb{C}$ be its solutions; then one has two cases:

(1) if $r_1 - r_2 \notin \mathbb{Z}$ then (4.1.1) has two linearly independent solutions of the form $u_k(z) = z^{r_k} h_k(z)$ for $k = 1, 2$, where h_k is holomorphic on $\mathbb{D}$ and $h_k(0) \neq 0$;

(2) if $r_1 - r_2 \in \mathbb{Z}$ then (4.1.1) has two linearly independent solutions of the form $u_1(z) = z^{r_1} h_1(z)$ and $u_2(z) = z^{r_2} h_2(z) + C u_1(z) \log(z)$ where $C \in \mathbb{C}$, and h_k is holomorphic on $\mathbb{D}$ with $h_k(0) \neq 0$ for $k = 1, 2$.

An analogous dichotomy for solutions of (4.1.2) is easier to state if we write the leading coefficient in the form $a = \frac{1}{2}(1 - \theta^2)$, where $\theta = \pm\sqrt{1-2a}$ will be called the *reduced exponent* of q at $z = 0$. With respect to the terminology used in [Allegretti and Bridgeland 2020], the *exponent* of q at $z = 0$ is $r = \pm 2\pi i \sqrt{1-2a} = 2\pi i \theta$. For the reader's convenience, we remark that in [Allegretti and Bridgeland 2020] a slightly different form of the Schwarzian derivative is used, leading to a different normalization for constants in the correspondence between differentials and monodromy of solutions. Observing that $\pm\theta = r_1 - r_2$, and recalling the relation $f = u_1/u_2$, one has the following:

(1) if $\theta \notin \mathbb{Z}$ then (4.1.2) has a solution of the form $f(z) = z^\theta M(z)$, where M is holomorphic at $z = 0$, $M(0) \neq 0$;

(2) if $\theta \in \mathbb{Z}$ then (4.1.2) has a solution of the form $f(z) = z^\theta M(z) + C \log(z)$, where $C \in \mathbb{C}$, and M is holomorphic at $z = 0$, $M(0) \neq 0$.

For each q one can regard a solution to (4.1.2) as a developing map for a projective structure on $\mathbb{D}^*$, equivariant with respect to the monodromy group of the equation. Notice that the holonomy of this structure (ie the monodromy of (4.1.2)) is a representation $\rho: \pi_1(\mathbb{D}^*) \to \mathrm{PSL}_2\mathbb{C}$ which is just the projectivization of the monodromy $\tilde{\rho}: \pi_1(\mathbb{D}^*) \to \mathrm{SL}_2\mathbb{C}$ of (4.1.1). If γ denotes a simple loop in $\mathbb{D}^*$ around $z = 0$, then the action of the monodromy is given by the linear fractional transformation $\rho(\gamma) \cdot z = e^{2\pi i \theta} z + 2\pi i C$.

A direct computation using the above description of solutions to (4.1.2) leads to the following statement. Here continuous extensions to the origin should be thought in the sense of the end-extension topology introduced in Section 3.1.

Lemma 4.1.1 *In the above notation, the following hold*:

(1) *if $\theta = 0$ then $\rho(\gamma)$ is parabolic (necessarily $C \neq 0$)*;

(2) *if $\theta \in \mathbb{Z} \setminus \{0\}$, then $\rho(\gamma)$ is trivial (if $C = 0$) or parabolic (if $C \neq 0$)*;

(3) *if $\theta \in \mathbb{R} \setminus \mathbb{Z}$, then $\rho(\gamma)$ is elliptic*;

(4)   *if $\theta \in \mathbb{Z} \oplus i\mathbb{R}$, then $\rho(\gamma)$ is hyperbolic;*

(5)   *if $\theta \in \mathbb{C} \setminus (\mathbb{Z} \oplus i\mathbb{R})$, then $\rho(\gamma)$ is purely loxodromic.*

Moreover, if $\theta \in \mathbb{R} \setminus \mathbb{Z}$, then a solution f of (4.1.2) extends continuously to $z = 0$.

As the reader might expect, projective structures in $\mathscr{P}^{\odot}(\Sigma)$ relate to the elliptic case in the above statement. On the other hand, a solution f of (4.1.2) does not extend continuously to $z = 0$ when $\theta = n \in \mathbb{Z}$ and $C \neq 0$ (ie when f is of the form $f(z) = z^n M(z) + C \log(z)$), which can be seen by inspecting the behavior of f along appropriately chosen sequences that spiral into the singularity. A similar phenomenon occurs when $\theta \notin \mathbb{R}$.

4.2 Meromorphic projective structures

We now recall how to construct projective structures in terms of meromorphic quadratic differentials, and discuss its relationship with our space $\mathscr{P}^{\odot}(\Sigma)$ of tame, relatively elliptic, and nondegenerate structures, introduced in Section 3.1. This is analogous to the classical parametrization of complex projective structures on closed surfaces by holomorphic quadratic differentials (see [Dumas 2009, Section 3] for an expository account). This section includes the proof of Theorem E.

Let us fix a complex structure $\overline{X}$ on the closed surface $\overline{\Sigma}$, and let $\bar{\sigma}_0$ be the $\mathbb{CP}^1$–structure on $\overline{X}$ defined by the Poincaré uniformization, ie the unique conformal metric of constant curvature -1, 0 or 1, the exact value depending on the genus g of $\overline{X}$. Let X be the induced complex structure on $\Sigma = \overline{\Sigma} \setminus \{x_1, \ldots, x_n\}$; notice X is a punctured Riemann surface, ie each x_j has a neighborhood biholomorphic to $\mathbb{D}^*$. We consider the space $\mathscr{Q}_2(X)$ of *meromorphic quadratic differentials* with at worst double poles at the punctures of X; these are meromorphic sections of the line bundle K_X^2, where K_X denotes the canonical bundle of X. More concretely, by slight abuse of notation, in suitable local complex coordinates around the puncture these differentials can be written as

$$q(z) = \left(\frac{a}{z^2} + O\left(\frac{1}{z} \right) \right) dz^2.$$

The leading coefficient at a double pole is a well-defined invariant of a quadratic differential, ie does not depend on the chosen coordinates (see [Strebel 1984, Section 4.2]). In particular, the local analysis developed in Section 4.1 applies, and provides a definition of exponents and reduced exponents of q at a puncture.

Moreover the properties of the Schwarzian derivative ensure that the Schwarz equation $\mathscr{S}f = q$ is well-defined on X, as soon as a background projective structure has been fixed, and we choose the Poincaré uniformization $\bar{\sigma}_0$. Local solutions are in general multivalued, ie they should be considered as functions on the universal cover, equivariant with respect to some representation $\rho_q : \pi(X) \to \mathrm{PSL}_2\mathbb{C}$, which is called the *monodromy* of q. We say a puncture is an *apparent singularity* if $\rho_q(\gamma)$ is trivial for a peripheral loop γ around the puncture. It is a theorem of Luo [1993] that differentials without apparent singularities are locally determined by their monodromy. The analogous results for holomorphic quadratic differentials is due to Hejhal [1975].

Following [Allegretti and Bridgeland 2020, Section 3; Gupta and Mj 2021, Section 3.1], we define a *meromorphic projective structure* to be the structure σ_q induced by a meromorphic quadratic differentials $q \in \mathcal{Q}_2(X)$ as follows: a developing map dev_q for σ_q is given by taking a local solution to $\mathcal{S}f = q$ and considering its analytic continuation as a function on the universal cover; the monodromy of the differential provides the holonomy ρ_q of the structure. The differential q is recovered from σ_q by computing the Schwarzian derivative of dev_q with respect to the background projective structure $\bar{\sigma}_0$.

For the sake of clarity, we emphasize that this correspondence between meromorphic quadratic differentials and meromorphic projective structures is not canonical, and does depend on the choice of a background projective structure. Changing this choice only translates the differentials by the vector space of holomorphic differentials; hence orders and leading coefficients of poles are well-defined invariant for the projective structure.

We are now ready to provide a proof of the following correspondence. Here $\mathscr{P}^\odot(\Sigma)$ is the space of tame, relatively elliptic and nondegenerate structures introduced in Section 3.1.

Theorem E *Let $\sigma \in \mathscr{P}(\Sigma)$ and let $X \in \mathcal{T}(\Sigma)$ be the underlying complex structure. Then $\sigma \in \mathscr{P}^\odot(\Sigma)$ if and only if X is a punctured Riemann surface and σ is represented by a meromorphic quadratic differential on X with double poles and reduced exponents in $\mathbb{R} \setminus \mathbb{Z}$.*

Proof We prove the backward direction first. Let X be a punctured Riemann surface structure on Σ, and let $\sigma = \sigma_q$ for some meromorphic quadratic differential $q \in \mathcal{Q}_2(X)$ with reduced exponents $\theta_i \in \mathbb{R} \setminus \mathbb{Z}$. By Lemma 4.1.1, since the θ_i are real but not integers, the developing map for σ extends continuously to the punctures (ie σ is tame), and the peripheral holonomy of σ is elliptic at every puncture. In particular, the holonomy representation is known to be nondegenerate by [Allegretti and Bridgeland 2020, Theorem 6.1], as there are no apparent singularities. Therefore $\sigma \in \mathscr{P}^\odot(\Sigma)$.

We now prove the forward direction. Let $\sigma \in \mathscr{P}^\odot(\Sigma)$, and let U be a neighborhood of a puncture x, which is some conformal annulus. We claim that its modulus is infinite. Let $E \in \mathcal{E}(\widetilde{\Sigma})$ be an end covering x, and let $\widetilde{U}$ be the lift of U around E. By Theorem D we can choose U so that $\mathrm{dev}: \widetilde{U} \to D^* = \mathrm{dev}(\widetilde{U})$ is a conformal covering map onto a punctured disk. The family of curves Γ in D^* joining the boundary to the puncture has infinite extremal length, lifts to a family of curves in $\widetilde{U}$ joining $\partial\widetilde{U}$ to E, and projects to a family of curves in U joining the boundary to the puncture x. Since extremal length is conformally invariant, this family has infinite extremal length in U, hence the modulus of U is infinite. This shows that the complex structure X underlying σ is that of a punctured Riemann surface.

Finally let us check the conditions on the differential are satisfied. Let $\tilde{q} = \mathcal{S}(\mathrm{dev})$; recall we have fixed the Poincaré uniformization $\bar{\sigma}_0$ as a reference projective structure on $\overline{\Sigma}$, and we are taking Schwarzian derivatives with respect to the induced structure on Σ. Since dev is a conformal immersion, possibly branching only at the ends, $\tilde{q}$ is holomorphic on $\widetilde{\Sigma}$, possibly with double poles at the ends. By the classical cocycle property of the Schwarzian, $\tilde{q}$ descends to a meromorphic quadratic differential q with at worst

double poles on Σ. By Lemma 4.1.1, since the peripheral holonomy is elliptic, the reduced exponents must be in $\mathbb{R} \setminus \mathbb{Z}$.   □

For completeness, with respect to the list of cases in Lemma 4.1.1, we observe the following. Differentials with zero reduced exponents at all punctures correspond to parabolic projective structures (see [Deroin and Dujardin 2017; Hussenot Desenonges 2019; Kra 1969; 1971a; 1971b]). Differentials with integer nonzero reduced exponents and trivial holonomy at the punctures (apparent singularities) correspond to branched projective structures (see [Calsamiglia et al. 2014a; 2019; Francaviglia and Ruffoni 2021; Mandelbaum 1972]). The next lemma implies that for structures in $\mathscr{P}^{\odot}(\Sigma)$ the absolute value of the exponent at a puncture coincides with the value of the index, as defined in Section 3.4.

Lemma 4.2.1 If $q \in \mathfrak{Q}_2(X)$ has reduced exponent $\pm\theta \in \mathbb{R} \setminus \mathbb{Z}$ at a puncture x, then the index of σ_q at that puncture is $\mathrm{I}_\sigma(x) = 2\pi|\theta|$.

Proof Let z be a coordinate around the puncture, let η be a simple closed positively oriented peripheral loop around the puncture. Up to normalizing by a Möbius transformation, we can assume that a local determination of the developing map is given by $w = f(z) = \mathrm{dev}_{q_\sigma}(z) = z^\theta M(z)$, for $\theta > 0$ and for some M holomorphic and nonzero at $z = 0$ (see Section 4.1). Then the statement follows from the following computation in local coordinates:

$$\int_{f(\eta)} \frac{dw}{w} = \int_\eta \frac{\theta z^{\theta-1} M(z) + z^\theta M'(z)}{z^\theta M(z)} dz = \theta \int_\eta \frac{dz}{z} + \int_\eta \frac{M'(z)}{M(z)} dz = 2\pi i \theta,$$

where the second integral vanishes, because M is holomorphic, and η can be chosen to be small enough to enclose $z = 0$ but no zero of M.   □

Remark 4.2.2 When the exponent (equivalently the reduced exponent) is not zero, a choice of a sign is called a *signing* of the projective structure at that puncture, and can be used to define a framing from the holonomy representation (see [Allegretti and Bridgeland 2020; Gupta 2021]). This is in general an arbitrary choice. However, as observed in Corollary 3.1.5, continuously extending the developing map to the punctures always provides a canonical framing for structures in $\mathscr{P}^{\odot}(\Sigma)$.

5 Structures on the thrice-punctured sphere

In this chapter we prove Theorems A and B about grafting structures on the *thrice-punctured sphere* $S := \mathbb{S}^2 \setminus \{x_\alpha, x_\beta, x_\gamma\}$. This is the oriented topological space obtained from the 2–dimensional unit sphere $\mathbb{S}^2$ by removing three distinct points $\{x_\alpha, x_\beta, x_\gamma\} \subset \mathbb{S}^2$. The points $\{x_\alpha, x_\beta, x_\gamma\}$ are the *punctures* of S. (For the easier case of the twice-punctured sphere we refer the reader back to Remark 3.3.8.) The *fundamental group* $\pi_1(S)$ of S is isomorphic to the free group on two generators $\mathbb{F}_2$. Once and for all we fix the presentation

$$\pi_1(S) = \langle \alpha, \beta, \gamma \mid \alpha\beta\gamma = 1 \rangle \cong \mathbb{F}_2,$$

where each generator $\delta \in \{\alpha, \beta, \gamma\}$ can be represented by a peripheral loop (also denoted by δ) around x_δ, oriented to travel around the puncture in the anticlockwise direction. Furthermore, we denote by $E_\delta \in \mathcal{E}(\widetilde{S})$ the end in the end-extended universal cover $\widetilde{S}^{\#}$ of S, that is fixed by δ.

In this setting, we observe that $\mathcal{P}^\bullet(S)$ is the space of complex projective structures whose underlying conformal structure is that of $\mathbb{CP}^1 \setminus \{0, 1, \infty\}$. The $\mathrm{PSL}_2\mathbb{C}$–character variety can be explicitly described (see [Heusener and Porti 2004, Remark 4.4] for details). A conjugacy class of representations is said to be *nondegenerate relatively elliptic* if it is the class of a nondegenerate relatively elliptic representation. It follows from Theorem E and [Gupta 2021, Theorem 1.1] that any nondegenerate relatively elliptic conjugacy class arises from the holonomy of a structure in $\mathcal{P}^\odot(S)$. We will see that the structure can be chosen to be of a special type (see Corollary 5.1.4).

Remark 5.0.1 A relatively elliptic representation of $\pi_1(S)$ is degenerate if and only if its image is a subgroup of rotations around two fixed points, ie a group of coaxial rotations (see [Gupta 2021, Section 2.4]).

The main result of this chapter is a complete description of $\mathcal{P}^\odot(S)$. We begin in Section 5.1 by constructing some structures in $\mathcal{P}^\odot(S)$, called *triangular structures*, which will be our key examples. Then in Section 5.2 we show that $\mathcal{P}^\odot(S)$ is precisely the space of complex projective structures obtained by grafting triangular structures.

5.1 Triangular structures

In this section we construct a family of structures in $\mathcal{P}^\odot(S)$ which will be the main reference example for the rest of the paper.

First, we fix the following *ideal triangulation* $\mathfrak{T}$ of S (see Figure 14). For every distinct pair $\delta, \delta' \in \{\alpha, \beta, \gamma\}$, let $e_{\delta\delta'}$ be a simple arc on S from x_δ to $x_{\delta'}$. The collection of arcs $\{e_{\alpha\beta}, e_{\beta\gamma}, e_{\alpha\gamma}\}$ are the *ideal edges* of $\mathfrak{T}$, and subdivide S into two *ideal triangles* t_S and $\bar{t}_S$. The orientation of S induces an orientation on t_S (resp. $\bar{t}_S$) such that the punctures are ordered as $(x_\alpha, x_\beta, x_\gamma)$ (resp. $(x_\alpha, x_\gamma, x_\beta)$) on its boundary. The ideal triangulation $\mathfrak{T}$ lifts to a triangulation $\widetilde{\mathfrak{T}}$ of $\widetilde{S}^{\#}$. We notice that the restriction of $\widetilde{\mathfrak{T}}$ to $\widetilde{S}$ is an ideal triangulation of $\widetilde{S}$. We denote by $\tilde{t}_S$ the unique triangle in $\widetilde{\mathfrak{T}}$ with vertices $\{E_\alpha, E_\beta, E_\gamma\}$, and by $\tilde{t}_S^\delta$ the unique triangle adjacent to $\tilde{t}_S$ that does not have E_δ as its vertex. It is easy to check that $\tilde{t}_S$ projects onto t_S, while $\{\tilde{t}_S^\alpha, \tilde{t}_S^\beta, \tilde{t}_S^\gamma\}$ all project onto $\bar{t}_S$.

Recall that $\triangle \subset \mathbb{R}^3$ is the standard 2–simplex (see Section 2.3). Let $\tau : \triangle \to \mathbb{CP}^1$ be a nondegenerate triangular immersion, with vertices (V_a, V_b, V_c) and angles (a, b, c). Let $\mathfrak{C}_\tau = (\mathscr{C}_{ab}, \mathscr{C}_{bc}, \mathscr{C}_{ac})$ be the configuration of circles determined by τ, defined such that $V_x, V_y \in \mathscr{C}_{xy}$, for all distinct pairs $x, y \in \{a, b, c\}$. From Corollary 2.2.7 we have a relatively elliptic representation associated to $\mathfrak{C}_\tau$ given by

$$\rho_\tau := \rho_{\mathfrak{C}_\tau} : \pi_1(S) \to \mathrm{PSL}_2\mathbb{C},$$

$$\rho_\tau(\alpha) := J_{ac} J_{ab}, \quad \rho_\tau(\beta) := J_{ab} J_{bc}, \quad \rho_\tau(\gamma) := J_{bc} J_{ac},$$
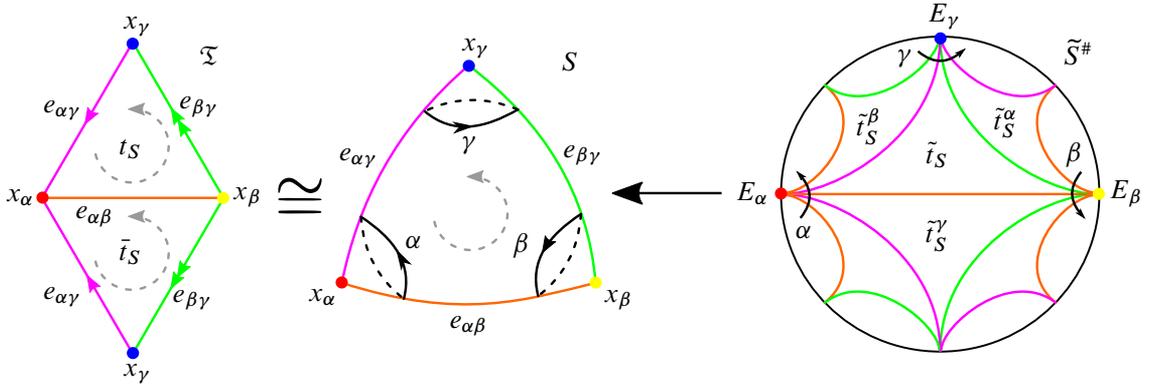
Figure 14: The ideal triangulation $\mathfrak{T}$ of the thrice-punctured sphere S, and its lift to the end-extended universal cover $\widetilde{S}^{\#}$.

where J_{xy} denotes the reflection of $\mathbb{CP}^1$ in $\mathscr{C}_{xy}$. Notice that, if τ embeds onto a Euclidean, hyperbolic or spherical triangle with angles rational multiples of π, then the image of this representation is a discrete Euclidean, hyperbolic or spherical group; however a generic choice of τ results in a nondiscrete subgroup of $\mathrm{PSL}_2\mathbb{C}$.

The *triangular structure* $\sigma_\tau \in \mathcal{P}(S)$ associated to the triangular immersion $\tau: \triangle \to \mathbb{CP}^1$ is the structure defined by the developing pair $(\mathrm{dev}_\tau, \rho_\tau)$, where the developing map is constructed as follows. Recall that (V_1, V_2, V_3) are the vertices of $\triangle$. Consider the following maps:

(1) $\varphi: \tilde{t}_S \to \triangle$, the unique simplicial map mapping $(E_\alpha, E_\beta, E_\gamma)$ to (V_1, V_2, V_3);

(2) $\varphi^\gamma: \tilde{t}_S^\gamma \to \triangle$, the unique simplicial map mapping $(E_\beta, E_\alpha, E_{\beta\gamma\beta^{-1}})$ to (V_1, V_2, V_3);

(3) $\iota: \triangle \to \triangle$, the unique (orientation reversing) simplicial map mapping (V_1, V_2, V_3) to (V_2, V_1, V_3);

(4) $\tau^\gamma := J_{ab} \circ \tau \circ \iota$, the *triangular immersion conjugate to* τ, mapping (V_1, V_2, V_3) to $(V_b, V_a, J_{ab}(V_c))$.

Then we define

$$(\mathrm{dev}_\#)_\tau|_{\tilde{t}_S} := \tau \circ \varphi \quad \text{and} \quad (\mathrm{dev}_\#)_\tau|_{\tilde{t}_S^\gamma} := \tau^\gamma \circ \varphi^\gamma.$$

Since this defines $(\mathrm{dev}_\#)_\tau$ on a fundamental domain for the action of $\pi_1(S)$ on $\widetilde{S}^{\#}$, we can then extend it by equivariance with respect to the representation ρ_τ to obtain a global $(\mathrm{dev}_\#)_\tau: \widetilde{S}^{\#} \to \mathbb{CP}^1$. The developing map dev_τ is the restriction of $(\mathrm{dev}_\#)_\tau$ to $\widetilde{S}$. Notice that, when τ is an embedding, this is the pillowcase structure obtained by doubling $\tau(\triangle)$.

By construction, triangular structures are nondegenerate, tame and their holonomy representations are relatively elliptic. We record this in the following lemma.

Lemma 5.1.1 *Let τ be a nondegenerate triangular immersion and let σ_τ be the associated triangular structure. Then $\sigma_\tau \in \mathcal{P}^{\circ}(S)$.*

Triangular immersions that are especially simple, eg embeddings, carry some obvious curves that one can graft along, namely the edges $e_{\delta\delta'}$ of the triangulation $\mathfrak{T}$. Other graftable curves are those joining one
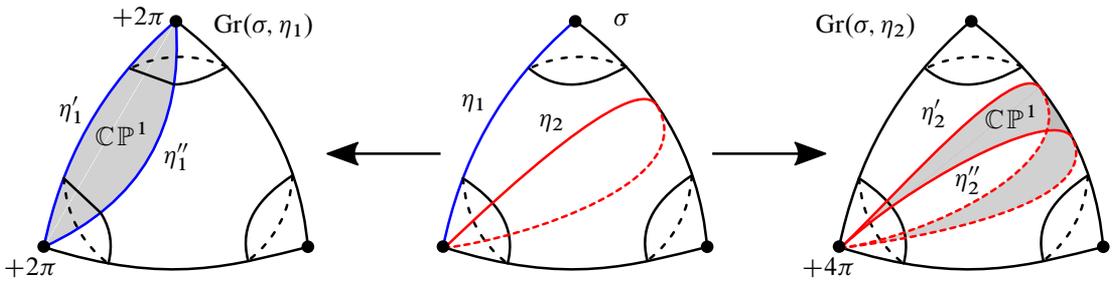
Figure 15: An edge-grafting and a core-grafting on a structure σ.

puncture to itself by crossing the triangle. We introduce the following terminology, motivated by these observations (see Section 3.1 for the general definition of this surgery). Let $\sigma \in \mathscr{P}^{\circ}(S)$ and let $\eta: I \to S$ be a graftable curve. The grafting along η will be called an *edge-grafting* if η joins two different punctures, and a *core-grafting* if it starts and ends at the same puncture and separates S into two punctured disks. The inverse surgery will be called *edge-degrafting* and *core-degrafting* respectively (see Figure 15).

Example 5.1.2 Some embedded triangular structures allow for an easy description of edge-grafting. Let $\tau, \tau': \triangle \to \mathbb{CP}^1$ be two triangular embeddings such that $\sigma_{\tau'}$ differs from σ_τ by the insertion of a disk D along one of the edges (see the first two pictures of Figure 16). Then $\sigma_{\tau'}$ is isomorphic to the structure obtained by edge-grafting σ_τ along that edge. Indeed reflecting in the edges of $\tau'(\triangle)$ we obtain a copy of $\mathbb{CP}^1$ obtained by the union of D and its complement. Since D is included in $\tau'(\triangle)$, its complement is contained in a suitable reflection of it; the union of D and its complement gives precisely a grafting region on $\sigma_{\tau'}$. This grafting procedure can be iterated by thinking of immersions as membranes spread over $\mathbb{CP}^1$, obtained by including additional disks across the edges that are being grafted. This is a particularly concrete way of thinking about edge-grafting triangular structures.

A triangular structure is said to be *Euclidean/hyperbolic/spherical atomic* if it comes from a Euclidean/hyperbolic/spherical atomic triangular immersion (see the end of Section 2.3). The terminology is motivated by the main theorem (Theorem B), which states that every tame and relatively elliptic $\mathbb{CP}^1$–structure is obtained by grafting an atomic structure.
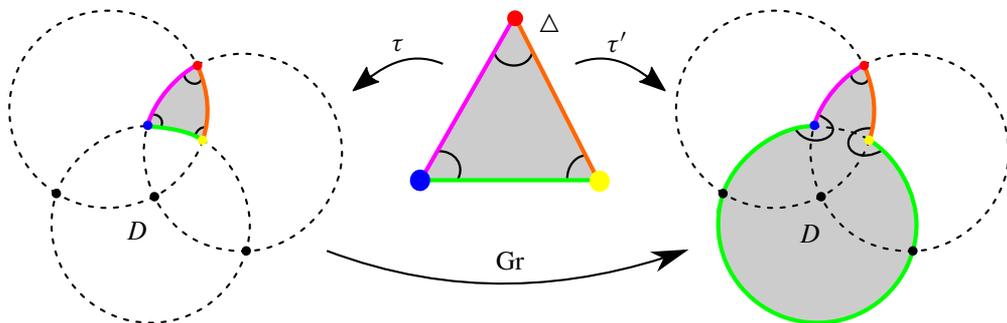

Figure 16: An edge-grafting on an embedded structure σ.
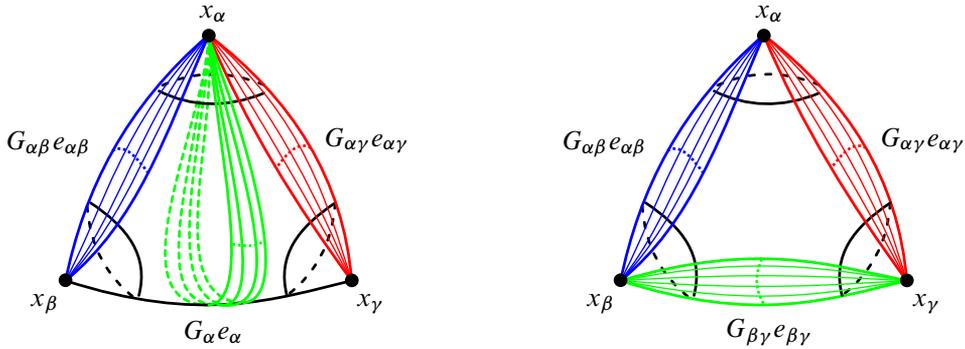
Figure 17: The multicurves η (on the left), and η' (on the right).

Lemma 5.1.3 *Let σ be an atomic triangular structure with indices $I_\sigma := (2a, 2b, 2c)$. Let $e_{\delta\delta'}$ be the edge of the triangle of $\mathfrak{T}$ in S connecting the two distinct punctures x_δ and $x_{\delta'}$. Let e_δ be a simple ideal arc in S connecting the puncture x_δ to itself by crossing the edge opposite to x_δ. For $G_{\alpha\beta}, G_{\alpha\gamma}, G_\alpha, G_{\beta\gamma} \in \mathbb{N}$, consider the formal sums*

$$\eta := G_{\alpha\beta} e_{\alpha\beta} + G_{\alpha\gamma} e_{\alpha\gamma} + G_\alpha e_\alpha \quad \text{and} \quad \eta' := G_{\alpha\beta} e_{\alpha\beta} + G_{\alpha\gamma} e_{\alpha\gamma} + G_{\beta\gamma} e_{\beta\gamma}.$$

If σ is spherical or hyperbolic, then σ is graftable along both η and η', up to small deformations. If σ is Euclidean and we further assume that $a \in (0, 3\pi)$ while $b, c \in (0, \pi)$, then

(1) *if $a \in (0, \pi)$ and $-a + b + c = \pi$, then σ is graftable along η', but not along any arc isotopic to e_α;*

(2) *if $a \in (\pi, 2\pi)$ and $a - b - c = \pi$, then σ is graftable along η, but not along any arc isotopic to $e_{\beta\gamma}$;*

(3) *if $a \in (2\pi, 3\pi)$, then σ is graftable along η, but not along any arc isotopic to $e_{\beta\gamma}$;*

(4) *otherwise σ is graftable along both η and η'.*

Proof We begin by noticing that η (and η') can be realized as a group of pairwise disjoint arcs in S (see Figure 17); therefore we only need to check that σ is graftable once along each arc (see Remark 3.1.6).

If σ comes from a triangular immersion τ supported by a spherical configuration, then σ is graftable along both η and η' because the triangular immersion τ is an embedding (see Figures 5 (right) and 7 (right)); hence each simple ideal arc develops injectively into $\mathbb{CP}^1$.

Similarly, if τ is supported by a hyperbolic configuration, then τ is an embedding unless it is as in Figure 8(2)(i). These are immersions where one angle is in $(\pi, 2\pi)$, say for example a, and $a - b - c > \pi$. In these situations, the edge $e_{\beta\gamma}$ (opposite to the large angle a) is not graftable on the nose, as the developing map develops it surjectively to a circle. However any arbitrarily small deformation of it is graftable (see Figure 18).

Finally, suppose that τ is supported by a Euclidean configuration. Here we further assume $a \in (0, 3\pi)$ while $b, c \in (0, \pi)$, namely that if there is an angle larger than π, then it is a. Here we have an issue only when a puncture is mapped to the common intersection point y of the Euclidean configuration. If $a \in (0, \pi)$
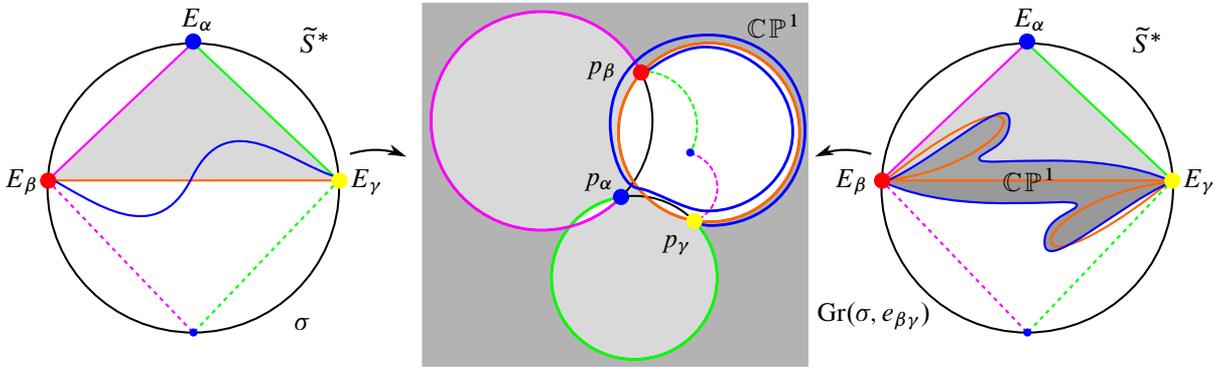
Figure 18: An edge-grafting on the hyperbolic atomic structure coming from a hyperbolic atomic triangular immersion as in Figure 8(2)(i).

and $-a + b + c = \pi$ (case (1)), then the puncture x_α develops to y and it is not possible to core-graft along any arc isotopic to e_α (see Figure 4, right). On the other hand, every edge is injectively developed, and therefore σ is graftable along η'. If $a \in (\pi, 2\pi)$ and $a - b - c = \pi$ (case (1) and Figure 9(2)(ii)), then both x_β, x_γ are mapped to y, thus σ is not graftable along any arc isotopic to $e_{\beta\gamma}$, and in particular along η'. However, e_α is injectively developed, hence σ is graftable along η. Case (3) is similar to the previous one (see Figure 10). The remaining Euclidean cases are embeddings where x_α never maps to y, hence all relevant arcs are injectively developed. $\qquad\square$

We conclude this section with a simple consequence of Lemma 5.1.1, namely that almost every nondegenerate framed relatively elliptic representation is the framed holonomy representation of an atomic triangular structure. Recall that a framing of a representation ρ is a ρ–equivariant map $\mathscr{F} : \mathscr{E}(\tilde{\Sigma}) \to \mathbb{CP}^1$ from the space of ends to $\mathbb{CP}^1$, and that for structures in $\mathscr{P}^{\odot}(S)$ there is a canonical framing given by a continuous extensions of the developing map (Corollary 3.1.5). We remark that [Gupta 2021, Theorem 1.2] states that a nondegenerate framed representation is the holonomy of a signed meromorphic projective structure with respect to some framing, while here we realize these framed representations with respect to this canonical framing (compare the discussion in Remark 4.2.2). To simplify the statement of the following result, we say that a framing $\mathscr{F}$ is *pathological* if $\mathscr{F}$ maps the entire set of ends to a single point. In our context, the holonomy representation of a triangular structure is pathological if and only if the underlying configuration of circles is Euclidean and the framing consists only of the point at infinity. Therefore the holonomy representation of an atomic triangular structure is never pathological. Note that a pathological framing is not considered degenerate according to the definition in Section 3.1.

Corollary 5.1.4 *Every nondegenerate framed relatively elliptic representation that is not pathological is the framed holonomy representation of an atomic triangular structure. In particular, $\mathscr{R}^{\odot}(S) = \mathrm{Hol}(\mathscr{P}^{\odot}(S))$.*

Proof Suppose ρ is a nondegenerate relatively elliptic representation, with a nonpathological framing $\mathscr{F}$. Then $(\rho(\alpha), \rho(\beta), \rho(\gamma))$ is an ordered triple of elliptic transformations with trivial product. As ρ is

nondegenerate, $(\rho(\alpha), \rho(\beta), \rho(\gamma))$ share at most one common fixed point. By Corollary 2.2.7, there is a unique nondegenerate configuration of circles $\mathfrak{C} := (\mathscr{C}_{ab}, \mathscr{C}_{bc}, \mathscr{C}_{ac})$ associated to $(\rho(\alpha), \rho(\beta), \rho(\gamma))$. By construction,

$$p_\alpha := \mathscr{F}(E_\alpha) \in \mathscr{C}_{ab} \cap \mathscr{C}_{ac}, \quad p_\beta := \mathscr{F}(E_\beta) \in \mathscr{C}_{ab} \cap \mathscr{C}_{bc}, \quad p_\gamma := \mathscr{F}(E_\gamma) \in \mathscr{C}_{bc} \cap \mathscr{C}_{ac}.$$

We are going to show that there is an atomic triangular immersion τ supported by $\mathfrak{C}$, with vertices $(p_\alpha, p_\beta, p_\gamma)$. As a consequence, the framed holonomy representation of its associated triangular structure σ_τ is $(\rho, \mathscr{F})$, proving the first part of the corollary. If $\mathfrak{C}$ is a spherical configuration, the points $(p_\alpha, p_\beta, p_\gamma)$ are the vertices of a unique triangular region R in $\mathbb{CP}^1 \setminus \mathfrak{C}$. Depending on the cyclic order of $(p_\alpha, p_\beta, p_\gamma)$ on the boundary of R, we either take τ to map onto R, or to map onto the complement of R in a disk (see Figure 7, right). If $\mathfrak{C}$ is a hyperbolic configuration, we refer to Table 1 to check that any framing is realized by at least one triangular immersion τ. Finally, Table 3 shows that any framing that is not pathological, namely $(-, -, -)$ and $(-, -, -)^*$, can be realized by at least one triangular immersion τ.

The last statement of the corollary follows from the observation that every nondegenerate relatively elliptic conjugacy class $[\rho]$ has a class representative ρ that can be framed with a nondegenerate and nonpathological framing. $\qquad\square$

5.2 Grafting Theorems A and B

We are now ready to prove the main results about the Grafting Conjecture. A key step will be being able to recognize structures based on their indices, which we are able to do thanks to the description of $\mathscr{P}^\odot(S)$ in terms of meromorphic differentials (Theorem E).

Up to isomorphism, there is a unique complex structure on the thrice-punctured sphere, namely that of $\mathbb{CP}^1 \setminus \{0, 1, \infty\}$. The space of meromorphic quadratic differentials with double poles at 0, 1 and ∞ can be described as

$$\left\{ q_\Theta = \left(\frac{1 - \theta_1^2}{2z^2} + \frac{1 - \theta_2^2}{2(z-1)^2} + \frac{\theta_1^2 + \theta_2^2 - \theta_3^2 - 1}{2z(1-z)} \right) dz^2 \,\Big|\, \Theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{C}^3 \right\}.$$

A direct computation shows that q_Θ has double poles at 0, 1 and ∞ with reduced exponents θ_1, θ_2 and θ_3, respectively. In particular, the indices of the structure defined by the differential q_Θ are $(2\pi|\theta_1|, 2\pi|\theta_2|, 2\pi|\theta_3|)$ (see Lemma 4.2.1). Therefore we obtain the following statement.

Proposition 5.2.1 *If $\sigma, \sigma' \in \mathscr{P}^\odot(S)$ have the same indices, then $\sigma = \sigma'$.*

Proof By Theorem E we know that $\sigma = \sigma_q$ and $\sigma' = \sigma_{q'}$ for some meromorphic differentials $q, q' \in \mathscr{Q}_2(S)$, with real noninteger reduced exponents at each puncture. Since the index at each puncture is the same, by Lemma 4.2.1 the exponent at each puncture is also the same (up to sign). So q and q' have the same leading coefficient at each puncture, but this determines them completely, so $q = q'$. $\qquad\square$

Notice that the developing maps of structures obtained with $\theta_i \in (0, 1)$ correspond to Schwarz triangle maps. The special cases in which $\theta_i = 1/p_i$, for $p_i \in \mathbb{Z}$, correspond to the classic uniform tilings of

the sphere, Euclidean or hyperbolic plane. In the general case $\theta_i \in \mathbb{R} \setminus \mathbb{Z}$, the associated holonomy representations are not discrete, and the groups are not isomorphic to triangle groups.

A direct application of Proposition 5.2.1 to Lemmas 2.3.3 and 2.3.4 allows us to easily characterize atomic structures through their indices.

Lemma 5.2.2 *Let* $\sigma \in \mathscr{P}^\odot(S)$ *with indices* $(2a, 2b, 2c)$. *Then* σ *is atomic if and only if* (*up to relabeling the punctures*) *either*

(1) $a \in (0, 2\pi)$ *and* $b, c \in (0, \pi)$, *or*

(2) $a \in (2\pi, 3\pi)$ *and* $b, c \in (0, \pi)$ *and* $a - b - c = \pi$.

Proof Atomic structures are defined in such a way that their indices satisfy the above conditions (see Lemmas 2.3.3 and 2.3.4). But more importantly, every triple of numbers $(2a, 2b, 2c)$ satisfying those conditions is the triple of indices of an atomic structure; see for example Tables 1, 2 and 3. The fact that there are no other structures with those indices follows by Proposition 5.2.1. □

As observed in Corollary 3.1.5, the holonomy representation of a structure in $\mathscr{P}^\odot(S)$ carries a natural framing, given by the extension of the developing map to the punctures. Edge-grafting and core-grafting do not change the holonomy representation, nor this framing (see Lemma 3.1.7).

Theorem B *Every* $\sigma \in \mathscr{P}^\odot(S)$ *is obtained by a sequence of edge- and core-graftings on an atomic triangular structure with the same framed holonomy.*

Proof Let $\sigma \in \mathscr{P}^\odot(S)$, and let $2a := \mathrm{I}_\sigma(x_\alpha)$, $2b := \mathrm{I}_\sigma(x_\beta)$ and $2c := \mathrm{I}_\sigma(x_\gamma)$ be its indices. Without loss of generality we can assume that $a \geq b \geq c$. Indeed we can rename the punctures so that $\mathrm{I}_\sigma(x_\alpha)$ is the largest index, and the case where $a \geq c \geq b$ follows by a similar argument.

Let $k_a = \lfloor a/\pi \rfloor$, $k_b = \lfloor b/\pi \rfloor$, $k_c = \lfloor c/\pi \rfloor \in \mathbb{N}$. We are going to reduce the triple (a, b, c) to a triple (a', b', c') by subtracting as many integer multiple of π as possible in a certain controlled way, until (a', b', c') satisfies the conditions of Lemma 2.3.3, that is

$$(5.2.1) \qquad a' \in (0, \pi) \cup (\pi, 2\pi) \quad \text{and} \quad b', c' \in (0, \pi).$$

We distinguish two cases:

(i) If $k_a \geq k_b + k_c$, let

$$G_{\alpha\gamma} := k_c, \quad G_{\alpha\beta} := k_b, \quad G_\alpha := \left\lfloor \tfrac{1}{2}(k_a - (k_b + k_c)) \right\rfloor, \quad G_{\beta\gamma} := 0.$$

(ii) If $k_a < k_b + k_c$, let $L := k_a - k_b$, $L' := k_c + k_b - k_a$ and

$$G_{\alpha\gamma} := L + \left\lfloor \tfrac{1}{2}L' \right\rfloor, \quad G_{\alpha\beta} := k_b - \left\lceil \tfrac{1}{2}L' \right\rceil, \quad G_\alpha := 0, \quad G_{\beta\gamma} := \left\lceil \tfrac{1}{2}L' \right\rceil.$$

Either way, let

$$a' := a - \pi(G_{\alpha\gamma} + G_{\alpha\beta} + 2G_\alpha), \quad b' := b - \pi(G_{\beta\gamma} + G_{\alpha\beta}), \quad c' := c - \pi(G_{\alpha\gamma} + G_{\beta\gamma}).$$

It is easy to check that $G_{\alpha\gamma}, G_{\alpha\beta}, G_\alpha, G_{\beta\gamma} \geq 0$, and

$$G_{\beta\gamma} + G_{\alpha\beta} = k_b, \quad G_{\alpha\gamma} + G_{\beta\gamma} = k_c, \quad G_{\alpha\gamma} + G_{\alpha\beta} + 2G_\alpha \in \{k_a, k_a - 1\};$$

therefore (5.2.1) is satisfied, and by Lemma 2.3.3 there is a triangular immersion τ with angles (a', b', c'). Let σ_τ be the associated triangular structure. By construction σ_τ is atomic with indices $(2a', 2b', 2c')$, thus it is left to check if σ_τ grafts to σ.

Recall we have fixed an ideal triangulation $\mathfrak{T}$ of S. Let $e_{\delta\delta'}$ be the edges of $\mathfrak{T}$ connecting the two distinct punctures x_δ and $x_{\delta'}$. Let e_δ be a simple ideal arc in S connecting the puncture x_δ to itself by crossing the edge opposite to x_δ. Consider the multicurve

$$\mu := G_{\alpha\beta}e_{\alpha\beta} + G_{\alpha\gamma}e_{\alpha\gamma} + G_\alpha e_\alpha + G_{\beta\gamma}e_{\beta\gamma}.$$

If μ is graftable then grafting σ_τ along μ would yield a structure with indices $(2a, 2b, 2c)$ and the same framed holonomy as σ_τ (Lemma 3.1.7). It follows from Proposition 5.2.1 that $\sigma = \mathrm{Gr}(\sigma_\tau, \mu)$, so it is left to check if σ_τ is graftable along μ.

Depending on the above cases, we remark that at least one of $G_{\beta\gamma}$ and G_α is 0; hence μ is either η or η' in the notation of Lemma 5.1.3.

If $G_{\beta\gamma} = G_\alpha = 0$ then $\mu = \eta = \eta'$ and every atomic triangular structure σ_τ is graftable along μ.

If $G_\alpha > 0$ then $G_{\beta\gamma} = 0$ and $\mu = \eta$. Lemma 5.1.3 covers every case except the Euclidean case where $a' \in (0, \pi)$ and $-a' + b' + c' = \pi$. In this case we must consider a different atomic structure σ'_τ and curve μ', as σ_τ is not graftable along e_α. Let

$$a'' := a' + 2\pi, \quad b'' := b', \quad c'' := c',$$

$$G'_\alpha := G_\alpha - 1, \quad \mu' := G_{\alpha\beta}e_{\alpha\beta} + G_{\alpha\gamma}e_{\alpha\gamma} + G'_\alpha e_\alpha.$$

By construction $a'' \in (2\pi, 3\pi)$, $b'', c'' \in (0, \pi)$ and $a'' - b'' - c'' = \pi$; therefore there is an atomic triangular structure σ'_τ with indices $(2a'', 2b'', 2c'')$ (see Lemma 2.3.4). Furthermore, the structure σ'_τ is graftable along μ' (Lemma 5.1.3). Grafting σ'_τ yields a structure with indices $(2a, 2b, 2c)$, which must be σ by Proposition 5.2.1, concluding this case.

Lastly, suppose that $G_{\beta\gamma} > 0$. This time $G_\alpha = 0$ and $\mu = \eta'$. Recall that $a' < 2\pi$, hence the only case that is not covered by Lemma 5.1.3 is the Euclidean case where $a' \in (\pi, 2\pi)$ and $a' - b' - c' = \pi$. We are once again forced to consider a different atomic structure as σ_τ is not graftable along $e_{\beta\gamma}$. Let

$$a'' := a' - \pi, \quad b'' := b' + \pi, \quad c'' := c',$$

$$G'_{\alpha\gamma} := G_{\alpha\gamma} + 1, \quad G'_{\beta\gamma} := G_{\beta\gamma} - 1, \quad \mu' := G_{\alpha\beta}e_{\alpha\beta} + G'_{\alpha\gamma}e_{\alpha\gamma} + G'_{\beta\gamma}e_{\beta\gamma}.$$

By construction $b'' \in (\pi, 2\pi)$, $a'', c'' \in (0, \pi)$ and $-a'' + b'' + c'' = \pi$; therefore there is an atomic triangular structure σ'_τ with indices $(2a'', 2b'', 2c'')$ (see Lemma 2.3.3). The structure σ'_τ is graftable along μ' according to Lemma 5.1.3 part (5) applied to the triple (b'', c'', a''). Once again, grafting σ'_τ along μ' yields a structure with indices $(2a, 2b, 2c)$, which must be σ by Proposition 5.2.1, concluding the proof. $\qquad\square$
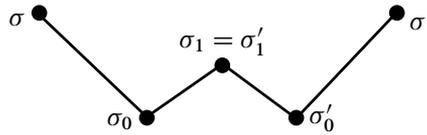
Figure 19: To prove Theorem A we find a path of graftings and degraftings from σ to σ', passing through atomic structures.

Theorem B has two interesting consequences. The first is the promised characterization of atomic structures in terms of grafting.

Corollary 5.2.3 *A structure $\sigma \in \mathcal{P}^{\odot}(S)$ is atomic if and only if it is not degraftable.*

Proof For one implication, let σ be a structure which cannot be degrafted. Then by Theorem B it must be atomic.

For the reverse implication, let σ be atomic. Suppose by contradiction that σ was degraftable to some structure σ'. Recall that core-grafting increases one index by 4π and edge-grafting increases two indices by 2π. Then σ cannot be one of the atomic structures coming from the atomic triangular immersions of Lemma 2.3.3, as its indices would be too small. It follows that σ is the atomic triangular structure associated to an atomic triangular immersion τ from Lemma 2.3.4. Without loss of generality we may assume that the largest index of σ is at x_α, while the other two are less than 2π, so that

$$\mathrm{I}_\sigma(x_\alpha) \in (4\pi, 6\pi), \quad \mathrm{I}_\sigma(x_\beta), \mathrm{I}_\sigma(x_\gamma) \in (0, 2\pi), \quad \mathrm{I}_\sigma(x_\alpha) - \mathrm{I}_\sigma(x_\beta) - \mathrm{I}_\sigma(x_\gamma) = 2\pi.$$

Then σ cannot be obtained by edge-grafting σ', and the only option is that σ' is a core-degrafting at x_α on σ. In particular $\mathrm{I}_{\sigma'}(x_\alpha) = \mathrm{I}_\sigma(x_\alpha) - 4\pi \in (0, 2\pi)$ and

$$\mathrm{I}_{\sigma'}(x_\alpha), \mathrm{I}_{\sigma'}(x_\beta), \mathrm{I}_{\sigma'}(x_\gamma) \in (0, 2\pi) \quad \text{and} \quad -\mathrm{I}_{\sigma'}(x_\alpha) + \mathrm{I}_{\sigma'}(x_\beta) + \mathrm{I}_{\sigma'}(x_\gamma) = 2\pi.$$

It follows that σ' is an atomic triangular structure (Lemma 5.2.2), coming from a triangular immersion τ' enclosed in a Euclidean configuration (Lemma 2.3.1). But this is impossible because σ' is not core-graftable at x_α (Lemma 5.1.3 part (1)), giving the desired contradiction. □

Next, we obtain that edge-grafting and core-grafting (together with the inverse operations) account for all the possible deformations that preserve the holonomy as a framed representation.

Theorem A *Two structures in $\mathcal{P}^{\odot}(S)$ have the same framed holonomy if and only if it is possible to obtain one from the other by some combination of graftings and degraftings along ideal arcs.*

Proof One direction is clear by Lemma 3.1.7. For the reverse implication, suppose $\sigma, \sigma' \in \mathcal{P}^{\odot}(S)$ have the same framed holonomy. By Theorem B, the structure σ (resp. σ') can be degrafted to an atomic structure σ_0 (resp. σ_0') having the same framed holonomy.

Let τ_0 and τ_0' be the atomic triangular immersions defining σ_0 and σ_0', with angles (a_0, b_0, c_0) and (a_0', b_0', c_0'), respectively. Since these structures have the same framed holonomy, up to conjugation we

can assume that τ_0 and τ_0' are supported by the same configuration of circles $\mathfrak{C}$ (see Corollary 2.2.7), and that $\tau_0(V_j) = \tau_0'(V_j)$, for $j = 1, 2, 3$. By Corollary 2.3.7 we are in one of the following two cases:

(1) $(a_0, b_0, c_0) = (a_0', b_0', c_0')$;

(2) $(a_0 - a_0', b_0 - b_0', c_0 - c_0') = (\pi, -\pi, 0)$ up to permutation.

In the first case σ_0 and σ_0' have the same indices; hence $\sigma_0 = \sigma_0'$ by Proposition 5.2.1, and we are done. For the second case, let us fix the permutation $(a_0 - a_0', b_0 - b_0', c_0 - c_0') = (\pi, -\pi, 0)$, as the other cases are similar. Then in particular $a_0, b_0' \in (\pi, 2\pi)$ while $a_0', b_0, c_0, c_0' \in (0, \pi)$. Let σ_1 (resp. σ_1') be the triangular structure obtained by grafting σ_0 along $e_{\beta\gamma}$ (resp. σ_0' along $e_{\alpha\gamma}$). These structures exist by Lemma 5.1.3 (with respect to η'), and they both have indices

$$(2a_0, 2b_0 + 2\pi, 2c_0 + 2\pi) = (2a_0' + 2\pi, 2b_0', 2c_0' + 2\pi).$$

We explicitly observe that Lemma 5.1.3 has only two cases in which η' is not graftable, and a direct inspection of Table 3 shows that those two structures are covered by the case $(a_0, b_0, c_0) = (a_0', b_0', c_0')$ above (see Remark 2.3.9). It follows that $\sigma_1 = \sigma_1'$ by Proposition 5.2.1, completing the proof. $\qquad\square$

Appendix   Tables of atomic triangular immersions

angles range					target angles		
a	b	c	conditions	type	$(\hat{a}, \hat{b}, \hat{c})$	signs	figure
$(0, \pi)$	$(0, \pi)$	$(0, \pi)$	$a + b + c < \pi$	H	(a, b, c)	$(+, +, +)$	Figure 5, left
$(0, \pi)$	$(0, \pi)$	$(0, \pi)$	$a + \pi < b + c$	H	$(a, \pi - b, \pi - c)$	$(+, -, -)$	Figure 7, left
$(0, \pi)$	$(0, \pi)$	$(0, \pi)$	$b + \pi < a + c$	H	$(\pi - a, b, \pi - c)$	$(-, +, -)$	Figure 7, left
$(0, \pi)$	$(0, \pi)$	$(0, \pi)$	$c + \pi < a + b$	H	$(\pi - a, \pi - b, c)$	$(-, -, +)$	Figure 7, left
$(\pi, 2\pi)$	$(0, \pi)$	$(0, \pi)$	$a + b + c > 3\pi$	H	$(2\pi - a, \pi - b, \pi - c)$	$(-, -, -)$	Figure 8(1)
$(0, \pi)$	$(\pi, 2\pi)$	$(0, \pi)$	$a + b + c > 3\pi$	H	$(\pi - a, 2\pi - b, \pi - c)$	$(-, -, -)$	Figure 8(1)
$(0, \pi)$	$(0, \pi)$	$(\pi, 2\pi)$	$a + b + c > 3\pi$	H	$(\pi - a, \pi - b, 2\pi - c)$	$(-, -, -)$	Figure 8(1)
$(\pi, 2\pi)$	$(0, \pi)$	$(0, \pi)$	$a - b - c > \pi$	H	$(2\pi - a, b, c)$	$(-, +, +)$	Figure 8(2)
$(0, \pi)$	$(\pi, 2\pi)$	$(0, \pi)$	$-a + b - c > \pi$	H	$(a, 2\pi - b, c)$	$(+, -, +)$	Figure 8(2)
$(0, \pi)$	$(0, \pi)$	$(\pi, 2\pi)$	$-a - b + c > \pi$	H	$(a, b, 2\pi - c)$	$(+, +, -)$	Figure 8(2)
$(\pi, 2\pi)$	$(0, \pi)$	$(0, \pi)$	$a - b + c < \pi$	H	$(a - \pi, \pi - b, c)$	$(+, -, +)$	Figure 8(3)(i)
$(0, \pi)$	$(\pi, 2\pi)$	$(0, \pi)$	$a + b - c < \pi$	H	$(a, b - \pi, \pi - c)$	$(+, +, -)$	Figure 8(3)(i)
$(0, \pi)$	$(0, \pi)$	$(\pi, 2\pi)$	$-a + b + c < \pi$	H	$(\pi - a, b, c - \pi)$	$(-, +, +)$	Figure 8(3)(i)
$(\pi, 2\pi)$	$(0, \pi)$	$(0, \pi)$	$a + b - c < \pi$	H	$(a - \pi, b, \pi - c)$	$(+, +, -)$	Figure 8(3)(ii)
$(0, \pi)$	$(\pi, 2\pi)$	$(0, \pi)$	$-a + b + c < \pi$	H	$(\pi - a, b - \pi, c)$	$(-, +, +)$	Figure 8(3)(ii)
$(0, \pi)$	$(0, \pi)$	$(\pi, 2\pi)$	$a - b + c < \pi$	H	$(a, \pi - b, c - \pi)$	$(+, -, +)$	Figure 8(3)(ii)

Table 1: Table of atomic triangular immersions of hyperbolic type.

| angles range | | | | | target angles | | |
a	b	c	conditions	type	$(\hat{a},\hat{b},\hat{c})$	signs	figure
$(0,\pi)$	$(0,\pi)$	$(0,\pi)$	$a+b+c>\pi$ $a+\pi>b+c$ $b+\pi>a+c$ $c+\pi>a+b$	S	(a,b,c)	$(+,+,+)$	Figure 5, right
$(\pi,2\pi)$	$(0,\pi)$	$(0,\pi)$	$3\pi>a+b+c$ $a+b>\pi+c$ $a+c>\pi+b$ $\pi>a-b-c$	S	$(2\pi-a,\pi-b,\pi-c)$	$(-,-,-)$	Figure 7, right
$(0,\pi)$	$(\pi,2\pi)$	$(0,\pi)$	$3\pi>a+b+c$ $a+b>\pi+c$ $b+c>\pi+a$ $\pi>-a+b-c$	S	$(\pi-a,2\pi-b,\pi-c)$	$(-,-,-)$	Figure 7, right
$(0,\pi)$	$(0,\pi)$	$(\pi,2\pi)$	$3\pi>a+b+c$ $b+c>\pi+a$ $a+c>\pi+b$ $\pi>-a-b+c$	S	$(\pi-a,\pi-b,2\pi-c)$	$(-,-,-)$	Figure 7, right

Table 2: Table of atomic triangular immersions of spherical type.

| angles range | | | | | target angles | | |
a	b	c	conditions	type	$(\hat{a},\hat{b},\hat{c})$	signs	figure
$(0,\pi)$	$(0,\pi)$	$(0,\pi)$	$a+b+c=\pi$	E	(a,b,c)	$(+,+,+)$	Figure 4, left
$(0,\pi)$	$(0,\pi)$	$(0,\pi)$	$-a+b+c=\pi$	E	$(a,\pi-c,\pi-b)$	$(-,+,+)^*$	Figure 4, right
$(0,\pi)$	$(0,\pi)$	$(0,\pi)$	$a-b+c=\pi$	E	$(\pi-a,\pi-c,b)$	$(+,-,+)^*$	Figure 4, right
$(0,\pi)$	$(0,\pi)$	$(0,\pi)$	$a+b-c=\pi$	E	$(\pi-a,c,\pi-b)$	$(+,+,-)^*$	Figure 4, right
$(\pi,2\pi)$	$(0,\pi)$	$(0,\pi)$	$a+b+c=3\pi$	E	$(2\pi-a,\pi-c,\pi-b)$	$(+,+,+)^*$	Figure 9(1)
$(0,\pi)$	$(\pi,2\pi)$	$(0,\pi)$	$a+b+c=3\pi$	E	$(\pi-a,\pi-c,2\pi-b)$	$(+,+,+)^*$	Figure 9(1)
$(0,\pi)$	$(0,\pi)$	$(\pi,2\pi)$	$a+b+c=3\pi$	E	$(\pi-a,2\pi-c,\pi-b)$	$(+,+,+)^*$	Figure 9(1)
$(\pi,2\pi)$	$(0,\pi)$	$(0,\pi)$	$a-b-c=\pi$	E	$(2\pi-a,c,b)$	$(+,-,-)^*$	Figure 9(2)
$(0,\pi)$	$(\pi,2\pi)$	$(0,\pi)$	$-a+b-c=\pi$	E	$(a,c,2\pi-b)$	$(-,+,-)^*$	Figure 9(2)
$(0,\pi)$	$(0,\pi)$	$(\pi,2\pi)$	$-a-b+c=\pi$	E	$(a,2\pi-c,b)$	$(-,-,+)^*$	Figure 9(2)
$(\pi,2\pi)$	$(0,\pi)$	$(0,\pi)$	$a-b+c=\pi$	E	$(a-\pi,\pi-b,c)$	$(+,-,+)$	Figure 9(3)(i)
$(0,\pi)$	$(\pi,2\pi)$	$(0,\pi)$	$a+b-c=\pi$	E	$(a,b-\pi,\pi-c)$	$(+,+,-)$	Figure 9(3)(i)
$(0,\pi)$	$(0,\pi)$	$(\pi,2\pi)$	$-a+b+c=\pi$	E	$(\pi-a,b,c-\pi)$	$(-,+,+)$	Figure 9(3)(i)
$(\pi,2\pi)$	$(0,\pi)$	$(0,\pi)$	$a+b-c=\pi$	E	$(a-\pi,b,\pi-c)$	$(+,+,-)$	Figure 9(3)(ii)
$(0,\pi)$	$(\pi,2\pi)$	$(0,\pi)$	$-a+b+c=\pi$	E	$(\pi-a,b-\pi,c)$	$(-,+,+)$	Figure 9(3)(ii)
$(0,\pi)$	$(0,\pi)$	$(\pi,2\pi)$	$a-b+c=\pi$	E	$(a,\pi-b,c-\pi)$	$(+,-,+)$	Figure 9(3)(ii)
$(2\pi,3\pi)$	$(0,\pi)$	$(0,\pi)$	$a-b-c=\pi$	E	$(a-2\pi,\pi-b,\pi-c)$	$(+,-,-)$	Figure 10
$(0,\pi)$	$(2\pi,3\pi)$	$(0,\pi)$	$-a+b-c=\pi$	E	$(\pi-a,b-2\pi,\pi-c)$	$(-,+,-)$	Figure 10
$(0,\pi)$	$(0,\pi)$	$(2\pi,3\pi)$	$-a-b+c=\pi$	E	$(\pi-a,\pi-b,c-2\pi)$	$(-,-,+)$	Figure 10

Table 3: Table of atomic triangular immersions of Euclidean type.

References

[Allegretti and Bridgeland 2020] **D G L Allegretti**, **T Bridgeland**, *The monodromy of meromorphic projective structures*, Trans. Amer. Math. Soc. 373 (2020) 6321–6367 MR Zbl

[Baba 2010] **S Baba**, *A Schottky decomposition theorem for complex projective structures*, Geom. Topol. 14 (2010) 117–151 MR Zbl

[Baba 2012] **S Baba**, *Complex projective structures with Schottky holonomy*, Geom. Funct. Anal. 22 (2012) 267–310 MR Zbl

[Baba 2015] **S Baba**, *2π–grafting and complex projective structures, I*, Geom. Topol. 19 (2015) 3233–3287 MR Zbl

[Baba 2017] **S Baba**, *2π–grafting and complex projective structures with generic holonomy*, Geom. Funct. Anal. 27 (2017) 1017–1069 MR Zbl

[Baba 2020] **S Baba**, *On Thurston's parameterization of $\mathbb{C}P^1$–structures*, from "In the tradition of Thurston—geometry and topology" (K Ohshika, A Papadopoulos, editors), Springer (2020) 241–254 MR Zbl

[Bridson and Haefliger 1999] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[Calsamiglia et al. 2014a] **G Calsamiglia**, **B Deroin**, **S Francaviglia**, *Branched projective structures with Fuchsian holonomy*, Geom. Topol. 18 (2014) 379–446 MR Zbl

[Calsamiglia et al. 2014b] **G Calsamiglia**, **B Deroin**, **S Francaviglia**, *The oriented graph of multi-graftings in the Fuchsian case*, Publ. Mat. 58 (2014) 31–46 MR Zbl

[Calsamiglia et al. 2019] **G Calsamiglia**, **B Deroin**, **V Heu**, **F Loray**, *The Riemann–Hilbert mapping for $\mathfrak{sl}_2$ systems over genus two curves*, Bull. Soc. Math. France 147 (2019) 159–195 MR Zbl

[Chenakkod et al. 2022] **S Chenakkod**, **G Faraco**, **S Gupta**, *Translation surfaces and periods of meromorphic differentials*, Proc. Lond. Math. Soc. 124 (2022) 478–557 MR Zbl

[Deroin and Dujardin 2017] **B Deroin**, **R Dujardin**, *Complex projective structures: Lyapunov exponent, degree, and harmonic measure*, Duke Math. J. 166 (2017) 2643–2695 MR Zbl

[Dumas 2009] **D Dumas**, *Complex projective structures*, from "Handbook of Teichmüller theory, II" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 13, Eur. Math. Soc., Zürich (2009) 455–508 MR Zbl

[Eremenko 2004] **A Eremenko**, *Metrics of positive curvature with conic singularities on the sphere*, Proc. Amer. Math. Soc. 132 (2004) 3349–3355 MR Zbl

[Faraco 2020] **G Faraco**, *Distances on the moduli space of complex projective structures*, Expo. Math. 38 (2020) 407–429 MR Zbl

[Faraco and Ruffoni 2019] **G Faraco**, **L Ruffoni**, *Complex projective structures with maximal number of Möbius transformations*, Math. Nachr. 292 (2019) 1260–1270 MR Zbl

[Francaviglia and Ruffoni 2021] **S Francaviglia**, **L Ruffoni**, *Local deformations of branched projective structures: Schiffer variations and the Teichmüller map*, Geom. Dedicata 214 (2021) 21–48 MR Zbl

[Gallo et al. 2000] **D Gallo**, **M Kapovich**, **A Marden**, *The monodromy groups of Schwarzian equations on closed Riemann surfaces*, Ann. of Math. 151 (2000) 625–704 MR Zbl

[Goldman 1987] **W M Goldman**, *Projective structures with Fuchsian holonomy*, J. Differential Geom. 25 (1987) 297–326 MR Zbl

[Gunning 1967]   **R C Gunning**, *Special coordinate coverings of Riemann surfaces*, Math. Ann. 170 (1967) 67–86 MR Zbl

[Gupta 2021]   **S Gupta**, *Monodromy groups of $\mathbb{C}\mathrm{P}^1$–structures on punctured surfaces*, J. Topol. 14 (2021) 538–559 MR Zbl

[Gupta and Mj 2020]   **S Gupta**, **M Mj**, *Monodromy representations of meromorphic projective structures*, Proc. Amer. Math. Soc. 148 (2020) 2069–2078 MR Zbl

[Gupta and Mj 2021]   **S Gupta**, **M Mj**, *Meromorphic projective structures, grafting and the monodromy map*, Adv. Math. 383 (2021) art. no. 107673 MR Zbl

[Hejhal 1975]   **D A Hejhal**, *Monodromy groups and linearly polymorphic functions*, Acta Math. 135 (1975) 1–55 MR Zbl

[Hensel 2011]   **S W Hensel**, *Iterated grafting and holonomy lifts of Teichmüller space*, Geom. Dedicata 155 (2011) 31–67 MR Zbl

[Heusener and Porti 2004]   **M Heusener**, **J Porti**, *The variety of characters in* $\mathrm{PSL}_2(\mathbb{C})$, Bol. Soc. Mat. Mexicana 10 (2004) 221–237 MR Zbl

[Hille 1969]   **E Hille**, *Lectures on ordinary differential equations*, Addison-Wesley, Reading, MA (1969) MR Zbl

[Hussenot Desenonges 2019]   **N Hussenot Desenonges**, *Heijal's theorem for projective structures on surfaces with parabolic punctures*, Geom. Dedicata 200 (2019) 93–103 MR Zbl

[Ince 1944]   **E L Ince**, *Ordinary differential equations*, Dover, New York (1944) MR Zbl

[Kamishima and Tan 1992]   **Y Kamishima**, **S P Tan**, *Deformation spaces on geometric structures*, from "Aspects of low-dimensional manifolds" (Y Matsumoto, S Morita, editors), Adv. Stud. Pure Math. 20, Kinokuniya, Tokyo (1992) 263–299 MR Zbl

[Kapovich 2020]   **M Kapovich**, *Periods of abelian differentials and dynamics*, from "Dynamics: topology and numbers" (P Moree, A Pohl, Ľ Snoha, T Ward, editors), Contemp. Math. 744, Amer. Math. Soc., Providence, RI (2020) 297–315 MR Zbl

[Kra 1969]   **I Kra**, *Deformations of Fuchsian groups*, Duke Math. J. 36 (1969) 537–546 MR Zbl

[Kra 1971a]   **I Kra**, *Deformations of Fuchsian groups, II*, Duke Math. J. 38 (1971) 499–508 MR Zbl

[Kra 1971b]   **I Kra**, *A generalization of a theorem of Poincaré*, Proc. Amer. Math. Soc. 27 (1971) 299–302 MR Zbl

[Kulkarni and Pinkall 1994]   **R S Kulkarni**, **U Pinkall**, *A canonical metric for Möbius structures and its applications*, Math. Z. 216 (1994) 89–129 MR Zbl

[Luo 1993]   **F Luo**, *Monodromy groups of projective structures on punctured surfaces*, Invent. Math. 111 (1993) 541–555 MR Zbl

[Mandelbaum 1972]   **R Mandelbaum**, *Branched structures on Riemann surfaces*, Trans. Amer. Math. Soc. 163 (1972) 261–275 MR Zbl

[Maskit 1969]   **B Maskit**, *On a class of Kleinian groups*, Ann. Acad. Sci. Fenn. Ser. A I 442 (1969) 8 MR Zbl

[Mondello and Panov 2016]   **G Mondello**, **D Panov**, *Spherical metrics with conical singularities on a 2–sphere: angle constraints*, Int. Math. Res. Not. 2016 (2016) 4937–4995 MR Zbl

[Poincaré 1908]   **H Poincaré**, *Sur l'uniformisation des fonctions analytiques*, Acta Math. 31 (1908) 1–63 MR Zbl

[Ratcliffe 2006]   **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, 2nd edition, Graduate Texts in Math. 149, Springer (2006)   MR Zbl

[Ruffoni 2019]   **L Ruffoni**, *Multi(de)grafting quasi-Fuchsian complex projective structures via bubbles*, Differential Geom. Appl. 64 (2019) 158–173   MR Zbl

[Ruffoni 2021]   **L Ruffoni**, *Bubbling complex projective structures with quasi-Fuchsian holonomy*, J. Topol. Anal. 13 (2021) 843–887   MR Zbl

[Strebel 1984]   **K Strebel**, *Quadratic differentials*, Ergebnisse der Math. 5, Springer (1984)   MR Zbl

Department of Mathematics, Florida State University
Tallahassee, FL, United States

Department of Mathematics, Florida State University
Tallahassee, FL, United States

Department of Mathematics, Florida State University
Tallahassee, FL, United States

Department of Mathematics, Tufts University
Medford, MA, United States

ballas@math.fsu.edu,   bowers@math.fsu.edu,   alex.casella.usyd@gmail.com,
lorenzo.ruffoni2@gmail.com

Shadows of 2–knots and complexity

HIRONOBU NAOE

We introduce a new invariant for a 2–knot in S^4, called the shadow-complexity, based on the theory of Turaev shadows, and we give a characterization of 2–knots with shadow-complexity at most 1. Specifically, we show that the unknot is the only 2–knot with shadow-complexity 0 and that there exist infinitely many 2–knots with shadow-complexity 1.

1 Introduction

A 2–*knot* is a smoothly embedded 2–sphere in S^4. The first example of a nontrivial 2–knot was given by Artin [1925], called a spun knot. The nontriviality is of fundamental interest in knot theory. In the classical case, namely the 1–knot case, the trefoil knot can be said to be the simplest nontrivial knot in terms of some numerical invariants: the crossing number, the bridge number, the tunnel number, and so forth. Then one naturally wonders which 2–knot is the simplest nontrivial one. The answer to this question should be based on as many criteria as possible that measures, in some sense, how different a given 2–knot is from the trivial one.

There are several studies on numerical invariants for 2–knots. For examples, we refer the reader to [Satoh 2000; Satoh and Shima 2004; Yajima 1964] for the triple point number and to [Saito and Satoh 2005; Satoh 2009] for the sheet number. These two invariants are defined with broken surface diagrams [Carter

et al. 1997; Roseman 1998], which is a natural analogue of classical knot diagrams. Yoshikawa [1994] introduced the ch-index by using ch-diagrams (also called marked graph diagrams), and he gave the table of 2–knots with ch-index up to 10. These invariants measure how complicated the descriptions of a given 2–knot must be, like the crossing numbers for 1–knots. On the one hand, for example, the bridge number and the tunnel number for classical knots seem to be the complexity of the embedding or the complement. Recently, bridge positions of knotted surfaces, called bridge trisections, were introduced by Meier and Zupan [2017; 2018] using a trisection (see [Gay and Kirby 2016] for the details) of the ambient 4–manifold, and the bridge numbers for knotted surfaces were defined by using this notion. Incidentally, Kirby and Thompson introduced an integer-valued invariant, called the $\mathscr{L}$–invariant or the Kirby–Thompson length, for closed 4–manifolds, and this notion was adapted to the knotted surfaces in [Blair et al. 2022].

In the present paper, we propose to study 2–knots using a Turaev *shadow*; a notion introduced by Turaev [1994]. A shadow is a simple polyhedron embedded in a closed 4–manifold as a 2–skeleton, in other words, a simple polyhedron such that the complement of its neighborhood is diffeomorphic to a 4–dimensional 1–handlebody. Turaev showed that regions of a shadow are naturally equipped with half-integers such as "relative self-intersection numbers", which has sufficient information to reconstruct the ambient 4–manifold. It is known as Turaev's reconstruction theorem. Thus, a shadow can be treated as a description of a 4–manifold, which brings about several studies: Stein structures, spin^c structures and complex structures, stable maps and hyperbolic structures on 3–manifolds, Lefschetz fibrations, and so on; see [Costantino 2006b; 2008; Costantino and Thurston 2008; Ishikawa and Koda 2017; Ishikawa and Naoe 2020] for examples. Moreover, as another benefit of the theory of shadows, we can define a complexity for 4–manifolds, called the *shadow-complexity*. Costantino [2006a] introduced this notion. The shadow-complexity of a 4–manifold is defined as the minimum number of specific points called true vertices contained in a shadow of the 4–manifold. He also gave the classification of closed 4–manifolds with complexity up to 1 in a special case. Martelli [2011] gave a characterization of all the closed 4–manifolds with shadow-complexity 0. He also showed that a closed simply connected 4–manifold with shadow-complexity 0 is diffeomorphic to S^4 or the connected sum of some copies of $\mathbb{CP}^2$, $\overline{\mathbb{CP}}^2$ and $S^2 \times S^2$. This implies that the shadow-complexity is a diffeomorphism invariant. Note that exotic smooth structures on $\mathbb{CP}^2 \# k\overline{\mathbb{CP}}^2$ have been found for some k. Koda, Martelli and the present author [Koda et al. 2022] introduced a new complexity, called the connected shadow-complexity, and gave a characterization of all the closed 4–manifolds with connected shadow-complexity at most 1.

If a surface F is embedded in a shadow of a 4–manifold, then such an F in the 4–manifold is smoothly embedded and generally knotted. In view of this situation, we define a shadow of a 2–knot K as a shadow X of the ambient 4–manifold S^4 such that K is embedded in X. Of course, we can define a shadow for a knotted surface as well (see Remark 4.5), but the focus in this paper is 2–knots.

Every closed 4–manifold admits a shadow, which follows from the existence of a handle decomposition. We can also show the following.

Theorem 3.2 *Every 2–knot admits a shadow.*

The above theorem allows us to define a complexity for 2–knots

$$\{2\text{–knots in } S^4\} \to \mathbb{Z}_{\geq 0}$$

as the minimum number of true vertices in a shadow of K. We call this number the *shadow-complexity* of K and write it as $\mathrm{sc}(K)$.

The aim of this paper is to give the classification of 2–knots with shadow-complexity at most 1. First, one might expect that the unknot is the only 2–knot with shadow-complexity 0 as well as some other numerical invariants for 1– or 2–knots, which is indeed true.

Theorem 6.4 *A 2–knot has shadow-complexity* 0 *if and only if it is unknotted.*

The unknotted 2–knot admits a special shadow with no true vertices, which implies that the unknot is also the only 2–knot with special shadow-complexity 0. The condition for a shadow to be special is fairly strong, and any special polyhedron with one true vertex cannot be a shadow of any 2–knot except for the unknot.

Theorem 7.2 *There are no 2–knots with special shadow-complexity* 1.

It is noteworthy that the special shadow-complexity for closed 4–manifolds is a finite-to-one invariant [Martelli 2005, Corollary 2.7]. However, that for 2–knots is not finite-to-one as noted in Remark 8.11, where we find infinitely many 2–knots with special shadow-complexity at most 5. We actually have not determined the special shadow-complexity of any nontrivial 2–knot. Note that all the special polyhedra with complexity up to 2 are listed in [Koda and Naoe 2020, Appendix B], so we already have possible candidates of shadows of 2–knots with special shadow-complexity 2 if such a 2–knot exists.

Before stating the theorem on the complexity 1 case, we introduce a series of 2–knots. For $n \in \mathbb{Z}$, let K_n be a 2–knot presented by a banded unlink diagram shown in Figure 1. See [Hughes et al. 2020] and Section 4 for the definition and the details of banded unlink diagrams. Note that K_0 is trivial. The knot K_{-1} is the spun trefoil, and K_1 is the knot 9_1 in Yoshikawa's table [1994]. For any $n \in \mathbb{Z}$, the knot K_n is a ribbon 2–knots. As stated in Proposition 8.9, two 2–knots K_n and $K_{n'}$ are not equivalent unless $n = n'$, which can be distinguished by their Alexander polynomials.

The following is the main theorem for 2–knots with shadow-complexity 1.

Theorem 8.10 *A 2–knot K whose knot group is not infinite cyclic has shadow-complexity* 1 *if and only if K is diffeomorphic to K_n for some nonzero integer n.*

The unknotting conjecture says that a 2–knot is unknotted if its knot group is infinite cyclic, which is known to be true in the topological category in [Freedman and Quinn 1990] and is also studied in the
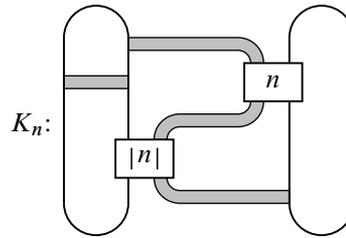
Figure 1: Banded unlink diagram of K_n for $n \in \mathbb{Z}$. Note that the number written in the lower-left box is the absolute value $|n|$ of n.

smooth category in [Kawauchi 2021] (see also [Kawauchi 2023]). Supposing the unknotting conjecture is true in the smooth category, we obtain the complete classification of all 2–knots with shadow-complexity at most 1.

One interpretation of the shadow-complexity for 2–knots is an analogue to the tunnel number for 1–knots. We recall that the tunnel number of a 1–knot is the minimum number of 1–cells such that the complement of the neighborhood of the union of the 1–knot and the 1–cells is a 3–dimensional 1–handlebody. The procedure to make the complement trivial is similar to a construction of a shadow from a 2–knot. By definition, a shadow of a 2–knot K can be obtained from K by attaching 1–cells and 2–cells so that the complement of the neighborhood is diffeomorphic to a 4–dimensional 1–handlebody. In this sense, the shadow-complexity can seem to measure the nontriviality of the complement of a given 2–knot.

Organization

In Section 2, we review the theory of Turaev shadows and encoding graphs. In Section 3, we define a shadow of a 2–knot, and give a presentation of the knot group using a shadow. In Section 4, we introduce a banded unlink diagram, from which we construct a shadow of a 2–knot. In Section 5, we introduce two important modifications: compressing disk addition and connected-sum and reduction. In Section 6, we give the proof for the complexity zero case. A large part of Section 7 is devoted to compute the knot groups of shadows having one true vertex. In Section 8, we give the proofs for the complexity one case by describing the 2–knots in banded unlink diagrams.

Acknowledgements

2 Preliminaries

For polyhedral spaces $A \subset B$, let $\mathrm{Nbd}(A; B)$ denote a regular neighborhood of A in B. If B collapses onto A, we use the notation $B \searrow A$.
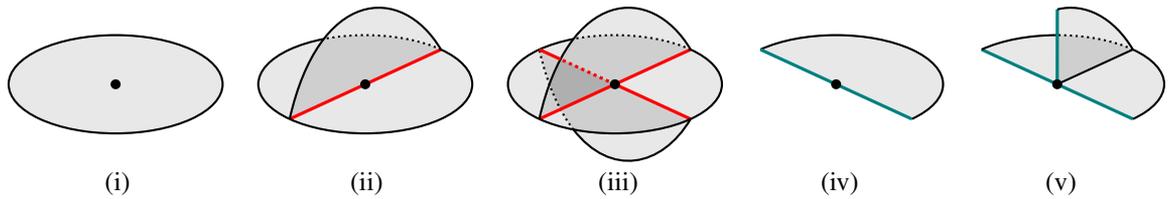
Figure 2: Local models of simple polyhedra.

For integers $0 \le k \le n$, an *n–dimensional k–handlebody* is an *n*–dimensional manifold admitting a handle decomposition consisting of handles whose indices are at most k.

Throughout this paper, we assume that any manifold is compact, connected, oriented and smooth unless otherwise noted.

2.1 Simple polyhedra and shadows of 4–manifolds with boundary

A *simple polyhedron* is a compact connected space whose every point has a regular neighborhood homeomorphic to one of (i)–(v) in Figure 2. Let X be a simple polyhedron. The *singular set* $S(X)$ of X is the set of points of type (ii), (iii) or (v) in Figure 2. The point of type (iii) is called a *true vertex*. Let $c(X)$ denote the number of true vertices contained in X, and this number is called the *complexity* of X. A connected component consisting of points of type (ii) is called a *triple line*. The *boundary* ∂X of X is the set of points of type (iv) or (v). It is clear that the boundary of a simple polyhedron is a (possibly nonconnected) trivalent graph. If $\partial X = \varnothing$, we say that X is *closed*. A *region* of X is a connected component of $X \setminus S(X)$. A region is called a *boundary region* if it contains a point of ∂X (or equivalently, a point of type (iv)), otherwise it is called an *internal region*. If every region is simply connected, then X is called a *special polyhedron*.

We then define a shadow of a 4–manifold with boundary.

Definition 2.1 Let M be a 4–manifold with boundary. A simple polyhedron X properly embedded in M is called a *shadow* of M if X is locally flat in M and $M \searrow X$.

Note that X is *locally flat* in M if for any $x \in X$, there exists a local chart (U_x, ϕ_x) around x in M such that $X \cap U_x$ is contained in a smooth 3–ball in U_x.

The notion of a shadow was introduced by Turaev [1994], and he proved the following.

Theorem 2.2 [Turaev 1994] *Any 4–dimensional 2–handlebody admits a shadow.*

We next define *gleams* of regions of a shadow. Let R be an internal region of X, and set $X_S = \mathrm{Nbd}(S(X); X)$. Then there exists a (possibly nonorientable) 3–dimensional 1–handlebody H_S in M such that X_S is properly embedded in H_S and $H_S \searrow X_S$. Set $\bar{R} = R \setminus \mathrm{Int}\, H_S$. Its boundary $\partial \bar{R} = R \cap \partial H_S$ forms a disjoint union of circles in the surface ∂H_S. Set $B = \mathrm{Nbd}(\bar{R}; \partial H_S)$, which is a disjoint union

of some annuli or Möbius bands. Let $\overline{R}'$ be a small perturbation of $\overline{R}$ in M such that $\partial \overline{R}' \subset B$ and $|\overline{R} \cap \overline{R}'| < \infty$. Then the *gleam* $\mathfrak{gl}(R)$ is defined as

(1)
$$\mathfrak{gl}(R) = \#(\text{Int}\,\overline{R} \cap \text{Int}\,\overline{R}') + \tfrac{1}{2}\#(\partial \overline{R} \cap \partial \overline{R}'),$$

where # is the algebraic intersection number.

The number of the Möbius bands in B is actually determined only by the combinatorial structure of X, that is, it does not depends on the embedding $X \subset M$. If it is even, the region R is said to be *even*, otherwise *odd*. The gleam $\mathfrak{gl}(R)$ is an integer if and only if R is even.

2.2 Shadowed polyhedra and shadows of closed 4–manifolds

For a simple polyhedron X, we assign a half integer to each internal region R of X so that it is an integer if and only if R is even. Such a polyhedron is called a *shadowed polyhedron*.

The following theorem is known as Turaev's reconstruction theorem.

Theorem 2.3 [Turaev 1994] *There is a canonical way to construct a 4–manifold M_X with boundary from a shadowed polyhedron X such that X is a shadow of M_X. Moreover, the gleam of an internal region of X coincides with that coming from* (1).

The gleam is a kind of "local self-intersection number" as one can see in (1). Indeed, the intersection form for the 4–manifold M_X, reconstructed from a shadowed polyhedron X, can be calculated with the gleam (see [Turaev 1994] and the next subsection). Especially, if a closed surface F is embedded in a shadowed polyhedron X, the sum of all the gleams of regions contained in F coincides with the self-intersection number of F in M_X.

We then define a shadow for a closed 4–manifold.

Definition 2.4 Let W be a closed 4–manifold. A simple polyhedron X embedded in W is called a *shadow* of W if it is locally flat in W and $W \setminus \text{Int}\,\text{Nbd}(X; W)$ is diffeomorphic to a 4–dimensional 1–handlebody.

By definition, $\partial\,\text{Nbd}(X; W)$ must be diffeomorphic to the connected-sum of some copies of $S^1 \times S^2$. Thus, if X is a shadowed polyhedron and ∂M_X is diffeomorphic to the connected-sum of some copies of $S^1 \times S^2$, then X is a shadow of some closed 4–manifold by [Laudenbach and Poénaru 1972].

We then define the complexities of 4–manifold that was introduced by Costantino [2006a].

Definition 2.5 For a 4–manifold W, the *shadow-complexity* $\text{sc}(W)$ and the *special shadow-complexity* $\text{sc}^{\text{sp}}(W)$ of W are the minimum number of true vertices of all shadows of W and that of all special shadows of W, respectively.
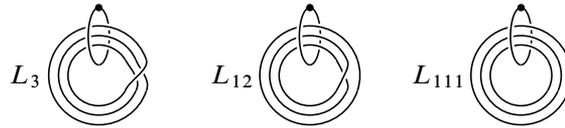
Figure 3: Links in $S^1 \times B^3$. The polyhedra Y_3, Y_{12} and Y_{111} can be constructed from these links.

2.3 Intersection forms

Let X be a shadow of a closed 4–manifold W. Since X can be considered as a 2–skeleton of W, the inclusion $\iota: X \hookrightarrow W$ induces an epimorphism $\iota_*: H_2(X) \to H_2(W)$.

We equip each orientable region R_i of X with an orientation arbitrarily. Then any element in $H_2(X)$ is represented by a sum $\sum a_i R_i$ of oriented regions $R_1, \ldots, R_n$ with integer coefficients $a_1 \ldots, a_n \in \mathbb{Z}$. Defining the intersection form Q_X on $H_2(X)$ as

$$Q_X \left(\sum a_i R_i, \sum b_i R_i \right) = \sum a_i b_i \, \mathfrak{gl}(R_i),$$

we can calculate the intersection form Q_W on $H_2(W)$ as

$$Q_W(\alpha, \beta) = Q_X \left(\sum a_i R_i, \sum b_i R_i \right),$$

where α and β are the images of $\sum a_i R_i$ and $\sum b_i R_i$ by ι_*, respectively. See [Turaev 1994] for the details.

2.4 Topological types of neighborhoods of singular sets with $c \leq 1$

Let X be a simple polyhedron and S be a connected component of $S(X)$. Here we review the topological types of $\mathrm{Nbd}(S; X)$ in the cases $c(\mathrm{Nbd}(S; X)) = 0$ and 1. Note that $\mathrm{Nbd}(S; X)$ itself is a simple polyhedron.

First suppose $c(\mathrm{Nbd}(S; X)) = 0$, that is, S is a circle. There are three possibilities, Y_3, Y_{12} and Y_{111}, for topological types of $\mathrm{Nbd}(S; X)$. These simple polyhedra are interpreted as follows. Let $\pi: S^1 \times B^3 \to S^1$ be the canonical projection, and let L_3, L_{12} and L_{111} be the links in $S^1 \times B^3$ given in Figure 3. Then Y_3, Y_{12} and Y_{111} are the mapping cylinders of π restricted to L_3, L_{12} and L_{111}, respectively.

Next we suppose $c(\mathrm{Nbd}(S; X)) = 1$. Then S is an 8–shaped graph, that is, the wedge sum $S^1 \vee S^1$ of two circles. In this case, there are 11 possible topological types $X_1, \ldots, X_{11}$ of $\mathrm{Nbd}(S_1; X)$, which are explained as follows. Let π be a natural projection from $(S^1 \times B^3) \natural (S^1 \times B^3)$ to $S^1 \vee S^1$, and let L_i be the link in $(S^1 \times B^3) \natural (S^1 \times B^3)$ given in Figure 4 for $i \in \{1, \ldots, 11\}$. Then X_i is the mapping cylinder of π restricted to L_i.

Note that the over/under information of the links in Figures 3 and 4 does not matter for defining the polyhedra $Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$ since they are links in $S^1 \times B^3$ or $(S^1 \times B^3) \natural (S^1 \times B^3)$.
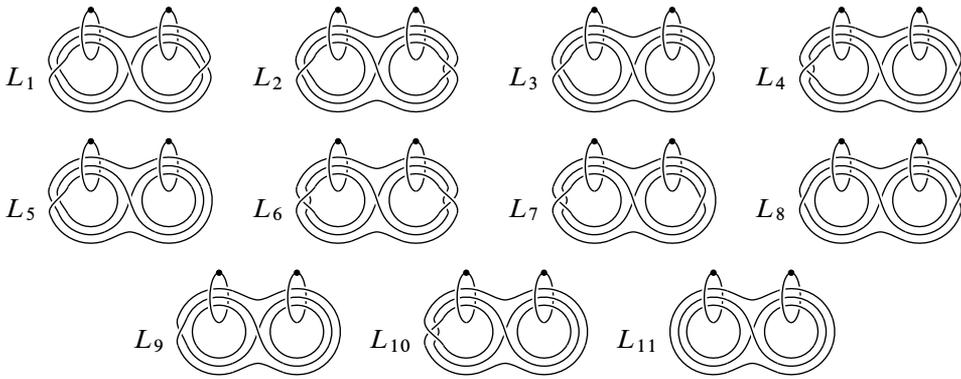
Figure 4: Links in $(S^1 \times B^3) \natural (S^1 \times B^3)$. The polyhedra $X_1, \ldots, X_{11}$ can be constructed from these links.

2.5 Encoding graphs

Here we explain a presentation of a simple polyhedron X using a graph consisting of some special vertices. This notion was introduced by Martelli [2011] for the case $c(X) = 0$ and generalized in [Koda et al. 2022] to the case where each connected component of $S(X)$ contains at most one true vertex.

Let X be a simple polyhedron whose boundary consists of circles and suppose that $c(X) \leq 1$. We first give a decomposition of X into some fundamental portions. Each connected component of $S(X)$ is homeomorphic to S^1 or $S^1 \vee S^1$. As reviewed in the previous subsection, a connected component of $\mathrm{Nbd}(S(X); X)$ is homeomorphic to one of $Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$. Then each component of $X \setminus \mathrm{Int}\,\mathrm{Nbd}(S(X); X)$ is a compact surface corresponding to a region of X. Note that such a surface is possibly nonorientable, and hence it is decomposed into some disks, pair of pants and Möbius bands. Thus, we conclude that X is decomposed (along circles contained in regions) into some copies of $D, P, Y_2, Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$, where D is a 2–disk, P is a pair of pants, and Y_2 is a Möbius band.

The above decomposition induces an *encoding graph G of X* that has one vertex for each portion $D, P, Y_2, Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$ or boundary component of X. Two vertices are connected by an edge if the corresponding portions in X are adjacent. Hence each edge e of G corresponds to a simple closed curve contained in a region of X, which is determined up to isotopy. This simple closed curve
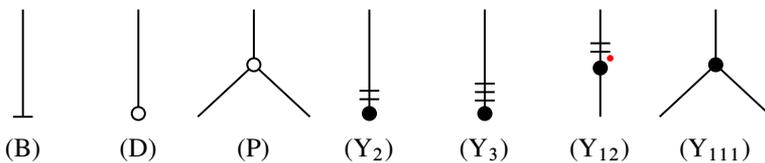

Figure 5: The vertices of types (B), (D), (P), (Y_2), (Y_3), (Y_{12}), and (Y_{111}), which correspond to boundary components of X and portions D, P, Y_2, Y_3, Y_{12} and Y_{111}, respectively.
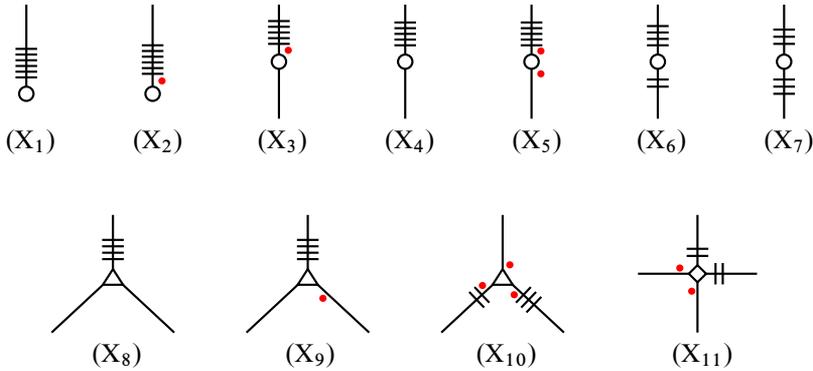
Figure 6: The vertices of types $(X_1), \ldots, (X_{11})$, which correspond to portions $X_1, \ldots, X_{11}$, respectively.

is called a *lift* of e. The vertices corresponding to $D, P, Y_2, Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$ are said to be of types (D), (P), $(Y_2), (Y_3), (Y_{12}), (Y_{111}), (X_1), \ldots, (X_{11})$, respectively. These are shown in Figures 5 and 6, where the vertex of type (B) indicates a boundary component of X. See also [Koda et al. 2022; Martelli 2011].

If an encoding graph G is a tree, the polyhedron X is uniquely reconstructed from G up to homeomorphism. In such a case, we say that G *encodes* X. Hence, in this case, the fundamental group of X can be computed from G by using van Kampen's theorem. The necessary information is summarized in Tables 1 and 2, which exhibit encoding graphs of the portions $D, P, Y_2, Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$, presentations of their fundamental groups and the homotopy classes of their boundaries. Here each vertex of type (B) is denoted by v_i for some $i \in \{1, \ldots, 4\}$, and γ_i is the corresponding component of the boundary of a portion.

portion	graph	π_1	boundary classes in π_1
D		$\{1\}$	$\gamma_1 = 1$
P		$\langle x, y, z \mid xyz \rangle$	$\gamma_1 = x, \gamma_2 = y, \gamma_3 = z$
Y_2		$\langle x \rangle$	$\gamma_1 = x^2$
Y_3		$\langle x \rangle$	$\gamma_1 = x^3$
Y_{12}		$\langle x \rangle$	$\gamma_1 = x, \gamma_2 = x^2$
Y_{111}		$\langle x \rangle$	$\gamma_1 = \gamma_2 = \gamma_3 = x$

Table 1: Encoding graphs of D, P, Y_2, Y_3, Y_{12} and Y_{111}, their fundamental groups, and the homotopy classes of the boundary components.

portion	graph	π_1	boundary classes in π_1
X_1		$\langle x, y \rangle$	$\gamma_1 = xyx^{-2}y^{-2}$
X_2		$\langle x, y \rangle$	$\gamma_1 = xyx^2y^{-2}$
X_3		$\langle x, y \rangle$	$\gamma_1 = y,\ \gamma_2 = xyx^{-2}y^{-1}$
X_4		$\langle x, y \rangle$	$\gamma_1 = y,\ \gamma_2 = xyx^{-2}y$
X_5		$\langle x, y \rangle$	$\gamma_1 = y,\ \gamma_2 = xyx^2y^{-1}$
X_6		$\langle x, y \rangle$	$\gamma_1 = xy,\ \gamma_2 = x^2y^{-2}$
X_7		$\langle x, y \rangle$	$\gamma_1 = xy^2,\ \gamma_2 = x^2y^{-1}$
X_8		$\langle x, y \rangle$	$\gamma_1 = xyx^{-1}y^{-1},\ \gamma_2 = x,\ \gamma_3 = y$
X_9		$\langle x, y \rangle$	$\gamma_1 = xyxy^{-1},\ \gamma_2 = x,\ \gamma_3 = y$
X_{10}		$\langle x, y \rangle$	$\gamma_1 = y,\ \gamma_2 = xy,\ \gamma_3 = x^2y^{-1}$
X_{11}		$\langle x, y \rangle$	$\gamma_1 = x,\ \gamma_2 = y,\ \gamma_3 = xy,\ \gamma_4 = xy^{-1}$

Table 2: Encoding graphs of $X_1, \ldots, X_{11}$, their fundamental groups, and the homotopy classes of the boundary components.

Each boundary component γ of $Y_3, Y_{12}, Y_{111}, X_1, \ldots, X_{11}$ is represented by a word in $\langle x \rangle$ or $\langle x, y \rangle$ as in Tables 1 and 2, and its length coincides with the number of triple lines along which γ goes, counted with multiplicity. This number is called the *length* of γ as well.

Several edges of G are decorated with some dashes and red dots near the vertices; see Figure 5 and 6. The number of dashes indicates the length of the corresponding boundary, and a red dot indicates the parity of the corresponding region of X. Note that the length for a Möbius band Y_2 has not defined, but the incident edge of a vertex of type (Y_2) is also decorated with two dashes for consistency with the other notations. We also note that a red dot of a vertex of type (Y_{12}) is sometimes omitted if no confusion arises.

Notice that an encoding graph is not uniquely determined from X. Two moves on encoding graphs are shown in Figure 7: the left is a *YV-move* and the right is an *IH-move*. These moves correspond to giving another decomposition of a region, so they do not change the homeomorphism types of the corresponding polyhedra.

Figure 7: YV-move and IH-move.

Suppose that G is a tree and let $G' \subsetneq G$ be a subgraph. Let $N(G')$ denote the neighborhood of G', that is, $N(G') \subset G$ is obtained from G' by adding all the vertices adjacent to vertices of G' and all the edges between them. Then we replace each vertex in $N(G') \setminus G'$ with a vertex of type (D), and the obtained graph is called the (D)–*closure* of G', denoted by $\hat{G}'$. See Figure 8 for an example. The left of the figure shows an encoding graph G and subgraphs $G_0, \dots, G_4$ of G, and the right one shows the (D)–closure $\hat{G}_0$ of G_0.

3 Shadows of 2–knots and knot groups

3.1 Shadows of 2–knots

A smoothly embedded surface K in a 4–manifold W is called a *knotted surface*. If K and W are diffeomorphic to S^2 and S^4, respectively, then K is called a 2–*knot*. A 2–knot is said to be *unknotted* (or *trivial*) if it bounds a smooth 3–ball in S^4.

We now define a shadow of a 2–knot as follows.

Definition 3.1 Let K be a 2–knot. A shadow X of S^4 is called a *shadow* of K if K is embedded in X.

We can define a shadow also for a knotted surface in a similar manner, but it is not our focus in this paper.

Note that an unknotted 2–knot is a shadow of itself with gleam 0. In general, a 2–knot is unknotted if and only if it admits a shadow without true vertices, which will be shown in Theorem 6.4.

Theorem 3.2 *Every* 2–*knot admits a shadow.*

By considering the handle decomposition relative to $\mathrm{Nbd}(K; S^4)$, we can prove the above as an application of Theorem 2.2. In Section 4, we will give a recipe for making a shadow of a 2–knot from a *banded unlink diagram*, which gives an alternative proof of Theorem 3.2.

Then we define complexities for 2–knots as well as for 4–manifolds.
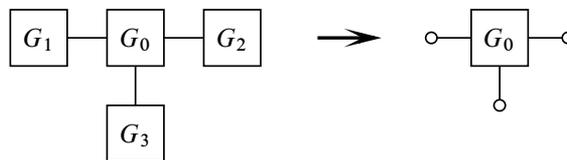
Figure 8: The (D)–closure of a subgraph G_0.

Definition 3.3 For a 2–knot K, the *shadow-complexity* $\mathrm{sc}(K)$ and the *special shadow-complexity* $\mathrm{sc}^{\mathrm{sp}}(K)$ of K are defined as the minimum number of true vertices of all shadows of K and that of all special shadows of K, respectively.

3.2 Knot groups

Let K be a 2–knot. The *knot group $G(K)$* of K is the fundamental group of the complement of K. We will give a presentation of $G(K)$ in Proposition 3.4. To state this proposition, we first give some notation.

Let X be a shadow of K and X_K be a regular neighborhood $\mathrm{Nbd}(K; X)$ of K in X. Choose a regular neighborhood $\mathrm{Nbd}(K; S^4)$ such that X_K is proper in $\mathrm{Nbd}(K; S^4)$, and let it be denoted by M_K. Set

$$X' = X \setminus \mathrm{Int}\, X_K,$$

$$T = \partial X' \cap \partial X_K \subset \partial X',$$

$$M_{X'} = \mathrm{Nbd}(X'; S^4 \setminus \mathrm{Int}\, M_K).$$

We assume that X' and T are connected for simplicity. Note that this can always be assumed by applying some $(0 \rightarrow 2)$–moves (see [Costantino 2004; Turaev 1994]) in advance; see also Remark 3.5. The gluing map $\partial M_{X'} \cap M_K \rightarrow \partial M_K$ will be written as f. Note that T is a graph and the valency of each vertex of T is 3. By the definition of shadows of 2–knots, the knot complement $S^4 \setminus K$ admits a decomposition

$$(M_K \setminus K) \cup_f M_{X'} \cup (3\text{– and } 4\text{–handles}).$$

We can easily see that $M_K \setminus K \cong (S^2 \times D^2) \setminus (S^2 \times \{0\})$ and it retracts onto ∂M_K ($\cong S^2 \times S^1$). We also see that $M_{X'}$ retracts onto X' with keeping T. Thus, the knot group $G(K)$ can be computed from $\partial M_K \cup_{f|_T} X'$.

Choose a basepoint $t \in T$ and a presentation

$$\pi_1(X', t) = \langle S \mid R \rangle.$$

Since T is a graph, its fundamental group is freely generated by some loops

$$w_1, \ldots, w_m \in \pi_1(X', t).$$

We assign an orientation to K arbitrarily. Then the fundamental group of ∂M_K has a presentation

$$\pi_1(\partial M_K, t) = \langle \mu \rangle,$$

where μ is the meridian of K whose orientation agrees with those of K and S^4.

For each $i \in \{1, \ldots, m\}$, there is a 2–chain $D_i = \sum a_{i,j} R_j$ in K with

$$\partial D_i = [w_i] \in H_1(\mathrm{Cl}(X_K \setminus K)) \cong H_1(S(X) \cap K),$$

where R_j is a region contained in K with an orientation induced from that of K. Set

(2) $$\mathrm{gl}(w_i) = \sum a_{i,j}\, \mathfrak{gl}(R_j).$$

This number is equal to the algebraic intersection number of $\tilde{D}_i$ and K in M_K, where $\tilde{D}_i$ is a smooth oriented surface bounded by $f(w_i)$ in M_K. Hence we have $f(w_i) = \mu^{\mathrm{gl}(w_i)}$ in $\pi_1(\partial M_K, *)$. Therefore, by van Kampen's theorem, we obtain the following.

Proposition 3.4 *Under the above settings,*

$$G(K) \cong \langle S, \mu \mid R, w_1 \mu^{-\mathrm{gl}(w_1)}, \ldots, w_m \mu^{-\mathrm{gl}(w_m)} \rangle.$$

Remark 3.5 The assumption that X' and T are connected is not essential. The case where X' and T are not connected is as follows. Let $X'_1, \ldots, X'_k$ and $T_1, \ldots, T_k$ denote the connected components of X' and T, respectively. Then the knot group $G(K)$ is presented as

$$G(K) \cong \langle S_1, \ldots, S_k, \mu \mid R_1, \ldots, R_k, w_1 \mu^{-\mathrm{gl}(w_1)}, \ldots, w_{m'} \mu^{-\mathrm{gl}(w_{m'})} \rangle,$$

where $\pi_1(X'_j) \cong \langle S_j \mid R_j \rangle$ for $j \in \{1, \ldots, k\}$ and $w_1, \ldots, w_{m'}$ are loops in $\pi_1(X'_1) * \cdots * \pi_1(X'_k)$ such that they generate $\pi_1(T_1) * \cdots * \pi_1(T_k)$.

4 Banded unlink diagrams

In this section, we give a review of a description called a *banded unlink diagram* for a knotted surface. See [Hughes et al. 2020] for the details. We start with the definition of banded links in a 3–manifold.

4.1 Banded links

For a link L in a 3–manifold N, the image $b = \mathrm{Im}\beta$ of an embedding $\beta \colon [0,1] \times [0,1] \to N$ with $L \cap b = \beta(\{0,1\} \times [0,1])$ is called a *band*. The *core* of b is defined as $\beta([0,1] \times \{1/2\})$. The pair $(L, \boldsymbol{b})$ of L and mutually disjoint bands $\boldsymbol{b} = b_1 \cup \cdots \cup b_n$ is called a *banded link*. The *negative resolution* and the *positive resolution* of $(L, \boldsymbol{b})$, respectively, are defined as the links L and $L_{\boldsymbol{b}}$, where

$$L_{\boldsymbol{b}} = \left(L \setminus \left(\bigcup_{i=1}^n \beta_i(\{0,1\} \times [0,1]) \right) \right) \cup \left(\bigcup_{i=1}^n \beta_i([0,1] \times \{0,1\}) \right).$$

4.2 Banded unlink diagrams

Let W be a closed 4–manifold, and fix a handle decomposition of W having a single 0–handle. For $i \in \{0, \ldots, 4\}$, let W_i denote the handlebody consisting of all the handles with indices at most i. Clearly, $W_0 \cong B^4$ and $W_4 = W$. Suppose that this handle decomposition is described by a Kirby diagram $\mathcal{K} = \mathcal{L}_1 \sqcup \mathcal{L}_2 \subset S^3$, where $\mathcal{L}_1$ is a dotted unlink indicating the 1–handles and $\mathcal{L}_2$ is a framed link indicating the 2–handles. Note that S^3 in which $\mathcal{K}$ is drawn is considered as the boundary ∂W_0 of the 0–handle, and we can identify the complement $S^3 \setminus \nu\mathcal{K}$ of a tubular neighborhood $\nu\mathcal{K}$ of $\mathcal{K}$ with a subset in ∂W_2.
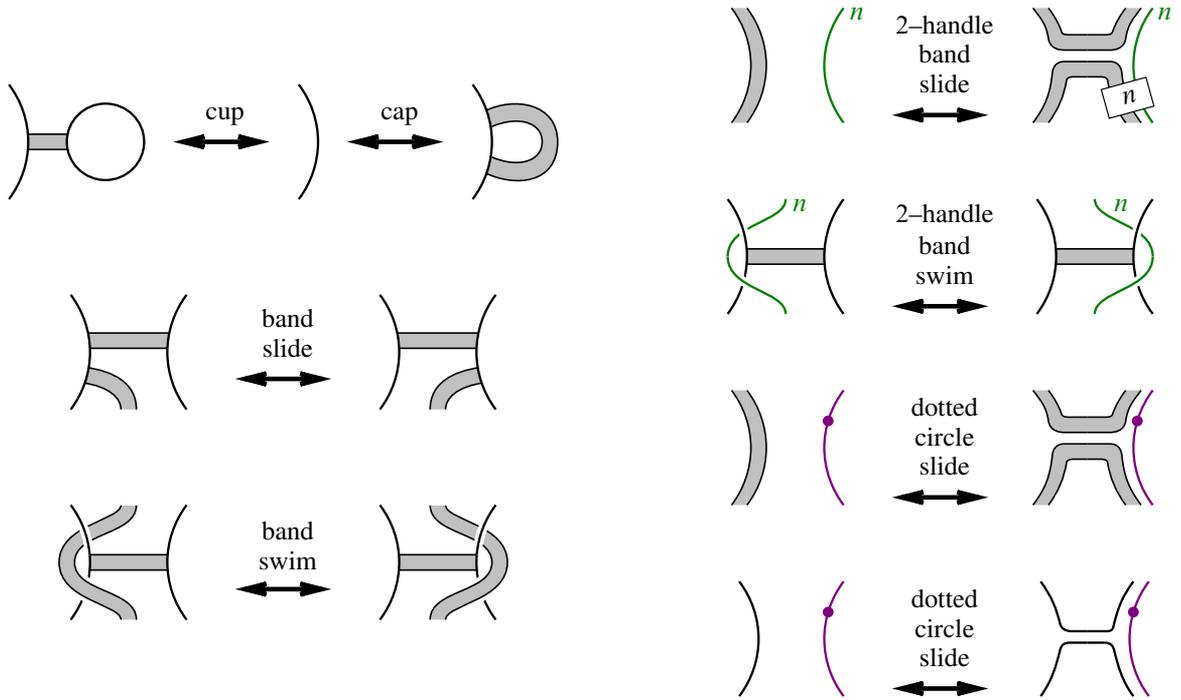
Figure 9: Band moves of banded unlink diagrams.

Let $(L, \boldsymbol{b})$ be a banded link in $S^3 \setminus \nu\mathcal{K}$. Note that $(L, \boldsymbol{b})$ is considered as a banded link in ∂W_2 and also in ∂W_1. Suppose that the negative resolution L and the positive resolution $L_{\boldsymbol{b}}$ are unlinks in ∂W_1 and ∂W_2, respectively. Then we call the triple $(\mathcal{K}, L, \boldsymbol{b})$ a *banded unlink diagram* in $\mathcal{K}$.

A banded unlink diagram $(\mathcal{K}, L, \boldsymbol{b})$ can be interpreted as a presentation of a knotted surface in the 4–manifold W in the following way. By the definition of a banded unlink diagram, there exist collections $\mathcal{D}_1 \subset \partial W_1$ and $\mathcal{D}_2 \subset \partial W_2$ of 2–disks bounded by L and $L_{\boldsymbol{b}}$, respectively. We push the interiors of $\mathcal{D}_1$ and $\mathcal{D}_2$ into Int W_1 and $W \setminus W_2$, respectively, with keeping the boundaries. Then set $K = \mathcal{D}_1 \cup \boldsymbol{b} \cup \mathcal{D}_2$. It forms a knotted surface in W, and $(\mathcal{K}, L, \boldsymbol{b})$ is also said to be a *banded unlink diagram* for K.

If a Kirby diagram for S^4 consists of no dotted circles and no framed knots, we will say that such a diagram is the *trivial Kirby diagram*. By Kawauchi, Shibuya and Suzuki [Kawauchi et al. 1982] and also by Lomanaco [1981], it was shown that any 2–knot admits a banded unlink diagram in the trivial Kirby diagram. In the general case, Hughes, Kim and Miller [Hughes et al. 2020] showed that any knotted surface in any closed 4–manifold admits a banded unlink diagram.

See Figure 9. The three moves shown in the left of the figure (in the trivial Kirby diagram) were introduced by Yoshikawa [1994], and it was shown that these moves are sufficient to relate any two banded unlink diagrams describing the same knotted surface by Swenton [2001] and also by Kearton and Kurlin [2008]. The other moves in Figure 9 were introduced by Hughes, Kim and Miller [Hughes et al. 2020] for the general case. The seven kinds of moves exhibited in the figure are called *band moves*.
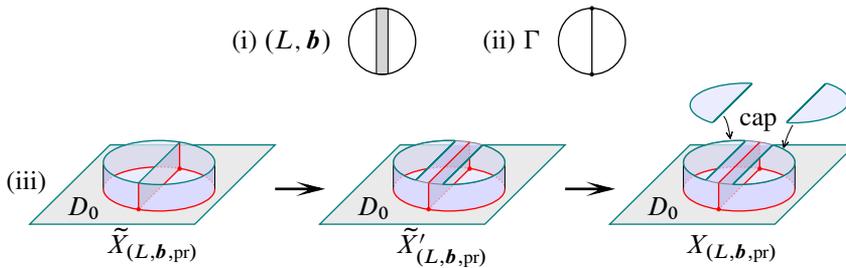
Figure 10: An example of a banded unlink diagram $(L, \boldsymbol{b})$, Γ, and polyhedra $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$, $\widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})}$ and $X_{(L,\boldsymbol{b},\mathrm{pr})}$.

Theorem 4.1 (Hughes, Kim and Miller [Hughes et al. 2020]) *Two banded unlink diagrams* $(\mathcal{K}, L, \boldsymbol{b})$ *and* $(\mathcal{K}, L', \boldsymbol{b}')$ *representing the same knotted surface are related by a finite sequence of band moves.*

4.3 Shadows from banded unlink diagrams

We again focus on the case of 2–knots. In this subsection, we give a construction of a shadow of a 2–knot from a banded unlink diagram.

Let K be a 2–knot and $(\mathcal{K}, L, \boldsymbol{b})$ a banded unlink diagram for K, where $L = L_1 \sqcup \cdots \sqcup L_m$ and $\boldsymbol{b} = b_1 \sqcup \cdots \sqcup b_n$. For simplicity, we suppose $\mathcal{K}$ is the trivial Kirby diagram, and we will write $(L, \boldsymbol{b})$ instead of $(\mathcal{K}, L, \boldsymbol{b})$. In this case, the ambient 4–manifold S^4 is decomposed into one 0–handle W_0 and one 4–handle, and $(L, \boldsymbol{b})$ is assumed to be in ∂W_0. As explained in Section 4.2, the 2–knot K lies in S^4 so that $K \cap \partial W_0 = L \cup \boldsymbol{b}$, and we recall the notations $\mathcal{D}_1 = (K \cap \operatorname{Int} W_0) \cup L$ and $\mathcal{D}_2 = (K \cap S^4 \setminus W_0) \cup L_{\boldsymbol{b}}$.

Step 1 Let Γ be the union of L and the cores of $b_1, \ldots, b_n$, which is a trivalent graph in $S^3 = \partial B^4$. Let π be a regular projection from Γ to a 2–disk D_0 such that the images of $L_1, \ldots, L_m$ bound mutually disjoint 2–disks $D_1, \ldots, D_m \subset D_0$, respectively. Then consider (abstractly) the mapping cylinder

$$\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})} = \big(D_0 \cup (\Gamma \times [0, 1])\big)/\sim$$

of π, where $\sim$ is defined by $\pi(x) \sim (x, 0)$ for $x \in \Gamma$. Since π is chosen so that the diagram is regular, $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ is a simple polyhedron. This polyhedron can be embedded in the 4–ball W_0 as a shadow since $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ can collapses onto the disk D_0. Actually, there is a natural gleam on $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ determined from the diagram of Γ such that it corresponds to a 4–ball in which $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ is embedded as a shadow. See Remark 4.3. Then $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ will be considered as a shadow of W_0, and we can naturally identify

- $\partial \widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})} \setminus \partial D_0$ with $\Gamma \subset \partial W_0$, and
- $\bigcup_{i=1}^m \big(D_i \cup (L_i \times [0, 1])\big)$ with $\mathcal{D}_1 \subset W_0$.

See Figure 10 for an example. The banded link $(L, \boldsymbol{b})$ shown in Figure 10(i) consists of one unknot L and one band $\boldsymbol{b} = b$, and it is a presentation of the trivial 2–knot. Figure 10(ii) shows a diagram of the graph $\Gamma = L \cup (\text{core of } b)$. Then $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$ is a polyhedron as shown in the leftmost of Figure 10(iii).
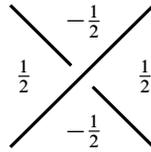
Figure 11: The local contribution to the gleams.

Step 2 The graph Γ lies in ∂W_0 as the boundary of $\widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})}$, and the whole banded link $(L,\boldsymbol{b})$ is also embedded in ∂W_0. Set

$$\widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})} = \widetilde{X}_{(L,\boldsymbol{b},\mathrm{pr})} \cup \boldsymbol{b}$$

and then push $\mathrm{Nbd}(\boldsymbol{b}; \widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})})$ into $\mathrm{Int}\, W_0$ so that $\widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})}$ is properly embedded in W_0. See the center of Figure 10(iii). Note that $\widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})}$ is also a shadow of W_0.

Step 3 The boundary $\partial \widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})} \subset \partial W_0$ is the positive resolution $L_{\boldsymbol{b}}$ of $(L,\boldsymbol{b})$, which is the m'–component unlink, where $m' = 2 + n - m$. We attach m' 2–handles to W_0 along $L_{\boldsymbol{b}}$ with 0–framing, and we set

$$X_{(L,\boldsymbol{b},\mathrm{pr})} = \widetilde{X}'_{(L,\boldsymbol{b},\mathrm{pr})} \cup D'_1 \cup \cdots \cup D'_{m'},$$

where $D'_1, \ldots, D'_{m'}$ are the core disks of the 2–handles. See the right of Figure 10(iii). These disks $D'_1, \ldots, D'_{m'}$ correspond to $\mathscr{D}_2$. The 4–manifold W_0 with the 2–handles attached is diffeomorphic to $\natural_{m'}(S^2 \times D^2)$, and $X_{(L,\boldsymbol{b},\mathrm{pr})}$ is its shadow. We can obtain S^4 from this 4–manifold by attaching m' 3–handles and a 4–handle in a canonical way [Laudenbach and Poénaru 1972]. Hence $X_{(L,\boldsymbol{b},\mathrm{pr})}$ is a shadow of S^4, and the 2–knot K is realized in S^4 as

$$\left(\bigcup_{i=1}^{m} \left(D_i \cup (L_i \times [0,1]) \right) \right) \cup \boldsymbol{b} \cup \left(\bigcup_{j=1}^{m'} D'_j \right)$$

in $X_{(L,\boldsymbol{b},\mathrm{pr})}$. Thus we have the following.

Proposition 4.2 *The polyhedron $X_{(L,\boldsymbol{b},\mathrm{pr})}$ is a shadow of K.*

Remark 4.3 The gleams of regions of $X_{(L,\boldsymbol{b},\mathrm{pr})}$ can be easily calculated. For regions on the disk D_0, we can use the rule shown in Figure 11: the gleam of an internal region contained in D_0 is given as the sum of the local contribution shown in the figure at each crossing of the diagram of Γ adjacent to the region [Costantino and Thurston 2008; Turaev 1994]. The gleam of the region forming (core of b_i) $\times [0,1]$ is given as the number of times b_i twists with respect to D_0 on the diagram. Each of the remaining regions is a part of K and contains a core disk D'_j of a 2–handle. The gleams of them are the minus of the writhe number of L'_j on D_0, where L'_j is the component of $L_{\boldsymbol{b}}$ to which D'_j is attached.

Remark 4.4 All the true vertices of $X_{(L,\boldsymbol{b},\mathrm{pr})}$ lie on D_0. Each of them derives from a crossing of the diagram of Γ or a trivalent vertex of Γ. Thus, we can estimate the shadow-complexity of K from the diagram of Γ. Examples will be studied in the next subsection.
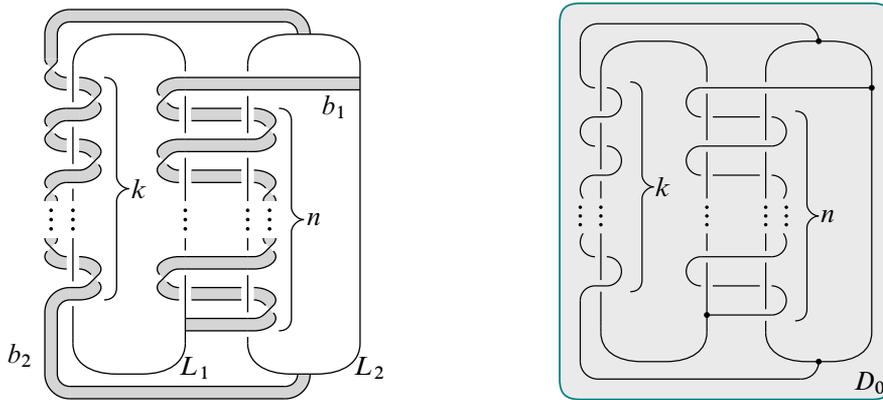
Figure 12: Left: a banded unlink diagram of the k–twist spun knot of $T(2, 2n + 1)$. Right: a diagram of the graph Γ that is the union of L_1, L_2 and the cores of b_1, b_2.

Remark 4.5 Even if $\mathcal{K}$ is not trivial, we also can construct a shadow of K by considering a shadow of W_2 instead of the disk D_0.

4.4 Examples

Let $T(2, 2n + 1)^k$ denote the k–twist spun of the classical torus knot $T(2, 2n + 1)$. Figure 12, left, shows a banded unlink diagram $(L, \boldsymbol{b}) = (L_1 \sqcup L_2, b_1 \cup b_2)$ for $T(2, 2n + 1)^k$ that was given in [Jabłonowski 2016]. Considering a natural projection $\mathrm{pr} \colon \Gamma \to D_0$ from the graph $\Gamma = L_1 \cup L_2 \cup$ (cores of b_1, b_2) to a 2–disk D_0, we draw a regular diagram of Γ as in Figure 12, right. This diagram has $4n + 2k$ crossings, and Γ has 4 trivalent vertices. Hence $X_{(L, \boldsymbol{b}, \mathrm{pr})}$, a shadow of $T(2, 2n + 1)^k$, has $4n + 2k + 4$ true vertices. The polyhedron $X_{(L, \boldsymbol{b}, \mathrm{pr})}$ has a single boundary region, which is adjacent to $2k + 3$ true vertices. These true vertices are eliminated by collapsing, and the resulting polyhedron is still a shadow of $T(2, 2n + 1)^k$. Therefore, we obtain an upper bound of the shadow-complexity of the twist spun knot $T(2, 2n + 1)^k$.

Proposition 4.6 $$\mathrm{sc}(T(2, 2n + 1)^k) \leq 4n + 1.$$

Remark 4.7 The 1–twist spun of any 1–knot is trivial [Zeeman 1965]. As we will see in Theorem 6.4, the shadow-complexity of an unknotted 2–knot is 0. The 2–knot $T(2, 2n + 1)^k$ with $n = 1$ and $k = 0$ is the spun trefoil, which is K_{-1} in our notation; see Figure 1. We will show that its shadow-complexity is 1 in Theorem 8.10.

5 Modifications and fundamental groups

A subspace Y of a simple polyhedron X is called a *subpolyhedron* if there exist simple closed curves $\gamma_1 \sqcup \cdots \sqcup \gamma_n$ in $X \setminus (S(X) \cup \partial X)$ such that Y is the closure of a connected component of $X \setminus (\gamma_1 \sqcup \cdots \sqcup \gamma_n)$. It is obvious that Y itself is a simple polyhedron. Each simple closed curve γ_i is a boundary component

of Y, and it is called a *cut end* of Y in X. In other words, a boundary component of Y but not of X is called a cut end. If X is a shadowed polyhedron, Y can also be assigned with gleams canonically [Naoe 2017].

Henceforth, we fix the following notation:

- K is a 2–knot, and

- X is a shadow of K.

Note that X is simply connected since it is a shadow of S^4.

5.1 Compressing disk addition

Let γ be a simple closed curve contained in a region of X. Let $\pi : M_X \to X$ be the projection, where M_X is the 4–manifold obtained from X by Turaev's reconstruction. Since X is also a shadow of S^4, we have $\partial M_X \cong \#_h(S^1 \times S^2)$ for some $h \in \mathbb{Z}_{\geq 0}$. Hence $(\pi|_{\partial M_X})^{-1}(\gamma)$ is an embedded torus in $\#_h(S^1 \times S^2)$. Every embedded torus in $\#_h(S^1 \times S^2)$ has a compressing disk by Dehn's lemma, so let D_γ be such a disk for the torus $(\pi|_{\partial M_X})^{-1}(\gamma)$. We consider a 2–disk D'_γ enlarged from D_γ such that $D'_\gamma \setminus D_\gamma \subset \pi^{-1}(\gamma)$ and $\partial D'_\gamma \subset X$, and then modify the disk D'_γ near its boundary so that $\partial D'_\gamma$ is a generically immersed curve in $\mathrm{Nbd}(\gamma; X)$ by a small perturbation. This can be done without creating self-intersections of $\mathrm{Int}\, D'_\gamma$. We thus obtain a new simple polyhedron $X \cup D'_\gamma$ embedded in S^4.

Proposition 5.1 [Koda et al. 2022] *Under the above settings, $X \cup D'_\gamma$ is a shadow of S^4 and also of K.*

The disk D'_γ is called a *compressing disk* of γ. The addition of D'_γ corresponds to attaching a 2–handle that is canceled with a 3–handle.

Note that the image of ∂D_γ by π is contained in γ. Then we can define a map $\rho : \partial D_\gamma \to \gamma$ by $\rho(x) = \pi(x)$. There are two important cases:

(i) $\deg(\rho) = 0$, and

(ii) $|\deg(\rho)| = 1$.

In other words,

(i) $\partial D'_\gamma$ is null-homotopic in $\mathrm{Nbd}(\gamma; X)$, and

(ii) $\partial D'_\gamma$ is homotopic to γ in $\mathrm{Nbd}(\gamma; X)$.

The disk D_γ is said to be *vertical* if (i), and *horizontal* if (ii). Figure 13 shows the modification of X to $X \cup D'_\gamma$ in the cases (i) and (ii).

Remark 5.2 If γ is a small circle bounding a disk in a region and has a vertical compressing disk, then we often say that the region has a vertical compressing disk. Clearly, any such γ has a horizontal compressing disk.
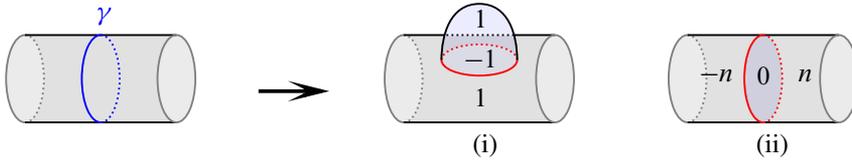
Figure 13: A vertical compressing disk (i) and a horizontal compressing disk (ii).

5.2 Connected-sum and reduction

Suppose that X has a disk region D such that $\mathrm{Nbd}(\partial D; X)$ is homeomorphic to Y_{111} and $\mathfrak{gl}(D) = 0$. The neighborhood $\mathrm{Nbd}(D; X)$ is shown in the left of Figure 14. Note that there exists a smooth 3–ball B_D in S^4 such that $\mathrm{Nbd}(D; X) \subset B_D$ since $\mathfrak{gl}(D) = 0$. We consider a modification as shown in Figure 14. Precisely, we first remove $\mathrm{Int}\,\mathrm{Nbd}(D; X)$ from X and then cap each of the resulting boundary circles off with a 2–disk. This modification can be performed locally in B_D as in the figure. Since X is simply connected, this modification produces two new simple polyhedra X_1' and X_2'. We suppose that X_1' contains the whole K, and we say that X_1' is obtained from X by the *connected-sum reduction* along D.

Proposition 5.3 *Suppose $c(X) \le 1$ and that X_1' is obtained from X by the connected-sum reduction along a disk region D. Then X_1' is a shadow of K.*

Proof Let W denote the 4–sphere in which K and X are embedded. By [Martelli 2011, Proposition 4.1], the 4–sphere W can be decomposed as $W_1 \# W_2$, where W_1 and W_2 are closed 4–manifolds admitting shadows X_1' and X_2', respectively. Since either X_1' or X_2' has no true vertices by $c(X) \le 1$, the corresponding 4–manifold, namely W_1 or W_2, is diffeomorphic to S^4 by [Martelli 2011, Corollary 1.8]. Thus, both W_1 and W_2 are diffeomorphic to S^4. Then $W \setminus \mathrm{Int}\,\mathrm{Nbd}(X_1'; W)$ is diffeomorphic to a 4–dimensional 1–handlebody, and hence X_1' is a shadow of K. □

5.3 Lemmas on encoding graphs

Here we prepare some lemmas about the shape and the types of the vertices of an encoding graph of X.

Lemma 5.4 [Martelli 2011, Claim 1] *Suppose that a loop $\gamma \subset X$ separates X into two connected components V_1 and V_2. Then both V_1' and V_2' are simply connected, where $V_1' = V_1 \cup_\gamma D^2$ and $V_2' = V_2 \cup_\gamma D^2$.*

Proof The quotient space X/γ is homeomorphic to the wedge sum $V_1' \vee V_2'$. Then we have a surjection from $\pi_1(X) = \{1\}$ onto $\pi_1(V_1') * \pi_1(V_2')$. □

Figure 14: Connected sum.

Figure 15: Graphs that are not contained in a graph encoding a simply connected polyhedron.

We rephrase the above lemma in terms of encoding graphs as below.

Lemma 5.5 *Let G be a tree encoding X and $G' \subsetneq G$ be a subgraph. Then a simple polyhedron encoded by the* (D)*–closure $\hat{G}'$ is also simply connected.*

Lemma 5.6 *Suppose $c(X) \leq 1$. Let G be a graph encoding X.*

(1) *G is a tree.*

(2) *G does not have a vertex of type (Y_2), (Y_3), (X_1), (X_2), (X_5), (X_6) or (X_7).*

(3) *G does not contain either subgraph shown in Figure 15.*

Proof   (1) We can embed G in X as a retract, and hence there is a surjection $\pi_1(X) \to \pi_1(G)$.

(2) Assume that G has a vertex v of type (Y_2), (Y_3), (X_1), (X_2), (X_5), (X_6) or (X_7). Then the polyhedron encoded by the (D)–closure $\hat{v}$ is not simply connected, which is a contradiction to Lemma 5.5.

(3) The proof is similar to that of (2).   □

Remark 5.7   (1) The fundamental groups of the special polyhedra encoded by the (D)–closures of vertices of types (Y_2), (Y_3), (X_1), (X_2), (X_5), (X_6) and (X_7) are isomorphic to $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z}$, $\langle x, y \mid xyx^{-2}y^{-2} \rangle$, $\langle x, y \mid xyx^2y^{-2} \rangle$, $\mathbb{Z}/3\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$, respectively.

(2) The two polyhedra encoded by the (D)–closures of the subgraphs shown in Figure 15 are homeomorphic, and their fundamental groups are isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

5.4 Lemmas on the fundamental groups of subpolyhedra

In this subsection, we present four lemmas on the fundamental group of subpolyhedron in X. Lemmas 5.9 and 5.10 treat a subpolyhedron with one cut end, and Lemmas 5.11 and 5.12 treat a subpolyhedron with two cut ends. Note that we will assume that X is closed in these lemmas, which actually does not matter for our main theorems due to Lemma 6.1.

Definition 5.8 Let G be a tree graph encoding a simple polyhedron. Let v and v' be vertices of G and v' is of type (Y_{12}). If the edge incident to v' with no dashes is contained in the shortest path from v to v', then v' is said to be *one-sided to v*. Otherwise we say that v' is *two-sided to v*.

Lemma 5.9 *Suppose X is closed, and let V be a subpolyhedron of X with a single cut end γ and $c(V) = 0$. Then there exists a nonnegative integer k such that $\pi_1(V) \cong \langle \gamma \mid \gamma^{2^k} \rangle$.*
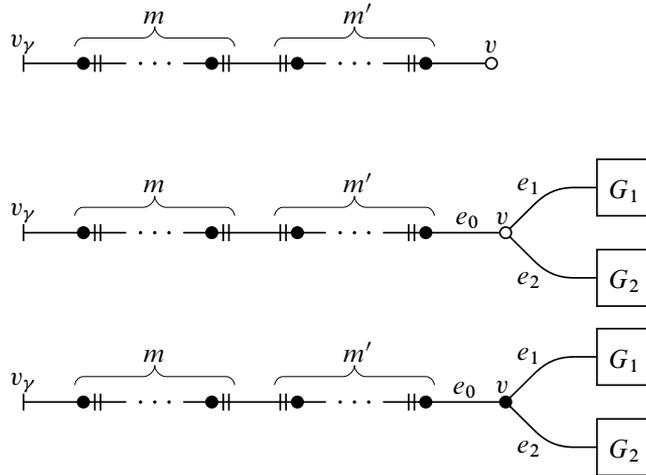
Figure 16: The possible cases of a graph G encoding V.

Proof Let G be an encoding graph of V. From Lemma 5.4, the polyhedron $V \cup_\gamma D^2$ is simply connected, and G is a tree from Lemma 5.6(1). This graph G has exactly one vertex of type (B) corresponding to γ; let v_γ denote it. Other vertices in G are of types (D), (Y_{12}), (P) or (Y_{111}) by Lemma 5.6(2). Note that the valencies of types (D), (Y_{12}), (P) and (Y_{111}) are 1, 2, 3 and 3, respectively.

We assign an orientation to γ arbitrarily. Let v be the nearest vertex from v_γ among those of types (D), (P) and (Y_{111}). By Lemma 5.6(3), the possible cases are shown in Figure 16, where $m, m' \in \mathbb{Z}_{\geq 0}$. We prove the argument by induction on the number of vertices of types (D), (P) and (Y_{111}) in G, so it is enough to consider the following (i)–(iii).

(i) Suppose that G is as shown in Figure 16, top. Then we easily obtain $\pi_1(V) \cong \langle \gamma \mid \gamma^{2^m} \rangle$.

(ii) Suppose that G is as shown in Figure 16, middle, and that $\pi_1(V_i) \cong \langle \gamma_i \mid \gamma_i^{2^{k_i}} \rangle$ for $i \in \{1, 2\}$, where V_i is the polyhedron encoded by the subgraph G_i and $\gamma_i = \partial V_i$. Let γ_0 be a lift of e_0. Note that γ_i is a lift of e_i for $i \in \{1, 2\}$. By Lemma 5.5 applied to the subgraph $v \cup G_1 \cup G_2$, the group $\langle \gamma_1, \gamma_2 \mid \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_1 \gamma_2 \rangle$ must be trivial, and hence we have $k_1 = 0$ or $k_2 = 0$. Suppose $k_2 = 0$. Then

$$
\begin{aligned}
\pi_1(V) &\cong \langle \gamma, \gamma_0, \gamma_1, \gamma_2 \mid \gamma^{2^m} \gamma_0^{2^{m'}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 \gamma_1 \gamma_2 \rangle \\
&\cong \langle \gamma, \gamma_0, \gamma_1 \mid \gamma^{2^m} \gamma_0^{2^{m'}}, \gamma_1^{2^{k_1}}, \gamma_0 \gamma_1 \rangle \\
&\cong \langle \gamma, \gamma_0 \mid \gamma^{2^m} \gamma_0^{2^{m'}}, \gamma_0^{2^{k_1}} \rangle.
\end{aligned}
$$

It becomes trivial by adding a relation $\gamma = 1$ since $\pi_1(V \cup_\gamma D^2) = \{1\}$. Hence $m' = 0$ or $k_1 = 0$, and the lemma follows in either case.

(iii) Suppose that G is as shown in Figure 16, bottom, and that $\pi_1(V_i) \cong \langle \gamma_i \mid \gamma_i^{2^{k_i}} \rangle$ for $i \in \{1, 2\}$, where V_i is the polyhedron encoded by the subgraph G_i and $\gamma_i = \partial V_i$. Let γ_0 be a lift of e_0. We have

$$
\pi_1(V) \cong \langle \gamma, \gamma_0, \gamma_1, \gamma_2 \mid \gamma^{2^m} \gamma_0^{2^{m'}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = \gamma_1 = \gamma_2 \rangle \cong \langle \gamma, \gamma_0 \mid \gamma^{2^m} \gamma_0^{2^{m'}}, \gamma_0^{2^{k_3}} \rangle,
$$

where $k_3 = \min\{k_1, k_2\}$. It becomes trivial by adding a relation $\gamma = 1$ since $\pi_1(V \cup_\gamma D^2) = \{1\}$. Hence $m' = 0$ or $k_3 = 0$, and the lemma follows in either case. $\qquad\square$

For each vertex v of type (D), let $k(v)$ denote the number of vertices of type (Y$_{12}$) one-sided to v_γ contained in the geodesic from v_γ to v. By the proof of Lemma 5.9, the integer k stated in the lemma is given as the minimum of $k(v)$ for any vertex v of type (D). Thus we also have proved the following.

Lemma 5.10 *Under the same notation as in Lemma 5.9 and its proof, if $\pi_1(V) = \{1\}$, then there exists a leaf such that all the vertices of type (Y$_{12}$) contained in the geodesic (in G) from v_γ to the leaf are two-sided to v_γ.*

Lemma 5.11 *Suppose X is closed, and let U be a subpolyhedron of X with exactly two cut ends $\gamma_1 \sqcup \gamma_2$ and $c(U) = 0$. Suppose that $[\gamma_1]$ is not a torsion element in $H_1(U)$.*

(1) *$[\gamma_2]$ is also not a torsion element in $H_1(U)$.*

(2) *$\pi_1(U) \cong \langle \gamma_1, \gamma_2 \mid (\gamma_1^{2^m} \gamma_2^{2^l})^{2^k} \rangle$ for some $k, l, m \in \mathbb{Z}_{\geq 0}$.*

Proof (1) We have $c(U \cup_{\gamma_2} D^2) = 0$. Hence $H_1(U \cup_{\gamma_2} D^2)$ is a finite cyclic group generated by $[\gamma_1]$ by Lemma 5.9. Suppose, to derive a contradiction, that $[\gamma_2]$ is a torsion element in $H_1(U)$. Then the normal subgroup $\langle [\gamma_2] \rangle$ generated by $[\gamma_2]$ is contained in the torsion subgroup $t(H_1(U))$ of $H_1(U)$. Since the free part $H_1(U)/t(H_1(U))$ has the nontrivial element $[\gamma_1]$ and $H_1(U \cup_{\gamma_2} D^2) \cong H_1(U)/\langle [\gamma_2] \rangle$, the free part of $H_1(U \cup_{\gamma_2} D^2)$ is also nontrivial, which is a contradiction.

(2) Let G be a tree encoding U, and let v_1 and v_2 be the vertices of type (B) corresponding to γ_1 and γ_2, respectively. Let ℓ be the geodesic from v_1 to v_2. The vertices in G other than v_1 or v_2 are of types (D), (Y$_{12}$), (P) or (Y$_{111}$)

We assume that ℓ contains a vertex v of type (Y$_{111}$) and lead to a contradiction. Recall that there are three edges incident to v. See Figure 17(i-1). Let e be the edge such that it is incident to v and between v and v_2, and let e' be the edge incident to v and not on ℓ. Let G_1, G_2 and G_3 be subgraphs of G as indicated in Figure 17(i-1). Now let U' be one of the components containing γ_1 obtained by cutting U along a lift of e, which is encoded in Figure 17(i-2). This subpolyhedron U' has two boundary components: one is γ_1 and the other, namely a lift of e, will be denoted by γ_2'. Since U' itself satisfies the assumption of the lemma, the cycle $[\gamma_2']$ is not a torsion element in $H_1(U')$ by (1). It is homologous to a cycle represented by a lift of e', which is a torsion element in the subpolyhedron encoded by G_3 by Lemma 5.9. It is a contradiction. Therefore, the vertices between v_1 and v_2 are of types (Y$_{12}$) or (P).

If ℓ does not contain vertices of type (P), the graph G is as shown in Figure 17(ii) by Lemma 5.6(3). We then have $\pi_1(U) \cong \langle \gamma_1, \gamma_2 \mid \gamma_1^{2^m} \gamma_2^{2^l} \rangle$.

We then suppose that G has a vertex v of type (P), and G is as shown in Figure 17(iii-1). We can assume that each vertex of type (Y$_{12}$) in ℓ is two-sided to v by Lemma 5.6(3). Suppose that there is a subgraph of
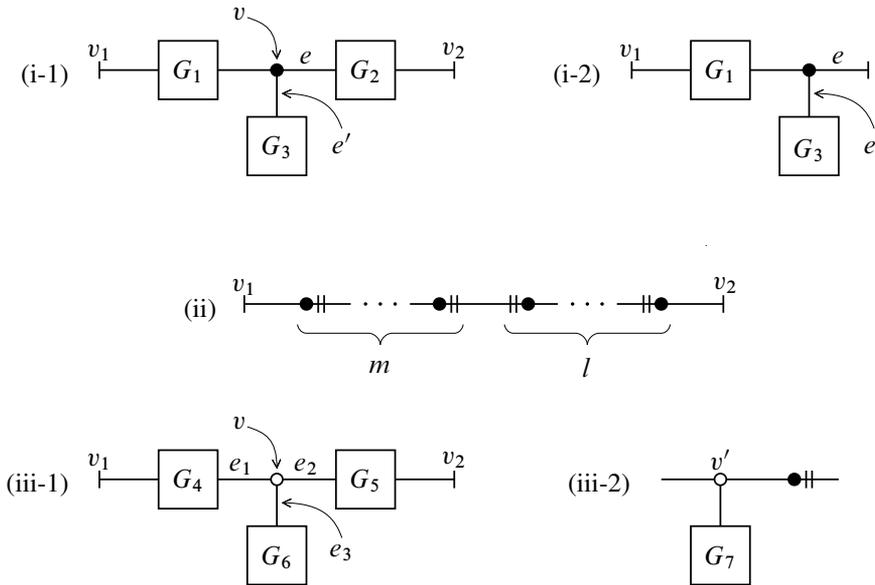
Figure 17: Encoding graphs of U used in the proof of Lemma 5.11.

G_4 as shown in Figure 17(iii-2), where v' is a vertex of type (P) contained in ℓ. The fundamental group of the subpolyhedron encoded by G_7 is isomorphic to $\langle \gamma' \mid \gamma'^{2^{k'}} \rangle$ for some $k' \in \mathbb{Z}_{\geq 0}$ by Lemma 5.9, where γ' is the boundary of the subpolyhedron. The simple polyhedron encoded by the (D)–closure of the graph shown in Figure 17(iii-2) is simply connected by Lemma 5.5. Hence the subpolyhedron encoded by G_7 must be simply connected as well. Therefore, the graph G_4 in Figure 17(iii-1) encodes a polyhedron whose fundamental group is presented by $\langle \gamma_1, u_1 \mid u_1 \gamma_1^{2^m} \rangle$, where u_1 is a lift of e_1 and m is the number of vertices of type (Y$_{12}$) in $G_4 \cap \ell$. Similarly, the graph G_5 in Figure 17(iii-1) encodes a polyhedron whose fundamental group is presented by $\langle \gamma_2, u_2 \mid u_2 \gamma_2^{2^l} \rangle$, where u_2 is a lift of e_2 and l is the number of vertices of type (Y$_{12}$) in $G_5 \cap \ell$. By Lemma 5.9, the graph G_6 in Figure 17(iii-1) encodes a polyhedron whose fundamental group is presented by $\langle u_3 \mid u_3^{2^k} \rangle$ for some $k \in \mathbb{Z}_{\geq 0}$, where u_3 is a lift of e_3. Thus, we obtain a presentation

$$\pi_1(U) \cong \langle \gamma_1, \gamma_2, u_1, u_2, u_3 \mid u_1 \gamma_1^{2^m}, u_2 \gamma_2^{2^l}, u_3^{2^k}, u_1 u_2 u_3 \rangle \cong \langle \gamma_1, \gamma_2 \mid (\gamma_1^{2^m} \gamma_2^{2^l})^{2^k} \rangle. \qquad \square$$

From the above proof, we immediately obtain the following lemma, which will be used in the proof of Theorem 8.6.

Lemma 5.12 *Under the same notation as in Lemma 5.11 and its proof, if $\pi_1(U) \cong \langle \gamma_1, \gamma_2 \mid \gamma_1^{2^m} \gamma_2 \rangle$, then*

- *any vertex lying in ℓ is of type (Y$_{12}$) or (P),*
- *m is the number of the vertices of type (Y$_{12}$) lying in ℓ, all of which are one-sided to v_1, and*
- *the subpolyhedron corresponding to each connected component of $G \setminus \ell$ is simply connected.*

6 2–Knots with complexity zero

From now on, we discuss the classification of 2–knots according to the shadow-complexity, and we give the proof of the theorem for the case $sc(K) = 0$ in this section.

Let us start with the following lemma. It is an analogue of [Koda and Naoe 2020, Lemma 1.3], and the original statement in the paper is for shadows of "4–manifolds". Roughly speaking, the shadow-complexity of any 2–knot is always attained by a closed shadow.

Lemma 6.1 *If $sc(K) = n$, then K admits a closed shadow with complexity exactly n.*

Proof The proof is almost the same as that of [Koda and Naoe 2020, Lemma 1.3], so we only sketch the proof.

Let X be a shadow of K with $c(X) = n$ and $\partial X \neq \varnothing$. Then X collapses onto an almost-simple polyhedron Y (see [Matveev 2003] for the definition and the details) that is minimum with respect to collapsing and has at most n true vertices. Note that the collapsing is done in a regular neighborhood $\mathrm{Nbd}(X; S^4)$, which is also a regular neighborhood of Y in S^4. Since K is a 2–sphere embedded in X, it is kept by collapsing, that is, K is also embedded in Y. The polyhedron Y is either a closed simple polyhedron or the union of a closed simple polyhedron and a graph. If the former, this Y is what we required. If the latter, as in the proof of [Koda and Naoe 2020, Lemma 1.3], Y can be modified to a closed simple polyhedron Y' such that Y and Y' have the same regular neighborhood in S^4 and also that no new true vertices are created by the modification. $\square$

As well as in Lemma 5.9, we consider a subpolyhedron having one cut end in the next lemma. However, unlike in Lemma 5.9, Lemma 6.2 gives a homological condition, and a subpolyhedron can have true vertices and boundary components other than the cut end. Recall the notation $gl(\gamma)$ defined in formula (2).

Lemma 6.2 *Set $X_K = \mathrm{Nbd}(K; X)$, and suppose that ∂X_K has a circle component γ. Let X' be a connected component of $X \setminus \mathrm{Int}\, X_K$ with $X_K \cap \partial X' = \gamma$. Give orientations to K and γ arbitrarily. Then at least one of the following holds:*

- *$H_1(X')$ is an infinite cyclic group generated by $[\gamma]$, or*
- *$gl(\gamma) = 0$.*

Proof We have $H_1(\gamma) \cong \mathbb{Z}\langle\gamma\rangle$ and $H_1(X) \cong 0$. From the Mayer–Vietoris sequence

$$H_1(\gamma) \to H_1(X \setminus \mathrm{Int}\, X') \oplus H_1(X') \to H_1(X),$$

$H_1(X')$ admits a surjection from $\mathbb{Z}$ generated by $[\gamma]$.

If $H_1(X') = \mathbb{Z}/k\mathbb{Z}\langle\gamma\rangle$ for some $k \in \mathbb{Z}_{>0}$, there is a 2–chain C in X' such that $\partial C = k[\gamma]$. Let

$$D = \sum a_j R_j$$

be a 2–chain in X_K with $\partial D = [\gamma]$, where R_j is a region contained in K. Then $C - kD$ is a homology cycle in X, and we have

$$Q([K], C - kD) = -k \sum a_j \mathfrak{gl}(R_j) = -k \mathrm{gl}(\gamma),$$

where Q is the intersection form on $H_2(X)$. Since the intersection form of S^4 is trivial, $\mathrm{gl}(\gamma) = 0$. $\quad\square$

Lemma 6.3 *Suppose X is closed and that $S(X) \cap K$ has a circle component γ bounding a disk region D on K. Let X' be a connected component of $X \setminus K$ with $\partial X' = \gamma$. If X' does not contain true vertices, then $X \setminus X'$ is a shadow of K.*

Proof From Lemma 5.9, $H_1(X') \cong \mathbb{Z}/2^k\mathbb{Z}$ for some $k \in \mathbb{Z}_{\geq 0}$. Then $\mathfrak{gl}(D) = 0$ by Lemma 6.2, and $X \setminus X'$ is a shadow of K by Proposition 5.3. $\quad\square$

Theorem 6.4 *A 2–knot K is unknotted if and only if $\mathrm{sc}(K) = 0$.*

Proof Let X be a shadow of a 2–knot K with $c(X) = 0$. By Lemma 6.1, we can assume $\partial X = \varnothing$. Since $c(X) = 0$, each component of $S(X) \cap K$ is a circle, and each component of $X \setminus K$ does not contain true vertices. By iterating Lemma 6.3, K admits itself as a shadow. Hence K is unknotted. The converse is obvious. $\quad\square$

7 Existence of 2–knots with complexity one

From here, we always assume that

- X is a closed shadow of a 2–knot K, and
- $c(X) = 1$.

Note that $K \cap S(X)$ is not empty. Then there are two cases:

(i) the true vertex is contained in a component of $K \cap S(X)$, which is an 8–shaped graph, and

(ii) the true vertex does not lie on K and every component of $K \cap S(X)$ is a circle.

Therefore, we can also assume the following by Lemma 6.3:

- $K \cap S(X)$ is connected, and it is a circle or an 8–shaped graph.

Moreover, we fix the notation

- $X' = X \setminus \mathrm{Int}\,\mathrm{Nbd}(K; X)$,

which is also connected by the above assumption.

7.1 True vertex lies on K

Suppose that the true vertex of X lies on K. Let us first consider the case of special shadow-complexity 1. We need the following lemma.

Lemma 7.1 *Let K be a 2–knot. Suppose that S^4 admits a decomposition consisting of $\mathrm{Nbd}(K; S^4)$, one 2–handle, two 3–handles and one 4–handle. Then K is unknotted.*

Proof We regard $\mathrm{Nbd}(K; S^4)$ as the union of a 0–handle h_0 and a 2–handle h_2. This 2–handle h_2 is attached along the 0–framed unknot L lying on the boundary of h_0 since $\mathrm{Nbd}(K; S^4) \cong S^2 \times D^2$. Let h_2' be the other 2–handle in the decomposition, and let L' be its attaching circle. This framed knot L' is contained in $\partial h_0 \setminus \nu L$, and hence $L \sqcup L'$ forms a 2–component link in $\partial h_0 \cong S^3$. By [Gompf et al. 2010, Proposition 3.2], we can modify the link $L \sqcup L'$ into the unlink only by using handle-slides of L' over L. Then L' is contained in a small 3–ball in the boundary of $h_0 \cup h_2$ $(= \mathrm{Nbd}(\mathcal{K}; S^4))$, and the 2–sphere K can be pushed to the boundary of $h_0 \cup h_2 \cup h_2'$ by isotopy. The resulting 2–sphere plays a role of the attaching sphere of a 3–handle by [Laudenbach and Poénaru 1972]. Hence K bounds a 3–ball in S^4, which is the definition of K to be unknotted. $\square$

Theorem 7.2 *There are no 2–knots with special shadow-complexity 1.*

Proof Let us suppose that X is a special shadow X of K with $c(X) = 1$. Then $S(X)$ is connected and $S(X) \subset K$. Moreover, it is homeomorphic to an 8–shaped graph, and $\mathrm{Nbd}(S(X); K)$ is homeomorphic to a pair of pants. Hence $\mathrm{Nbd}(S(X); X)$ is homeomorphic to X_{11} (see Section 2.4), and X' is a 2–disk. Thus, S^4 is decomposed as in Lemma 7.1, and K is unknotted and $\mathrm{sc^{sp}}(K) = 0$. $\square$

We next consider the nonspecial case.

Proposition 7.3 *If the true vertex of X lies on K, then $G(K)$ is an infinite cyclic group.*

Proof Let S be the component of $K \cap S(X)$ containing the true vertex. As well as in the proof of Theorem 7.2, $\mathrm{Nbd}(S; X)$ is homeomorphic to X_{11}, and $\partial X' = X' \cap \mathrm{Nbd}(K; X)$ is a circle. By Lemma 5.9, $\pi_1(X') \cong \langle \gamma \mid \gamma^{2^k} \rangle$ for some $k \in \mathbb{Z}_{\geq 0}$, where $\gamma = \partial X'$. Then we have $G(K) \cong \langle \gamma, \mu \mid \gamma^{2^k}, \gamma \mu^{-\mathrm{gl}(\gamma)} \rangle$ by Proposition 3.4. Since $\mathrm{gl}(\gamma) = 0$ by Lemma 6.2, $G(K)$ is an infinite cyclic group. $\square$

7.2 True vertex does not lie on K

Hereafter we suppose that the true vertex of X does not lie on K, that is, it is contained in X'. In this subsection, we investigate the knot group of K.

The part $S(X) \cap K$ of the singular set separates K into two disk regions, and their gleams are g and $-g$ for some $g \in \mathbb{Z}_{\geq 0}$ since the self-intersection number of K is 0.

If $g = 0$, K is a shadow of itself by Proposition 5.3. Then K is unknotted.

We henceforth suppose $g > 0$. Let us orient K arbitrarily, and then an oriented meridian μ of K is defined. Set $\gamma = \partial X'$, and let G be a graph encoding X'. This graph G has exactly one vertex of type (B), which corresponds to γ and will be denoted by v. By Lemma 5.6(2), G have exactly one vertex v_0 of type (X$_3$),
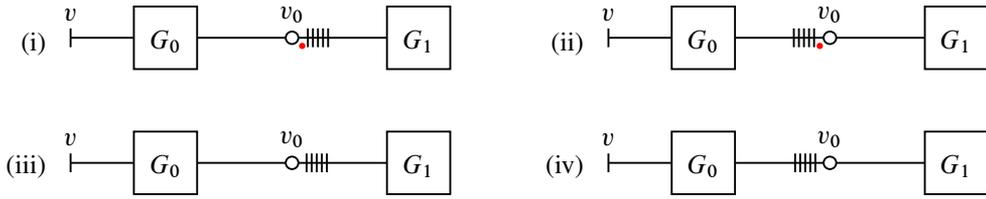
Figure 18: The possible cases of encoding graph G of X such that v_0 is of type (X_3) or (X_4).

(X_4), (X_8), (X_9), (X_{10}) or (X_{11}), and hence G is one of those shown in Figures 18, 19 and 20. Note that X_8 and X_{11} have symmetries that interchange two boundary components with the same length. Let X_{v_0} be the connected component of $\mathrm{Nbd}(S(X); X)$ corresponding to v_0, which is homeomorphic to one of X_3, X_4, X_8, X_9, X_{10} or X_{11}.

Let $\mathfrak{a}$ denote the abelianization map of a group. In the following, we discuss what kind of 2–knot admits a shadow encoded by a graph in Figures 18, 19 and 20.

Case 1 (v_0 is of type (X_3) or (X_4)) The graph G is one of those shown in Figure 18. Let G_0 and G_1 be subgraphs of G as indicated in the figure. The subpolyhedron X' is decomposed into U, X_{v_0} and V, where U and V are the subpolyhedron corresponding to G_0 and G_1, respectively. Note that U has two cut ends $\gamma \sqcup \gamma_0$, and also note that V has one cut end γ_1. Then we have $\pi_1(V) \cong \langle \gamma_1 \mid \gamma_1^{2^{k_1}} \rangle$ by Lemma 5.9 for some $k_1 \in \mathbb{Z}_{\geq 0}$. We also have $H_1(X') \cong \mathbb{Z}\langle \gamma \rangle$ by Lemma 6.2. Then we can apply Lemma 5.11 to U, and we have $\pi_1(U) \cong \langle \gamma, \gamma_0 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}} \rangle$ for some $k_0, l, m \in \mathbb{Z}_{\geq 0}$.

Lemma 7.4 *The following hold.*

(1) *If G is shown in Figure 18(i), then $\pi_1(U) \cong \langle \gamma, \gamma_0 \mid \gamma^{2^m} \gamma_0 \rangle$ and V is simply connected. Moreover, $G(K) \cong \langle x, \mu \mid x^2 \mu^n x^{-1} \mu^{-n} \rangle$, where $n = 2^m g$.*

(2) *If G is shown in Figure 18(ii), then $G(K)$ is an infinite cyclic group.*

(3) *If G is shown in Figure 18(iii), then $\pi_1(U) \cong \langle \gamma, \gamma_0 \mid \gamma^{2^m} \gamma_0 \rangle$ and V is simply connected. Moreover, $G(K) \cong \langle x, \mu \mid x^2 \mu^n x^{-1} \mu^n \rangle$, where $n = 2^m g$.*

(4) *If G is shown in Figure 18(iv), then $G(K)$ is an infinite cyclic group.*

Proof (1) The polyhedron X' is decomposed as $U \cup X_3 \cup V$, and U and V are glued with X_3 along the boundary components of X_3 with length 1 and 5, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_0 = y, \gamma_1 = xyx^{-2}y^{-1} \rangle$$
$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} y^{2^l})^{2^{k_0}}, (xyx^{-2}y^{-1})^{2^{k_1}} \rangle$$
$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}y \rangle \oplus \mathbb{Z}\langle 2^{k_1} x \rangle$$
$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \mu \mid (\gamma^{2^m} y^{2^l})^{2^{k_0}}, (xyx^{-2}y^{-1})^{2^{k_1}}, \gamma\mu^{-g} \rangle$$
$$\cong \langle x, \mu \mid x^2\mu^n x^{-1}\mu^{-n} \rangle,$$

where $n = 2^m g$.

(2) The polyhedron X' is decomposed as $U \cup X_3 \cup V$, and U and V are glued with X_3 along the boundary components of X_3 with length 5 and 1, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1 \mid y = \gamma_1, xyx^{-2}y^{-1} = \gamma_0, (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}} \rangle$$
$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}(xyx^{-2}y^{-1})^{2^l})^{2^{k_0}}, y^{2^{k_1}} \rangle$$
$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}x \rangle \oplus \mathbb{Z}\langle 2^{k_1}y \rangle$$
$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \mu \mid y = \gamma_1, xyx^{-2}y^{-1} = \gamma_0, (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma\mu^{-g} \rangle$$
$$\cong \langle \mu \rangle.$$

(3) The polyhedron X' is decomposed as $U \cup X_4 \cup V$, and U and V are glued with X_4 along the boundary components of X_4 with length 1 and 5, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1 \mid y = \gamma_0, xyx^{-2}y = \gamma_1, (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}} \rangle$$
$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}y^{2^l})^{2^{k_0}}, (xyx^{-2}y)^{2^{k_1}} \rangle$$
$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}y \rangle \oplus \mathbb{Z}\langle 2^{k_1}(2y - x) \rangle$$
$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \mu \mid y = \gamma_0, xyx^{-2}y = \gamma_1, (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma\mu^{-g} \rangle$$
$$\cong \langle x, \mu \mid x^2\mu^n x^{-1}\mu^n \rangle,$$

where $n = 2^m g$.

(4) The polyhedron X' is decomposed as $U \cup X_4 \cup V$, and U and V are glued with X_4 along the boundary components of X_4 with length 5 and 1, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1 \mid y = \gamma_1, xyx^{-2}y = \gamma_0, (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}} \rangle$$
$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}(xyx^{-2}y)^{2^l})^{2^{k_0}}, y^{2^{k_1}} \rangle$$
$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}(2y - x) \rangle \oplus \mathbb{Z}\langle 2^{k_1}y \rangle$$
$$\cong H_1(X').$$
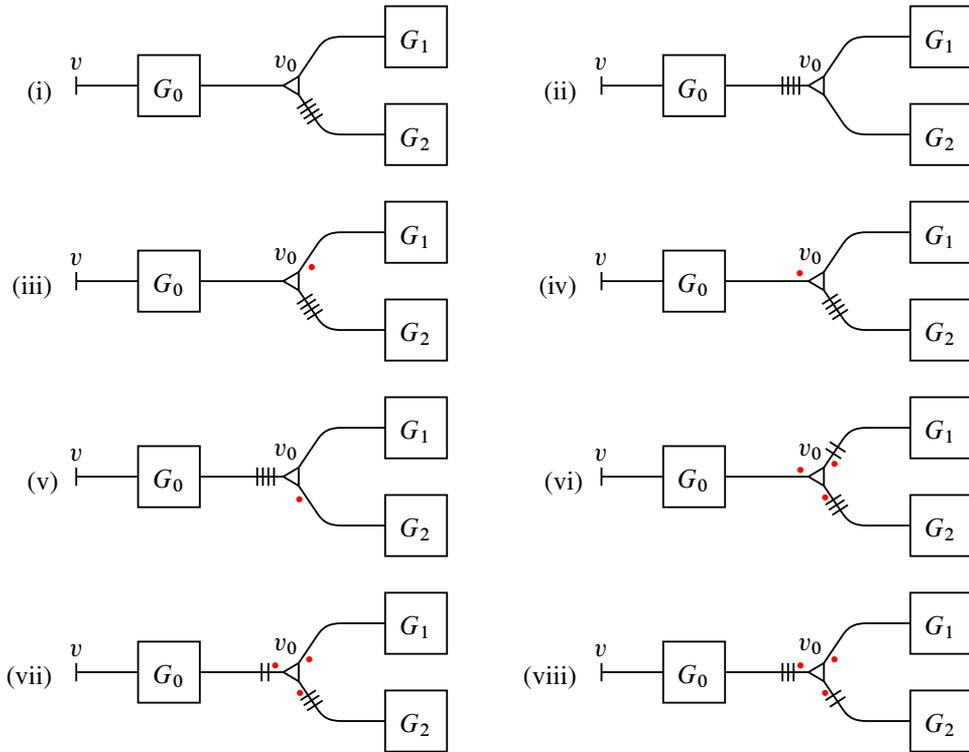
Figure 19: The possible cases of encoding graph G of X such that v_0 is of type (X_8), (X_9) or (X_{10}).

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \mu \mid (\gamma^{2^m}(xyx^{-2}y)^{2^l})^{2^{k_0}}, y^{2^{k_1}}, \gamma\mu^{-g} \rangle \cong \langle \mu \rangle. \qquad \square$$

Case 2 (v_0 is of type (X_8), (X_9) or (X_{10})) The graph G is one of those shown in Figure 19. Let G_0, G_1 and G_2 be subgraphs of G as indicated in the figure. The subpolyhedron X' is decomposed into U, X_{v_0}, V_1 and V_2, where U, V_1 and V_2 are the subpolyhedron corresponding to G_0, G_1 and G_2, respectively. Note that U has two cut ends $\gamma \sqcup \gamma_0$, and also note that V_i has one cut end γ_i for $i \in \{1, 2\}$. Then we have $\pi_1(V_i) \cong \langle \gamma_i \mid \gamma_i^{2^{k_i}} \rangle$ by Lemma 5.9 for some $k_i \in \mathbb{Z}_{\geq 0}$. We also have $H_1(X') \cong \mathbb{Z}\langle \gamma \rangle$ by Lemma 6.2. Then we can apply Lemma 5.11 to U, and we have $\pi_1(U) \cong \langle \gamma, \gamma_0 \mid (\gamma^{2^m}\gamma_0^{2^l})^{2^{k_0}} \rangle$ for some $k_0, l, m \in \mathbb{Z}_{\geq 0}$.

Lemma 7.5 *The following hold.*

(1) *If G is shown in Figure 19(i), then $G(K)$ is an infinite cyclic group.*

(2) *Figure 19(ii) does not encode a shadow of any 2–knot.*

(3) *Figure 19(iii) does not encode a shadow of any 2–knot.*

(4) *If G is shown in Figure 19(iv), then $G(K)$ is an infinite cyclic group.*

(5) *Figure 19(v) does not encode a shadow of any 2–knot.*

(6) *Figure 19*(vi) *does not encode a shadow of any 2–knot.*

(7) *Figure 19*(vii) *does not encode a shadow of any 2–knot.*

(8) *Figure 19*(viii) *does not encode a shadow of any 2–knot.*

Proof (1) Suppose that G is as shown in Figure 19(i). The polyhedron X' is decomposed as

$$U \cup X_8 \cup V_1 \cup V_2,$$

and U, V_1 and V_2 are glued with X_8 along the boundary components of X_8 with length 1, 1 and 4, respectively. The fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = x, \gamma_1 = y, \gamma_2 = xyx^{-1}y^{-1} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} x^{2^l})^{2^{k_0}}, y^{2^{k_1}}, (xyx^{-1}y^{-1})^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}x \rangle \oplus \mathbb{Z}\langle 2^{k_1} y \rangle$$

$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \mu \mid (\gamma^{2^m} x^{2^l})^{2^{k_0}}, y^{2^{k_1}}, (xyx^{-1}y^{-1})^{2^{k_2}}, \gamma\mu^{-g} \rangle \cong \langle \mu \rangle.$$

(2) Suppose G is as shown in Figure 19(ii). The polyhedron X' is decomposed as $U \cup X_8 \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_8 along the boundary components of X_8 with length 4, 1 and 1, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = xyx^{-1}y^{-1}, \gamma_1 = x, \gamma_2 = y \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}(xyx^{-1}y^{-1})^{2^l})^{2^{k_0}}, x^{2^{k_1}}, y^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma \rangle \oplus \mathbb{Z}\langle 2^{k_1} x \rangle \oplus \mathbb{Z}\langle 2^{k_2} y \rangle$$

$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.

(3) Suppose G is as shown in Figure 19(iii). The polyhedron X' is decomposed as $U \cup X_9 \cup V_1 \cup V_2$. One of the boundary components of X_9 has length 4 and the other two have 1. Note that, however, X_9 does not have a symmetry such as X_8. The boundary components of X_9 are represented by words $xyxy^{-1}$, x and y. Here U, V_1 and V_2 are glued with X_9 along the boundary components of X_9 corresponding to x, y and $xyxy^{-1}$, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = x, \gamma_1 = y, \gamma_2 = xyxy^{-1} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} x^{2^l})^{2^{k_0}}, y^{2^{k_1}}, (xyxy^{-1})^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}x \rangle \oplus \mathbb{Z}\langle 2^{k_1} y \rangle \oplus \mathbb{Z}\langle 2^{k_2+1} x \rangle$$

$$\cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.

(4)   Suppose that G is as shown in Figure 19(iv). The polyhedron X' is decomposed as $U \cup X_9 \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_9 along the boundary components of X_9 corresponding to y, x and $xyxy^{-1}$, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = y, \gamma_1 = x, \gamma_2 = xyxy^{-1} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} y^{2^l})^{2^{k_0}}, x^{2^{k_1}}, (xyxy^{-1})^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}y \rangle \oplus \mathbb{Z}\langle 2^{k_1}x \rangle \oplus \mathbb{Z}\langle 2^{k_2+1}x \rangle \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$. Hence $k_0 = k_1 = l = 0$. Thus,

$$G(K) \cong \langle x, y, \gamma, \mu \mid (\gamma^{2^m} y^{2^l})^{2^{k_0}}, x^{2^{k_1}}, (xyxy^{-1})^{2^{k_2}}, \gamma\mu^{-g} \rangle \cong \langle \mu \rangle.$$

(5)   Suppose G is as shown in Figure 19(v). The polyhedron X' is decomposed as $U \cup X_9 \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_9 along the boundary components of X_9 corresponding to $xyxy^{-1}$, x and y, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = xyxy^{-1}, \gamma_1 = x, \gamma_2 = y \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} (xyxy^{-1})^{2^l})^{2^{k_0}}, x^{2^{k_1}}, y^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle / \mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0+1}x \rangle \oplus \mathbb{Z}\langle 2^{k_1}x \rangle \oplus \mathbb{Z}\langle 2^{k_2}y \rangle \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.

(6)   Suppose G is as shown in Figure 19(vi). The polyhedron X' is decomposed as $U \cup X_{10} \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_{10} along the boundary components of X_{10} with length 1, 2 and 3, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = x, \gamma_1 = xy, \gamma_2 = xy^{-2} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} x^{2^l})^{2^{k_0}}, (xy)^{2^{k_1}}, (xy^{-2})^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \frac{\mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle}{\mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}x \rangle \oplus \mathbb{Z}\langle 2^{k_1}(x + y) \rangle \oplus \mathbb{Z}\langle 2^{k_2}(x - 2y) \rangle} \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.

(7)   Suppose G is as shown in Figure 19(vii). The polyhedron X' is decomposed as $U \cup X_{10} \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_{10} along the boundary components of X_{10} with length 2, 1 and 3, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = xy, \gamma_1 = x, \gamma_2 = xy^{-2} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} (xy)^{2^l})^{2^{k_0}}, x^{2^{k_1}}, (xy^{-2})^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \frac{\mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle}{\mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}(x + y) \rangle \oplus \mathbb{Z}\langle 2^{k_1}x \rangle \oplus \mathbb{Z}\langle 2^{k_2}(x - 2y) \rangle} \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.
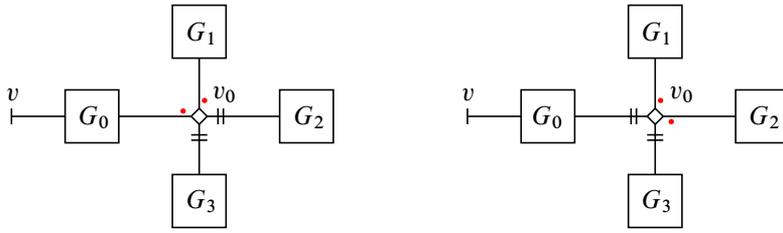
Figure 20: The possible cases of encoding graph G of X such that v_0 is of type (X_{11}).

(8) Suppose G is as shown in Figure 19(viii). The polyhedron X' is decomposed as $U \cup X_{10} \cup V_1 \cup V_2$, and U, V_1 and V_2 are glued with X_{10} along the boundary components of X_{10} with length 3, 1 and 2, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_0 = xy^{-2}, \gamma_1 = x, \gamma_2 = xy \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}(xy^{-2})^{2^l})^{2^{k_0}}, x^{2^{k_1}}, (xy)^{2^{k_2}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \frac{\mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle}{\mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}(x - 2y) \rangle \oplus \mathbb{Z}\langle 2^{k_1} x \rangle \oplus \mathbb{Z}\langle 2^{k_2}(x + y) \rangle} \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible. $\qquad \square$

Case 3 (v_0 is of type (X_{11})) The graph G is one of those shown in Figure 20. Let G_0, G_1, G_2 and G_3 be subgraphs of G as indicated in the figure. The subpolyhedron X' is decomposed into U, X_{v_0}, V_1, V_2 and V_3, where U, V_1, V_2 and V_3 are the subpolyhedron corresponding to G_0, G_1, G_2 and G_3. Note that U has two cut ends $\gamma \sqcup \gamma_0$ as a subpolyhedron of X, and also note that V_i has one cut end γ_i for $i \in \{1, 2, 3\}$. Then we have $\pi_1(V_i) \cong \langle \gamma_1 \mid \gamma_1^{2^{k_i}} \rangle$ by Lemma 5.9 for some $k_i \in \mathbb{Z}_{\geq 0}$. We also have $H_1(X') \cong \mathbb{Z}\langle \gamma \rangle$ by Lemma 6.2. Then we can apply Lemma 5.11 to U, and we have $\pi_1(U) \cong \langle \gamma, \gamma_0 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}} \rangle$ for some $k_0, l, m \in \mathbb{Z}_{\geq 0}$.

Lemma 7.6 *The following hold.*

(1) *Figure 20, left, does not encode a shadow of any 2–knot.*

(2) *Figure 20, right, does not encode a shadow of any 2–knot.*

Proof (1) Suppose G is as shown in Figure 20, left. The polyhedron X' is decomposed as $U \cup X_{11} \cup V_1 \cup V_2 \cup V_3$, and U, V_1, V_2 and V_3 are glued with X_{11} along the boundary components of X_{11} with length 1, 1, 2 and 2, respectively. The fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2, \gamma_3 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_3^{2^{k_3}}, \gamma_0 = x, \gamma_1 = y, \gamma_2 = xy, \gamma_3 = xy^{-1} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m} x^{2^l})^{2^{k_0}}, y^{2^{k_1}}, (xy)^{2^{k_2}}, (xy^{-1})^{2^{k_3}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \frac{\mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle}{\mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0} x \rangle \oplus \mathbb{Z}\langle 2^{k_1} y \rangle \oplus \mathbb{Z}\langle 2^{k_2}(x + y) \rangle \oplus \mathbb{Z}\langle 2^{k_3}(x - y) \rangle} \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible.
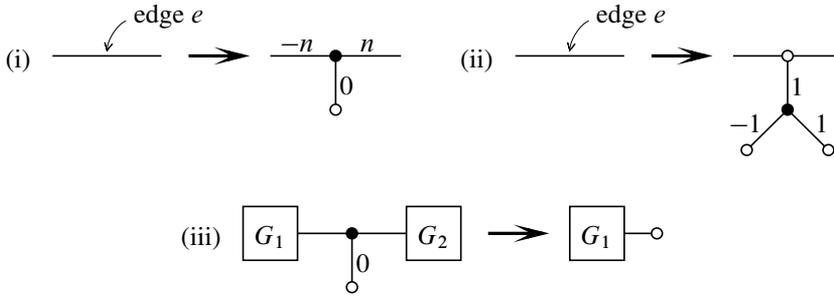
Figure 21: Adding a horizontal compressing disk (i), adding a vertical compressing disk (ii), and connected-sum reduction (iii).

(2) Suppose G is as shown in Figure 20, right. The polyhedron X' is decomposed as

$$U \cup X_{11} \cup V_1 \cup V_2 \cup V_3,$$

and U, V_1, V_2 and V_3 are glued with X_{11} along the boundary components of X_{11} with length 2, 1, 1 and 2, respectively. Then the fundamental group of X' and its abelianization are obtained as follows:

$$\pi_1(X') \cong \langle x, y, \gamma, \gamma_0, \gamma_1, \gamma_2, \gamma_3 \mid (\gamma^{2^m} \gamma_0^{2^l})^{2^{k_0}}, \gamma_1^{2^{k_1}}, \gamma_2^{2^{k_2}}, \gamma_3^{2^{k_3}}, \gamma_0 = xy, \gamma_1 = x, \gamma_2 = y, \gamma_3 = xy^{-1} \rangle$$

$$\cong \langle x, y, \gamma \mid (\gamma^{2^m}(xy)^{2^l})^{2^{k_0}}, x^{2^{k_1}}, y^{2^{k_2}}, (xy^{-1})^{2^{k_3}} \rangle$$

$$\xrightarrow{\mathfrak{a}} \frac{\mathbb{Z}\langle x \rangle \oplus \mathbb{Z}\langle y \rangle \oplus \mathbb{Z}\langle \gamma \rangle}{\mathbb{Z}\langle 2^{m+k_0}\gamma + 2^{l+k_0}(x+y) \rangle \oplus \mathbb{Z}\langle 2^{k_1}x \rangle \oplus \mathbb{Z}\langle 2^{k_2}y \rangle \oplus \mathbb{Z}\langle 2^{k_3}(x-y) \rangle} \cong H_1(X').$$

It must be an infinite cyclic group generated by $[\gamma]$, which is impossible. $\qquad\square$

8 Classification of 2–knots with complexity one

8.1 Lemmas on decorated graphs

We now define a *decoration* of an edge e of an encoding graph G as a half-integer such that it is an integer if and only if the number of red dots appended to e is even (actually, zero or two). If every edge of G is assigned with a decoration, G is called a *decorated graph*. A decoration corresponds to a gleam, and a decorated tree encodes a shadowed polyhedron.

We can easily describe how a decorated graph G changes by adding a compressing disk and a connected-sum reduction. See Figure 21. If a lift of an edge e of G has a horizontal (resp. vertical) compressing disk, we can replace the edge e as shown in Figure 21(i) (resp. (ii)). If a decorated graph is as shown in the left of Figure 21(iii) and if the subpolyhedron corresponding to the subgraph G_1 contains K, we can adopt a decorated graph shown in the right of the figure.

In this subsection, we provide some modifications of shadows and decorated graphs not changing a 2–knot.
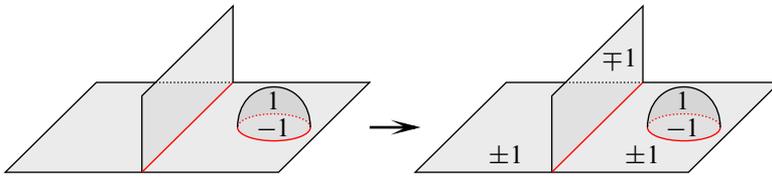
Figure 22: A move on a shadow: if a region R has a vertical compressing disk, then we can modify the gleams of R, R' and R'' as in the figure, where R' and R'' are regions adjacent to a common triple line with R.

Lemma 8.1 *Suppose that X' has a part as shown in the left of Figure 22. Then the move shown in the figure and its inverse modify X to another shadow of K.*

See [Koda et al. 2022] for the proof of the above lemma.

We next introduce eight moves on decorated graphs as shown in Figure 23; moves (a), (b), (c), (d), (e), (f), (g) and (h). Note that the decoration r in a move (g) is not $\pm\frac{1}{2}$.

Lemma 8.2 *Let G be a decorated graph of X and G' be the subgraph corresponding to X'. Then the moves shown in Figure 23 that are performed on G' modify G to another decorated graph encoding a shadow of K.*

Proof Moves (a) and (b) are obtained by a move in Figure 22.

A move (c) is explained in [Martelli 2011, Figure 34(7)].

A move (d) is a obtained by a move (a), a connected-sum reduction, and a YV-move.

A move (e) is a kind of *propagation principle* [Koda et al. 2022]: if two of the three regions adjacent to a triple line have vertical compressing disks, then the other also has.

A move (f) is explained in [Martelli 2011, Figure 34(4)].

A move (g) is explained in Figure 24. Let $\pi\colon M_X \to X$ be a natural projection, where $M_X = \mathrm{Nbd}(X; S^4)$. Then the preimage of the subpolyhedron corresponding to the leftmost graph of Figure 24 by $\pi|_{\partial M_X}$ is homeomorphic to the complement of the $(2, 2r)$–torus knot in S^3; see [Ishikawa and Koda 2017, Figure 11]. Recall that $r \neq \frac{1}{2}$. Let e be an edge as indicated in the figure, then a lift of e has a vertical compressing disk by Property P and Property R. Hence we can add a vertical compressing disk as in Figure 24(i). The move in Figure 24(ii) is obtained by performing a move (b) as many times as necessary. The move in Figure 24(iii) is done by a move (f) and a YV-move.

A move (h) is explained in Figure 25. The move in Figure 25(i) is obtained by performing moves (c) as many times as necessary. The move in Figure 25(ii) can be done by using [Martelli 2011, Figure 34(3)]. Here we need the following claim:

Claim 1 *A lift of e' has a vertical compressing disk, where e' is an edge indicated in the figure.*
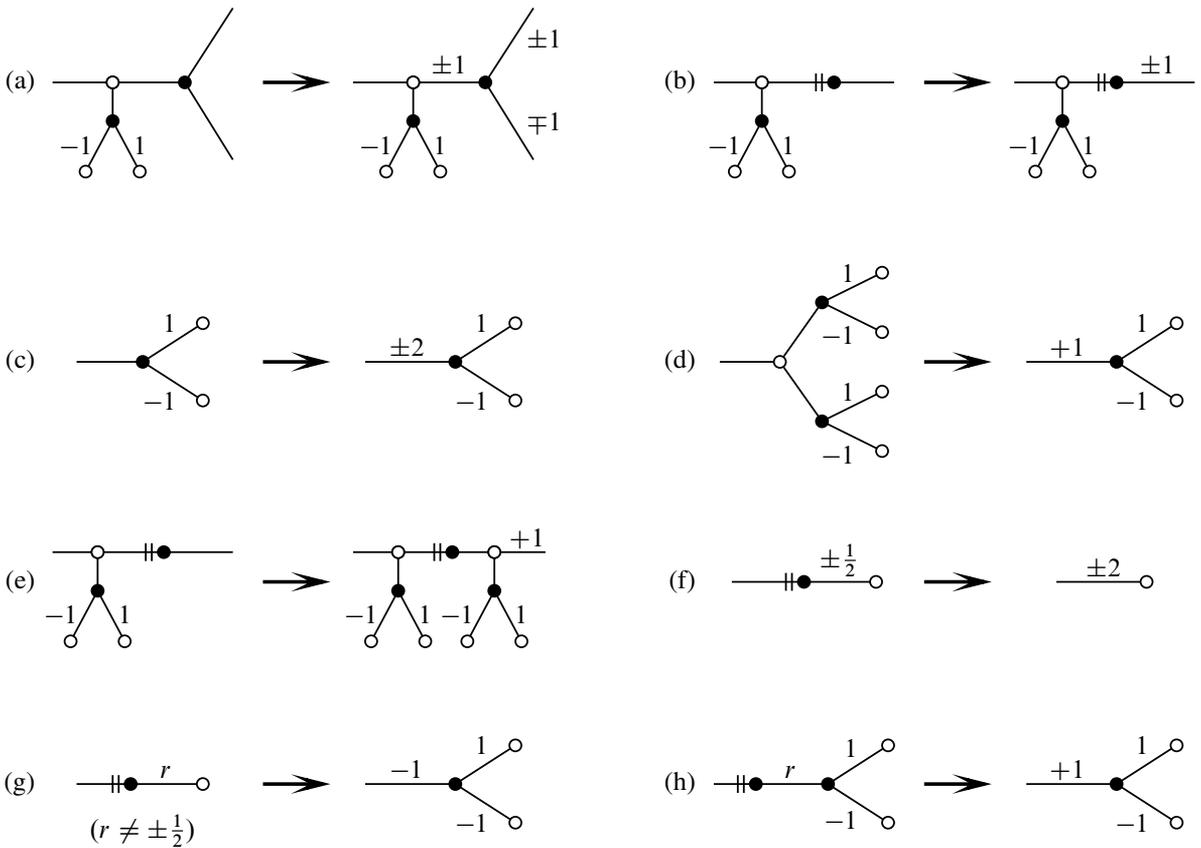
Figure 23: Eight moves on decorated graphs. The decoration r in (g) is not $\pm\frac{1}{2}$.

Proof Let $\pi\colon M_X \to X$ be a natural projection. The preimage of a subpolyhedron homeomorphic to Y_{111} by $\pi|_{\partial M_X}$ is homeomorphic to $P \times S^1$. Hence the subgraph after the move in Figure 24(ii) corresponds to a 3–manifold homeomorphic to the Seifert fibered space $(D^2; (2, 1), (2, -1))$, which has one torus boundary. The Dehn filling of this manifold along the (p, q)–slope is $(S^2; (2, 1), (2, -1), (p, q))$. Note that the slope with $(p, q) = (p, 1)$ is sent to a lift of the edge e' by π injectively, and the slope with $(p, q) = (1, 0)$ is sent to one point by π. The 3–manifold $(S^2; (2, 1), (2, -1), (p, q))$ is not homeomorphic to $\#_h(S^1 \times S^2)$ for any $h \in \mathbb{Z}_{\geq 0}$ unless $q = 0$. If $q = 0$, then we have $p = 1$ and $(S^2; (2, 1), (2, -1), (1, 0)) \cong (S^2; (2, 1), (2, -1))$.
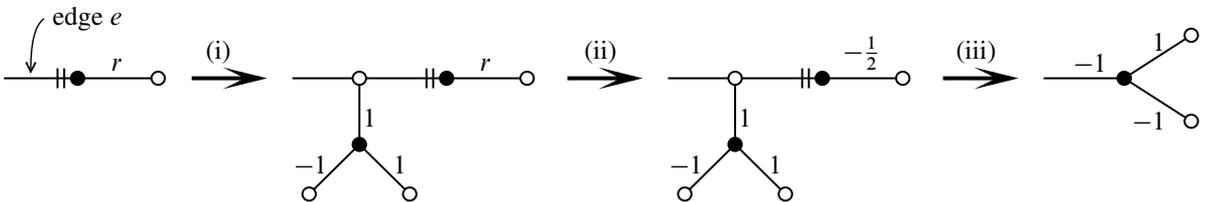


Figure 24: The proof of a move (g).
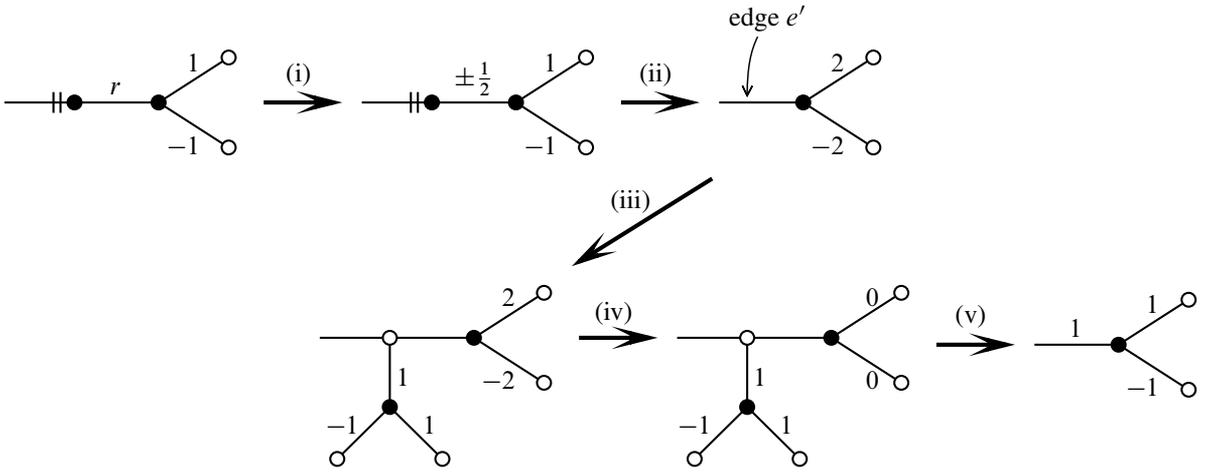
Figure 25: The proof of a move (h).

This space can be understood as a 3–manifold obtained by gluing two solid tori along their boundaries so that the meridians are identified, and hence it is homeomorphic to $S^1 \times S^2$. It follows that a lift of e' has a vertical compressing disk.   □

We then continue the proof for the move (h). The move of Figure 25(iii) is the addition of a compressing disk of a lift of e'. The move in Figure 25(iv) is done by applying move (a) twice. The move in Figure 25(v) is done by a connected-sum reduction and a YV-move.   □

Lemma 8.3 *Let V be a subpolyhedron of X with a single cut end γ and $c(V) = 0$. Let G be a graph encoding V and v_γ be the vertex of type (B) corresponding to γ. Suppose that G has a vertex v of type (D) that is adjacent to a vertex v' of type (Y_{111}). Let R be the disk region of X corresponding to v. If G has another vertex of type (Y_{111}) between v_γ and v', that is, if G is as shown in Figure 26, then $\mathfrak{gl}(R) = 0$.*

Proof We give an orientation to R arbitrarily, and we define edges e_1, e_2, e_2' and e_3 and subgraphs G_1, G_2 and G_3 of G as indicated in Figure 26. Let V_1, U_2 and V_3 be subpolyhedra of V encoded by G_1, G_2 and G_3, respectively. Note that each of V_1 and V_3 has one cut end and U_2 has two. Set $\gamma_i = \partial V_i$ for $i \in \{1, 3\}$, and note that it is a lift of e_i. Let γ_2 and γ_2' denote the cut ends of U_2, and also note that γ_2 and γ_2' are lifts of e_2 and e_2', respectively. By Lemma 5.9, we have $\pi_1(V_i) \cong \langle \gamma_i \mid \gamma_i^{2^{k_i}} \rangle$ for some
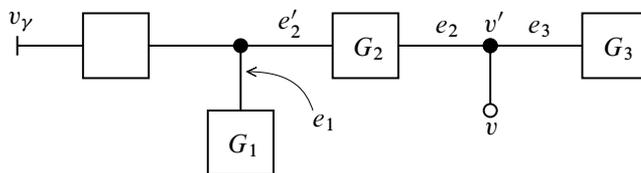

Figure 26: An encoding graph restricting the gleam of a disk region.
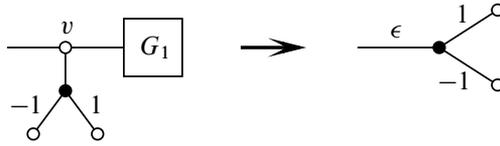
Figure 27: The modification of a decorated graph as in Lemma 8.4.

$k_i \in \mathbb{Z}_{\geq 0}$, and hence there is a 2–chain c_i in $V_i \subset V$ such that $\partial c_i = 2^{k_i}[\gamma_i]$ for $i \in \{1, 3\}$. We then define a 2–chain C_1 according to the order of $[\gamma_2]$;

- in the case $p_2[\gamma_2] = 0$ in $H_1(U_2)$ for some $p_2 \in \mathbb{Z}_{>0}$, there exists a 2–chain c_2 in U_2 with $\partial c_2 = p_2[\gamma_2]$, and then set $C_1 = c_2 - p_2[R]$;

- in the case where $[\gamma_2]$ is not a torsion element in $H_1(U_2)$, we have $p_2'[\gamma_2'] = p_2[\gamma_2]$ in $H_1(U_2)$ for some $p_2, p_2' \in \mathbb{Z} \setminus \{0\}$ by Lemma 5.11, and then set $C_1 = p_2'c_1 - p_2[R]$.

Define another 2–chain as $C_3 = c_3 - 2^{k_3}[R]$. These 2–chains C_1 and C_3 are homology cycles in $H_2(V)$ since $[\gamma_1] = [\gamma_2']$ and $[\gamma_2] = [\gamma_3] = \partial[R]$. Then we have $Q(C_1, C_3) = p_2 2^{k_3} \mathfrak{gl}(R)$, which must be 0. Hence $\mathfrak{gl}(R) = 0$. □

Lemma 8.4 *Suppose that X' contains a simply connected subpolyhedron V with one cut end such that a lift of the cut end has a vertical compressing disk. Then K admits a shadow obtained from X by replacing V with a 2–disk, and the gleam of the newly formed region containing the 2–disk is one of $-\frac{1}{2}, 0, \frac{1}{2}$ or 1.*

Proof We give the proof by using decorated graphs. The assumption in the statement implies that a decorated graph of X' has a subgraph G as shown in the left of Figure 27, where the subgraph G_1 in the figure corresponds to V. It is enough to modify the graph as in the figure. Note that the decoration ϵ in the right of the figure can be replaced with $\epsilon \pm 2$ by applying move (c), and hence the gleam of the corresponding region can be chosen to be one of $-\frac{1}{2}, 0, \frac{1}{2}$ or 1 by applying the move as many times as necessary.

Let v be the vertex of type (P) shown in the left of Figure 27. By Lemma 5.10, there exists a leaf in G_1 such that all the vertices of type (Y$_{12}$) contained in the geodesic from v to the leaf are two-sided to v. Taking the union of all such geodesics $\ell_1, \ldots, \ell_m$, which we denote by $T_\ell = \bigcup_{i=1}^m \ell_i$, is a subtree of G whose leaves are of type (D) except for v. Note that a vertex of type (P) contained in $T_\ell \setminus v$ is a trivalent vertex even in T_ℓ. We divide the proof into the following three cases:

(1) $T_\ell \setminus v$ does not contain a vertex of type (Y$_{111}$) or (P);

(2) $T_\ell \setminus v$ contains vertices of type (Y$_{111}$) or (P), and farthest one from v among them is of type (Y$_{111}$);

(3) $T_\ell \setminus v$ contains vertices of type (Y$_{111}$) or (P), and farthest one from v among them is of type (P).

(1) In this case, T_ℓ is a line, and all the vertices between v and the leaf are of type (Y$_{12}$). We can modify G as in Figure 27 by using moves (f), (g) and (h).
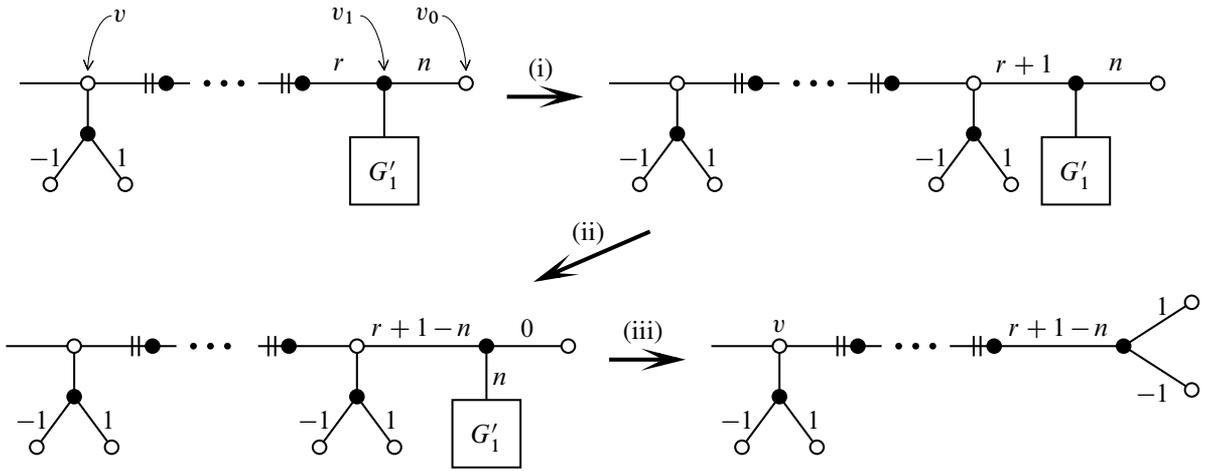
Figure 28: The case (2) in the proof of Lemma 8.4.

(2) Let v_1 be the vertex of type (Y_{111}) farthest from v. Then there is a geodesic ℓ_i containing v_1, and let v_0 be the other endpoint than v. The vertices between v_0 and v_1 are of type (Y_{12}). If the edge incident to v_0 is decorated by $r \neq \frac{1}{2}$, a move (g) can be applied, which is contrary to Lemma 8.3. Hence we can eliminate all the vertices of type (Y_{12}) between v_0 and v_1 using only moves (f), and then we can assume that v_0 and v_1 are connected by one edge. Let e denote this edge.

If ℓ_i has a vertex of type (Y_{111}) other than v_1, the edge e is decorated with 0 by Lemma 8.3. Then the vertex v_1 can be eliminated by a connected-sum reduction.

If ℓ_i has no vertices of type (Y_{111}) other than v_1, then G is as shown in the upper left of Figure 28. The modifications in Figure 28(i) and (ii) are done by moves (e) and (a), respectively. The move in Figure 28(iii) is a connected-sum reduction and a YV-move. The lower right graph in Figure 28 can be modified as required by moves (h), (d) and (c).

(3) Let v_2 be the vertex of type (P) farthest from v. Then G is as the uppermost graph in Figure 29. Let G_1', G_1'' and G_1''' be subgraphs of G_1 as defined in the figure. The subgraphs G_1'' and G_1''' do not contain vertices of type (Y_{111}) or (P). Then we can apply moves (f), (g) or (h) as well as in (1) to these subgraphs, and G is modified as shown in one of Figure 29(i), (ii) or (iii). Moreover, the moves of Figure 29(iv) and (v) are obtained by YV-moves, and the move of Figure 29(vi) is done by a move (d). In either case, the vertex v_2 is eliminated, and we obtain the modification in Figure 27 inductively. $\square$

8.2 Existence of compressing disks

For $i \in \{1, \ldots, 11\}$, the polyhedron X_i can be embedded in $\natural_2(S^1 \times B^3)$ as a shadow, and the complement of ∂X_i in $\partial(\natural_2(S^1 \times B^3))$ $(= \#_2(S^1 \times S^2))$ is a 3–manifold with tori boundary. Note that this 3–manifold $\#_2(S^1 \times S^2) \setminus \partial X_i$ admits a complete hyperbolic structure with finite volume [Costantino and Thurston
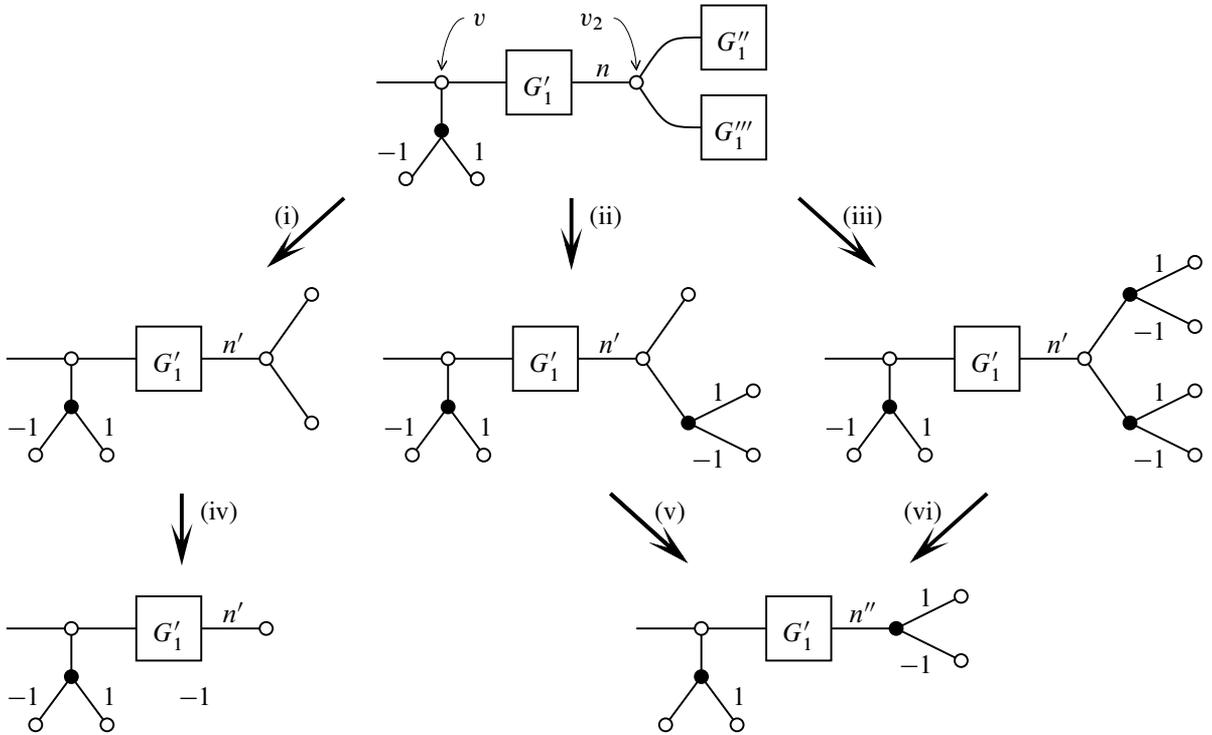
Figure 29: The case (3) in the proof of Lemma 8.4.

2008]. In [Koda et al. 2022], Dehn fillings on this 3–manifold giving $\#_h(S^1 \times S^2)$ for some $h \in \mathbb{Z}_{\geq 0}$ are studied, and it leads to the following.

Lemma 8.5 [Koda et al. 2022] *Suppose that X contains a subpolyhedron Y homeomorphic to X_3 or X_4. Then at least one of the following holds*:

(1) *both of the components of ∂Y have vertical compressing disks*; *or*

(2) *the component of ∂Y with length 1 has a horizontal compressing disk.*

8.3 Banded unlink diagram of 2–knot with complexity one

Recall that, for $n \in \mathbb{Z}$, K_n is a 2–knot defined by the banded unlink diagram shown in Figure 1. We first prove the essential part of Theorem 8.10:

Theorem 8.6 *If a 2–knot K with $G(K) \not\cong \mathbb{Z}$ has shadow-complexity 1, then K is diffeomorphic to K_n for some nonzero integer n.*

Proof Let X be a shadow of K with $c(X) = 1$, and let G be a decorated tree graph for X. By Lemmas 7.4, 7.5 and 7.6, G has exactly one vertex v_0 of type (X_3) or (X_4). Here we suppose that v_0 is of type (X_3). Then G is as shown in Figure 30 by Lemma 7.4.
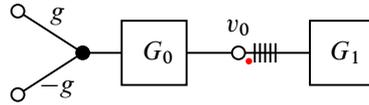
Figure 30: A decorated graph encoding a shadow of K having a subpolyhedron homeomorphic to X_3.

If Lemma 8.5(2) holds, then the graph G can be assumed as shown in the left of Figure 31. Applying a connected-sum reduction, we obtain the right graph. This graph encodes a simple polyhedron without true vertex, which is not our focus here.

We then assume that Lemma 8.5(1) holds, and the graph G is as shown in the top of Figure 32. Let U and V be the subpolyhedra corresponding to the subgraphs G_0 and G_1, respectively. By Lemmas 7.4, we have $\pi_1(U) \cong \langle \gamma, \gamma_1 \mid \gamma^{2^m} \gamma_1 \rangle$ and V is simply connected. We apply Lemmas 5.10 and 5.12 to U and V, respectively, and then the graph G can be assumed to be the second graph in Figure 32. Note that subgraphs $G_{0,0}, G_{0,1}, \ldots, G_{0,m}$ in the figure encode simply connected subpolyhedra by Lemma 5.12. The move in Figure 32(ii) is done by iterating moves (e), and the move in Figure 32(iii) is done by that in Figure 27 (cf Lemma 8.4).

From the bottom graph in Figure 32, we obtain a banded unlink diagram shown in the top of Figure 33. We refer the reader to [Koda and Naoe 2020] for a translation of a shadow into a Kirby diagram; see also Remark 8.7 and [Costantino and Thurston 2008; Martelli 2005]. Though the framings $\epsilon_0, \epsilon_1, \ldots, \epsilon_m$ and ϵ are determined from the gleams, each of them can be assumed to be 0 or 1 since there is a 0–framed knot as a meridian. The first move of Figure 33(i) is obtained by handle-slides and a cancellation of a 1-2 pair. We iterate the same process in Figure 33(ii). The move (iii) is obtained by handle-slides and a cancellation of a 1-2 pair, and we set $n = 2^m g$ and $\epsilon' = 0$ or 1. The move (iv) is done by a cup and 2–handle band swims. The move (v) is an isotopy, and (iv) is obtained by a 2–handle band slide and a cancellation of a 1-2 pair. The move (vii) is done by an isotopy if $\epsilon' = 1$, and we also need 2–handle band slides if $\epsilon' = 0$. The move (viii) is obtained by a cap and 2–handle band slides. Finally, applying a 2–handle band swim and a cancellation of a 2-3 pair, we obtain the diagram shown in Figure 1.

One can show the case where v_0 is of type (X_4) in a similar way to the above, so we skip the details. □

Remark 8.7 The method of a translation of a shadow only to a Kirby diagram is treated in [Koda and Naoe 2020], and that to a banded unlink diagram is actually not discussed. However, we can draw a diagram as shown in Figure 33 by considering a decomposition $X = \mathrm{Nbd}(K; X) \cup X'$ and using [Koda
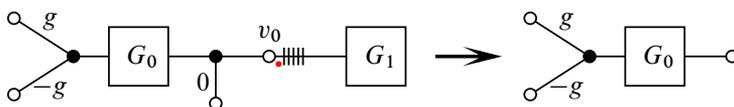

Figure 31: A decorated graph encoding a shadow of K having a subpolyhedron homeomorphic to X_3 such that the boundary component of the subpolyhedron with length 1 has a horizontal compressing disk.
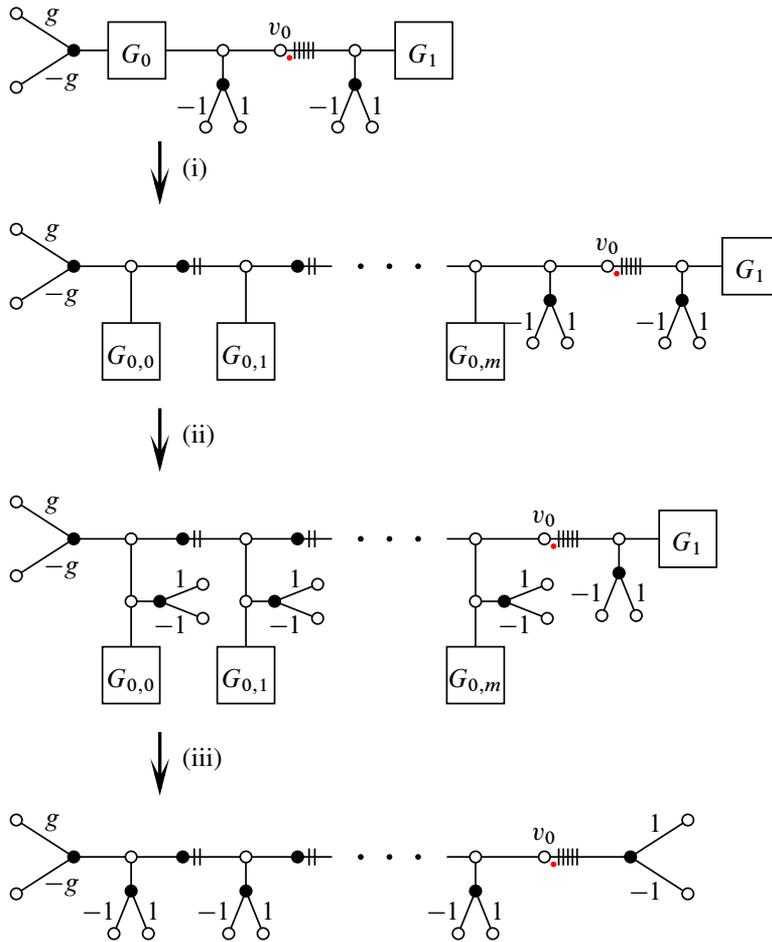
Figure 32: A decorated graph encoding a shadow of K having a subpolyhedron homeomorphic to X_3 such that the two boundary components of the subpolyhedron has vertical compressing disks.

and Naoe 2020, Lemmas 1.1 and 1.2]. Note that $\mathrm{Nbd}(K; X)$ is a shadow of $\mathrm{Nbd}(K; S^4) \cong S^2 \times D^2$ and $\partial \mathrm{Nbd}(K; X)$ is a knot in $\partial \mathrm{Nbd}(K; S^4) \cong S^2 \times S^1$ such that it winds g times along $\{\mathrm{pt}\} \times S^1$.

Remark 8.8 If v_0 is of type (X_3), the 2–knot K is diffeomorphic to K_n with $n = 2^m g > 0$. On the other hand, if v_0 is of type (X_4), the 2–knot K is diffeomorphic to K_n with $n = -2^m g < 0$.

The following implies that there exist infinitely many 2–knots with shadow-complexity 1.

Proposition 8.9 *The 2–knots K_n and $K_{n'}$ are not equivalent unless $n = n'$.*

Proof From Lemma 7.4 and Remark 8.8, we have

$$G(K_n) \cong \langle x, y \mid x^2 y^{|n|} x^{-1} y^{-n} \rangle,$$
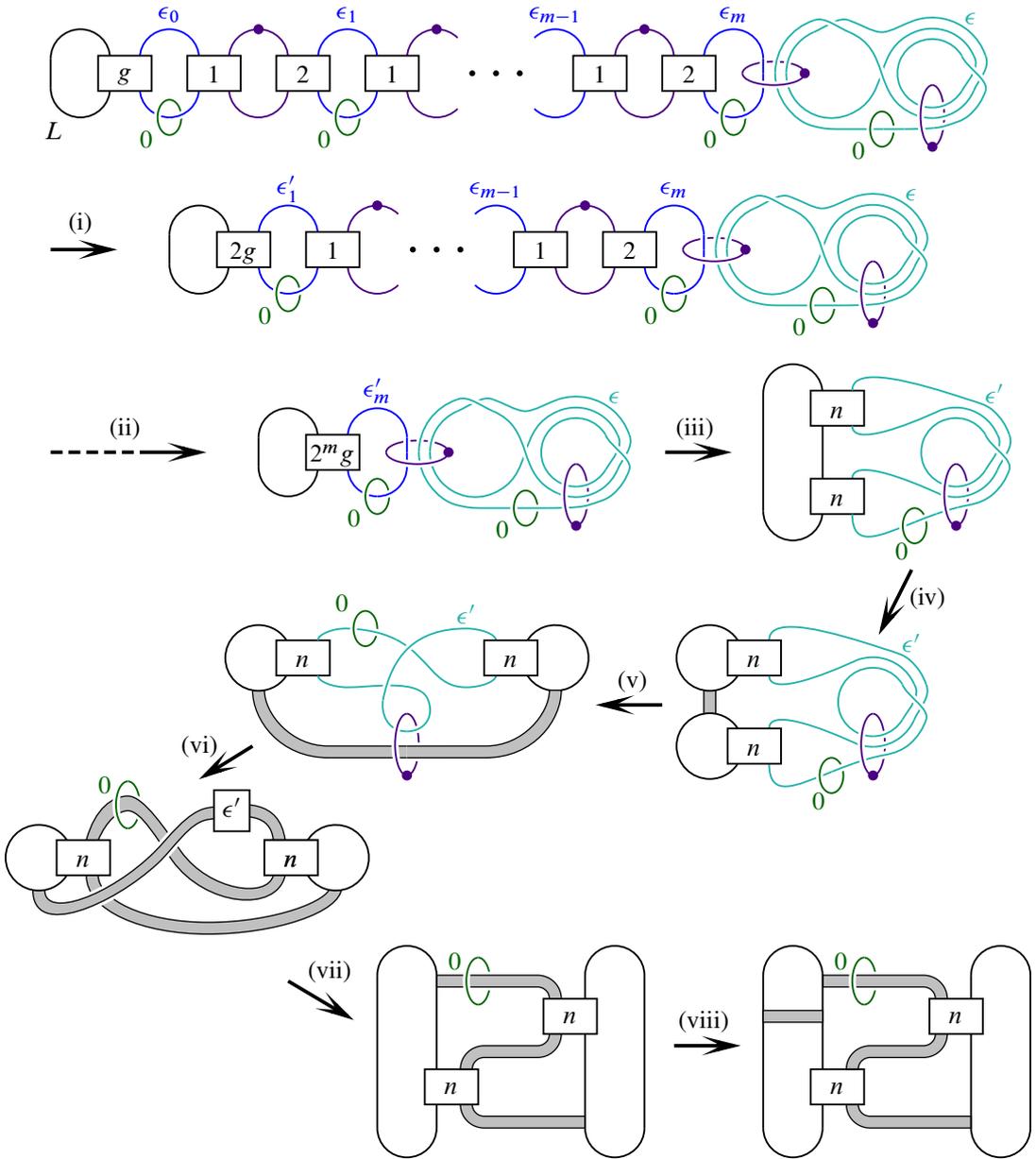
Figure 33: A banded unlink diagram of a 2–knot with complexity one and its modifications.

and its Alexander polynomial is

$$\Delta_{K_n}(t) = \begin{cases} 2 - t^n & \text{if } n \geq 0, \\ 1 - t^n + t^{2n} & \text{if } n < 0. \end{cases}$$

If $n \neq n'$, two polynomials $\Delta_{K_n}(t)$ and $\Delta_{K_{n'}}(t)$ are distinct. □

At last, we give the proof of the complexity one case.

Figure 34: A decorated graph of a shadow of a 2–knot with complexity one.

Theorem 8.10 *A 2–knot K whose knot group is not infinite cyclic has shadow-complexity 1 if and only if K is diffeomorphic to K_n for some nonzero integer n.*

Proof The only if part has been already discussed in Theorem 8.6.

Let n be an arbitrary nonzero integer. The banded unlink diagram of K_n shown in Figure 1 can be obtained from a shadow encoded in Figure 34, left or right, in the same way as in the proof of Theorem 8.6. Therefore, $\mathrm{sc}(K_n) \leq 1$, and hence $\mathrm{sc}(K_n) = 1$ by Proposition 8.9 and Theorem 6.4. □

Remark 8.11 Let X be a shadow of K_n encoded by a decorated graph as shown in Figure 34. Its singular set $S(X)$ has 3 connected components: two circles and one 8–shaped graph. Then we can obtain a special shadow of K_n from X by applying $(0 \to 2)$–moves twice [Costantino 2004; Turaev 1994], and hence we have $\mathrm{sc}^{\mathrm{sp}}(K_n) \leq 5$. This implies that the special shadow-complexity for 2–knots is not a finite-to-one invariant, while that for closed 4–manifolds is finite-to-one [Martelli 2005, Corollary 2.7].

References

[Artin 1925] **E Artin**, *Zur Isotopie zweidimensionaler Flächen im R_4*, Abh. Math. Sem. Univ. Hamburg 4 (1925) 174–177 MR Zbl

[Blair et al. 2022] **R Blair**, **M Campisi**, **S A Taylor**, **M Tomova**, *Kirby–Thompson distance for trisections of knotted surfaces*, J. Lond. Math. Soc. 105 (2022) 765–793 MR Zbl

[Carter et al. 1997] **J S Carter**, **J H Rieger**, **M Saito**, *A combinatorial description of knotted surfaces and their isotopies*, Adv. Math. 127 (1997) 1–51 MR Zbl

[Costantino 2004] **F Costantino**, *Shadows and branched shadows of 3 and 4–manifolds*, PhD thesis, Scuola Normale Superiore (2004)

[Costantino 2006a] **F Costantino**, *Complexity of 4–manifolds*, Exp. Math. 15 (2006) 237–249 MR Zbl

[Costantino 2006b] **F Costantino**, *Stein domains and branched shadows of 4–manifolds*, Geom. Dedicata 121 (2006) 89–111 MR Zbl

[Costantino 2008] **F Costantino**, *Branched shadows and complex structures on 4–manifolds*, J. Knot Theory Ramifications 17 (2008) 1429–1454 MR Zbl

[Costantino and Thurston 2008] **F Costantino**, **D Thurston**, *3–Manifolds efficiently bound 4–manifolds*, J. Topol. 1 (2008) 703–745 MR Zbl

[Freedman and Quinn 1990] **M H Freedman**, **F Quinn**, *Topology of 4–manifolds*, Princeton Math. Ser. 39, Princeton Univ. Press (1990) MR Zbl

[Gay and Kirby 2016] **D Gay**, **R Kirby**, *Trisecting 4–manifolds*, Geom. Topol. 20 (2016) 3097–3132 MR Zbl

[Gompf et al. 2010] **R E Gompf**, **M Scharlemann**, **A Thompson**, *Fibered knots and potential counterexamples to the property* 2*R and slice-ribbon conjectures*, Geom. Topol. 14 (2010) 2305–2347 MR Zbl

[Hughes et al. 2020] **M C Hughes**, **S Kim**, **M Miller**, *Isotopies of surfaces in* 4*–manifolds via banded unlink diagrams*, Geom. Topol. 24 (2020) 1519–1569 MR Zbl

[Ishikawa and Koda 2017] **M Ishikawa**, **Y Koda**, *Stable maps and branched shadows of* 3*–manifolds*, Math. Ann. 367 (2017) 1819–1863 MR Zbl

[Ishikawa and Naoe 2020] **M Ishikawa**, **H Naoe**, *Milnor fibration, A'Campo's divide and Turaev's shadow*, from "Singularities" (M Ishikawa, S Yokura, editors), World Sci., Hackensack, NJ (2020) 95–116 MR Zbl

[Jabłonowski 2016] **M Jabłonowski**, *On a banded link presentation of knotted surfaces*, J. Knot Theory Ramifications 25 (2016) art. id. 1640004 MR Zbl

[Kawauchi 2021] **A Kawauchi**, *Ribbonness of a stable-ribbon surface-link, I: A stably trivial surface-link*, Topology Appl. 301 (2021) art. id. 107522 MR Zbl

[Kawauchi 2023] **A Kawauchi**, *Uniqueness of an orthogonal* 2*–handle pair on a surface-link*, Contemp. Math. 4 (2023) 182–188

[Kawauchi et al. 1982] **A Kawauchi**, **T Shibuya**, **S Suzuki**, *Descriptions on surfaces in four-space*, *I*: *Normal forms*, Math. Sem. Notes Kobe Univ. 10 (1982) 75–125 MR Zbl

[Kearton and Kurlin 2008] **C Kearton**, **V Kurlin**, *All* 2*–dimensional links in* 4*–space live inside a universal* 3*–dimensional polyhedron*, Algebr. Geom. Topol. 8 (2008) 1223–1247 MR Zbl

[Koda and Naoe 2020] **Y Koda**, **H Naoe**, *Shadows of acyclic* 4*–manifolds with sphere boundary*, Algebr. Geom. Topol. 20 (2020) 3707–3731 MR Zbl

[Koda et al. 2022] **Y Koda**, **B Martelli**, **H Naoe**, *Four-manifolds with shadow-complexity one*, Ann. Fac. Sci. Toulouse Math. 31 (2022) 1111–1212 MR Zbl

[Laudenbach and Poénaru 1972] **F Laudenbach**, **V Poénaru**, *A note on* 4*–dimensional handlebodies*, Bull. Soc. Math. France 100 (1972) 337–344 MR Zbl

[Lomonaco 1981] **S J Lomonaco, Jr**, *The homotopy groups of knots, I: How to compute the algebraic* 2*–type*, Pacific J. Math. 95 (1981) 349–390 MR Zbl

[Martelli 2005] **B Martelli**, *Links, two-handles, and four-manifolds*, Int. Math. Res. Not. 2005 (2005) 3595–3623 MR Zbl

[Martelli 2011] **B Martelli**, *Four-manifolds with shadow-complexity zero*, Int. Math. Res. Not. 2011 (2011) 1268–1351 MR Zbl

[Matveev 2003] **S Matveev**, *Algorithmic topology and classification of* 3*–manifolds*, Algor. Computat. Math. 9, Springer (2003) MR Zbl

[Meier and Zupan 2017] **J Meier**, **A Zupan**, *Bridge trisections of knotted surfaces in* S^4, Trans. Amer. Math. Soc. 369 (2017) 7343–7386 MR Zbl

[Meier and Zupan 2018] **J Meier**, **A Zupan**, *Bridge trisections of knotted surfaces in* 4*–manifolds*, Proc. Natl. Acad. Sci. USA 115 (2018) 10880–10886 MR Zbl

[Naoe 2017] **H Naoe**, *Shadows of* 4*–manifolds with complexity zero and polyhedral collapsing*, Proc. Amer. Math. Soc. 145 (2017) 4561–4572 MR Zbl

[Roseman 1998] **D Roseman**, *Reidemeister-type moves for surfaces in four-dimensional space*, from "Knot theory", Banach Center Publ. 42, Polish Acad. Sci. Inst. Math., Warsaw (1998) 347–380 MR Zbl

[Saito and Satoh 2005]   **M Saito**, **S Satoh**, *The spun trefoil needs four broken sheets*, J. Knot Theory Ramifications 14 (2005) 853–858   MR   Zbl

[Satoh 2000]   **S Satoh**, *On non-orientable surfaces in 4–space which are projected with at most one triple point*, Proc. Amer. Math. Soc. 128 (2000) 2789–2793   MR   Zbl

[Satoh 2009]   **S Satoh**, *Triviality of a 2–knot with one or two sheets*, Kyushu J. Math. 63 (2009) 239–252   MR   Zbl

[Satoh and Shima 2004]   **S Satoh**, **A Shima**, *The 2–twist-spun trefoil has the triple point number four*, Trans. Amer. Math. Soc. 356 (2004) 1007–1024   MR   Zbl

[Swenton 2001]   **F J Swenton**, *On a calculus for 2–knots and surfaces in 4–space*, J. Knot Theory Ramifications 10 (2001) 1133–1141   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

[Yajima 1964]   **T Yajima**, *On simply knotted spheres in R^4*, Osaka Math. J. 1 (1964) 133–152   MR   Zbl

[Yoshikawa 1994]   **K Yoshikawa**, *An enumeration of surfaces in four-space*, Osaka J. Math. 31 (1994) 497–522   MR   Zbl

[Zeeman 1965]   **E C Zeeman**, *Twisting spun knots*, Trans. Amer. Math. Soc. 115 (1965) 471–495   MR   Zbl

Department of Mathematics, Tokyo Institute of Technology
Tokyo, Japan

naoe@math.titech.ac.jp

Automorphisms of some variants of fine graphs

Frédéric Le Roux

Maxime Wolff

Recently, Bowden, Hensel and Webb defined the *fine curve graph* for surfaces, extending the notion of curve graphs for the study of homeomorphism or diffeomorphism groups of surfaces. Later, Long, Margalit, Pham, Verberne and Yao proved that for a closed surface of genus $g \geqslant 2$, the automorphism group of the fine graph is naturally isomorphic to the homeomorphism group of the surface. We extend this result to the torus case $g = 1$; in fact our method works for more general surfaces, compact or not, orientable or not. We also discuss the case of a smooth version of the fine graph.

37E30, 57K20

1 Introduction

1.1 Context and results

For a connected, compact surface Σ_g of genus $g \geqslant 1$, Bowden, Hensel and Webb [2] recently introduced the *fine curve graph* $\mathscr{C}^\dagger(\Sigma)$, as the graph whose vertices are all the essential closed curves on Σ, with an edge between two vertices a and b whenever $a \cap b = \varnothing$, if $g \geqslant 2$, and whenever $|a \cap b| \leqslant 1$ if $g = 1$. They proved that for every $g \geqslant 1$, the graph $\mathscr{C}^\dagger(\Sigma)$ is hyperbolic, and derived a construction of an infinite dimensional family of quasimorphisms on $\mathrm{Homeo}_0(\Sigma)$, thereby answering long standing questions of Burago, Ivanov and Polterovich.

The ancestor of the fine graph is the usual curve complex of a surface Σ, ie the complex whose vertices are the isotopy classes of essential curves, with an edge (or a simplex, more generally) between some vertices if and only if they have disjoint representatives. Since its introduction by Harvey [5], the curve complex of a surface has been an extremely useful tool for the study of the mapping class group $\mathrm{Mod}(\Sigma)$ of that surface, as it acts on it naturally. In particular, the fact that this complex is hyperbolic, discovered by Masur and Minsky [13; 14], has greatly improved the understanding of the mapping class groups. The result of Bowden, Hensel and Webb, promoting the hyperbolicity of the curve complex to that of the fine curve graph, opens the door both to the study of what classical properties of usual curve complexes have counterparts in the fine curve graph, and to the use of this graph to derive properties of homeomorphism groups. A first step in this direction was taken by Bowden, Hensel, Mann, Militon and Webb [1], who explored the metric properties of the action of $\mathrm{Homeo}(\Sigma)$ on this hyperbolic graph.

A classical theorem by Ivanov (see [7], as well as Korkmaz [8] and Luo [12] for other surfaces) states that, when Σ is a closed surface of genus $g \geqslant 2$, every automorphism of $\mathscr{C}(\Sigma)$ is realized by a homeomorphism. Recently, Long, Margalit, Pham, Verberne and Yao [11] proved the following natural counterpart of Ivanov's theorem for fine graphs: provided Σ is a compact orientable surface of genus $g \geqslant 2$, the natural map

$$\mathrm{Homeo}(\Sigma) \to \mathrm{Aut}(\mathscr{C}^{\dagger}(\Sigma))$$

is an isomorphism. They also suggested that this map (with the appropriate version of $\mathscr{C}^{\dagger}$) may also be an isomorphism when $g = 1$, and conjectured that the automorphism group of the fine curve graph of smooth curves should be nothing more than $\mathrm{Diff}(\Sigma)$.

In this article, we address both these questions. Our motivation originates from the case of the torus: excited by [1], we wanted to understand more closely the relation between the rotation set of homeomorphisms isotopic to the identity and the metric properties of their actions on the fine graph. This subject will be treated in another article, joint with Passeggi and Sambarino [10]. The methods developed in the present article are valid not only for the torus but for a large class of surfaces.

We work on nonspherical surfaces (ie surfaces not embeddable in the 2–sphere, or equivalently, containing at least one nonseparating simple closed curve), orientable or not, compact or not. We consider the graph $\mathscr{NC}^{\dagger}_{\pitchfork}(\Sigma)$, whose vertices are the nonseparating simple closed curves, and with an edge between two vertices a and b whenever they are either disjoint, or have exactly one topologically transverse intersection point (see the beginning of Section 2 for more detail). Our first result answers a problem raised in [11].

Theorem 1 *Let Σ be a connected, nonspherical surface, without boundary. Then the natural map* $\mathrm{Homeo}(\Sigma) \to \mathrm{Aut}(\mathscr{NC}^{\dagger}_{\pitchfork}(\Sigma))$ *is an isomorphism.*

Our second result concerns the smooth version of fine graphs. We consider the graph $\mathscr{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$ whose vertices are the smooth nonseparating curves in Σ, with an edge between a and b if they are disjoint or have one transverse intersection point, in the differentiable sense (in particular, $\mathscr{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$ is not the subgraph of $\mathscr{NC}^{\dagger}_{\pitchfork}(\Sigma)$ induced by the vertices corresponding to smooth curves; it has fewer edges). The following result partially confirms the conjecture of [11]; here we restrict to the case of orientable surfaces for simplicity.

Theorem 2 *Let Σ be a connected, orientable, nonspherical surface, without boundary. Then all the automorphisms of $\mathscr{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$ are realized by homeomorphisms of Σ.*

In other words, if we denote by $\mathrm{Homeo}_{\infty\pitchfork}(\Sigma)$ the subgroup of $\mathrm{Homeo}(\Sigma)$ preserving the collection of smooth curves and preserving transversality, then the natural map

$$\mathrm{Homeo}_{\infty\pitchfork}(\Sigma) \to \mathrm{Aut}(\mathscr{NC}^{\dagger\infty}_{\pitchfork}(\Sigma))$$

is an isomorphism. We were surprised to realize however that $\mathrm{Homeo}_{\infty\pitchfork}(\Sigma)$ is strictly larger than $\mathrm{Diff}(\Sigma)$.
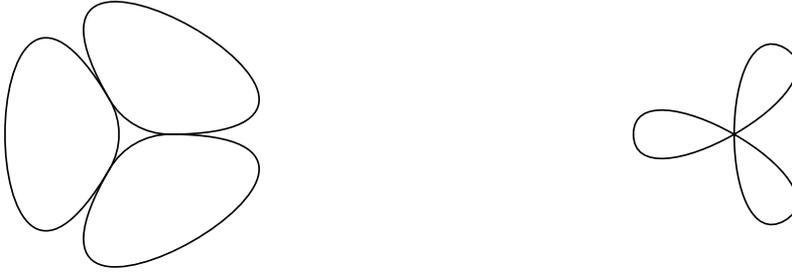
Figure 1: A necklace (left) and a bouquet (right) of three circles.

Proposition 3 *Every surface* Σ *admits a homeomorphism* f *such that* f *and* f^{-1} *preserve the set of smooth curves, and preserve transversality, but such that neither* f *nor* f^{-1} *is differentiable. In particular, the natural map*

$$\mathrm{Diff}(\Sigma) \to \mathrm{Aut}(\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger\infty}(\Sigma))$$

is not surjective.

1.2 Idea of the proof of Theorem 1

The main step in this proof is the following.

Proposition 4 *If* $\{a, b\}$ *or* $\{a, b, c\}$ *is a 2–clique or a 3–clique of* $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$ *then, from the graph structure of* $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$, *we can tell the type of the clique.*

If $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are two n–cliques in the graph $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$, we say they have the same *type* if the unions $\bigcup_{j=1}^{n} a_j$ and $\bigcup_{j=1}^{n} b_j$ are homeomorphic. Note that we do not require that such a homeomorphism extends to a homeomorphism of the surface (as we do not consider how the cliques are embedded, Figure 1 does not show the transversality at the intersection points). Let us describe the different types of 2– and 3–cliques. A 2–clique $\{a, b\}$, ie an edge of the graph $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$, may be of two distinct types: the intersection $a \cap b$ may be empty or not. For a 3–clique $\{a, b, c\}$, up to permuting the curves a, b and c, the cardinals of the intersections $a \cap b$, $a \cap c$ and $b \cap c$, respectively, may be $(1, 1, 1)$, $(1, 1, 0)$, $(1, 0, 0)$, or $(0, 0, 0)$. This determines the type of the 3–clique, except in the case $(1, 1, 1)$, where the intersection points $a \cap b$, $a \cap c$ and $b \cap c$ may be pairwise distinct, in which case we will speak of a 3–clique of type *necklace*, or these intersection points may be equal, in which case we will speak of a 3–clique of type *bouquet*; see Figure 1.

The main bulk of the proof of Proposition 4 consists in distinguishing the 3–cliques of type necklace from any other 3–clique of $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$. Here, the key is that among all the 3–cliques, the cliques $\{a, b, c\}$ of type necklace are exactly those such that the union $a \cup b \cup c$ contains nonseparating simple closed curves other than a, b and c. In terms of the graph structure, this leads to the following property, denoted by $N(a, b, c)$, which turns out to characterize these cliques:

There exists a finite set F of at most 8 vertices of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, all distinct from a, b and c, such that every vertex d connected to a, b and c in this graph, is connected to at least one element of F.

From this, we will easily characterize all the configurations of 2–cliques and 3–cliques in terms of similar statements in the first order logic of the graph $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$.

Now, let $T(\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma))$ denote the set of edges $\{a, b\}$ of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$ satisfying $|a \cap b| = 1$. Then we have a map

$$\mathrm{Point} \colon T(\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)) \to \Sigma,$$

which to each edge $\{a, b\}$ of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, associates the intersection point $a \cap b$. The next step in the proof of Theorem 1 now consists in characterizing the equality $\mathrm{Point}(a, b) = \mathrm{Point}(c, d)$ in terms of the structure of the graph. This characterization shows that every automorphism of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$ is realized by some bijection of Σ; then we prove that such a bijection is necessarily a homeomorphism (see Proposition 20).

In order to characterize the equality $\mathrm{Point}(a, b) = \mathrm{Point}(c, d)$, we introduce on $T(\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma))$ a relation $\Diamond$. This relation is generated, essentially, by $(a, b) \Diamond (b, c)$ if (a, b, c) is a 3–clique of type bouquet (see Section 3.2 for details). If $(a, b) \Diamond (c, d)$ then $\mathrm{Point}(a, b) = \mathrm{Point}(c, d)$. Interestingly, the converse is false, but we can still use this idea in order to characterize the points of Σ in terms of the graph structure of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$.

This subtlety between the relation $\Diamond$ and the equality of points is related to the nonsmoothness of the curves involved, and more precisely, to the fact that a curve may spiral infinitely with respect to another curve in a neighborhood of a common point. We think that this phenomenon is of independent interest and we investigate it in Section 4. In particular, we can easily state, in terms of the graph structure of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, an obstruction for a homeomorphism to be conjugate to a C^1–diffeomorphism; see Section 4.6.

1.3 Ideas of the proof of Theorem 2

In the smooth case, the adaptation of our proof of Theorem 1 fails from the start: indeed, the closed curves contained in the union $a \cup b \cup c$ of a necklace, and distinct from a, b and c, are not smooth. This suggests the idea to use sequences of curves (at the expense of losing the characterizations of configurations in terms of first order logic).

This time it is easiest to first characterize disjointness of curves (see Lemma 43), and then recover the different types of 3–cliques. Then the strategy follows the C^0 case.

Once we start to work with sequences, it is natural to say that a sequence (f_n) of curves not escaping to infinity converges to a in some weak sense, if for every vertex d such that $\{a, d\}$ is an edge of the graph, $\{f_n, d\}$ is also an edge for all n large enough. As it turns out, this property implies convergence in the C^0–sense to a, and is implied by convergence in the C^1–sense. But it is not equivalent to the convergence in the C^1–sense, and it is precisely this default of C^1–convergence that enables us to distinguish between disjoint or transverse pairs of curves.

Interestingly, this simple criterion for disjointness has no counterpart in the C^0–setting. Indeed, in that setting, no (noneventually constant) sequence of curves converges in this weak sense: given a curve a, and a sequence (f_n) of curves with, say, some accumulation point x in a, we can build a curve d intersecting a once transversally (topologically) at x, but oscillating so much that it intersects several times each f_n that does not contain x. From this perspective, none of our approaches in the C^0–setting and in the C^∞–setting are directly adaptable to the other.

1.4 Further comments

We can imagine many variants of fine graphs. For example, in the arXiv version of [2], for the case of the torus they worked with the graph $\mathcal{NC}^\dagger_{\pitchfork}(\Sigma)$ on which we are working here, whereas in the published version, they changed to a fine graph in which two curves a and b are still related by an edge when they have one intersection, not necessarily transverse.

More generally, in the spirit of Ivanov's metaconjecture, we expect that the group of automorphisms should not change from any reasonable variant to another. And indeed, using the ideas of [11, Section 2] and those presented here, we can navigate between various versions of fine graphs, and recover, from elementary properties of one version, the configurations defining the edges in another version, thus proving that their automorphism groups are naturally isomorphic. From this perspective, it seems satisfying to recover the group of homeomorphisms of the surface as the automorphism group of any reasonable variant of the fine graph. In this vein, we should mention that the results of [11, Section 2] directly yield a natural map $\mathrm{Aut}(\mathcal{C}^\dagger(\Sigma)) \to \mathrm{Aut}(\mathcal{NC}^\dagger_{\pitchfork}(\Sigma))$, and from there, our proof of Theorem 1 may be used as an alternative proof of their main result.

All reasonable variants of the fine graphs should be quasi-isometric, and a unifying theorem (yet out of reach today, as it seems to us) would certainly be a counterpart of the theorem by Rafi and Schleimer [15], which states that every quasi-isometry of the usual curve graph is bounded distance from an isometry.

1.5 Organization of the article

Section 2 is devoted to the recognition of the 3–cliques in the C^0–setting, and of some other configurations regarding the nonorientable case. We encourage the reader to skip, at first reading, everything that concerns the nonorientable case; these points should be easily identified, and this halves the length of the proof. In Section 3 we prove Theorem 1. In Section 4 we characterize, from the topological viewpoint, the relation $\Diamond$ introduced above in terms of the graph structure, and deduce our obstruction to differentiability. Finally in Section 5 we prove Theorem 2 and Proposition 3.

Acknowledgements

2   Recognizing configurations of curves

2.1   Standard facts and notation

We will use, often without mention, the following easy or standard facts for curves on surfaces.

The first is the classification of connected, topological surfaces with boundary (not necessarily compact). In particular, every topological surface admits a smooth structure. Given a closed curve a in a surface Σ, we can apply this classification to $\Sigma \smallsetminus a$ and understand all possible configurations of simple curves; this is the so-called *change of coordinates principle* in the vocabulary of the book of Farb and Margalit [3].

In particular, every closed curve a has a neighborhood homeomorphic to an annulus or a Möbius strip in which a is the "central curve". Very often in this article, we will consider the curves a' obtained by deforming a in such a neighborhood, so that a' is disjoint from a in the first case, or intersects it once, transversely, in the second, as in Figure 2. We will say that a' is obtained by *pushing a aside*.

If two curves a and b have a unique intersection point, we say that the intersection is *transverse*, or *essential*, if there is a homeomorphism of the surface which maps a and b to smooth curves intersecting transversely in the usual smooth sense. Otherwise we say the intersection is *inessential*.

The change of coordinates principle also applies to finite graphs embedded in Σ: there is a homeomorphism of Σ that sends any given graph to a smooth graph, such that all edges connected to a given vertex leave it in distinct directions. In the simple case when the graph is the union of two or three simple closed curves that pairwise intersect at most once, this observation justifies the description of the possible configurations of cliques in the introduction.

Here are two other useful facts.

Fact 5   *A simple closed curve a in a surface is nonseparating if and only if there exists a closed curve b such that $a \cap b$ is a single point and this intersection is transverse.*

Fact 6   *Let p and q be two distinct points, and x, x' and x'' three simple arcs, each with endpoints p and q, such that*

$$x \cap x' = x' \cap x'' = x \cap x'' = \{p, q\}.$$

If two of the three curves $x \cup x'$, $x' \cup x''$, $x'' \cup x$ are separating, then the third one is also separating.

Proof   Denote $y = x \smallsetminus \{p, q\}$, the arc x without its ends, and similarly, define y' and y''. Suppose $x \cup x'$ and $x \cup x''$ are separating. Denote by Σ_1 and Σ_2 (resp. Σ_3 and Σ_4) the components of $\Sigma \smallsetminus (x \cup x')$ (resp. $\Sigma \smallsetminus (x \cup x'')$) where Σ_2 contains y'' and Σ_4 contains y'. By looking at neighborhoods of p and q

Figure 2:   Left: pushing a two-sided curve a. Right: pushing a one-sided curve a.
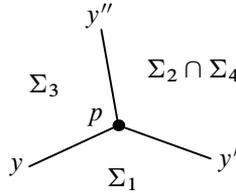
Figure 3: A neighborhood of p.

(see Figure 3), we see that $\Sigma' = \Sigma_2 \cap \Sigma_4$ is nonempty, and that the arc y bounds Σ_1 on one side, and Σ_3 on the other, so $\Sigma'' = \Sigma_1 \cup y \cup \Sigma_3$ is a surface. Now, $\Sigma \smallsetminus (x' \cup x'') = \Sigma' \cup \Sigma''$, and Σ' and Σ'' are disjoint by construction. □

2.2 Properties characterizing geometric configurations

Now we list the properties, in terms of the graph $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$, that will be used as characterizations of certain configurations of curves. This allows us to specify the statement of Proposition 4, which will be proved in the next paragraph, and define the relation $\diamondsuit$ in terms of the graph $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$.

In the following, the letters N D, T and B respectively stand for necklace, disjoint, transverse and bouquet. If a, b and c are vertices of this graph, we will denote by

- $N(a, b, c)$ the property that $\{a, b, c\}$ is a 3–clique of $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$ and there exists a finite set F of at most 8 vertices of $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$, all distinct from a, b and c, such that for every vertex d such that $\{a, b, c, d\}$ is a 4–clique, there is an edge from d to at least one element of F;

- $D(a, b)$ the property that $\{a, b\}$ is an edge of $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$ and there does not exist a vertex d such that $N(a, b, d)$ holds,

- $T(a, b)$ the property that $\{a, b\}$ is an edge of $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$ and $D(a, b)$ does not hold;

- $B(a, b, c)$ the property that $T(a, b)$, $T(a, c)$ and $T(b, c)$ all hold but $N(a, b, c)$ does not.

The following proposition is the main part of Proposition 4.

Proposition 7 *Let a, b and c be vertices of the graph $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$. Property $N(a, b, c)$ holds if and only if $\{a, b, c\}$ is a 3–clique of type necklace of $\mathcal{NC}^\dagger_\text{fh}(\Sigma)$.*

The proof of Proposition 7 will occupy Section 2.3 below. The following corollary complements Proposition 7 and provides a precise version of Proposition 4.

Corollary 8   • *Property $D(a, b)$ holds if and only if the curves a and b are disjoint.*

- *Property $T(a, b)$ holds if and only if a and b have a unique intersection point and the intersection is transverse.*

- *Property $B(a, b, c)$ holds if and only if $\{a, b, c\}$ is a 3–clique of type bouquet.*

Figure 4: Completing (a, b) to a 3–clique of type necklace.

Proof Let a and b be neighbors in the graph $\mathcal{N}\mathcal{C}^{\dagger}_{\pitchfork}(\Sigma)$. Of course, if a and b are disjoint, then $D(a, b)$ holds: there does not exist a curve d such that $N(a, b, d)$ holds, since this would mean that $\{a, b, d\}$ is of type necklace and then, by definition, a and b would intersect. Conversely, suppose that a and b are not disjoint, and let us prove that $D(a, b)$ does not hold, ie let us find a curve c such that $\{a, b, c\}$ is a 3–clique of type necklace. In the case when one of a or b is one-sided, up to exchanging the two, suppose a is one-sided. Then we may push a in order to find a curve c which makes a 3–clique of type necklace with a and b; see Figure 4, left. In the case when both a and b are two-sided, then by the change of coordinates principle, a regular neighborhood of $a \cup b$ is homeomorphic to a one-holed torus, embedded in Σ, with a choice of meridian and longitude coming from a and b. In this torus, a curve c with slope 1 will form a 3–clique of type necklace with a and b; see Figure 4, right. This proves the first point.

The second point is a straightforward consequence of the first, and the third simply follows from the second point together with Proposition 7. □

2.3 Proof of Proposition 7

The following lemma is a key step in the proof of the direct implication in Proposition 7.

Lemma 9 *Let $\{a, b, c\}$ be a 3–clique of $\mathcal{N}\mathcal{C}^{\dagger}_{\pitchfork}(\Sigma)$ which is not of type necklace. Then there exists a vertex d of $\mathcal{N}\mathcal{C}^{\dagger}_{\pitchfork}(\Sigma)$ such that $\{a, b, c, d\}$ is a 4–clique of $\mathcal{N}\mathcal{C}^{\dagger}_{\pitchfork}(\Sigma)$, and such that d meets every connected component of $\Sigma \smallsetminus (a \cup b \cup c)$.*

Before entering the proof, we note that we cannot remove the hypothesis that $\{a, b, c\}$ is not of type necklace. Indeed, in the flat torus $\Sigma = \mathbb{R}^2 / \mathbb{Z}^2$, consider three closed geodesics a, b and c respectively directed by $(1, 0)$, $(0, 1)$ and $(1, 1)$. By pushing c aside if necessary, we obtain a 3–clique of type necklace. The complement of $a \cup b \cup c$ in Σ has three connected components, and there is no curve d satisfying the conclusion of the lemma.

Proof Throughout the proof, we will denote $\Sigma' = \Sigma \smallsetminus (a \cup b \cup c)$. Up to permuting the curves a, b and c, we may suppose that the triple of cardinals of intersections, $(|a \cap b|, |a \cap c|, |b \cap c|)$, equals $(1, 1, 1)$, $(1, 1, 0)$, $(1, 0, 0)$, or $(0, 0, 0)$. We will deal with these cases separately.
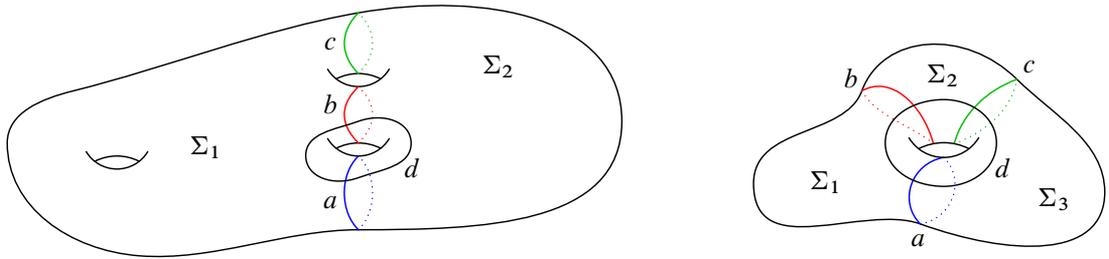
Figure 5: Finding d in the case $(0, 0, 0)$.

Let us begin with the case $(0, 0, 0)$. If Σ' is connected, then any curve d making a 4–clique with (a, b, c) satisfies the lemma. Such a curve can be found, for example, by pushing a aside. If Σ' has two connected components, denote them by Σ_1 and Σ_2. Since a, b and c are each nonseparating, at least two of the curves a, b, c (say, a and b) correspond to boundary components of both Σ_1 and Σ_2. Choose one point x_a in a and one point x_b in b. For $i = 1, 2$, there is an arc γ_i connecting x_a to x_b in Σ_i, and disjoint from the boundary of Σ_i except at its ends. Then the curve $d = \gamma_1 \cup \gamma_2$ satisfies the lemma (see Figure 5, left). It may also happen that Σ' has three connected components, in which case we find a curve d exactly in the same way; see Figure 5, right.

Next we deal with the case of intersections $(1, 0, 0)$. In this case, a and b intersect transversely, once, and c is disjoint from $a \cup b$. By hypothesis, the curve c is (globally) nonseparating. Consider the union $a \cup b$. If a or b is two-sided, then $a \cup b$ does not disconnect its regular neighborhoods. This is seen by traveling along a small band on one side of $a \cup b$ (see Figure 6, left). In this case, Σ' cannot have more connected components than $\Sigma \smallsetminus c$; hence Σ' is connected, and any curve d obtained by pushing c, as in the preceding case, satisfies the lemma. If both a and b are one-sided, then $a \cup b$ is locally disconnecting, so Σ' may have up to two connected components. In this case, a curve d obtained by pushing a satisfies the lemma (see Figure 6, right).

Now assume we are in the case $(1, 1, 1)$ or $(1, 1, 0)$. Since $\{a, b, c\}$ is not of type necklace, in these cases b and c do not meet outside a. We first treat the subcase when a is two-sided. For this we consider any curve d obtained by pushing a aside, and we claim that d meets every connected component of Σ'. Indeed, let C be such a component. Of course the closure of C meets a, b or c. Since both b and c meet a, it actually has to meet a, as we can see by traveling along b or c in C. More precisely, by following b or c in both directions, we see that C meets any neighborhood of a from both sides. Thus it meets d.

Figure 6: Finding d in case $(1, 0, 0)$.

case $(1, 1, 1)$ b and c two-sided b two-sided, c one-sided b and c one-sided
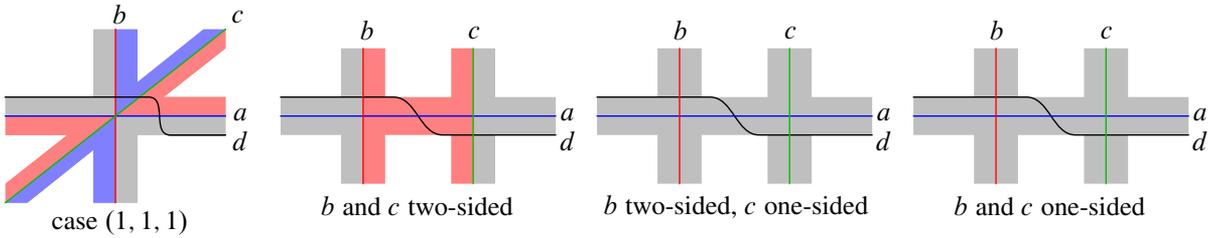
Figure 7: Finding d in cases $(1, 1, 1)$ and $(1, 1, 0)$.

It remains to treat the subcase when a is one-sided, first for the $(1, 1, 1)$ case, and then for the $(1, 1, 0)$ case. In the $(1, 1, 1)$ case, the curves a, b and c play symmetric roles, and by the above argument it just remains to consider the case when they are all one-sided. Then the situation is depicted on Figure 7, left: $a \cup b \cup c$ disconnects its regular neighborhoods into three connected components, and the figure shows a curve d, obtained by pushing a aside, which intersects all three components and such that $\{a, b, c, d\}$ is a 4–clique. In the remaining case the curves b and c play symmetric roles, and there are three different cases to consider, regarding whether b and c are one or two-sided. These three cases are pictured in Figure 7, and in each case, we obtain d by pushing a aside. □

We deduce the following.

Lemma 10 Let $\{a, b, c\}$ be a 3–clique of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$, not of type necklace. Let $(\alpha_1, \ldots, \alpha_n)$ be a finite family of vertices of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$, all distinct from a, b and c. Then there exists a vertex d of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$, such that

- $\{a, b, c, d\}$ is a 4–clique of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$;
- for all $j \in \{1, \ldots, n\}$, the intersection $d \cap \alpha_j$ is infinite; in particular, $\{d, \alpha_j\}$ is not an edge of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$.

As a corollary, we get the direct implication in Proposition 7.

Corollary 11 If $\{a, b, c\}$ is a 3–clique not of type necklace, then $N(a, b, c)$ does not hold.

Proof of Lemma 10 The hypotheses that $\{a, b, c\}$ is not of type necklace and $\alpha_j \notin \{a, b, c\}$, impose that for every j, the curve α_j is not contained in the union $a \cup b \cup c$. Hence, there exists a small subarc $\beta_j \subset \alpha_j$ lying in the complement of $a \cup b \cup c$, and we may further suppose that these n arcs are pairwise disjoint, and choose a point x_j in β_j for each j.

Now, let d_0 be a vertex of $\mathcal{NC}^{\dagger}_{\pitchfork}(\Sigma)$ as from Lemma 9. Since d_0 meets every component of

$$\Sigma' = \Sigma \smallsetminus (a \cup b \cup c),$$

we may perform a surgery on d_0, far from $a \cup b \cup c$, to obtain a new curve d_1 such that $\{a, b, c, d_1\}$ is still a 4–clique, and d_1 still meets every component of Σ', and d_1 passes through x_1. We may iterate this

process, to get a curve d_n which passes through x_j for every j, and such that $\{a, b, c, d_n\}$ is a 4–clique. Finally, we may perform a last surgery on d_n, in the neighborhood of all x_j, in order to obtain a curve d such that for each j, $|\beta_j \cap d|$ is infinite. □

It remains to prove the converse implication in Proposition 7, which we restate as Lemma 12.

Lemma 12 *Let $\{a, b, c\}$ be a 3–clique of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$ of type necklace. Then there exists a finite set F, of at most 8 vertices of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$ all distinct from a, b and c, and such that every d such that $\{a, b, c, d\}$ is a 4–clique of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$ is connected by an edge to some element of F.*

In fact this set F can be chosen explicitly, with cardinal at most 8, as follows. If $\{a, b, c\}$ is a 3–clique of type necklace, then there exists a family of arcs (x, X, y, Y, z, Z), all embedded in Σ, such that $a = x \cup X$, $b = y \cup Y$, $c = z \cup Z$, and such that these six arcs pairwise intersect at most at their ends. The union $a \cup b \cup c$ may be viewed as a graph embedded in Σ, and these six arcs are the edges of this embedded graph. We let F be the set of nonseparating curves, among the 8 curves $X \cup Y \cup Z$, $X \cup Y \cup z$, $X \cup y \cup Z$, etc (there is one choice of upper/lower case for each letter). In the course of the proof of Lemma 12, we will see that F is nonempty, and satisfies the lemma.

Proof Let $\{a, b, c\}$ be a 3–clique of type necklace. Let F be the set of nonseparating curves, as above, among all the 8 curves $x \cup y \cup z$, $X \cup y \cup z$, $X \cup Y \cup z$, etc. Let d be such that $\{a, b, c, d\}$ is a 4–clique. Up to permuting the curves a, b and c, we may suppose that $(|a \cap d|, |b \cap d|, |c \cap d|)$ equals $(1, 1, 1)$, $(1, 1, 0)$, $(1, 0, 0)$ or $(0, 0, 0)$; our proof proceeds case by case.

The easiest case is $(1, 0, 0)$. In this case, up to exchanging the arcs X and x, we may suppose that d intersects a at an interior point of X, and is disjoint from all the other arcs. Consider $f = X \cup y \cup z$. This curve intersects d at a unique point, transversely. It follows that f is nonseparating. Hence $f \in F$ and f satisfies the conclusion of the lemma.

Now let us deal simultaneously with the cases $(1, 1, 1)$ and $(1, 1, 0)$. Suppose first that the intersections of d with $a \cup b \cup c$ do not occur at the intersection points $a \cap b$, $a \cap c$ or $b \cap c$. Up to exchanging x with X, y with Y and z with Z, we may suppose that the intersections occur in the interior of the arcs X, Y and Z in the case $(1, 1, 1)$, and in the interior of the arcs X and Y in the case $(1, 1, 0)$. Now the curve $f = X \cup y \cup z$, for instance, satisfies the conclusion of the lemma.

Now suppose that d contains one of the points $a \cap b$, $a \cap c$ or $b \cap c$. In case $(1, 1, 1)$ we may suppose, up to permuting a, b and c, that d contains the point $a \cap b$, and in the case $(1, 1, 0)$, this is automatic, as d is disjoint from c. Now in any case, d cannot contain $a \cap c$ nor $b \cap c$, because it intersects a and b only once. Hence, up to exchanging z with Z, we may suppose $d \cap z = \varnothing$. In the neighborhood of the point $a \cap b$, up to homeomorphism, the configuration of our curves is as depicted in Figure 8, because all the intersections are supposed to be transverse. Then, up to exchanging X with x or Y with y, we can suppose that the arc $X \cup Y$ has a transverse intersection with d, and then the arc $f = X \cup Y \cup z$ satisfies the conclusion of the lemma.
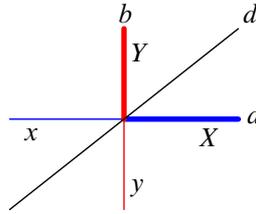
Figure 8: A suitable choice of X and Y.

We are left with the case $(0, 0, 0)$. In this case, any curve in F will satisfy the conclusion of the lemma, and hence all we have to do is to prove that F is nonempty. If $X \cup Y \cup Z$ and $X \cup Y \cup z$ were both separating, then so would be $c = z \cup Z$, by Fact 6. Hence, among these two curves, at least one is nonseparating, and F is nonempty (in fact, it contains at least 4 elements). □

2.4 One- or two-sided curves, and extra bouquets

In this last paragraph of this section, we will see how to recognize, from the graph structure of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$, some additional configurations. We insist that the work in this paragraph is useful only in the case when Σ is nonorientable; it is needed in order to make our proof of Theorem 1 work in that case (see Remark 22 below).

We start with a simple characterization of one-sided and two-sided curves.

- Two(a), the property that for all b such that $T(a, b)$ holds, there exists c such that $T(b, c)$ and $D(a, c)$ both hold,

- One(a), the negation of Two(a): there exists b such that $T(a, b)$ and such that there does not exist c satisfying $T(b, c)$ and $D(a, c)$.

Observation 13 *Let a be a vertex of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$. Then the curve a is one-sided, if and only if* One(a) *holds.*

Proof If a is two-sided, and b satisfies $T(a, b)$, by pushing a aside we find another curve c as in the definition of Two(a). This proves the reverse implication.

If a is one-sided, let b be a curve obtained by pushing a aside. We have $T(a, b)$, and a and b bound a disk. Any curve c disjoint from a, and with $T(b, c)$, has to enter this disk; but then, it has to get out, which is impossible without touching a and without intersecting b another time. □

Our next objective is to characterize when two one-sided curves a and b meet exactly once, nontransversely. We will do this in several steps.

Lemma 14 *Let a and b be one-sided simple curves of Σ. Suppose the intersection $a \cap b$ is not connected. Then there exists a vertex c of $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$, distinct from a and b, such that for every neighbor d of both a and b in this graph, and such that $D(a, d)$ or $D(b, d)$ (or both), the vertices c and d are neighbors in this fine graph.*

Proof Let x be a subarc of b, whose endpoints lie in a, and disjoint from a otherwise. Since $a \cap b$ is disconnected, such an arc exists, and has two distinct end points, p and q. Let x' and x'' be the two subarcs of a whose ends are p and q. From Fact 6, we know that $x \cup x'$ or $x \cup x''$ (or both) is a nonseparating curve; denote it by c. By construction, we have $c \neq a$ and $c \neq b$.

Now let d be a curve satisfying the hypothesis of the lemma. If d is disjoint from a and b, then it is disjoint from c; otherwise d intersects exactly one of a or b, far away from the other. So the intersection between d and c, if any, is still transverse, and d is a neighbor of c in $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$. $\square$

This contrasts with the situation we want to characterize, as we see now.

Lemma 15 *Let a and b be two one-sided curves, and suppose that $a \cap b$ consists in one, inessential intersection point. Then, for every nonseparating curve $c \notin \{a, b\}$, there exists d such that $T(a, d)$ and $D(b, d)$ hold but such that the intersection $c \cap d$ is infinite.*

Proof We first observe that $\Sigma' = \Sigma \smallsetminus (a \cup b)$ is connected. This is seen by following the curves a and b in both directions: the union $a \cup b$ does not disconnect its small neighborhoods. Let c be a curve as above. Then we may consider a first curve d_0, obtained by pushing a aside, in such a way that d_0 is disjoint from b (this is possible since the intersection $a \cap b$ is inessential). Since $c \notin \{a, b\}$, the curve c intersects Σ'. Since d_0 meets every component of Σ' (there is only one), we may deform it into a curve d which intersects c infinitely many times, exactly as in the proof of Lemma 10. $\square$

After these two lemmas, we have a simple sentence in terms of the graph $\mathcal{NC}^{\dagger}_{\text{fh}}(\Sigma)$, which holds when $a \cap b$ is a single inessential intersection point, and which guarantees that $a \cap b$ is connected. In order to upgrade this into a characterization of the first situation, we need to be able to exclude as well the cases when $a \cap b$ is a nondegenerate arc. These cases fall into two subcases: the intersection arc $a \cap b$ can be essential or inessential, exactly as an intersection point. One way to formalize this, is to say that the intersection $a \cap b$ is essential if a cuts a regular neighborhood of $a \cap b$ into two regions both containing a subarc of b, and inessential otherwise.

Lemma 16 *Let a and b be one-sided curves. Suppose that $a \cap b$ is a nondegenerate arc, and suppose this intersection is essential. Then there exist curves $\alpha, \alpha', \beta, \beta'$ obtained by pushing a aside, such that $B(a, \alpha, \beta)$, $B(b, \alpha, \beta)$, $B(a, \alpha', \beta')$, $B(b, \alpha', \beta')$, and $N(a, \alpha, \beta')$.*

Proof The curves α, β, α' and β' may be taken in a neighborhood of $a \cup b$, as pictured in Figure 9. $\square$

Finally, we deal with inessential arcs.

Lemma 17 *Let a and b be one-sided curves, such that $a \cap b$ is an inessential arc or intersection point. Let c be such that $T(a, c)$. Then there exists d such that $B(a, c, d)$ and $D(b, d)$, if and only if the intersection point $a \cap c$ does not belong to b.*
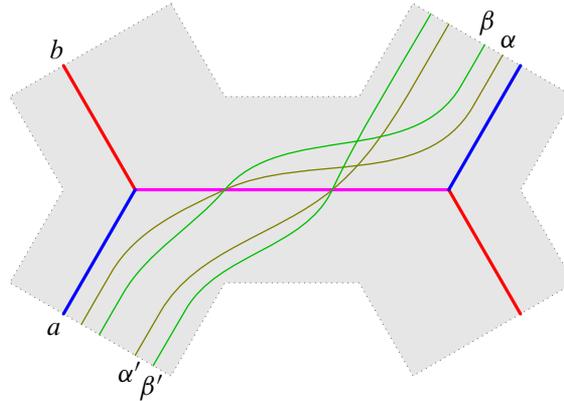
Figure 9: The curves α, β, α' and β'.

Proof If $a \cap c$ belongs to b, then any curve d satisfying $B(a, c, d)$ will meet b. Otherwise, we may push a aside, in order to obtain a curve d, as in Figure 10, left.   □

Putting all together, this yields the following characterization of inessential intersection points between one-sided curves.

Corollary 18 *Let a and b be one-sided curves. Then $a \cap b$ consists of one inessential intersection point if and only if the following conditions are satisfied:*

(1) *a and b are not neighbors in $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$;*

(2) *for every $c \notin \{a, b\}$, there exists d such that d is a neighbor of a and b in $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, and $D(a, d)$ or $D(b, d)$ (or both), but d is not a neighbor of c in that graph;*

(3) *there do not exist α, β, α' and β' such that $B(a, \alpha, \beta)$, $B(b, \alpha, \beta)$, $B(a, \alpha', \beta')$, $B(b, \alpha', \beta')$ and $N(a, \alpha, \beta')$ all hold;*

(4) *there do not exist c_1 and c_2 with $D(c_1, c_2)$ and with the property that, for $i = 1, 2$, we have $T(a, c_i)$ and for all d, $B(a, c_i, d)$ and $D(b, d)$ do not both hold.*

This enumeration of conditions expressed only in terms of the graph structure of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, with the addition of the conditions $\mathrm{One}(a)$ and $\mathrm{One}(b)$, will be also denoted by $I(a, b)$, for *inessential intersection* (of one-sided curves).

Proof First, let us check that if a and b have one inessential intersection point then $I(a, b)$ holds. Condition (1) holds by definition, and (2) follows from Lemma 15. The negation of condition (3) would imply that the cardinality of $a \cap b$ is at least 2. Indeed, $B(a, \alpha, \beta)$ implies that $\alpha \cap \beta$ is a point lying in a. Thus, the bouquet conditions imply that both $\alpha \cap \beta$ and $\alpha' \cap \beta'$ lie in $a \cap b$. And the condition $N(a, \alpha, \beta')$ then implies that $a \cap \alpha$ and $a \cap \beta'$ are disjoint; hence the two points $\alpha \cap \beta$ and $\alpha' \cap \beta'$ are distinct. Finally, condition (4) follows from Lemma 17. Indeed, this lemma implies that the two curves c_1 and c_2 should both contain a point of $a \cap b$; hence they cannot be disjoint.
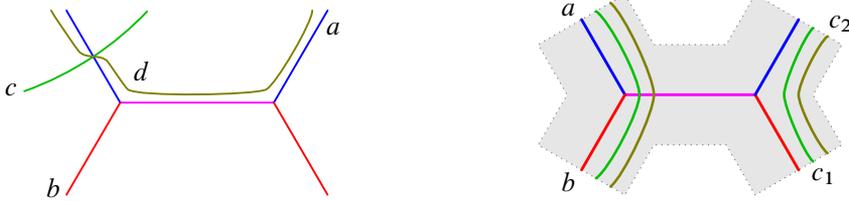
Figure 10: Some configurations of curves for properties I and xB.

Now, let a and b be any two nonseparating curves and suppose that $I(a, b)$ holds. By conditions (1) and (2), the intersection $a \cap b$ is nonempty, and connected. Along the lines of the proof of Lemma 16, we can see that condition (3) implies that $a \neq b$, so $a \cap b$ is an inessential intersection point, or an arc. Suppose for contradiction that it is a nondegenerate arc. By Lemma 16 and condition (3), this intersection arc cannot be essential. Now Figure 10, right, shows the desired contradiction with condition (4). □

Finally, we deal with extra bouquets. We denote by $xB(a, b, c)$ the property that $T(a, b)$, $T(a, c)$ and $I(b, c)$ all hold and, moreover, for all d such that $B(a, b, d)$ holds, $D(c, d)$ does not.

Lemma 19 *Let a, b and c be such that $T(a, b)$, $T(a, c)$ and $I(b, c)$, with b and c one-sided. Then $xB(a, b, c)$ holds if and only if the intersection points $a \cap b$, $a \cap c$ and $b \cap c$ coincide.*

Proof Of course if these points coincide, then property $xB(a, b, c)$ holds; every curve d such that $B(a, b, d)$ holds, must contain this point and hence cannot be disjoint from c.

Now suppose that these points do not coincide, hence, are three pairwise distinct points. Then, we may push b aside, in order to find a curve d which does not intersect c any more, as the intersection $b \cap c$ is not essential. This curve d, obtained by pushing b, can be made to satisfy $T(b, d)$, while crossing b precisely at the point $a \cap b$, and this intersection can be made transverse; the illustration of this situation is similar to Figure 10, left, and this time we leave it to the reader. This yields a curve d such that $B(a, b, d)$ holds and d disjoint from c. □

3 Proof of Theorem 1

Here as above, Σ is a connected surface admitting a nonseparating closed curve.

3.1 From bijections to homeomorphisms

In order to prove Theorem 1, it suffices to prove that every automorphism of $\mathcal{N\!C}^{\dagger}_{\hbar}(\Sigma)$ is supported by a bijection of the surface, in virtue of the following observation.

Proposition 20 *Let $f : \Sigma \to \Sigma$ be a bijection. We suppose that for every nonseparating simple closed curve $\alpha \subset \Sigma$, the sets $f(\alpha)$ and $f^{-1}(\alpha)$ are also nonseparating simple closed curves in Σ. Then f is a homeomorphism.*

Proof If our hypothesis was that f and f^{-1} are closed (ie send closed sets to closed sets), then f would be a homeomorphism. So our strategy is to use our hypothesis here in a similar fashion. We need only prove that f is continuous, the argument for f^{-1} is symmetric.

Let $x \in \Sigma$ and suppose that f is not continuous at x. Then there exists a sequence $(x_n)_{n \geqslant 0}$ of distinct points converging to x, and a neighborhood V of $f(x)$, such that for all n we have $f(x_n) \notin V$. Notice that in the open unit disk of the plane, up to homeomorphism, there is only one sequence of distinct points converging to the origin. With this in mind, we may construct an embedded arc, in Σ, with one end at x, and which contains all the points x_n. Then we may construct a nonseparating simple closed curve α containing this arc. By hypothesis, $f(\alpha)$ is a nonseparating closed curve in Σ, which contains $f(x)$. We may perform a surgery of $f(\alpha)$ inside V, to obtain a nonseparating simple closed curve β, which coincides with $f(\alpha)$ outside V but which does not contain $f(x)$. Now $f^{-1}(\beta)$ is, by hypothesis, a closed subset of Σ, which contains all the points x_n but not x. This is a contradiction. □

3.2 The adjacency relation ⬦

Let $E_T(\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma))$ denote the set of edges $\{a, b\}$ of $\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma)$ satisfying $T(a, b)$. Then we have a map

$$\mathrm{Point} \colon E_T(\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma)) \to \Sigma,$$

which to each edge $\{a, b\}$ associates the intersection point $a \cap b$. The main part of proof of Theorem 1 consists in showing that we can express the equality

$$\mathrm{Point}(a, b) = \mathrm{Point}(\alpha, \beta)$$

in terms of the graph. For this we introduce the equivalence relation ⬦ on $E_T(\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma))$ as follows. If $\{a, b, c\}$ is a 3–clique of $\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma)$ of type bouquet we set $\{a, b\} ⬦ \{a, c\}$. We also set $\{a, b\} ⬦ \{a, c\}$ if a, b and c satisfy the "extra bouquet" condition, denoted above by $xB(a, b, c)$; see Section 2.4 (this is void when Σ is orientable). Then ⬦ is defined as the equivalence relation generated by these relations. In other words, ⬦ is by definition the smallest equivalence relation that connects the pairs $\{a, b\}$ and $\{a, c\}$ whenever $B(a, b, c)$ or $xB(a, b, c)$ holds. We will gradually make this relation more explicit, in this section and in Section 4. When Σ is orientable, the relation ⬦ corresponds to the equivalence relation on triangles, generated by adjacency, in the subgraphs of $\mathcal{NC}^{\dagger}_{\mathrm{\uparrow}}(\Sigma)$ induced by curves passing through a common point. This is what motivates our notation.

The relation $\{a, b\} ⬦ \{a', b'\}$ obviously implies $\mathrm{Point}(a, b) = \mathrm{Point}(a', b')$. We will see that the converse is not true, and describe geometrically the equivalence classes in Section 4, but for now we will only need the following partial statement.

Proposition 21 *Let a, b, a' and b' be such that $T(a, b)$ and $T(a', b')$ hold. Suppose that they have the same intersection point, $x = \mathrm{Point}(a, b) = \mathrm{Point}(a', b')$, and suppose that the germs of a and a' coincide, ie there exists a neighborhood V of x such that $a \cap V = a' \cap V$. Then $\{a, b\} ⬦ \{a', b'\}$.*
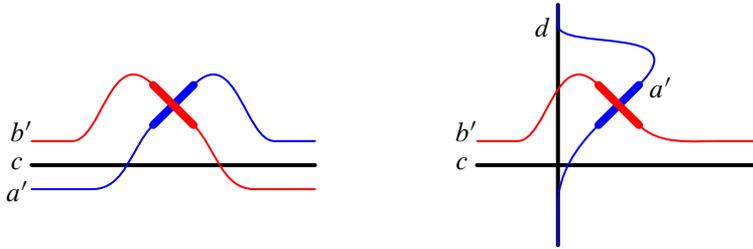
Figure 11: Completing the germs of a and b to form a nonbouquet 3–clique (a', b', c).

Remark 22 If Σ is a Klein bottle, there are no couples $\{a, b\}$ of two-sided curves such that $T(a, b)$ holds, and for every one-sided curve b, the curves c such that $T(b, c)$ holds fall into only two isotopy classes: that of b and that of a two-sided curve, prescribed by b. It follows that, without the extra bouquets in the definition of $\Diamond$, there would have been too many classes of $\Diamond$, as such a class would remember the isotopy class of a one-sided curve, and Proposition 21 would not be true in this special case. These extra bouquets will be used in the proof of Lemma 28 below.

We postpone the proof of the proposition to the end of this section; for now we will explain how it implies Theorem 1.

3.3 Proof of Theorem 1

If a, b and c are vertices of $\mathcal{N}\mathcal{C}^{\dagger}_{\text{⋔}}(\Sigma)$, we denote by $F(a, b, c)$ the property that $T(a, b)$ holds, and there exists an edge $\{a', b'\}$ with $\{a, b\} \Diamond \{a', b'\}$ such that $\{a', b', c\}$ is a 3–clique which is not of type bouquet. Note that this property $F(a, b, c)$ implies that c does not contain the point $\text{Point}(a', b') = \text{Point}(a, b)$. The next lemma asserts that $F(a, b, c)$ actually characterizes this geometric property, and the letter F stands for "$a \cap b$ is far from c".

Lemma 23 *Let a, b and c be vertices of $\mathcal{N}\mathcal{C}^{\dagger}_{\text{⋔}}(\Sigma)$, and suppose $T(a, b)$ holds. Then*

$$F(a, b, c) \iff \text{Point}(a, b) \notin c.$$

Proof The direct implication follows directly from the definitions; we have to prove the converse implication. Suppose $\text{Point}(a, b) \notin c$. Since Σ is connected, there exists a regular neighborhood of c containing the point $\text{Point}(a, b)$. Depending on whether c is one-sided or two-sided, up to homeomorphism, this leads to only two distinct situations. In Figure 11, we represent in bold the germs of the curves a and b near the point $a \cap b$, and show how to complete these germs to new curves a' and b' such that $\{a', b', c\}$ is a 3–clique not of type bouquet. When c is one-sided (see Figure 11, left), we may use two curves a' and b' obtained by pushing c, while when c is two-sided (see Figure 11, right), we have to use a curve d which meets c once transversely.

Now, by Proposition 21, we have $\{a, b\} \Diamond \{a', b'\}$, and $\{a', b', c\}$ is a nonbouquet 3–clique, so we have $F(a, b, c)$ by definition. □

Corollary 24 *Suppose $T(a, b)$ and $T(\alpha, \beta)$ hold. Then $\text{Point}(a, b) \neq \text{Point}(\alpha, \beta)$ if and only if there exists a nonseparating closed curve c such that $F(\alpha, \beta, c)$ holds but not $F(a, b, c)$.*

The corollary is a direct consequence of Lemma 23. It follows that the equality $\text{Point}(a, b) = \text{Point}(\alpha, \beta)$ can be expressed in terms of the graph structure of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, because, as a consequence of Corollary 8, being a 3–clique not of type bouquet is also characterized in terms of this graph structure. Now we can conclude the proof of Theorem 1, provided Proposition 21 holds.

Proof of Theorem 1 Let φ be an automorphism of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$. Given a point x in Σ, we choose two nonseparating simple closed curves a and b intersecting exactly once, transversely, at x, and set $\varphi_{\Sigma}(x) = \text{Point}(\varphi(a), \varphi(b))$. This formula is valid because, by Proposition 4, $\varphi(a)$ and $\varphi(b)$ are still nonseparating simple closed curves intersecting exactly once. The point $\varphi_{\Sigma}(x)$ does not depend on the choice of (a, b), because if (α, β) is another choice, the equalities $\text{Point}(a, b) = \text{Point}(\alpha, \beta)$ and $\text{Point}(\varphi(a), \varphi(b)) = \text{Point}(\varphi(\alpha), \varphi(\beta))$ can be all expressed in terms of the graph structure of $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$, thanks to Corollary 24. Thus, the map φ_{Σ} is well defined, and by following the definitions we observe that the map $(\varphi^{-1})_{\Sigma}$ is its inverse; hence φ_{Σ} is a bijection of Σ. Finally, it follows from Lemma 23 that for any nonseparating simple closed curve α, the curve $\varphi(\alpha)$ coincides with the set of points $\varphi_{\Sigma}(x)$ as x describes α. In other words, the automorphism φ is realized by the bijection φ_{Σ}. Proposition 20 concludes. $\square$

3.4 Connectedness of some arc graphs

In order to finally prove Proposition 21, we will first need a couple of elementary results on fine arc graphs.

Lemma 25 *Let S be a connected topological surface, with boundary, and let x and y be two distinct points of ∂S. Let $\mathcal{EA}^{\dagger}(S, x, y)$ be the graph whose vertices are the simple arcs joining x and y and which meet ∂S only at their ends, with an edge between two such arcs if and only if they are disjoint except at x and y. Then the graph $\mathcal{EA}^{\dagger}(S, x, y)$ is connected.*

Note that, when x and y are taken in the same connected component, we are not requiring that the arcs be nonseparating; this is the reason why we use the letter $\mathcal{E}$, for *extended*, in the same fashion as in [11].

Proof Let a and b be two vertices of this graph. As a first case we suppose that $a \cap b$ is made of a finite number of transverse intersection points; we will prove by induction on the cardinal of $a \cap b$ that, in this case, a and b are connected in $\mathcal{EA}^{\dagger}(S, x, y)$. If $a \cap b$ is as small as possible, ie is equal to $\{x, y\}$, then a and b are neighbors in this graph. Otherwise, if a and b intersect at other points than x and y, we may, in the spirit of [6], pick a *unicorn path* c, made of one subarc of a beginning at x, and one subarc of b ending at y. (For example, we may follow a until it first meets b after x, and then continue along b). Now we may push c aside while fixing its ends, at the appropriate side of c, to obtain a new arc c' with both $c' \cap a$ and $c' \cap b$ of cardinal strictly lower than that of $a \cap b$.

This proves that arcs intersecting at finitely many points are connected in $\mathcal{EA}^\dagger(S, x, y)$.

Now if the intersection $a \cap b$ is infinite, fix a differentiable structure on the surface S. We may consider a smooth curve a', neighbor of a, by pushing a aside while fixing its ends, and similarly, a smooth neighbor b' of b similarly. Up to perturbing b', we may suppose that b' is transverse to a'. By the step above, a' and b' are connected in $\mathcal{EA}^\dagger(S, x, y)$ and the lemma is proved. $\qquad\square$

We will also need a version for nonseparating arcs.

Lemma 26 *Let S be a connected surface containing a nonseparating curve. Suppose S has boundary, and let x and y be two distinct points in the same boundary component of S. Let $\mathcal{NA}^\dagger(S, x, y)$ be the set of nonseparating simple arcs connecting x to y, with an edge when they are disjoint away from x and y. Then $\mathcal{NA}^\dagger(S, x, y)$ is connected.*

The points x and y add some technicality; let us first prove the following simpler statement.

Lemma 27 *Let S be a connected surface containing a nonseparating curve, and with at least one boundary component, denoted by C. Let $\mathcal{NA}^\dagger(S, C)$ be the graph whose vertices are the nonseparating arcs joining two distinct points of C, and with an edge between two such vertices whenever they are disjoint. Then this graph is connected.*

This lemma is a variation on [11, Corollary 3.2]; here we additionally require that the arcs end at C. In fact, [11, Corollary 3.3] is stated for surfaces with $b > 0$ boundary components, but proved only in the case $b = 1$, which is the case needed in the proof of their main theorem. Lemma 27 may be used to extend this corollary to any $b > 0$.

Proof We begin with the observation that the graph $\mathcal{NA}^\dagger(S, C)$ has no isolated point. Indeed, if γ is a vertex of $\mathcal{NA}^\dagger(S, C)$, by definition it is nonseparating. So we may consider a simple closed curve u with one transverse intersection point with γ. This curve u is nonseparating; this follows from Fact 5, applied to u, and a curve v obtained by concatenation of γ with some arc of C. Now we can perform a surgery on u, and push its intersection point towards one end of γ until we hit C. This constructs an arc α, which is now disjoint from γ, and which is also nonseparating.

Next, we claim that we can suppose, without loss of generality, that the surface S is compact. Indeed, if γ_1 and γ_2 are vertices of $\mathcal{NA}^\dagger(S, C)$, and if u is a nonseparating curve intersecting γ_1 as above, consider the set $K = C \cup \gamma_1 \cup \gamma_2 \cup u$. This set is compact; hence there exists a compact topological subsurface S' of S containing K. This surface S' contains nonseparating curves, as it contains u and $\gamma_1 \cup C$, which may be used as above to find two simple closed curves u and v with one essential intersection. Now a path joining γ_1 to γ_2 in S' is also a path joining γ_1 to γ_2 in S. So, until the end of the proof, S is now supposed to be compact.

Next, observe that if two vertices γ_1 and γ_2 of $\mathcal{NA}^\dagger(S, C)$ are isotopic (ie there exists a continuous map $H : [0, 1]^2 \to S$ such that $\gamma_1(t) = H(0, t)$ and $\gamma_2(t) = H(1, t)$ for all t, $H(s, 0), H(s, 1) \in C$ for all s and

the curve $H_s : t \to H(s, t)$ is injective for all s), then γ_1 and γ_2 are in the same component of $\mathcal{NA}^\dagger(S, C)$. This argument is borrowed from [2]: for all s, the arc H_s has at least a neighbor α_s (by the first observation above), and the set of s' such that $H_{s'}$ is still a neighbor of α_s is open in $[0, 1]$. By compactness of $[0, 1]$, there exist a finite number of arcs $\alpha_1, \ldots, \alpha_n$, and a subdivision $0 = t_0 < t_1 < \cdots < t_n = 1$ such that α_j is disjoint from H_t for all $t \in [t_{j-1}, t_j]$ for all j, and now $(\gamma_1, \alpha_1, H_{t_1}, \alpha_2, \ldots, H_{t_{n-1}}, \alpha_n, \gamma_2)$ is a path of $\mathcal{NA}^\dagger(S, C)$ joining γ_1 to γ_2. As a result of this observation, we need only prove the connectedness of the graph $\mathrm{NA}(S, C)$, whose vertices are the isotopy classes of arcs between two distinct points of C, and with an edge between two vertices whenever the corresponding classes admit disjoint representatives.

We proceed with the observation that the graph $\mathrm{A}(S, C)$, defined exactly as $\mathrm{NA}(S, C)$ except we consider essential arcs, which may be separating, is connected. A simple way to do this is by using the idea of unicorn arcs exactly as in the proof of the preceding lemma: if two arcs a and b are in minimal position then their unicorn arcs are essential, and have fewer intersections with both a and b than the number of points of $a \cap b$.

We will promote the connectedness of $\mathrm{A}(S, C)$ to that of $\mathrm{NA}(S, C)$, by induction on the number of boundary components of S.

First, suppose that S has only one boundary component, C. Let γ_1 and γ_2 be two vertices of $\mathrm{NA}(S, C)$. We may connect them by a path $(\gamma_1, \alpha_1, \alpha_2, \ldots, \alpha_n, \gamma_2)$ in $\mathrm{A}(S, C)$, where each of the α_j may be separating; consider such a path with minimal number of separating arcs. For contradiction, and up to some relabeling, suppose α_1 is separating. Then it cuts S in two components; denote by S_1 the one containing γ_1 and S_2 the other. Then α_2 is also contained in S_1, otherwise we may delete α_1 from our path. Since S has no boundary component other than C and since the curve α_1 is essential, the surface S_2 contains a nonseparating arc, α'_1. This arc may be used instead of α_1 in our initial path from γ_1 to γ_2, contradicting the minimality of the number of separating arcs. This proves that $\mathrm{NA}(S, C)$ is connected if S has no other boundary component.

Now, we suppose, for inductive hypothesis, that $\mathrm{NA}(S', C')$ is connected for every surface S' with less boundary components than S. Let γ_1 and γ_2 be two vertices of $\mathrm{NA}(S, C)$. As before, consider a path $(\gamma_1, \alpha_1, \alpha_2, \ldots, \alpha_n, \gamma_2)$ in $\mathrm{A}(S, C)$ between them, with minimal number of separating arcs. For contradiction, and up to some relabeling, suppose α_1 is separating: it cuts S into two subsurfaces; let S_1 be the one containing γ_1, and, by hypothesis, must also contain α_2, and let S_2 be the other. If S_2 contains nonseparating arcs, we conclude as before. If not, then S_2 contains some of the boundary components of S; hence the surface with boundary $S' = S_1 \cup \alpha_1$ has strictly less boundary components than S. One is C', composed by an arc of C and the arc α_1, and there may be others.

If α_2 is nonseparating, then, by the induction hypothesis, there is a path $(\gamma_1, \beta_1, \ldots, \beta_k, \alpha_2)$ of $\mathrm{NA}(S', C')$ connected them. The arcs $\beta_1, \ldots, \beta_k$ may have end points in α_1, but we may perform a surgery in order to push all these points to C, and obtain arcs $\beta'_1, \ldots, \beta'_k$ which are also vertices of $\mathrm{NA}(S, C)$, and we are done in this case.

Finally, if α_2 is a separating arc (of S, or of S_1, equivalently), then we may find an arc α'_2 of S_1 which is nonseparating and disjoint from α_2. By following the last case above, there exists a path $(\gamma_1, \beta_1, \ldots, \beta_k, \alpha'_2)$ in NA(S, C); hence the path $(\gamma_1, \beta_1, \ldots, \beta_k, \alpha'_2, \alpha_2, \ldots, \alpha_n, \gamma_2)$ of A(S, C) has one less separating arc than the initial path. This contradiction ends the proof. $\square$

Proof of Lemma 26 Let γ_1 and γ_2 be two vertices of $\mathcal{NA}^\dagger(S, x, y)$. First, we may construct a neighbor γ'_2 in $\mathcal{NA}^\dagger(S, x, y)$ of γ_2, which, in a neighborhood of x (resp. y), touches γ_1 only at x (resp. y).

Indeed, there is a neighborhood U_x of x homeomorphic to the closed half unit disk

$$\{z \mid |z| \leqslant 1 \text{ and } \operatorname{Im}(z) \geqslant 0\},$$

where the middle ray ($\operatorname{Re}(z) = 0$) corresponds to the points of γ_1. On either side of this ray, we may find an arc disjoint from γ_1 and γ_2 except at 0, arbitrarily close to the boundary ($\operatorname{Im}(z) = 0$), and joining 0 to the unit circle, and then this small arc may be continued to construct a curve γ'_2 which consists of pushing γ_2 aside.

So we may suppose that γ_1 and γ_2, close to x and y, intersect only at these points, and we may now find neighborhoods U_x and U_y as above, such that their intersections with γ_1 and γ_2 are along rays in this disk, in distinct directions around 0. Let S' be the surface obtained by removing the interiors of U_x and U_y from S. Then the path given by applying Lemma 27 to S, yields a path from γ_1 to γ_2 in $\mathcal{NA}^\dagger(S, x, y)$, just by adding some rays in U_x and U_y to the corresponding arcs. $\square$

3.5 Proof of Proposition 21

Let us go back to the proof of Proposition 21. For the remainder of the section we fix a point $x \in \Sigma$. Let X denote the set of nonseparating simple closed curves passing through x.

Lemma 28 *Let* $a, b, c \in X$. *Suppose that* $T(a, b)$ *and* $T(a, c)$ *hold. Then* $(a, b) \Diamond (a, c)$.

Proof Let S be the surface obtained by cutting Σ along a; it is the surface with boundary obtained by gluing back two copies of the curve a to $\Sigma \smallsetminus a$. The point x of Σ yields two points, p and q, of ∂S, and the curves b and c define two arcs of S joining p and q. By Lemma 25, there exists a finite sequence $\gamma_0 = b, \ldots, \gamma_n = c$, of arcs of S joining p and q, with γ_i and γ_{i+1} disjoint except at p and q. For each i, the arc γ_i defines a closed curve in Σ, which has precisely one, transverse intersection with a; we will still denote it by γ_i, abusively.

For every i, if $T(\gamma_i, \gamma_{i+1})$ holds, then we have $(a, \gamma_i) \Diamond (a, \gamma_{i+1})$, by definition. If $T(\gamma_i, \gamma_{i+1})$ does not hold, then either γ_i or γ_{i+1} are both one-sided, or one of them is two-sided. In the first case, the condition $x B(a, \gamma_i, \gamma_{i+1})$ holds, by definition, and hence $(a, \gamma_i) \Diamond (a, \gamma_{i+1})$. In the second, up to reversing the notation suppose γ_i is two-sided. Figure 12 shows how to insert a curve δ such that $B(a, \delta, \gamma_i)$ and $B(a, \delta, \gamma_{i+1})$ both hold, and hence we still have $(a, \gamma_i) \Diamond (a, \gamma_{i+1})$ in this case.

By transitivity, we deduce that $(a, b) \Diamond (a, c)$. $\square$

Figure 12: Connecting the curves by common adjacency. The left shows the case when a is one-sided, and the right is when a is two-sided

The last ingredient for the proof of Proposition 21 is the following observation.

Observation 29 *Let a and a' be two nonseparating simple closed curves in Σ such that $a \cap a'$ is an arc. Then both sides of this arc lie in the same connected component of $\Sigma \smallsetminus (a \cup a')$.*

Proof A priori, the complement of $\Sigma \smallsetminus (a \cup a')$ may have up to four connected components, as suggested in Figure 13. Suppose first that the intersection $a \cap a'$ is essential. If a (resp. a') is one-sided, by following the curve a (resp. a') we see that $A = B$. If both a and a' are two-sided, by following a we see that $A = D$ and $C = B$, while by following a' we get $A = C$ and $B = D$, so $A = B$.

Now, suppose the intersection arc $a \cap a'$ is inessential. By following a, we see that $C = D$, regardless of a being one or two-sided. Thus, if $A \neq B$, then one of A or B, say A, is not connected from any of B, C or D. But this implies that a is separating, a contradiction.   □

We are now in a position to prove Proposition 21, but instead we will prove the following stronger statement, which will be more convenient later in this article.

Proposition 30 *Let a, b, a' and b' be such that $T(a, b)$ and $T(a', b')$ hold, with intersection point $x = \mathrm{Point}(a, b) = \mathrm{Point}(a', b')$, and suppose that a and a' locally "half coincide" near x, ie $a \cap a'$ contains a nondegenerate arc with endpoint x. Then $\{a, b\} \pitchfork \{a', b'\}$.*

Proof Suppose first that a and a' coincide along some arc with x as an endpoint, and are disjoint apart from this arc. By Observation 29, there exists a curve d passing through x such that $T(a, d)$ and $T(a', d)$. By Lemma 28, this implies $(a, d) \pitchfork (a', d)$, and by the same lemma we also have $(a, b) \pitchfork (a, d)$ and $(a', b') \pitchfork (a', d)$. Hence $(a, b) \pitchfork (a', b')$.
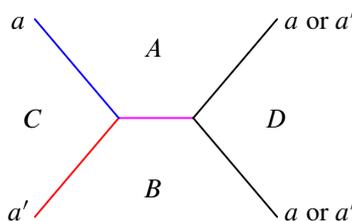

Figure 13: The arc $a \cap a'$ cannot disconnect.

Now we do not make the assumption any more that a and a' meet only along an arc. Still, thanks to the hypothesis of the proposition, we may choose a set V homeomorphic to a closed disk, with x on its boundary, and such that $a \cap V = a' \cap V$ is an arc whose endpoints are x and some other point y. Lemma 26, applied to the surface $\Sigma \smallsetminus \mathring{V}$, provides a sequence $a_0 = a, \ldots, a_n = a'$, of nonseparating curves such that for all i, the curves a_i and a_{i+1} intersect only along the arc $a \cap V$; hence we may conclude by applying iteratively the reasoning above. $\qquad\square$

4 Local subgraphs

In Section 3.2, we considered edges $\{a, b\}$ and $\{a', b'\}$ in the graph $\mathcal{NC}^{\dagger}_{\text{ⴔ}}(\Sigma)$ satisfying

$$|a \cap b| = |a' \cap b'| = 1,$$

and $\mathrm{Point}(a, b) = \mathrm{Point}(a', b')$. We defined and used the equivalence relation $\Diamond$. The aim of this section is to provide a geometric interpretation of the equivalence classes. The results here are not used anywhere else in the paper. In particular, this section is not used in the proof of our main results. Nevertheless, we think it may help the reader to get a clear picture of the situation.

4.1 The graph of germs

Let x be a marked point in the surface Σ. In this section the we will study the local geometry of curves near x, so we may assume that $(\Sigma, x) = (\mathbb{R}^2, 0)$ whenever this is convenient. Given two simple arcs $a, a' \colon [0, 1] \to \Sigma$ with $a(0) = a'(0) = x$, we say that a and a' *locally coincide* at x if there exists a neighborhood V of x such that $a([0, 1]) \cap V = a'([0, 1]) \cap V$. This is an equivalence relation, whose equivalence classes are called *germs of simple arcs at x*. The germ of a is denoted $[a]_x$. We say that a and a' *locally intersect only at x* if there exists a neighborhood V of x such that $a([0, 1]) \cap a'([0, 1]) \cap V = \{x\}$. This second relation depends only on the germs of a and a', and thus induces a relation on germs. Let us consider the graph $\mathcal{A}(x)$ whose vertices are the germs of simple arcs at x, with an edge between the germs of a and a' whenever a and a' locally intersect only at x.

This graph is *not* connected, in fact it has infinitely (uncountably) many connected components, as we will see below. We postpone the description of the connected components to explain the relation with the adjacency relation defined in Section 3.2. We say that two vertices α and α' of the graph $\mathcal{A}(x)$ are *comparable* if they belong to the same connected component of the graph.

4.2 Germs and adjacency

Given a point x in Σ and a simple closed curve a in Σ that contains x, we choose any one of the two germs of simple arc at x included in a and denote it by $\lfloor a \rfloor_x$. Which one of the two germs is chosen will not matter in what follows.

Proposition 31 *Let a, b, a' and b' be vertices in $\mathcal{NC}_{\pitchfork}^{\dagger}(\Sigma)$ such that $T(a, b)$ and $T(a', b')$ hold, and assume $\mathrm{Point}(a, b) = \mathrm{Point}(a', b')$. Then $\{a, b\} \Diamond \{a', b'\}$ if and only if the germs $\lfloor a \rfloor_x$ and $\lfloor a' \rfloor_x$ are comparable.*

Proposition 31 will be proved in Section 4.5 below. The aim of the next two sections is to provide a simple characterization of distance, and connected components, in the graph of germs; see Proposition 33 below.

4.3 Distance in local subgraphs

In this section, we give a geometric interpretation of the distance in three different graphs, which are very much like the graph of germs $\mathscr{A}(x)$.

Let Σ be one of the following surfaces:

(1)   the compact annulus $\mathbb{S}^1 \times [0, 1]$,

(2)   the open annulus $\mathbb{S}^1 \times \mathbb{R}$, or

(3)   the 2–torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$.

We consider nonoriented simple arcs in Σ; more precisely, simple curves connecting both sides of the annulus in case (1), properly embedded images of the real line connecting both ends of the open annulus in case (2), or simple closed curves in a fixed homotopy class, say homotopic to $\{0\} \times \mathbb{S}^1$ in case (3). Let $\mathscr{A}$ denote the graph whose vertices are one of the three above family of curves, with an edge between two curves whenever they are disjoint.

In order to express geometrically the distance in $\mathscr{A}$, let us consider the cyclic cover $p \colon \widetilde{\Sigma} \to \Sigma$, respectively in cases (1), (2) and (3),

$$p \colon \mathbb{R} \times [0, 1] \to \mathbb{R}/\mathbb{Z} \times [0, 1], \quad p \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}/\mathbb{Z} \times \mathbb{R}, \quad p \colon \mathbb{R} \times \mathbb{S}^1 \to \mathbb{R}/\mathbb{Z} \times \mathbb{S}^1$$

given by the formula $p(x, y) = (x \bmod 1, y)$. Let T be the deck transformation $(x, y) \to (x + 1, y)$.

Now consider two curves a, b which are vertices of the graph $\mathscr{A}$. Let $\tilde{a}$ and $\tilde{b}$ be respective lifts of a and b under the covering map p. Note that the set

$$\{k \in \mathbb{Z} \mid T^k(\tilde{a}) \cap \tilde{b} \neq \varnothing\}$$

is an interval of $\mathbb{Z}$, which is finite in the compact cases (1) and (3) but may be infinite in the open annulus case (2). We define the *relative width* $\mathrm{Width}(a, b)$ as the cardinal of this set. This is an element of $\{0, 1, \ldots, +\infty\}$. The reader may check easily that $\mathrm{Width}(a, b) = \mathrm{Width}(b, a)$.

Proposition 32 *For every vertices $a \neq b$ of the graph $\mathscr{A}$, the distance in the graph is given by*

$$d(a, b) = \mathrm{Width}(a, b) + 1.$$

In cases (1) and (3), the graph $\mathscr{A}$ is connected. In case (2), a and b are in the same connected component of $\mathscr{A}$ if and only if $\mathrm{Width}(a, b) < +\infty$.

Proof Let a and b be as in the statement, and denote $w = \mathrm{Width}(a, b)$. We first assume that $w < +\infty$. By Schoenflies' theorem (in case (2), applied in the two-point compactification of the annulus, which is a sphere), we may assume that a is a vertical curve whenever this makes our life easier. We first note that if $w = 0$ then a and b admit lifts that are disjoint from every T–translate of each other, which shows that a and b are disjoint, and thus $d(a, b) = 1$. Let us now assume $w > 0$, and prove the two following key properties.

(i) For every vertex a' of $\mathscr{A}$ such that $d(a, a') = 1$,
$$\mathrm{Width}(a', b) \geq w - 1.$$

(ii) There exists a vertex a' of $\mathscr{A}$ such that $d(a, a') = 1$ and
$$\mathrm{Width}(a', b) \leq w - 1.$$

To prove the first property, consider a' such that $d(a, a') = 1$. By definition of the width w, we may find lifts $\tilde{a}$ and $\tilde{b}$ of a and b such that $\tilde{b}$ is disjoint from $\tilde{a}$ and $T^{w+1}\tilde{a}$ but meets $T(\tilde{a}), \ldots, T^w(\tilde{a})$. Since a and a' are disjoint, there is a lift $\tilde{a}'$ of a' which is between $\tilde{a}$ and $T(\tilde{a})$. Then the curves
$$T(\tilde{a}'), \ldots, T^{w-1}(\tilde{a}')$$
are between the two curves $T(\tilde{a})$ and $T^w(\tilde{a})$, and those two curves are not in the same connected component of $\widetilde{\Sigma} \smallsetminus T^i(\tilde{a}')$, for $i = 1, \ldots, w - 1$. Since the curve $\tilde{b}$ is connected and meets the two curves $T(\tilde{a})$ and $T^w(\tilde{a})$, it must meet all the $T^i(\tilde{a}')$. This proves that $\mathrm{Width}(a', b) \geq w - 1$.

Let us prove the second property. We consider $\tilde{a}$ and $\tilde{b}$ as above. Let S denote the compact strip or annulus bounded by $\tilde{a} \cup T(\tilde{a})$. Remember that $\tilde{b}$ meets $T(\tilde{a})$ but not $\tilde{a}$. Thus $\tilde{b} \cap S$ is included in a (maybe infinite) family of *bigons*, ie topological disks bounded by a simple closed curve made of a segment of the curve $T(\tilde{a})$ and a segment of the curve $\tilde{b}$. Let S^+ denote the union of these bigons. Symmetrically, the curve $\tilde{b}' := T^{-w}(\tilde{b})$ meets $\tilde{a}$ but not $T(\tilde{a})$. Thus $T^{-w}(\tilde{b}) \cap S$ is included in a union S^- of bigons formed by the curves $\tilde{a}$ and $\tilde{b}'$. A key point is that the sets S^- and S^+ are disjoint, because the curves $\tilde{b}$ and $\tilde{b}'$ are disjoint, since b is simple. Thus we may construct a homeomorphism H supported in S such that $H(S^-)$ is included in an arbitrarily small neighborhood of $\tilde{a}$, and $H(S^+)$ is included in an arbitrarily small neighborhood of $T(\tilde{a})$. In particular, we may find a curve $\tilde{a}'$, which is a lift of some element a' of $\mathscr{A}$, included in the interior of S and disjoint from both S^- and S^+ (to be more explicit, take $\tilde{a}' = H^{-1}(\{1/2\} \times [0, 1])$ in the annulus case, in coordinates for which a is the vertical curve $\{0\} \times [0, 1]$). Note that $\tilde{a}'$ is disjoint from $\tilde{b}$ and $T^{-w}(\tilde{b})$, and separate both curves, ie the first one is on the right-hand side of $\tilde{a}'$, and the second one is on the left-hand side. Thus the set
$$\{k \in \mathbb{Z} \mid T^k(\tilde{b}) \cap \tilde{a}' \neq \varnothing\}$$
has cardinality at most $w - 1$. Which proves that $\mathrm{Width}(a', b) \leq w - 1$, as wanted.

Using (i) and (ii), an induction on n shows that $d(a, b) = n$ if and only if $\mathrm{Width}(a, b) + 1 = n$, which completes the proof in the case when $\mathrm{Width}(a, b)$ is finite. When $\mathrm{Width}(a, b) = +\infty$, an argument

analogous to property (i) above shows that $\mathrm{Width}(a', b) = +\infty$ for every a' such that $d(a, a') = 1$. This shows that a and b are not in the same connected component of the graph. □

4.4 Distance in the graph of germs

Let us go back to the graph of germs $\mathscr{A}(x)$. Assume $(\Sigma, x) = (\mathbb{R}^2, 0)$. Given two vertices a and b of $\mathscr{A}(x)$, let us define their *local relative width* $\mathrm{Width}(a, b)$ as the number of turns b does around x in arbitrary small neighborhoods of x, in a coordinate system in which a goes straight to x. More precisely, the plane minus the origin is identified with the open annulus $\mathbb{S}^1 \times \mathbb{R}$, and we consider the graph $\mathscr{A}$ from the previous section in the open annulus case. Then $\mathrm{Width}(a, b)$ is defined as the infimum of the quantity $\mathrm{Width}(A, B)$, where A and B are vertices of $\mathscr{A}$ whose germs respectively equal a and b. Here is a more practical definition, which is easily seen to be equivalent. Consider the universal cover $p \colon \widetilde{\Sigma} \to \Sigma$ as above. Abuse the definition by still denoting $a, b \colon [0, 1] \to \Sigma$ two curves with $a(0) = b(0) = 0$ whose germs respectively equal a and b. Let $\tilde{a}$ and $\tilde{b}$ denote lifts of (the restrictions to $(0, 1]$ of) a and b in $\widetilde{\Sigma}$. Then the number $\mathrm{Width}(a, b) = w$ is characterized by the two following properties:

   (i)   for every $t_0 \in (0, 1]$, the restriction of $\tilde{a}$ to $(0, t_0]$ meets at least w integer translates of $\tilde{b}$;

   (ii) there exists $t_0 \in (0, 1]$ such that the restriction of $\tilde{a}$ to $(0, t_0]$ meets exactly w integer translates of $\tilde{b}$;

Analogously to the previous section, the distance in the graph of germs is characterized by the local relative width.

Proposition 33 *Let $a \neq b$ be two vertices of the graph $\mathscr{A}(x)$. Then a and b are in the same connected component of $\mathscr{A}(x)$ if and only if $\mathrm{Width}(a, b) < +\infty$. In this case, the distance in the graph is given by*

$$d(a, b) = \mathrm{Width}(a, b) + 1.$$

The proof is very similar to the proof in the previous section. Details are left to the reader.

4.5 Proof of Proposition 31

Let a, b, a' and b' be vertices of $\mathscr{N}\mathscr{C}^{\dagger}_{\mathrm{fh}}(\Sigma)$ such that $T(a, b)$ and $T(a', b')$ hold, and assume that $\mathrm{Point}(a, b) = \mathrm{Point}(a', b')$. Denote by x the common intersection point.

If c is another vertex such that $\{a, b, c\}$ is a 3–clique of type bouquet or extra bouquet, then the germs $\lfloor a \rfloor_x$ and $\lfloor c \rfloor_x$ are disjoint, thus comparable. This entails the direct implication in Proposition 31.

Let us prove the converse implication. We assume that the germs $\lfloor a \rfloor_x$ and $\lfloor a' \rfloor_x$ are comparable. In other words, there exists arcs $\alpha_0, \ldots, \alpha_n$ with $\alpha_i(0) = x$ and whose sequence of corresponding germs is a path from $\lfloor a \rfloor_x$ to $\lfloor a' \rfloor_x$ in the graph of germs. Note that each germ α_i may be extended to a nonseparating curve a_i, and we can find another nonseparating curve b_i such that $T(a_i, b_i)$ holds. Thus the end of the proof is a direct consequence of the following lemma.

Lemma 34 *Let a, b, a' and b' be vertices in $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$ such that $T(a,b)$ and $T(a',b')$ hold. Assume that for some choices $\lfloor a \rfloor_x$ and $\lfloor a' \rfloor_x$ of arcs at x included respectively in a and a', the germs $\lfloor a \rfloor_x$ and $\lfloor a' \rfloor_x$ intersect only at x. Then $\{a,b\} \Diamond \{a',b'\}$.*

Proof Let c be an arc that contains x in its interior and locally coincides with $\lfloor a \rfloor_x \cup \lfloor a' \rfloor_x$. Extend c into a nonseparating closed curve, still denoted by c, and consider any other nonseparating curve d such that $T(c,d)$ holds. Since c locally "half coincides" near x with both a and a', we may apply [Proposition 30](#) twice, and get that $\{a,b\} \Diamond \{c,d\} \Diamond \{a',b'\}$. $\qquad\qquad\qquad\square$

4.6 Curves and diffeomorphisms

In this short subsection we explain how one can use the fine curve graph to detect fundamental nondifferentiability.

Let Φ be an automorphism of $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$. We introduce the following property $D(\Phi)$:

For all vertices a and b of $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$ such that $T(a,b)$ holds, if $\mathrm{Point}(\Phi(a),\Phi(b)) = \mathrm{Point}(a,b)$ then there exists a' and b' such that $T(a',b')$ holds, $\mathrm{Point}(a',b') = \mathrm{Point}(a,b)$ and $\{\Phi(a'),\Phi(b')\} \Diamond \{a',b'\}$.

Note that this property is invariant under conjugacy in the group of automorphisms; indeed, according to [Lemma 23](#), this property is defined entirely in terms of the graph structure of $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$. Let h be a homeomorphism of Σ, and denote $\Phi = \Phi_h$ the action of h on the graph $\mathcal{N}\mathcal{C}_{\pitchfork}^{\dagger}(\Sigma)$.

Observation 35 *If h is differentiable everywhere, then property $D(\Phi_h)$ holds.*

Indeed, the hypothesis $\mathrm{Point}(\Phi(a),\Phi(b)) = \mathrm{Point}(a,b)$ is equivalent to the fact that $x = \mathrm{Point}(a,b)$ is a fixed point of h. Since h is differentiable at x, it is easy to check that the germ of any smooth arc at x is comparable to its image. Take any two smooth curves a' and b' such that $T(a',b')$ and $\mathrm{Point}(a',b') = \mathrm{Point}(a,b)$; then [Proposition 31](#) tells us that $\{\Phi(a'),\Phi(b')\} \Diamond \{a',b'\}$.

Now consider a particular homeomorphism h of Σ and assume that h admits a fixed point where, for some local polar coordinates, h is defined by

$$(r,\theta) \mapsto \left(r, \theta + \frac{1}{r}\right).$$

Observation 36 *Property $D(\Phi_h)$ does not hold.*

An easy proof of this is obtained by considering the *local rotation interval* of h at x, as defined in [9, Section 2.3]. Indeed, the local rotation interval of h at x equals $\{+\infty\}$, which accounts for the fact that orbits turn faster and faster around x, in the positive direction, as we get nearer and nearer to x (the quickest way to check this is to show that the *local rotation set* of h at x is $\{+\infty\}$, and then to apply [9, théorème 3.9] that relates the local rotation set and the local rotation interval). We argue by contradiction to show that property $D(\Phi_h)$ does not hold. Assuming property $D(\Phi_h)$ holds, consider curves a and b such that $T(a,b)$ holds and $\mathrm{Point}(a,b) = x$. Let a' and b' be given by property $D(\Phi_h)$, such that

$\{\Phi(a'), \Phi(b')\} \Diamond \{a', b'\}$. The reverse direction of Proposition 31 tells us that the germs of $h(a')$ and a' are comparable at x. This entails easily, from the definition, that the local rotation interval of h at s is a bounded interval, a contradiction.

5 Fine graph of smooth curves

In this section we address the case of smooth curves, and prove Theorem 2 and Proposition 3. Throughout the section, Σ will be a connected, nonspherical surface without boundary, endowed with a smooth structure. In Section 5.3 we will restrict to the orientable case.

5.1 From bijections to higher regularity

One step in the proof of Theorem 1 was Proposition 20, in which we proved that if an automorphism of $\mathcal{NC}^{\dagger}_{\mathrm{fh}}(\Sigma)$ is supported by a bijection of Σ, then that bijection is a homeomorphism of Σ.

We may ask the same question about automorphisms of $\mathcal{NC}^{\dagger\infty}_{\mathrm{fh}}(\Sigma)$, and this paragraph is devoted to the proof of the following two statements. We denote by $\mathrm{Homeo}_{\infty\mathrm{fh}}(\Sigma)$ the group of bijections of Σ which preserve the family of smooth, nonseparating closed curves, and preserve transversality between such curves. The first statement below justifies this notation. Here, for simplicity we restrict to the case of orientable surfaces.

Proposition 37 *Let Σ be a connected, nonspherical orientable surface. The group $\mathrm{Homeo}_{\infty\mathrm{fh}}(\Sigma)$ is contained in $\mathrm{Homeo}(\Sigma)$.*

Proof Let $h \in \mathrm{Homeo}_{\infty\mathrm{fh}}(\Sigma)$. We will prove that the image under h of any open set is an open set. This is the continuity of h^{-1}, and by applying the argument to h we also get the continuity of h.

To do this, we only need to consider the images of a family of sets that generates the topology. Given three nonseparating curves a, b and c, we denote by $V(a; b, c)$ the union of all the nonseparating curves d that meet a and are disjoint from b and c.

Observation 38 *The set $V(a; b, c)$ is the union of some of the connected components of the complement of $b \cup c$ that meet a. In particular, it is an open set.*

Indeed, let x be a point of $V(a; b, c)$. By definition there is a nonseparating curve d passing through x and meeting a but not b nor c. Consider another point y that belongs to the connected component V_x of the complement of $b \cup c$ that contains x. By modifying d using an arc connecting x to y in V_x, we find another curve d', isotopic to d, still meeting a but not b nor c, and passing through y. This proves that $V(a; b, c)$ contains V_x, and the observation follows.

Now let a be a nonseparating curve. Let a^+ and a^- be obtained by pushing a to both sides. Then $V(a; a^+, a^-)$ is a neighborhood of a, and by making a^+ and a^- vary we get a basis of neighborhoods $\mathcal{B}(a)$ of the curve a. The union of all these families $\mathcal{B}(a)$ generates the topology of Σ.

Thus it suffices to check that the image under h of each set $V(a; b, c)$ is an open set. But since h is a bijection, we have

$$h(V(a; b, c)) = V(h(a); h(b), h(c)).$$

By hypothesis, $h(a)$, $h(b)$ and $h(c)$ are nonseparating closed curves, and by the observation this set is open. $\qquad\square$

We now prove Proposition 3 stated in the introduction, namely the existence of elements of $\mathrm{Homeo}_{\infty\textrm{h}}(\Sigma)$ that are not smooth.

Proof of Proposition 3 We will construct a homeomorphism $F\colon \mathbb{R}^2 \to \mathbb{R}^2$, which is not differentiable at the origin, but such that both F and F^{-1} send smooth curves to smooth curves. The construction can easily be modified to make F compactly supported, and then be transported on our surface Σ. It will be clear from the construction that this map preserves transversality.

Let $h\colon \mathbb{R} \to \mathbb{R}$ be a smooth diffeomorphism supported in the segment $[1/2, 2]$. That is to say, $h(x) = x$ for all x outside $[1/2, 2]$; we suppose however that $h(1) \neq 1$. We consider the map F defined by $F(x, y) = (x, xh(y/x))$ if $x \neq 0$, and $F(x, y) = (x, y)$ otherwise. We claim that this map has the desired property.

This map, as well as its inverse, is obviously smooth in restriction to $\mathbb{R}^2 \smallsetminus \{(0, 0)\}$. Direct computation shows that F has directional derivatives in all directions around the origin, but the "differential" fails to be linear: both partial derivatives are those of the identity, while the directional derivative in the direction $(1, 1)$ is not. So F is not differentiable at the origin.

Now, let $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ be a smooth, proper embedding. If $(0, 0)$ is not in the image of γ, then of course, $F \circ \gamma$ is still smooth. So suppose, say, that $\gamma(0) = (0, 0)$. If $\gamma'(0)$ lies outside the two (opposite) sectors of vectors of slopes between $1/2$ and 2, then $F \circ \gamma$ and γ have the same germ at 0. Otherwise, and up to reparametrization, we can write, near 0, $\gamma(t) = (t, \alpha(t))$ where α is a smooth map (satisfying $\alpha(0) = 0$)), from a neighborhood of 0, to $\mathbb{R}$. This yields the formula

$$F \circ \gamma(t) = \left(t, th\left(\frac{\alpha(t)}{t}\right)\right).$$

Now, the smoothness of $F \circ \gamma$ follows from the following elementary observation.

Claim *Let $\alpha\colon \mathbb{R} \to \mathbb{R}$ be a smooth map satisfying $\alpha(0) = 0$. Then the map $t \mapsto \alpha(t)/t$ when $t \neq 0$, and $\alpha'(0)$ when $t = 0$, is smooth.*

Indeed, by the fundamental theorem of calculus, for all $t \in \mathbb{R}^*$ we have

$$\frac{\alpha(t)}{t} = \int_0^1 \alpha'(ts)ds,$$

and this integral with parameter can be differentiated indefinitely.[1] $\qquad\square$

[1] We borrow this elegant argument from [4].
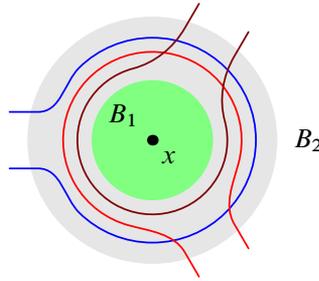
Figure 14: Three bottles. Any arc from B_1 to the outside of B_2 and disjoint from d_1 must meet d_2 and d_3.

5.2 A weak convergence for sequences of curves

In order to prove Theorem 2, we now explain how to recognize configurations of smooth curves. Given two vertices a and b of the graph $\mathcal{NC}_{\pitchfork}^{\dagger\infty}(\Sigma)$, we will denote by $a - b$ the property that they are neighbors in the graph. If $(f_n)_{n\in\mathbb{N}}$ is a sequence of vertices, we denote by $(f_n)_{n\in\mathbb{N}} - a$ the property that for all n large enough, $f_n - a$. The first property of sequences we may recover from the graph is the distinction of what curves go to infinity.

Lemma 39 *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of vertices of $\mathcal{NC}_{\pitchfork}^{\dagger\infty}(\Sigma)$. The following are equivalent*:

- *for all d, we have $(f_n)_{n\in\mathbb{N}} - d$;*
- *for every compact subset K of Σ, for every n large enough, $K \cap f_n = \varnothing$.*

Proof The second statement implies the first, as we may just take $K = d$. Let us prove the converse implication by contraposition. Suppose K intersects infinitely many f_n. Since K is compact, there is a point $x \in K$, such that every neighborhood of x intersects infinitely many f_n. We consider two open sets B_1 and B_2 with $x \in B_1$ and $\bar{B}_1 \subset B_2$, and three bottle-shaped arcs, as in Figure 14. These arcs may be continued to form three nonseparating closed curves, d_1, d_2 and d_3. Now, let n be such that f_n enters B_1. If f_n does not enter nor leave B_2 through the neck of the bottle corresponding to d_1, then we cannot have $f_n - d_1$, since f_n and d_1 have to intersect at least twice. Hence, f_n passes through the neck of d_1, and in order to impose that $f_n - d_2$, another arc of f_n has to get out of the bottle corresponding of d_2 through its neck. But then f_n has to meet d_3 twice, and we cannot have $f_n - d_3$. In other words, for all n such that f_n enters B_1, we can't have $f_n - d_1$ and $f_n - d_2$ and $f_n - d_3$, hence the first statement is not true, and our implication is proved. $\qquad\square$

Thus, we will say here that a sequence $(f_n)_{n\in\mathbb{N}}$ is *relevant* if it has no subsequence $(f_{\varphi(n)})_{n\in\mathbb{N}}$ such that for all d, $(f_{\varphi(n)})_{n\in\mathbb{N}} - d$. We now explore, for such sequences, the following notion of convergence. We say that a relevant sequence (f_n) *converges in a weak sense* to a curve a if for every d such that $a - d$, we have $(f_n) - d$. We denote this property by $W((f_n), a)$.

Lemma 40 *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of vertices, and a be a vertex of $\mathcal{NC}_{\text{⑪}}^{\dagger\infty}(\Sigma)$.*

- *If $W((f_n), a)$, then the sequence $(f_n)_{n\in\mathbb{N}}$ converges in the Hausdorff topology to a; for every neighborhood V of the curve a, for all n large enough, we have $f_n \subset V$.*

- *If the sequence $(f_n)_{n\in\mathbb{N}}$ of curves, with some appropriate parametrization, converges in the C^1–topology to a, then $W((f_n)_{n\in\mathbb{N}}, a)$.*

Proof We first prove the first point. Let us first mention that the statement we wrote is indeed equivalent to the Hausdorff convergence, because any essential curve in a small enough neighborhood of a must pass close to every point of a, and thus is Hausdorff-close to a. Suppose for contradiction that for some neighborhood V of a, we have $f_n \not\subset V$ infinitely often. Then we may find a point x, not in $\overline{V}$, such that every neighborhood of x is visited by infinitely many f_n. We may construct three bottle-shaped arcs around x exactly as in the proof of the preceding lemma, and complete these arcs to nonseparating simple closed curves d_1, d_2 and d_3, which can be requested to satisfy $a - d_i$ for $i = 1, 2, 3$. Then we cannot have $f_n - d_i$ for all $i \in \{1, 2, 3\}$ for the same reason as in this preceding proof, and this contradicts the hypothesis that $W((f_n), a)$ holds.

The second point is the well known stability of transversality in the C^1–topology. $\square$

Remark 41 In fact, the condition $W((f_n), a)$ implies C^0–convergence, in the following sense. Given a parametrization α of a, we can choose the parametrizations of the f_n's yielding a sequence of parametrized curves converging uniformly to α.

As we will not use this fact, we only sketch a quick argument. Let V be a small tubular neighborhood of a, and let $d_1, \ldots, d_N$ be simple closed nonseparating curves, each meeting a transversely at one point, and cutting V in small chunks $V_1, \ldots, V_N$ that are met by α in that cyclic order. Let n be large enough that $f_n \subset V$ and $f_n - d_i$ for each i. Then $f_n \cap V_i$ is connected, and f_n visits the pieces $V_1, \ldots, V_N$ in that order. Hence we may choose a parametrization of f_n, say F_n, in such a way that for all $i \in \{1, \ldots, N\}$ and all t, $F_n(t) \in V_i$ if and only if $\alpha(t) \in V_i$. This implies that F_n is uniformly close to α.

This notion is actually somewhere strictly in between C^0–convergence and C^1–convergence, as we remark in the following example. This construction will play a crucial role below in the proof of Theorem 2.

Example 42 Let a be a smooth nonseparating curve in Σ. We choose a point p in a, and a chart around one of its points, diffeomorphic to $\mathbb{R}^2$, in such a way that a corresponds to the axis of equation $y = 0$ in that plane. In this chart, we consider the functions $f_1 : x \mapsto 2/(1 + x^2)$, and for all $n \geq 2$, $f_n : x \mapsto f_1(nx)/n$. Abusively, we still denote their graphs by the same letters, and then, we may extend these arcs, viewed in Σ, to simple closed curves (consisting of pushing a aside), that converges C^1 to a outside of the point p. Abusively we still use the same letters f_n to denote these closed curves.

The sequence $(f_n)_{n\geq 1}$ does not converge C^1 to a, because f_n has slope -1 at the point $(1/n, 1/n)$.

Nonetheless, we claim that $W((f_n), a)$ holds. Indeed, let d be such that $a - d$. Since the sequence (f_n) converges C^1 to a everywhere except at the origin of this $\mathbb{R}^2$ chart, the only case in which it is not already clear that $(f_n) - d$ is when d meets a transversely at the origin. If d has a strictly positive slope there, then for n large enough, the intersection $f_n \cap d$ will be transverse because the slopes of f_n are all negative in the region $x > 0$. The case when d has negative slope is symmetric, and if d has vertical slope, it will be transverse with f_n since these have bounded slopes.

5.3 Recognizing configurations of smooth curves

In this last section we assume that our surface Σ is orientable.

If a, b are vertices of $\mathcal{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$, we will denote by $D_\infty(a, b)$ the condition that $a - b$ and for all sequences (f_n) and (g_m) such that $W((f_n), a)$ and $W((g_m), b)$, we have $f_n - g_m$ for all m, n large enough.

Lemma 43 *Let a and b be smooth nonseparating curves. Then $D_\infty(a, b)$ holds if and only if a and b are disjoint.*

Proof Suppose first that a and b are disjoint. Then they admit disjoint neighborhoods, V_1 and V_2. For any sequences (f_n) and (g_m) with $W((f_n), a)$ and $W((g_m), b)$, for all m and n large enough we have $f_n \subset V_1$ and $g_m \subset V_2$, by Lemma 40. Hence, $f_n - g_m$ for all m and n large enough, and $D_\infty(a, b)$ holds indeed.

Now, suppose that a and b are not disjoint. Since $a - b$, the curves a and b have a transverse intersection, and in an appropriate chart diffeomorphic to $\mathbb{R}^2$, the curves a and b correspond respectively to the axes $y = 0$ and $x = 0$.

Then we may form a sequence (f_n) such that $W((f_n), a)$ exactly as in Example 42, and for (g_n) we just exchange coordinates x and y. For all n, the curves f_n and g_n have a nontransverse intersection point (at $(1/n, 1/n)$ in the chart of Example 42); hence the condition $D_\infty(a, b)$ does not hold. $\qquad\square$

In the end of the proof, the curves f_n and g_n were tangent at their intersection point, hence not neighbors in the graph $\mathcal{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$. This tangency may look like an accidental reason for f_n and g_n not being neighbors in the graph $\mathcal{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$, but upon changing the formula of f_1 in Example 42 by a small C^1 perturbation we can get f_n and g_n to have, for example, infinitely many transverse intersection points.

From now on, we restrict ourselves to the case of orientable surfaces. One reason is that it would take more work to recover the extra bouquets and not only the bouquets; one other reason is that the next lemma works best when at least one of a, b or c is two-sided.

Lemma 44 *Suppose Σ is orientable. Let $\{a, b, c\}$ be a 3–clique of $\mathcal{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$, and suppose that these three curves pairwise intersect. Then the following are equivalent:*

(1) *This 3–clique is of type bouquet.*

(2) *There exists a relevant sequence (f_n) of vertices of $\mathcal{NC}^{\dagger\infty}_{\pitchfork}(\Sigma)$, which are all disjoint from a, and such that for all d disjoint from b and satisfying $c - d$, we have $(f_n) - d$.*

Proof Suppose $\{a, b, c\}$ is of type bouquet. Then the sequence $(f_n)_{n \in \mathbb{N}}$ can be constructed explicitly. Let $p = b \cap c$. Fix a (smooth) metric on Σ, we remove all points of the ball $B(p, 1/n)$ off the curves b and c, this gives two arcs. There is a natural way of adding smooth subarcs of $B(p, 1/n)$ in order to extend this union of two arcs, to a curve f_n which does not intersect a. In a one-holed torus neighborhood of $b \cup c$, with a choice of meridian and longitude coming from b and c, these curves f_n have slope 1, or -1; these are indeed nonseparating simple closed curves. Now if d is disjoint from b and satisfies $c - d$, then either d is disjoint from c, and then f_n is disjoint from d for all n large enough, or d has a transverse intersection with c at a point distinct from p, and we also have $f_n - d$ for all n large enough. Thus, (1) implies (2).

Conversely, suppose (2). We first claim that the sequence (f_n) then concentrates into neighborhoods of $b \cup c$. For contradiction, suppose that we can find a neighborhood V of $b \cup c$, such that $f_n \not\subset V$ for infinitely many n. Then, there exists a point x, with $x \notin b \cup c$, and such that every neighborhood of x meets infinitely many f_n. Then we may choose three bottle-shaped arcs around x, and complete them into curves d_1, d_2 and d_3 disjoint from b and satisfying $d_j - c$ for $j = 1, 2, 3$. Indeed, we may start with a curve d_0 obtained by pushing b aside, and then perform surgeries on d_0. The same reasoning as in the proof of Lemma 39 shows that $f_n \not\sim d_j$ for some $j \in \{1, 2, 3\}$ and for infinitely many n, contradicting the hypothesis (2).

Now, suppose for contradiction that $\{a, b, c\}$ is a necklace. Then, for a sufficiently small regular neighborhood V of $b \cup c$, we may observe that $V \smallsetminus a$ is contractible. Hence it cannot contain any nonseparating simple closed curve f_n, and the hypothesis (2) cannot be fulfilled. This proves that (2) implies (1). $\square$

Now the proof of Theorem 2 is a straightforward adaptation of the proof of Theorem 1. The statements about connectedness of complexes of arcs, for example, are equivalent to their counterparts with regularity, because of the argument of homotopy recalled in the proof of Lemma 27 and borrowed from [2].

References

[1] **J Bowden**, **S Hensel**, **K Mann**, **E Militon**, **R Webb**, *Rotation sets and actions on curves*, Adv. Math. 408 (2022) art. id. 108579 MR Zbl

[2] **J Bowden**, **S W Hensel**, **R Webb**, *Quasi-morphisms on surface diffeomorphism groups*, J. Amer. Math. Soc. 35 (2022) 211–231 MR Zbl

[3] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[4] **D Fischer**, *Smoothness of division of infinitely differentiable functions*, reply on Mathematics Stack Exchange board (2016) Available at https://math.stackexchange.com/questions/1606068/smoothness-of-division-of-infinitely-differentiable-functions

[5] **W J Harvey**, *Boundary structure of the modular group*, from "Riemann surfaces and related topics" (I Kra, B Maskit, editors), Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 245–251 MR Zbl

[6] **S Hensel**, **P Przytycki**, **R C H Webb**, *1–Slim triangles and uniform hyperbolicity for arc graphs and curve graphs*, J. Eur. Math. Soc. 17 (2015) 755–762 MR Zbl

[7] **N V Ivanov**, *Automorphisms of complexes of curves and of Teichmüller spaces*, Int. Math. Res. Not. 1997 (1997) 651–666 MR Zbl

[8] **M Korkmaz**, *Automorphisms of complexes of curves on punctured spheres and on punctured tori*, Topology Appl. 95 (1999) 85–111 MR Zbl

[9] **F Le Roux**, *L'ensemble de rotation autour d'un point fixe*, Astérisque 350, Soc. Math. France, Paris (2013) MR Zbl

[10] **F Le Roux**, **A Passeggi**, **M Sambarino**, **M Wolff**, *A note on weak conjugacy for homeomorphisms of surfaces*, preprint (2024) arXiv 2407.01042

[11] **A Long**, **D Margalit**, **A Pham**, **Y Verberne**, **C Yao**, *Automorphisms of the fine curve graph* (2021) arXiv 2108.04872 To appear in Trans. Amer. Math. Soc.

[12] **F Luo**, *Automorphisms of the complex of curves*, Topology 39 (2000) 283–298 MR Zbl

[13] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR Zbl

[14] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR Zbl

[15] **K Rafi**, **S Schleimer**, *Curve complexes are rigid*, Duke Math. J. 158 (2011) 225–246 MR Zbl

Institut de Mathématiques de Jussieu-Paris Rive Gauche, UMR 7586, CNRS, Univ. Paris Diderot, Sorbonne Paris, France

Institut de Mathématiques de Jussieu-Paris Rive Gauche, UMR 7586, CNRS, Univ. Paris Diderot, Sorbonne Paris, France

frederic.le-roux@imj-prg.fr, maxime.wolff@imj-prg.fr

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT 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 Algebraic & Geometric 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 AGT is available on the AGT 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.